SLAC–PUB–8026
Decmber 1998

Introduction to Microwave Linacs*

D.H. Whittum
Stanford Linear Accelerator Center, Stanford University
Stanford, California

Abstract

The elements of microwave linear accelerators are introduced starting with the principles of acceleration and accelerating structures. Considerations for microwave structure modelling and design are developed from an elementary point of view. Basic elements of microwave electronics are described for application to the accelerator circuit and instrumentation. Concepts of beam physics are explored together with examples of common beamline instruments. Charged particle optics and lattice diagnostics are introduced. Considerations for fixed-target and colliding-beam experimentation are summarized.

Contributed to Proceedings of the NATO Advanced Study Institute on Techniques and Concepts of High-Energy Physics St. Croix, US Virgin Islands, 16-29 June 1998

*Work supported by Department of Energy contract DE–AC03–76SF00515.

Table of Contents

1. Introduction

The microwave linear accelerator or *linac* is a hybrid of microwave electronics, mechanical engineering, metallurgy, and craftsmanship. The primary component is a *structure* consisting of cups cut from copper, stacked, brazed, baked, mounted and fed with a copper tube, a *waveguide*. Every eight milliseconds the structure is pulsed with microwave power, megawatts for microseconds. If the copper cups have been shaped and tuned to one part in ten-thousand, they resonate in concert, driven by the incident wave. Sparks or *beams* fly through at about the speed of light, surfing the electromagnetic wave. Beams are wayward things and *magnets* are used to guide them. At the end of the linac, high-energy physicists pan through the debris, wondering when the statistics will improve. In these notes we contemplate why a linac looks as it does, and how it works.

In Sec. 2 we start with the problem of acceleration and develop the logic behind the multi-cell structure, and the electromagnetic fields it can support. Detailed calculations aren't really necessary here, but may be found in [1]. Appreciating the character of the fields we may provide for acceleration, we determine what manner of beam may benefit from them. In Sec. 3 we develop the subject of beam-physics, consisting mostly of a view of a beam as a collection of single particles, communicating, if at all, through their collective fields. We include discussion of beamline instrumentation, magnetic optics, and an upstream view of fixed-target and collider experiments. The topography of the whole subject is sketched in Fig. 1. We will concentrate only on the "main linac", the region of the machine where the beam is highly relativistic. With electrons in mind, this is most of the machine. For definiteness, examples are drawn for the most part from the Two-Mile Accelerator [2, 3].

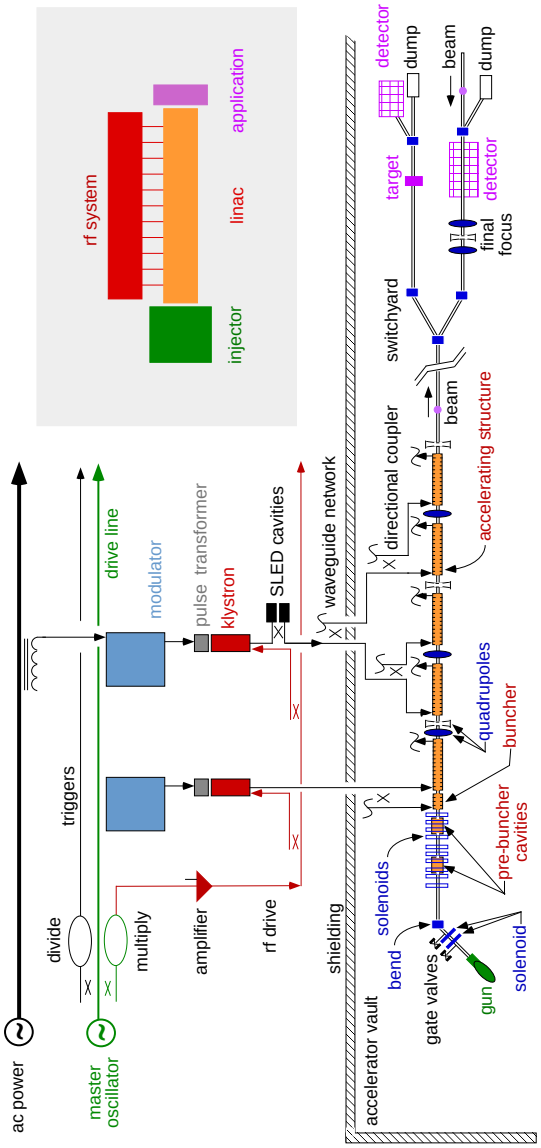

Figure 1. A microwave accelerator consists of a number of sub-systems working together. These include the microwave system—colloquially, the "rf" system—and the injector. These incorporate accelerating structures, requiring cooling water, mechanical support and alignment, and a vacuum system. The injector includes a gun, and may incorporate a laser system or an rf bunching system. Beam transport requires magnets, and their associated power supplies and cooling. Reliable operation requires rf and beam-monitoring by means of instrumentation circuits. In these notes we introduce the essential features of the linac, the beam, and the instrumentation.

2. Principles of Acceleration

2.1. ELECTRODYNAMICS

Classical electrodynamics is the foundation for the principles of acceleration, and is worth a short review. We consider a particle of charge q, and associate with it a position $\vec{R}$ and a velocity $\vec{V} = d\vec{R}/dt$. Mechanical momentum is given by $\vec{P} = m\gamma\vec{V}$, with m the particle mass,

$$\gamma = \left(1 - \frac{\vec{V}^2}{c^2}\right)^{-1/2}, \tag{1}$$

the Lorentz factor, and c the speed of light. Kinetic energy is

$$\varepsilon = mc^2\left(\gamma - 1\right) = \left(m^2c^4 + c^2\vec{P}^2\right)^{1/2} - mc^2. \tag{2}$$

Particle motion is governed by

$$\frac{d\vec{P}}{dt} = \vec{F}, \tag{3}$$

where the Lorentz force

$$\vec{F} = q\left(\vec{E} + \vec{V} \times \vec{B}\right), \tag{4}$$

is determined from $\vec{E}$, the electric field in units of V/m, and $\vec{B}$, the magnetic induction or flux density in units of $\mathrm{Wb/m^2}$ or T. This relation defines the fields, abstracts them from their sources, describes the motion of particles, and the response of media. It is half of electrodynamics.

Where the particle is relativistic, V is close to c, and the effect of the magnetic induction can be appreciable. It is easier to produce 1 T, than it is to produce the equivalent 3×10^8 V/m. For this reason magnets are used to control the motion of highly relativistic particles. At the same time, an electric field is needed to do work, to "accelerate". A linac includes accelerating structures to shape the electric field for acceleration, and magnets to shape the magnetic induction needed for beam guidance.

Associated with electric field and magnetic induction, are two constructs, the electric displacement $\vec{D}$ and the magnetic field $\vec{H}$. In vacuum, $\vec{D} = \varepsilon_0\vec{E}$ and $\vec{H} = \vec{B}/\mu_0$, where ε_0 and μ_0 define a choice of units, subject to $1/\sqrt{\mu_0\varepsilon_0} = c \approx 2.9979 \times 10^8$m/s. In practical units, $\sqrt{\mu_0/\varepsilon_0} = Z_0 \approx 376.7\ \Omega$. In media, and in the frequency domain, these expressions take the form $\vec{D} = \varepsilon\vec{E}$ and $\vec{H} = \vec{B}/\mu$. The quantity ε is the permittivity and μ is the permeability. These fields are governed by Maxwell's equations,

$$\vec{\nabla} \bullet \vec{D} = \rho, \tag{5}$$

$$\vec{\nabla} \bullet \vec{B} = 0, \tag{6}$$

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}, \tag{7}$$

$$\vec{\nabla} \times \vec{H} = \frac{\partial \vec{D}}{\partial t} + \vec{J}. \tag{8}$$

$\vec{J}$ is the external current density, that due to charged particles not already incorporated in μ, and ρ is the external charge density. It is rare in practice to drag out Maxwell's equations, as typically one employs a derived circuit-equivalent for accelerator structures. However, it is often helpful to refer to conservation of energy and momentum.

We consider a volume V in which we find total charge density ρ and current density $\vec{J}$. Employing Maxwell's equations one may show that

$$-\vec{J} \bullet \vec{E} = \vec{\nabla} \bullet \vec{S} + \frac{\partial u}{\partial t}, \tag{9}$$

where the Poynting flux is

$$\vec{S} = \vec{E} \times \vec{H}, \tag{10}$$

and energy density is

$$u = \frac{1}{2}\left(\varepsilon_0 \vec{E}^2 + \mu_0 \vec{H}^2\right). \tag{11}$$

In practice energy conservation is employed as a check of one's understanding of a circuit. In steady-state, power incident on an accelerator circuit should be balanced by what is reflected, transmitted, dissipated, and taken up by the beam.

Momentum conservation also figures prominently. The Lorentz force law may be expressed as

$$\frac{d\vec{P}_{mech}}{dt} = \int_V dV \left(\rho \vec{E} + \vec{J} \times \vec{B}\right),$$

where $\vec{P}_{mech}$ is the mechanical momentum of the constituent particles within the volume V. Applying Maxwell's equations one may cast this in the form

$$\left(\frac{d\vec{P}_{mech}}{dt} + \frac{d\vec{P}_{em}}{dt}\right)^a = \int_{\partial V} T^{ab} dS_b, \tag{12}$$

where summation is implied over repeated indices and $a, b = 1, 2, 3$ index Cartesian spatial coordinates. The momentum associated with the fields in the volume is

$$\vec{P}_{em} = \frac{1}{c^2} \int_V dV \, \vec{S},$$

and $\vec{S}$ is the Poynting vector of Eq. (10). The flux of momentum through ∂V, the bounding surface, is given by the electromagnetic stress tensor,

$$T^{ab} = \varepsilon_0 E^a E^b + \mu_0 H^a H^b - u\delta_{ab}, \tag{13}$$

with δ_{ab} the Kronecker delta. This result will be helpful for structure design.

2.2. MATERIALS NEEDED

As to the matter of acceleration, Eq. (4) implies that a particle's energy ε varies according to

$$\frac{d\varepsilon}{dt} = q\vec{V} \bullet \vec{E}, \tag{14}$$

and with this we may classify accelerators according to the method of producing $\vec{E}$: electrostatic, inductive or fully electromagnetic. The first two are prominent in the history of accelerators, and to the present-day. However, for high-energy particles there is no substitute for the last, the microwave accelerator. It is not obvious at first how to employ an electromagnetic wave for acceleration. In the 1930's this was considered a research project [4]. In free-space, we are familiar with the mechanism of radiation pressure. Particles jitter transversely in a passing wave, and their jitter velocity causes them to be deflected in the forward direction by the transverse magnetic field. In this way particles may be pushed along by the wave, accelerated. This is just Thomson scattering. One intuits however that high-energy particles, with large relativistic inertia, jitter little. Thus radiation pressure or "second order acceleration" doesn't work well at high-energy, at least not directly. One may in principle circumvent the problem by introducing another external field to enforce the requisite transverse motion. This could be another wave, or a static magnetic field, as is seen in the *inverse free-electron laser* and the *inverse cyclotron maser*. The problem is solved in *laser-plasma accelerators*, by arranging that the radiation pressure act on low-energy plasma electrons, radial currents from which then induce an electric field, parallel to the high-energy beam and adequate for acceleration.

Historically, however, and perhaps logically, there is a more direct approach, "first-order acceleration". To see what this entails, we consider first a wave in free-space, and first-order particle motion in the wave. The wave may be decomposed into plane-waves, and at first-order in the applied fields, the impulse received by the particle is just the sum of the impulse from each wave. Thus it is enough to analyze motion in a plane wave of a particular angular frequency ω. We denote the component of wavenumber parallel to the particle motion, $k_{||} = \hat{V} \bullet \vec{k}$, with $\vec{k}$ the wavevector, and $\vec{V}$ the particle velocity. Net energy gain of the particle is given by the integral

over the particle displacement s

$$\Delta\varepsilon = \int ds\ E_{||}\ \cos\Big(\omega\, t - k_{||} s\Big),$$

with $E_{||}$ the parallel component of the electric field. Energy gain depends on the phase $\psi = \omega\, t - k_{||} s$ witnessed by the particle. Notice however that

$$\frac{d\psi}{ds} = \frac{\omega}{V} - k_{||} > 0$$

since in free-space $\omega/|\vec{k}| = c > V$. Wave-crests continually flow past the particle alternately accelerating or decelerating it. On average $\Delta\varepsilon = 0$. Secular first-order acceleration in infinite free-space is not possible.

There are two solutions to this problem. One is to terminate the interaction of the particle with the wave, the other is to modify the dispersion characteristics of the wave to obtain synchronism with the particle. In the former case, the wave is trapped in the accelerator, forming a *standing-wave*; in the latter case, one has a *travelling-wave*. Both require the introduction of media. While dielectrics, plasmas, and other media are conceivable, and quite interesting as research projects, Hansen's original concept for a geometry employed *conducting* boundaries [4], and in fifty years has proved hard to beat. To see how microwaves within a conducting geometry may provide acceleration, it is enough at first to consider only a single cavity.

2.3. THE ACCELERATOR CAVITY

We consider how to describe and design a cavity useful for acceleration. First let us reduce Maxwell's equations to a simpler, pictorial form for a perfect conductor-vacuum geometry, as in Fig. 2. As we will see, structures are made from material of high-conductivity so as to minimize power loss. A high-conductivity material will quickly "short" any tangential electric field at its surface. Thus due to the ease with which charge redistributes itself over a conducting surface, electric field lines in a cavity terminate normal to a boundary, or not at all, closing on themselves. Similarly, currents are induced and flow to cancel the magnetic field within the conductor. There is no magnetic charge, however and so magnetic field lines never terminate and they can only close on themselves. This much we obtain from Gauss's Law, Eq. (5), and the solenoidal condition, Eq. (6). Faraday's Law, Eq. (7), and Ampere's Law, Eq. (8), combined describe the mutual excitation of electric and magnetic field, and dictate oscillation in a hollow conductor. In circuit terms, a region of space permeable to magnetic field presents an inductance, and a region permitting electric field presents a capacitance, and where both

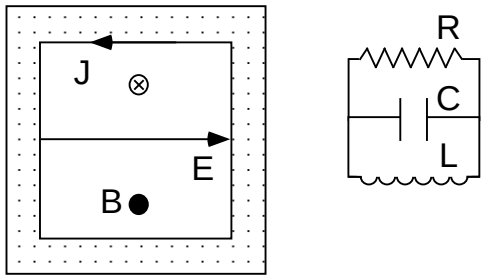

Figure 2. Maxwell's equations applied to a conductor-vacuum geometry reduce to a prescription for field lines, and wall currents. In the first approximation, the system behaves as an LC circuit. Equivalent resistance is low.

are present one expects to see behavior as in an LC circuit, oscillation at angular frequency $\omega = 1/\sqrt{LC}$. These considerations, summarized in Fig. 2, suggest that a cavity may be useful not only for field shaping, and thus first-order acceleration, but that the LC character of the system might be exploited for resonant excitation.

Yet conductors are not perfect. Wall currents flow through a lossy medium, resistance is present. Our LC circuit is really an RLC circuit. To quantify resistance, let us examine the matter of rf-dissipation for a normal conductor. Ohm's law takes the microscopic form $\vec{J} = \sigma\vec{E}$, where σ is the conductivity. Preferring higher conductivity, we might consult a listing of materials as in Table 1, and find that, considering cost, there is nothing

TABLE 1. DC electrical resistivity of some example materials in units of $1.7 \times 10^{-8}\Omega$-m.

Material	Resistivity	
stainless steel	42	(300-series)
free-cutting brass	4.0	(63% Cu, 34% Zn 3% Pb 0.15% Fe)
brass	2.2	(66% Cu, 34% Zn)
aluminium	$1.5 - 2.9$	
gold	1.4	
copper	1.0	(OFE)
silver	0.93	

better than the item labelled "OFE copper", the stuff of linacs. This is 99.99% pure, "oxygen-free electronic-grade" copper, "C10100" in the Unified Numbering System for Metals. Conductivity is $\sigma \approx 5.8 \times 10^7$ mho/m.

In the Drude model the figure σ arises from electron motion in the applied field, subject to drag due to collisions with the ions. With collisions,

electrons undergo a random-walk, and thus it is not surprising to find that the magnetic field, in the conductor satisfies a diffusion equation. In a random-walk process, diffusion depth is proportional to square-root of time, and in fact, the penetration depth, or *skin-depth*, for an rf-excitation takes the form $\delta = \sqrt{2/\mu\sigma\omega}$, with ω the angular frequency of the excitation. At 3 GHz ("S-Band") in copper, $\delta \approx 1\ \mu$m.

Knowing the skin-depth, one may determine power dissipation in a conducting boundary. The wall current density J should be sufficient to cancel the local magnetic field H within the conductor, on the order of $J \approx H/\delta$. This results in a power dissipation per unit volume J^2/σ within the conductor, so that net dissipated Poynting flux is $\delta\, J^2/\sigma \approx R_s\, H^2$. The surface resistance is, $R_s = 1/\sigma\delta$, 14 mΩ for copper at S-Band, and varying as $\omega^{1/2}$. Net dissipated power may be expressed then in terms of an integral over the conducting boundary, and quantified in terms of a *wall quality factor*, Q_w,

$$P_w = \frac{1}{2} R_s \int\limits_{wall} \vec{H}^2 dS \equiv \frac{\omega\, U}{Q_w}. \tag{15}$$

A factor of 1/2 arises from time-averaging, and $\vec{H}$ is the peak magnetic field. The energy stored in the cavity, U, may be expressed in terms of $\vec{H}$, from Eq. (11), with a volume integral, permitting us to solve for

$$\frac{1}{Q_w} = \frac{\delta}{2} \frac{\int\limits_{wall} \vec{H}^2 dS}{\int\limits_{volume} \vec{H}^2 dV}. \tag{16}$$

Roughly speaking, in each oscillation in a volume V, with a conducting surface area A, a fraction of the energy is dissipated, in proportion to the lossy volume $A\delta$. Quality factor is $Q_w \approx V/2A\delta$ in order of magnitude. Since cavity dimensions are of order the free-space wavelength λ one can see that $Q_w \approx \lambda/4\delta$ will be a large number. The corresponding decay-time for fields is

$$T_0 = \frac{2Q_w}{\omega}, \tag{17}$$

as one can check by solving $dU/dt = -P_w = -\omega U/Q_w$. For an S-Band accelerator $\lambda \approx 0.1$ m and $Q_w \approx 10^4$; the decay-time may be quite long, 1 μs, even while the oscillation period is short, 0.3 ns.

This discussion provides a description of a cavity as an RLC circuit. Next let us consider just what kind of accelerator our cavity can be. We picture an initially unexcited cavity, through which a beam-tube has been cut, as in Fig. 3. We suppose the beam tube is small enough that the fields cannot propagate into it, but remain confined. We consider the transit of a relativistic charge q through the cavity.

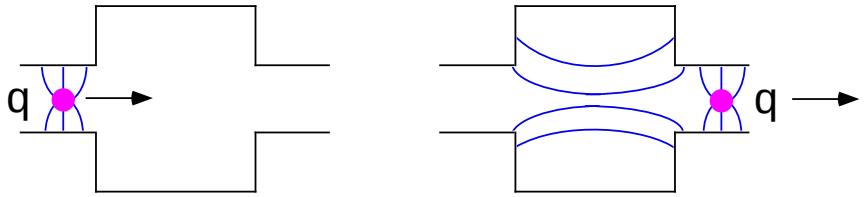

Figure 3. We consider the interaction of a charged particle with a cavity.

After the particle has left the cavity we suppose that energy U has been deposited in the the accelerating mode of the cavity. From the law of energy conservation, we expect a quadratic relation $U = k_l q^2$ for some choice of constant k_l characteristic of the cavity mode. This *loss-factor* has units of V/C. After the particle has left the cavity, the accelerating mode oscillates with some amplitude, V. To relieve the term "amplitude" of any ambiguity, we define V in units of voltage such as would be witnessed by a trailing particle. Notice that the energy loss by the charge q itself may be seen as arising through the action of this voltage. In transit through the cavity the charge witnesses an electric field that starts at 0 and rises to its maximum value. Thus the effective self-induced voltage is $V/2$, so that $U = qV/2$, or $V = 2k_l q$. Single-bunch beam-induced voltage is determined by the bunch charge and the loss-factor. A bit of algebra shows then that $U = V^2/4k_l$. This last relation, between energy and voltage must be independent of the means by which the voltage was produced. Thus if we wish to employ this cavity to accelerate, we must provide some energy, and the amount of energy is determined from the voltage required, and the loss-factor. Large loss-factor implies that the cavity is frugal with energy. It says though nothing about dissipation—that's a different kind of frugal that we'll come to shortly. It also implies that beam-induced voltage or "beam-loading" is large. A good accelerator is a good decelerator. This reciprocity between externally-supplied voltage, and beam-induced voltage is referred to as the fundmantal theorem of beam-loading.

If large loss-factor is good, how large can loss-factor be? We consider the energetics of the accelerator cavity in more detail. Suppose the length of the cavity is L and the cross-section is A. In terms of peak on-axis electric field E stored energy is roughly

$$U \approx \frac{1}{2}\varepsilon_0 E^2 \times \frac{1}{4} \times AL,$$

with a factor of 1/4 to account for the radial variation in E^2. Next, let us relate E to "gap-voltage" V. For an electron crossing a gap of length L,

with oscillating field of amplitude E, voltage gain is *not* $E \times L$, rather it is

$$V = \int ds E\left(t = t_0 + \frac{s}{V}\right)$$

for an electron entering at time t_0. For an electric field varying sinusoidally in time, uniformly across the gap, the result is that maximum voltage gain is $V \approx E \times L \times T$, where $T = \sin(\theta/2)/(\theta/2)$ is the *transit-time factor*, and $\theta = \omega L/V$ is the *transit-angle*. For a fixed field amplitude, maximum voltage gain corresponds to $\theta \approx \pi$. In this case, the electron enters the cavity when the field $E = 0$ and leaves just as $E = 0$ a half-cycle later.

With this we may compute

$$k_l = \frac{V^2}{4U} \approx \frac{L}{\varepsilon_0 A} T^2,$$

roughly the inverse of the capacitance of two parallel plates. At first one might think that large loss-factor would be easy to obtain by lengthening the cavity, or decreasing the plate area. However, loss-factor has a maximum at $L \approx 0.371\lambda$; greater lengths lead to lower k_l due to the transit-time factor. At the same time, cross-section cannot be arbitrarily small, since the electric field must vanish at the outer boundaries. For simple geometries, such as circular or rectangular pipe, this implies minimum $A \approx \lambda^2/2$. This corresponds to loss factor $k_l \approx 1/\varepsilon_0\lambda$, or 1 V/pC for an S-Band cavity. Shorter wavelength accelerators can make do with lower stored energy, but they exhibit heavier beam-loading, and thus must be operated with lower bunch charge. Associated with loss-factor is a second, more commonly employed quantity, $[R/Q] = V^2/\omega U = 4k_l/\omega$, "R-over-Q". This is an awkward notation, but it is conventional. Our estimate corresponds to

$$\left[\frac{R}{Q}\right] \approx \frac{8}{\pi^2} Z_0 \frac{\sin^2(\theta/2)}{(\theta/2)}$$

with a broad maximum in the vicinity of 220 Ω for cavity length in the range $L \approx \lambda/4 - \lambda/2$. $[R/Q]$ is a single thing, not two as the notation unfortunately seems to imply, and it is a function only of the *shape* of the cavity, not the scale.

Notice that $[R/Q]$ and k_l have no connection with wall conductivity or dissipation. They characterize energy-loss of a short bunch, one much shorter than a period, and very much shorter than a damping time. These quantities know nothing of dissipation. They determine the energy that must be stored to establish a prescribed gap voltage.

Understanding the geometric scalings for the energetics of the conductor-vacuum geometry forming an accelerator cavity, we can turn next to the

practical matter of putting energy into the cavity. For this we need a short excursion into the subject of waveguide.

2.4. WAVEGUIDE FOR ALL OCCASIONS

To move a signal from point A to point B, one needs "waveguide", two-conductor cable (coaxial, stripline) or hollow waveguide (rectangular, circular, elliptical). The two most common gometries are illustrated in Fig. 4, and we'll visit with these at length in Sec. 2.14. However, just now we are concerned with the *function* of the guide; independent of the particular guide geometry employed, this may be pictured quite simply. There are two directions in the guide, forward and reverse, or "+" and "−". We consider a single-frequency, steady-state excitation, and inspect first the case of a wave propagating in one direction, call it the "+" direction. We may describe the amplitude of the wave by considering the electric field between the conductors, and gauging its magnitude and phase by a phasor with units of voltage, call it V_+. There is also a magnetic field and we may describe it by a phasor with units of current I_+. Precise definitions of these phasors would refer to a choice of normalization, and one might be concerned that such a choice would be arbitrary. In fact, there is arbitrariness in the choice of normalization for such "circuit-equivalent" descriptions of wave systems. We can take a bit of the arbitrariness out by asking that the quantity $P_+ = \Re V_+ I_+^*/2$ should correspond to the power flow through the guide, i.e., the Poynting flux integrated across the waveguide cross-section. This still leaves one with the freedom to choose a normalization corresponding to something with units of impedance $Z_w = V_+/I_+$, referred to as the waveguide impedance or waveguide mode impedance.

Next we consider the more general situation consisting of both forward and reverse waves together on the line. The net voltage and current phasors at a point z along the guide are given by

$$\tilde{V}(z,\omega) = V_+(\omega)\, e^{-j\beta z} + V_-(\omega)\, e^{j\beta z}, \tag{18}$$

$$Z_w \tilde{I}(z,\omega) = V_+(\omega)\, e^{-j\beta z} - V_-(\omega)\, e^{j\beta z}, \tag{19}$$

where β is the wavenumber in the guide geometry. Let us abbreviate $R = V_-(\omega)/V_+(\omega)$, thinking of this as a reflection coefficient due to a device placed downstream on the cable. The actual impedance at the plane z is then

$$Z = \frac{\tilde{V}(z,\omega)}{\tilde{I}(z,\omega)} = Z_w \frac{1 + e^{2j\beta z} R}{1 - e^{2j\beta z} R}. \tag{20}$$

This result indicates that for a short length of cable, with $\beta z << 1$, spatial phase-shift through the cable is small, and the guide is a lumped element.

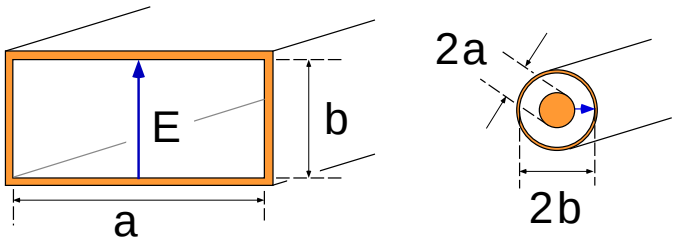

Figure 4. The two most common choices for waveguide are rectangular guide and coaxial line.

Longer cable constitutes a transmission line for waves. Generically we may call such a circuit, one comparable to a wavelength in size, a *microwave circuit*. Above 1 GHz most circuits are microwave circuits. A linac is a microwave circuit.

Returning to our forward and reverse waves, and the resulting standing-wave on the line, consider that if one probes the voltage associated with this standing-wave, as a function of z, one finds maximum voltage $|V_+| + |V_-|$, and minimum $|V_+| - |V_-|$. The ratio of maximum to minimum is

$$VSWR = \frac{1+|R|}{1-|R|}. \tag{21}$$

This *voltage standing-wave ratio* may be viewed as a property of the device to which the waveguide is attached, and it is a function of frequency. Typically microwave devices come with a specification for maximum VSWR over a range of frequencies. Let us consider then the microwave device foremost in our considerations just now, the accelerator cavity.

2.5. VIRTUES OF A RESONATOR

To power our cavity we consider a picture like that of Fig. 5. We make a small hole in the cavity wall and attach a waveguide to it. The guide transmits a wave through the hole or "coupling iris". This is just an electromagnetic version of the "ripple tank". The wave incident on the iris is initially mostly reflected. However, over time, the cavity gradually fills and as it does it begins to radiate; waves diffract through the coupling hole back into the external guide. Eventually, due to losses in the cavity walls, the system reaches steady-state, and the diffracted wave may partially or totally cancel the reflected wave. If the coupling iris geometry is such that the cancellation is total, the cavity is said to be "critically coupled". For a critically coupled cavity in steady-state, the incident power flows down the guide and into the cavity walls.

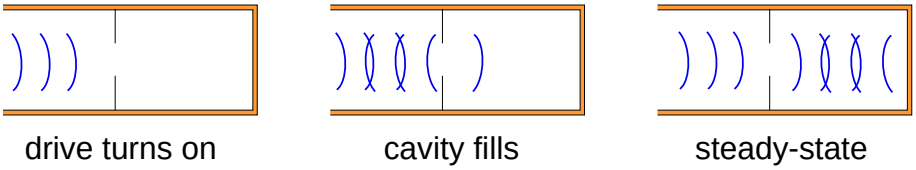

Figure 5. In the transient filling of a cavity by a forward wave in the connecting guide, the reverse wave is a superposition due to reflection and diffraction.

The effect of wall-losses, and external coupling may be quantified in terms of the wall quality factor, Q_w, and a "diffractive" or "external" quality factor, Q_e. If the energy stored in the cavity volume is U, then the power flowing into the walls is $P_w = \omega U/Q_w$. If the incident power is turned off, then the power flowing out of the cavity and down the waveguide is $P_e = \omega U/Q_e$. The net rate at which energy is leaving the cavity volume—with rf drive off— is the sum of these rates, and may be characterized by the "loaded Q",

$$\frac{1}{Q_L} = \frac{1}{Q_e} + \frac{1}{Q_w}. \tag{22}$$

Just as Q_e limits the rate at which energy may leave the cavity, it also limits the rate at which energy may be put into the cavity by means of external rf-drive. In general, the response-time of the cavity to external drive alone is the loaded fill-time

$$T_f = \frac{2Q_L}{\omega}. \tag{23}$$

If the cavity is critically coupled then in steady-state the rate of energy flow into the cavity should match the rate at which energy is absorbed in the walls, or $1/Q_w = 1/Q_e$ In this case, $T_f = T_0/2$. In general we may characterize coupling by the figure $\beta = Q_w/Q_e$, the "coupling parameter". Loaded fill time is then $T_f = T_0/(1+\beta)$.

Appreciating all this we can determine the power requirements for a critically-coupled cavity in steady-state. The forward power P_F in the connecting guide is just the power dissipated in the walls

$$P_w = \frac{\omega}{Q_w} U = \frac{\omega}{Q_w} \frac{V^2}{\omega\,[R/Q]} = \frac{V^2}{R_{shunt}}, \tag{24}$$

or $V = \sqrt{R_{shunt} P_w}$, where $R_{shunt} = Q_w\,[R/Q]$ is the *shunt-impedance*.

According to our calculations, a single normal-conducting S-Band cavity may have a shunt impedance of at most about 3 MΩ. Thus to achieve 1 MV of acceleration, we need a 0.3 MW source with a pulse length of about 1 μs. This steady-state power requirement is set by dissipation in copper. Thus if one had a less lossy material—a superconductor, for example— peak power

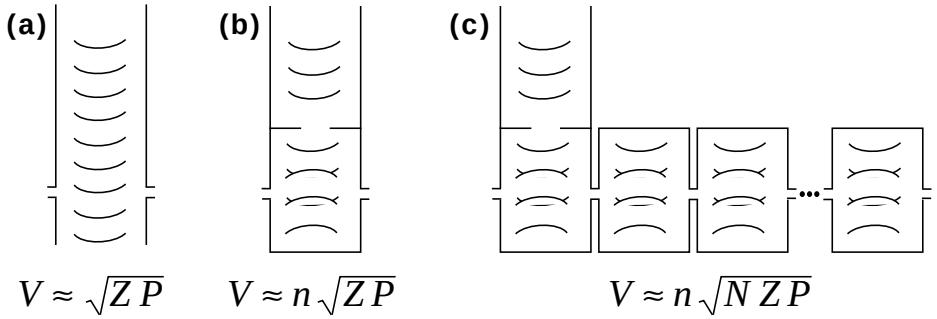

Figure 6. Summarizing the logic leading to the multi-cell linac. Accelerating voltage V is determined from the steady-state incident power P. The ratio V^2/P ("shunt-impedance") will be of order $Z \approx Z_0$ in free-space, as in (a). This can be improved greatly by use of resonant excitation, as in (b), as quantified by transformer ratio n, and by use of N cells as in (c). The beam travels to the right in this view.

requirements would be lower. To emphasize the point, consider that if one had a perfectly lossless cavity, and were running the machine at 120 Hz pulse repetition frequency, then with 8.3 ms between pulses, rf drive at the level of 30 W would be adequate to store the 1/4 J needed to establish 1 MV. What this means is that the shape, size, cost, and almost every other attribute of normal conducting microwave linacs is determined by losses in copper.

TABLE 2. Scaling of cavity parameters with frequency. No beam-tubes yet.

band	frequency	Q_w	T_0	$[R/Q]$	R_{shunt}
UHF	714 MHz	3.2×10^4	14 μs	$2.2\times10^2\Omega$	7 MΩ
S	2.856 GHz	1.6×10^4	1.8 μs	$2.2\times10^2\Omega$	3.5 MΩ
X	11.424 GHz	8.0×10^3	220 ns	$2.2\times10^2\Omega$	1.8 MΩ
W	91.392 GHz	2.8×10^3	9.8 ns	$2.2\times10^2\Omega$	0.6 MΩ

The resulting scalings for cavity parameters versus wavelength are summarized in Table 2; values listed there are for an idealized cylindrical cavity with negligibly small beam-tubes. To summarize our progress thus far, we have pursued the logic illustrated in Fig. 6, starting with a terminated interaction, and forming a cavity amenable to resonant excitation. We have arrived at the concept of shunt-impedance, something much bigger than the impedance of free-space, and limited only by losses, since it is derived from resonant energy storage. One sees in Fig. 6(c) the hint that we can

do better still, than a single-cavity, and we will come to this in Sec. 2.13. Before that, a few particulars of the single cavity bear elaboration.

2.6. SLATER'S THEOREM

Performance of a resonant circuit depends on tune and quality factor. The matter of tune relates directly to tolerances on cell-manufacture and assembly. Let us consider a closed lossless pillbox that has been excited in a particular mode. Let the mode frequency be ω, and the stored energy U. Next we slowly make a localized inward displacement of the conducting boundary by an amount $\delta\xi$—a *dimple*. Work done on the mode in the course of dimpling is given by

$$\delta U = -\int T\, dS\, \delta\xi,$$

where

$$T = \frac{1}{2}\varepsilon_0 \vec{E}^2 - \frac{1}{2}\mu_0 \vec{H}^2,$$

is the local electromagnetic stress on the surface, just Eq. (13), evaluated with the help of conducting boundary conditions. The signs and magnitudes in this expression are easily checked, for example, for the cases of a capacitor and an inductor. A charged capacitor feels compressive stress, and "wants" to collapse. An energized inductor wants to explode. One can also see this in the field line picture of the cavity mode, where electric field lines are strung from surface to surface pulling them inward, and magnetic field lines are coiled up inside, pushing out.

As we apply this displacement $\delta\xi$ we are perturbing a harmonic oscillator. This is like varying the length of a pendulum while it is swinging. In this case we have an adiabatic invariant,

$$\delta\left(\frac{U}{\omega}\right) = 0.$$

Adiabatic invariance determines the shift in mode frequency due to the perturbation to the boundary,

$$\frac{\delta\omega}{\omega} = \frac{\delta U}{U} = \frac{1}{2U}\int\left(\mu_0 \vec{H}^2 - \varepsilon_0 \vec{E}^2\right)\delta V. \tag{25}$$

The integral extends over the volume excluded from the cavity by the perturbation, and the perturbed fields are employed in the integral. This is Slater's theorem [5]. Pushing inward on a cavity wall where E is large does negative work, lowers the stored energy, and therefore the frequency.

Pushing inward on a wall where H is large requires that work be done, raises the stored energy, and therefore the frequency.

This result provides the quantitative basis for some practical accelerator bench-work, cavity tuning, and cavity field mapping. An S-Band cavity tunes at about 1 MHz/mil, where 1 mil $= 0.001'' = 25.4\ \mu$m. With $Q \approx 10^4$, allowable mistuning is a fraction of a MHz, i.e., a fraction of a mil. This is a constraint on machining precision and assembly, mostly assembly. The tolerance is sufficiently severe that it implies the need for post-assembly tuning in conjunction with bench measurements. Such microwave measurements also make use of Slater's theorem in various ways. For example one may "map" the electric-field profile in a structure by insertion of a small bead on the beam-axis. The bead perturbs the accelerating mode frequency by an amount proportional to E^2. Thus pulling the bead through the cavity and recording cavity frequency versus bead-position permits one to map the electric field. The external coupling of a cavity to a waveguide may be gauged by placing a shorting plunger on the input line, and plotting resonance frequency versus short position. Tune of a cavity with low Q_e will depend strongly on short position. An additional consequence of Slater's Theorem is the need for temperature regulation. The coefficient of thermal expansion for copper is $\alpha \approx 1.7 \times 10^{-5}$/K, and thus temperature control at the level of 1 K is required.

On this subject we could also mention the matter of *joints*. As seen in Fig. 2, the function of the copper boundary is to carry currents. If, in assembly, two copper boundaries should be joined so as to conduct current, then they should be *bonded* (brazed or diffusion bonded) so as to form a good current-carrying joint. Meanwhile, two flat-looking surfaces merely placed and held together may contact at as few as three points. Thus clamped structures, often tried, often don't work, exhibiting poor tune, and low Q_w. This is easy to understand from the picture of magnetic field lines coiled inside the copper, waiting for the chance to bulge out. A small crack looks like home to any nearby magnetic field, and it will move right in. Slater's theorem tells us that this lowers the mode frequency. Dissipation in the crack tends to lower Q_w. It is possible to make a clamped structure, however the cuts must not interrupt the flow of wall current.

2.7. EQUIVALENT CIRCUIT

To understand bench-measurements involving amplitude and phase, we need a bit more than the energetics we have considered thus far. Let us set down a mathematical description of the single-cavity system coupled to a beam and a waveguide. We concentrate on a single-mode.

The waveguide excitation may be described by the forward-going volt-

age in the connecting guide V_F and the reverse voltage V_R. These are coefficients of transverse electric field in the guide, i.e., wave amplitudes graced with units of voltage. Continuity of the field implies a condition $V_c = V_F + V_R$, with V_c the cavity voltage, where we choose units again such that this is the voltage witnessed by a relativistic particle traversing the cavity. Maxwell's equations may be expressed in the form

$$\left(\frac{d^2}{dt^2} + \omega_0^2\right) V_c = [lossy\ walls] + [waveguide] + [beam]\,,$$

where terms on the right are currents. The current corresponding to the excitation from the waveguide is just the current that is "missing" due to the coupling iris aperture. This in turn is proportional to the magnetic field at the reference plane. On the other hand, the magnetic field polarity is opposite for the forward and the reverse signals in the waveguide, since the direction of power flow is opposite for the two. Thus the virtual current associated with the waveguide excitation is proportional to $V_F - V_R$. The minus sign here is the same one appearing in Eq. (19). On general grounds then, our cavity-waveguide equation may be expressed as,

$$\left(\frac{d^2}{dt^2} + \omega_0^2\right) V_c = -\frac{\omega_0}{Q_w}\frac{dV_c}{dt} + \frac{\omega_0}{Q_e}\frac{d}{dt}\left(V_F - V_R\right) - 2k_l\frac{dI_b}{dt}$$

It is not obvious that the re-introduction of Q_w and Q_e in this way is consistent with our previous discussion, specifically Eqs. (15) and (22). However, this may be checked by considering steady-state energetics. Similarly one may check the beam-current term by considering a single bunch with reference to the principle of superposition.

Employing the continuity condition to eliminate reverse voltage, one may describe the cavity excitation as a driven, damped harmonic oscillator.

$$\left(\frac{d^2}{dt^2} + \frac{\omega_0}{Q_L}\frac{d}{dt} + \omega_0^2\right) V_c = 2\frac{\omega_0}{Q_e}\frac{dV_F}{dt} - 2k_l\frac{dI_b}{dt}. \tag{26}$$

This result permits one to compute transient and steady-state excitation by external drive through the waveguide, or by the beam. If one prefers, one may express this result in terms of an equivalent circuit, for example, as in Fig. 7. In steady-state, with beam off, and external drive at angular frequency ω, $V_F = \Re\tilde{V}_F e^{j\omega t}$, one may express the cavity voltage as $V_c = \Re\tilde{V}_c e^{j\omega t}$, with

$$\frac{\tilde{V}_c}{\tilde{V}_F} = \frac{2\beta}{1+\beta}\cos\psi\, e^{j\psi}.$$

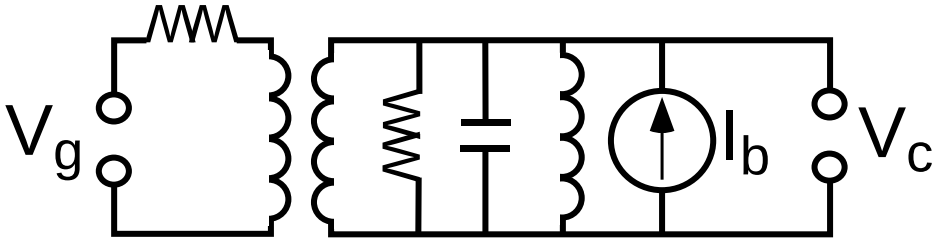

Figure 7. An equivalent circuit for a cavity coupled to a waveguide and a beam.

The *tuning-angle* ψ is given by

$$\tan\psi = Q_L\left(\frac{\omega_0}{\omega} - \frac{\omega}{\omega_0}\right). \tag{27}$$

Analysis of energy conservation in the steady-state provides one with the relation of the equivalent circuit quantities V_F and V_R to power. Forward power in the connecting guide is

$$P_F = \beta\frac{\left|\tilde{V}_F\right|^2}{R_{shunt}}. \tag{28}$$

This shows that the forward and reverse voltage variables we are using should be referred to an impedance $R_{shunt}/2\beta$. Since this will differ from the waveguide impedance Z_w, we put a transformer in the equivalent circuit model of Fig. 7. The equivalent turns ratio n is given by $n^2 = R_{shunt}/2\beta Z_w$. One can go on to express cavity voltage in terms of forward power,

$$\left|\tilde{V}_c\right| = \frac{2\beta^{1/2}}{(1+\beta)}\left|\cos\psi\right|\ \left(R_{shunt}P_F\right)^{1/2}. \tag{29}$$

This relation is to be contrasted with the expression in terms of dissipated power P_w, developed in Sec. 2.5, $V = \sqrt{R_{shunt}P_w}$. The two results are general, but we emphasize that $P_F \neq P_w$ in general. Optimum cavity voltage, at fixed power, with negligible beam-loading, occurs for critical coupling, $\beta = 1$ and perfect tune $\psi = 0$. To appreciate the effect of errors, consider that at $\psi = 26^\circ$ one has lost 10% of the gap voltage; this corresponds to fractional detuning $\delta = (\omega_0 - \omega)/\omega_0 \approx 1/4Q_L$. For a critically-coupled S-Band cavity this is a detuning of 100 kHz, equivalent to a dimension error at the 3-μm level.

For comparison with measurement, one may express the steady-state complex reflection coefficient as

$$S_{11} = \frac{\tilde{V}_R}{\tilde{V}_F} = \frac{2\beta}{1+\beta}\cos\psi e^{j\psi} - 1. \tag{30}$$

This quantity was referred to as R in the discussion of waveguide and both notations are common. The notation here refers to the *scattering-matrix* or "S-matrix". As we will see in Sec. 2.14, this is in general an $N \times N$ matrix for an N-port device. Here $N = 1$. This quantity is directly measurable as a function of frequency and in modern times special instruments, vector-network analyzers, are routinely employed for this purpose. A measurement of S_{11} permits one to infer ω_0, β, and Q_L, and to assess systematics associated with temperature regulation and vacuum. Thanks to Slater's theorem, repeated measurement of S_{11}, as a small dielectric bead is pulled along the beam-axis, permits one to map the electric field. Equation (30) represents the contribution to S_{11} from a single mode. In practice one may expect that measurement of S_{11} for a microwave cavity will exhibit both narrow and broad resonances. The latter are often resonances on the connecting guide, the former are the resonant modes residing within the cavity proper, and they are very narrow, since Q_L is very large.

Appreciating what a cavity mode and a waveguide can do to each other, let us consider the effect they may have on the beam.

2.8. ELECTROMAGNETIC MULTIPOLES

We consider a particle passing through a beam-tube with a structure of some kind enclosing the tube, as seen in Fig. 8. We approximate the particle trajectory as a straight line centered on transverse coordinate $\vec{r}_\perp$ and passing some reference plane at time t_0. The voltage experienced by the beam in transit through the geometry may be expressed in terms of axial electric field E_s as

$$V_c\left(\vec{r}_\perp, t_0\right) = \int_{-\infty}^{+\infty} ds\; E_s\left(\vec{r}_\perp, s, t_0 + \frac{s}{V}\right),$$

where V is the particle speed. We may express the electric field in terms of Fourier components,

$$E_s\left(\vec{r}_\perp, s, t\right) = \int_{-\infty}^{+\infty} \frac{d\omega}{\sqrt{2\pi}} e^{j\omega\, t} \tilde{E}_s\left(\vec{r}_\perp, s, \omega\right),$$

with a corresponding result for voltage,

$$V_c\left(\vec{r}_\perp, t_0\right) = \int_{-\infty}^{+\infty} \frac{d\omega}{\sqrt{2\pi}} e^{j\omega\, t_0} \tilde{V}_c\left(\vec{r}_\perp, \omega\right),$$

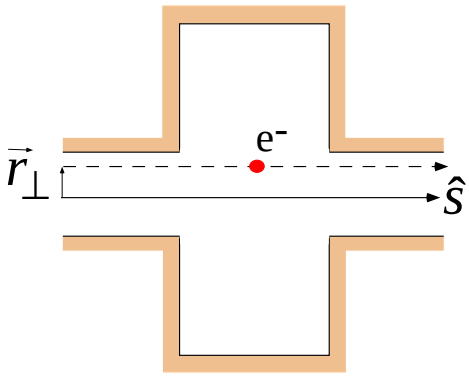

Figure 8. The interaction of a beam with a cavity depends on its transverse coordinate $\vec{r}_\perp$.

where

$$\tilde{V}_c\left(\vec{r}_\perp,\omega\right)=\int_{-\infty}^{+\infty}ds\;\tilde{E}_s\left(\vec{r}_\perp,s,\omega\right)e^{j\omega\,s/V}.$$

Applying Maxwell's equations in the vicinity of the beam-orbit,

$$\left(\nabla_\perp^2+\frac{\partial^2}{\partial s^2}+\frac{\omega^2}{c^2}\right)\tilde{E}_s\left(\vec{r}_\perp,s,\omega\right)=0,$$

we can, after an integration by parts, establish that

$$\left(\nabla_\perp^2-\Gamma^2\right)\tilde{V}_c\left(\vec{r}_\perp,\omega\right)=0, \tag{31}$$

where

$$\Gamma=\left(\frac{\omega^2}{V^2}-\frac{\omega^2}{c^2}\right)^{1/2}\approx\frac{\omega}{\gamma\,V}. \tag{32}$$

The voltage witnessed by a particle depends in general on its displacement from the cavity axis. For a relativistic beam $\Gamma\to 0$. In this limit, the voltage imparted by a cavity mode is a harmonic function of the transverse coordinate.

A harmonic function regular in the vicinity of the center-axis may be expressed in the form of a multipole expansion.

$$\tilde{V}_c\left(\vec{r}_\perp,\omega\right)=\sum_{m=0}^{\infty}r^m\left\{b_m\cos\left(m\phi\right)-a_m\sin\left(m\phi\right)\right\}, \tag{33}$$

introducing polar coordinates, $\vec{r}_\perp=(x,y)=r\left(\cos\phi,\sin\phi\right)$. Where the geometry respects circular symmetry, and for the case of a single excited

mode, the sum reduces to single terms, pure multipoles. In particular, the $m = 0$ term is independent of position. One prefers such a "monopole" mode for acceleration, since then gap voltage is independent of the beam orbit. On the other hand, real structures require a power feed, and this implies a deviation from cylindrical symmetry. Thus one is interested to quantify azimuthal mode purity in structures.

Generally one may classify modes based on their leading-order azimuthal harmonic. To appreciate the effects of such harmonics, it helps to consider not merely the longitudinal kicks associated with a structure excitation, but the transverse kicks as well. Transverse impulse per unit charge may be expressed as

$$\vec{P}_\perp(\vec{r}_\perp, t) = \int_{-\infty}^{+\infty} ds \left\{ \frac{1}{V}\vec{E}_\perp + \hat{s} \times \vec{B}_\perp \right\}_{\left(\vec{r}_\perp, s, t+\frac{s}{V}\right)}. \tag{34}$$

Employing Faraday's Law and the ballistic-trajectory assumption, one may show that

$$\frac{\partial \vec{P}_\perp}{\partial t} = -\vec{\nabla}_\perp V_c, \tag{35}$$

the Panofsky-Wenzel theorem [6].

To make these considerations more concrete, consider an accelerator structure respecting inversion symmetry in y, but not in x. The monopole-mode voltage will take the form, at leading order,

$$V_c(x, y, t_0) = \Re V_0 e^{j\psi_0} \left\{ 1 + B_1 x + B_2 \left(x^2 - y^2 \right) + \ldots \right\},$$

where V_0, ψ_0 are the amplitude and phase seen on-axis. An off-axis trajectory corresponds to a different amplitude and phase. Keeping only the dipole ($m = 1$) correction, and expressing $B_1 = |B_1| e^{j\psi_1}$, one can see that

$$V_c(x, y, t_0) \approx \Re V_0 e^{j\psi_0} e^{j\delta\psi} \left(1 + 2|B_1| x \cos\psi_1 + |B_1|^2 x^2 \right)^{1/2}$$

where

$$\tan\delta\psi = \frac{|B_1| x \sin\psi_1}{1 + |B_1| x \cos\psi_1}$$

Thus the absence of a reflection symmetry *permits* a phase and an amplitude asymmetry in the accelerating voltage. An additional consequence is the presence of transverse deflections associated with this mode,

$$\vec{P}_\perp = \vec{\nabla}_\perp \Re V_0 e^{j\psi_0} \frac{j}{\omega} \left\{ 1 + B_1 x + B_2 \left(x^2 - y^2 \right) + \ldots \right\} \tag{36}$$

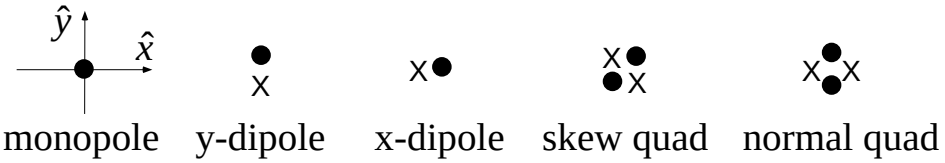

Figure 9. Electromagnetic multipoles depicted according to the symmetry of the characteristic voltage witnessed by a relativistic beam, travelling into the page.

or, at lowest order $P_x \approx -\left(V_0 \left|B_1\right| / \omega_0\right) \sin\left(\psi_0 + \psi_1\right)$. Thus particles phased "on-crest" ($\sin \psi_0 = 0$) experience a transverse deflection determined by the imaginary part of B_1, the *phase-asymmetry* of the mode. Such an asymmetric geometry permits acceleration at an angle. This is undesirable in an accelerator cavity. This problem appears in practice wherever structures are asymmetrically fed, as with a single waveguide feed, for example.

While the $m = 0$ accelerating mode represents the most important example of a multipole, there are others, as seen in Fig. 9. The dipole modes are essential in understanding the transverse dynamics of a high-current beam, and arise naturally in the context of *wakefields*.

2.9. WAKEFIELDS

We have seen in Sec. 2.3 that the accelerating mode is excited by the beam, with voltage determined by the loss-factor k_l. In fact, each mode of the structure is excited by the beam, to the extent that the mode has some parallel electric field component, specifically, non-zero loss factor evaluated on the beam orbit. This is not cause for immediate alarm, because it takes some effort actually to design a mode to interact well with the beam. Nevertheless, when beam intensity is high, one may be concerned. Beam-excitation of parasitic modes is described by an elegant formulation in terms of "wakefields" [7].

To illustrate, let us return to the single-cavity problem of Sec. 2.7, and solve explicitly for the beam-induced voltage in a particular mode, using Eq. (26), with no external drive $V_F = 0$. One can check by differentiation that

$$V_c\left(t\right) = -\int_{-\infty}^{t} dt'\, G\left(t - t'\right) \frac{dI_b}{dt}\left(t'\right).$$

where the Green's function is

$$G\left(\tau\right) = 2k_l \frac{\sin\left(\Omega\tau\right)}{\Omega} \exp\left(-\tfrac{1}{2}\nu\tau\right) H\left(\tau\right),$$

with H the step-function, $\Omega = (\omega_0^2 - \frac{1}{4}\nu^2)^{1/2}$, and $\nu = \omega_0/Q_L$. If we suppose cavities are placed along the beamline at separation L, we may translate this result into loss of voltage per unit length, or gradient,

$$\frac{d}{ds} mc^2\gamma = e \int_{-\infty}^{\tau} d\tau' W_{||}\left(\tau - \tau'\right) I_b\left(\tau'\right), \tag{37}$$

for an electron located at $\tau = t - z/c$. One can check, with an integration by parts, that $W_{||} = G'/L$. More generally the "longitudinal wakefield", $W_{||}$ includes contributions from other modes. In the frequency domain $\tilde{V}_c\left(\omega\right)/L = -Z_{||}\left(\omega\right)\tilde{I}_b\left(\omega\right)$, where $Z_{||}$ is the contribution to the "longitudinal impedance" per unit length from this mode, a Lorentzian line-shape with width set by Q_L,

$$Z_{||}\left(\omega\right) = \frac{2j\, k_l'\omega\, \omega_0}{\omega_0^2 - \omega^2 + j\,\omega\,\omega_0/Q_L}, \tag{38}$$

where $k_l' = k_l/L$. In general, impedance receives contributions from this and all other high-Q modes with longitudinal electric field non-zero on the beam-axis. In addition to such narrow-band terms in the impedance, there are broad-band terms at frequencies above cut-off in the beam-tube. At very high-frequencies quasi-optical phenomena appear, as nearby boundaries cause the space-charge fields of the beam to diffract. This effect is particularly severe when the bunch-length is short, for then high-frequencies are well-represented in the beam-distribution. In accounting for impedance one looks for any deviation from uniformity in the beam-tube. The most important constriction in an accelerator is the beam-passing aperture between accelerator cells. Other common pipe-irregularities include beam monitoring instrumentation, collimators, bellows sleeves, vacuum port shields, and transitions in beam-pipe dimension. Even for a smooth pipe, wall resistivity results in energy loss, and thus one is concerned wherever the beam-passing aperture is small and lossy. These effects are acute in circular machines, where the beam orbits through an astronomical length of pipe. However, even in a linac, longitudinal wakefields may account for a reduction in beam energy of several percent, when single-bunch charge is high, and bunch length is short. To characterize such effects in simple terms, one may refer to the total loss factor per unit length,

$$k_{tot}' = \frac{1}{Q_b^2} \int_{-\infty}^{+\infty} d\omega \, \left|\tilde{I}_b\left(\omega\right)\right|^2 \Re Z_{||}\left(\omega\right), \tag{39}$$

in terms of a form-factor quantifiying coherence at angular frequency ω, and the impedance describing the coupling to the beam-environment. So

for example, in an S-Band structure fundamental-mode loss factor would be 1 V/pC/cell, while total loss factor might be 4–8× larger, depending on the bunch length and beam-port radius [7].

A more baneful effect in a linac is the dipole wakefield excited by an off-axis beam. To illustrate, we return to the single cavity mode, and apply the Panofsky-Wenzel theorem, to the case of a dipole-mode excitation,

$$P_x = \Re\left(\frac{j}{\omega_0}\frac{\partial \tilde{V}_c}{\partial x}e^{j\omega_0 t}\right). \tag{40}$$

We may characterize the efficacy of this mode as a beam-deflector in much the same way that we characterized the monopole mode as a beam-accelerator. We define

$$\left[\frac{R_a}{Q}\right] = \frac{\left|\tilde{V}_c\left(x=a\right)\right|^2}{\omega_0 U},$$

just the $[R/Q]$ for this mode evaluated on an off-axis trajectory at $x = a$. In the limit $a \to 0$, we obtain the figure of merit

$$\left[\frac{R_\perp}{Q}\right] = \lim_{a\to 0}\left[\frac{R_a}{Q}\right]\frac{1}{a^2} = \frac{\left|\partial\tilde{V}_c/\partial x\right|^2}{\omega_0 U}. \tag{41}$$

For example, for the lowest dipole mode of a circular cavity of radius R one may show that maximum $[R_\perp/Q] \approx 3Z_0/R^2$.

In terms of $k_\perp = \omega_0\left[R_\perp/Q\right]/4$, one may express the energy deposited in this mode as $U = k_\perp Q_b^2 x_b^2$ for a beam offset x_b. One wouldn't mind a small amount of energy loss due to such a mode, particularly if the beam orbit were well-centered — except that with this mode comes a deflecting magnetic field. Consequently the head of a short bunch may deflect the tail. A rocking motion develops, and this motion is unstable, since the source of deflections is the beam offset itself. This effect was first observed in operation of linacs in multi-bunch mode, and appeared as a truncated current waveform on a current monitor. It was referred to at first as "pulse-shortening", and later, as "beam break-up" (BBU) when it was realized that transverse oscillations were the cause of beam loss [8]. Because of this wakefield, and the potential for bunch-to-bunch excitation, modern structure design entails elaborate measures to cause parasitic modes to damp and decohere [9]. Structure and lattice alignment have become subjects of the first importance. Ground motion, girder mechanical resonances, and thermal-bowing of girders have become subjects for research [10].

We can extract from our analysis a prescription for the transverse motion through the linac. For the case of the single cavity, we may describe

dipole-mode excitation according to

$$\left(\frac{\partial^2}{\partial t^2}+\frac{\omega_0}{Q_L}\frac{\partial}{\partial t}+\omega_0^2\right)\frac{\partial V_c}{\partial x}=-2k_\perp\frac{\partial I_b}{\partial t}x_b,$$

and the circuit parameters here refer to the dipole mode, not the accelerating mode. Applying the Panofsky-Wenzel theorem, Eq. (35), we may express this more directly in terms of the impulse,

$$\left(\frac{\partial^2}{\partial t^2}+\frac{\omega_0}{Q_L}\frac{\partial}{\partial t}+\omega_0^2\right)P_x=-2k_\perp I_b x_b.$$

The solution is

$$P_x\left(t\right)=-\int_{-\infty}^{t}dt'\,G\left(t-t'\right)I_b\left(t'\right)x_b\left(t'\right),$$

where the Green's function

$$G\left(\tau\right)=2k_\perp\frac{\sin\left(\Omega\tau\right)}{\Omega}\exp\left(-\tfrac{1}{2}\nu\tau\right)H\left(\tau\right).$$

Notational conventions differ on the subject of transverse wakefields. We will refer to the wakefield per unit length W_x, with the identification $G/L\rightarrow W_x$. The rate of change of momentum per unit length may then be expressed as

$$\frac{dp_x}{ds}=e\int_{-\infty}^{\tau}d\tau'\,W_x\left(\tau-\tau'\right)I_b\left(\tau'\right)x_b\left(\tau'\right), \tag{42}$$

in the absence of other transverse forces.

As with the longitudinal wakefield, the transverse wakefield picks up contributions due to any deviation from uniformity on the beamline, as well as wall resistivity. The narrow-band components of the corresponding transverse impedance account for the original beam break-up problems observed in multi-bunch mode on the linacs of old. These narrow-band terms, corresponding to the "long-range" wakefield, are diminished by detuning of successive cells, so as to arrange destructive interference between the dipole modes of the structure as a whole. This was achieved, accidentally, but happily, in the Two-Mile accelerator by virtue of constant-gradient design, followed by additional detuning *in situ* [11]. The broad-band terms, corresponding to the "short-range" wakefield, account for single-bunch beam break-up. This is controlled only by lowering bunch charge, or strengthening the magnetic lattice. At the design-stage one may reduce the short-range

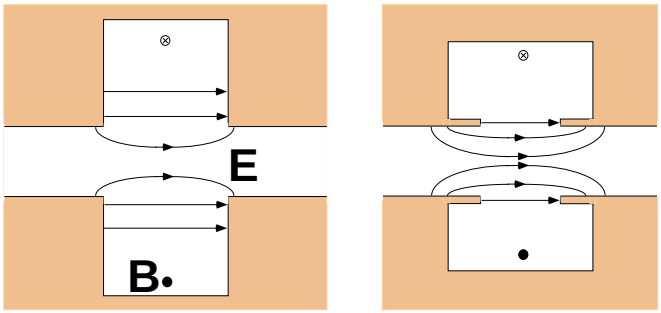

Figure 10. Cutting of the beam port in the cylindrical pillbox on the left reduces $[R/Q]$. This can be fixed with nose-cones as seen on the right.

wakefield by enlarging the beam port. However, a larger beam-port aperture easily lowers the $[R/Q]$. The conjunction of these two problems leads us directly to the matter of structure design.

2.10. CAVITY DESIGN

Thus far we have seen more or less what is desired of a cavity geometry, in terms of coupled-cavity circuit performance, and we can appreciate that there are consequences for beam dynamics. However, the parameters that determine the dynamics of the cavity and the beam are determined themselves by cuts in copper. Let us consider then how to accomplish the desired circuit parameters, and field-shape, by design, for an illustrative problem as illustrated in Fig. 10.

As we have seen, an accelerator cavity requires a beam-port. Cutting the port causes electric field lines to extend into the beam-tube, and develop transverse components. This reduces the integrated longitudinal field, and therefore $[R/Q]$. Reduction of the cavity radius and introduction of "nose-cones" attracts electric field lines to the gap, and improves the field-line shape. These considerations are illustrated with a simple estimate.

When the beam tube radius and gap are small, such a *re-entrant* cavity is approximately a coaxial line with a gap in the center conductor. Let us denote the beam tube radius a, inner conductor outer radius R_1, outer conductor inner radius R_2, gap length L_1, and cavity length L_2 as seen in Fig. 11. The resonant frequency of the circuit may be determined from the capacitance and inductance. Capacitance is determined by the energy stored in the electric field. If L_2 is short compared to a wavelength, most of this energy resides in the gap, and capacitance is $C \approx \varepsilon_0 \pi R_1^2 / L_1$. Inductance is determined by magnetic energy stored in the coaxial line, arising from displacement current $I = \varepsilon_0 \omega \pi R_1^2 E$, and this is $L \approx (\mu_0/2\pi) L_2 \ln(R_2/R_1)$. Resonant frequency is then $\omega = 1/\sqrt{LC}$. Using $[R/Q] = V^2/\omega U \approx 2/\omega C$,

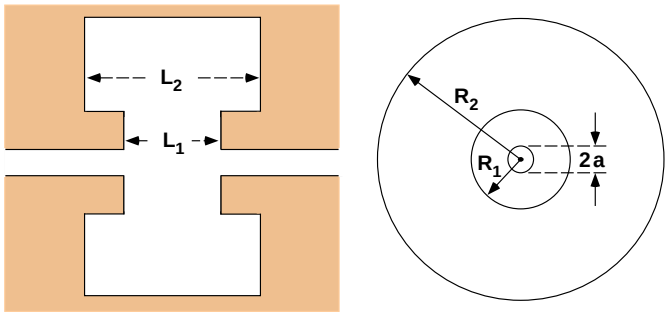

Figure 11. We analyze a reentrant cavity.

we find

$$\left[\frac{R}{Q}\right] \approx \frac{2}{\pi} Z_0 T^2 \sqrt{\frac{L_1 L_2}{2R_1^2} \ln\left(\frac{R_2}{R_1}\right)},$$

with T the transit angle factor for the gap. Inspection of these results for frequency and $[R/Q]$ suggests that one can in fact recover the high impedance of the closed pillbox by judicious design. This was accomplished, in the early days of linacs, by approximate calculations along the lines of our example [5]. In practice, the analytic approach was supplemented with "cut and try" on the bench, to improve on the guidance of the design equations. In modern times it is common to employ software to solve Maxwell's equations and thereby optimize cavity dimensions.

To illustrate, Fig. 12 shows an example of a re-entrant cavity, the phase-monitor cavity, designed by Altenmueller and Brunet for use on the Two-Mile Accelerator circa 1965 [12]. These still see use today, for fixed-target experiments [13]. The numerical geometry has been set-up with the code *GdfidL* [14]. In the notation of Fig. 11, dimensions are $L_1 = 1.0235''$, $L_2 = 1.7005''$, $2a = 0.8000''$, $2R_1 = 1.0000''$, $2R_2 = 2.6525''$. These dimensions were arrived at after iteration on the bench, and chosen to provide a frequency 2 MHz low prior to brazing, so that frequency could be adjusted by dimpling. Altenmueller and Brunet infer from their measurements $Q_w \approx 9600$, $Q_L \approx 1200$, coupling parameter $\beta \approx 7.0$, $[R/Q]/T^2 \approx 370\ \Omega$ (adjusting a factor of two conventional difference), $T \approx 0.819$, $[R/Q] \approx 248\ \Omega$. The corresponding numerical result for S_{11} is seen in Fig. 13. Numerical calculation with GdfidL indicates $[R/Q] \approx 236\ \Omega$, and $Q_w \approx 1.51 \times 10^4$. Observed temperature detuning coefficient is 25 kHz/°F, 1° RF for 0.5 °F.

Nose-cones are but one feature of interest in cavity design. Symmetric input and output couplers help to eliminate dipole kicks. To control long-range wakefields, slots may be added to cells to externally couple higher modes, effectively damping them. Wakefields can be caused to decohere

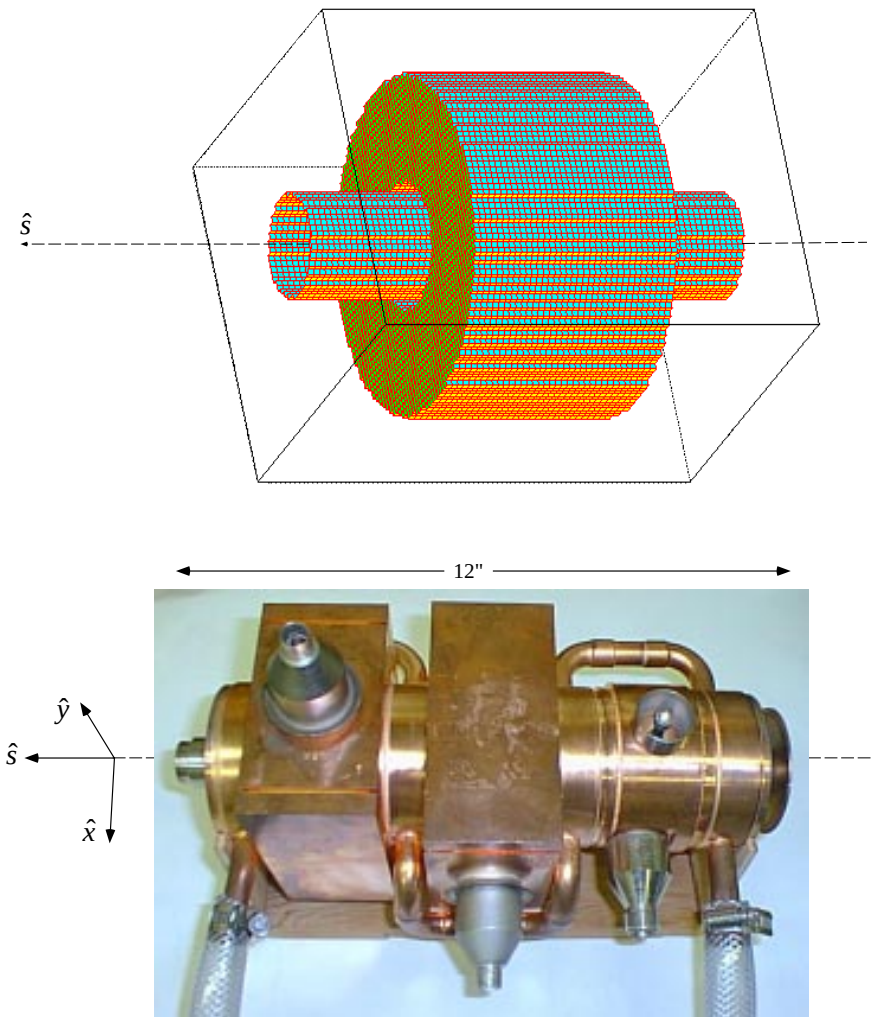

Figure 12. An example re-entrant cavity, the phase-monitor cavity of the original cavity beam-position monitor system on the Two-Mile Linac. Shown is the the numerical geometry employed for calculation and the actual cavity, part of a triplet.

by judicious detuning of cells. Such design work takes place under the constraint of good shunt-impedance for the accelerating mode. Interesting design problems also are found beyond the structures themselves, and include windows, couplers, mode-converters and other components of the linac rf system [15]. For high-power handling it is important to maintain low insertion loss, good match, and low peak field, everywhere but in the accelerating cavities themselves. Numerous codes are available and the designer

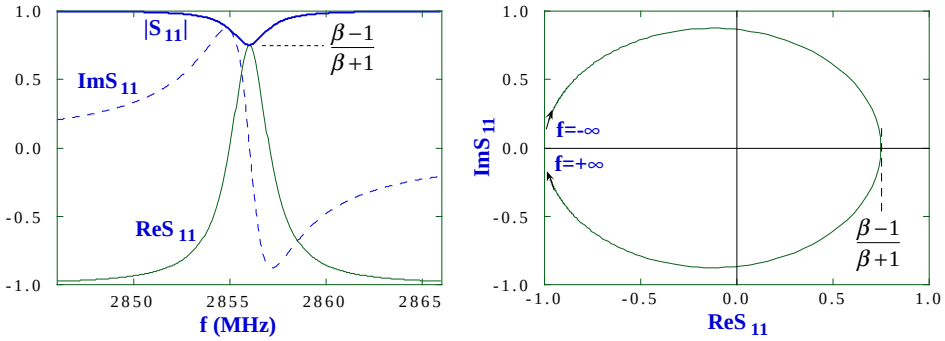

Figure 13. S_{11} for the phase cavity ($\beta \approx 7.0$) plotted versus frequency and in *Smith-Chart* form.

has quite a few tools to choose from [16].

2.11. TWO COUPLED CAVITIES

If we were to take the work of Secs. 2.3-2.5, on a single cavity with a single waveguide-feed, and extrapolate it to a 20 GeV machine, we would find that we need more than 10^4 power feeds, a lot of copper-waveguide. A more economical scheme turns out to be feasible and consists of a series of cavities placed along a common beam-tube. Power is fed by waveguide to a single-cavity, and finds its way to the other cavities, through the beam-tube. Such a multi-cell accelerator behaves as a chain of coupled pendula. Coupling constants, periods, and Q's are determined from the geometry sculpted in copper. The concept may be designed as a standing-wave accelerator, with a single feed. Or it may be implemented as a travelling-wave accelerator, with an input waveguide at one end, and output waveguide and load at the other end.

We consider first a two-cell structure as seen in Fig. 14. Recognizing the similarity with the problem of two coupled pendula we expect to find two modes of oscillation, a mode with $V_1 = V_2$ ("0-mode"), and a mode with $V_1 = -V_2$, ("π-mode"). For the 0-mode each field line from cell 1 connects to a field line from cell 2, and no field lines terminate on the coupling iris. Displacement current from cell 1 flows directly into cell 2 and no current flows through the shared common wall. Since no current flows, one could remove the common wall without disturbing the modal field pattern. In this case one has simply smooth pipe terminated at each end. This is just the original geometry of the uncoupled cells, but with twice the cavity length. However for the accelerating mode, as seen in Fig. 2, frequency depends only on radius. Thus the 0-mode frequency Ω_0 is just that of the uncoupled cells. In the π-mode field-lines terminate on the coupling iris,

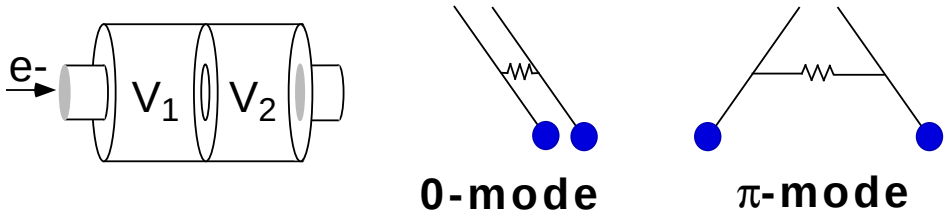

Figure 14. Two cells coupled through the beam-tube exhibit two modes of oscillation, like two coupled pendulums.

and we expect the tune to be sensitive to the iris geometry. From Slater's theorem, we expect that removal of copper in the vicinity of large electric field, and small magnetic field, will reduce the capacitance of the circuit, and therefore raise the frequency. Thus we expect the π-mode frequency $\Omega_\pi > \Omega_0$.

To treat the problem more quantitatively, let us describe the modes with amplitudes ξ_0 and ξ_π. That is to say, when the structure is excited in the zero mode, the electric field exhibits a certain spatial pattern that varies in time in proportion to $\xi_0(t)$. In general, fields may be a superposition of the two modes. According to our reasoning these mode amplitudes evolve according to

$$\left(\frac{d^2}{dt^2} + \Omega_0^2\right)\xi_0 = 0, \quad \left(\frac{d^2}{dt^2} + \Omega_\pi^2\right)\xi_\pi = 0.$$

We may express this in terms of the original cell voltages by identifying $\xi_0 = V_1 + V_2$, $\xi_\pi = V_1 - V_2$. These equalities assume merely a choice of units for ξ_0 and ξ_π. Substituting these expressions, and re-arranging, one finds

$$\left(\frac{d^2}{dt^2} + \Omega_0^2\right) V_1 = \frac{1}{2}\left(\Omega_\pi^2 - \Omega_0^2\right)(V_2 - V_1),$$

$$\left(\frac{d^2}{dt^2} + \Omega_0^2\right) V_2 = \frac{1}{2}\left(\Omega_\pi^2 - \Omega_0^2\right)(V_1 - V_2).$$

This result describes the time-evolution of two-coupled cavities in terms of the frequency shift between 0 and π modes. The strength of the coupling is quantified by the dimensionless figure

$$\kappa = 2\frac{\Omega_\pi^2 - \Omega_0^2}{\Omega_\pi^2 + \Omega_0^2}.$$

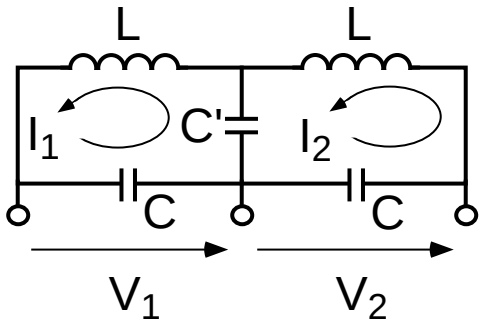

Figure 15. Equivalent circuit for two coupled cells.

In terms of this cell-to-cell coupling constant, and

$$\omega_0^2 = \frac{1}{2}\left(\Omega_\pi^2 + \Omega_0^2\right),$$

we may express our result as

$$\left(\frac{d^2}{dt^2} + \omega_0^2\right) V_1 = \frac{1}{2}\kappa\omega_0^2 V_2, \tag{43}$$

$$\left(\frac{d^2}{dt^2} + \omega_0^2\right) V_2 = \frac{1}{2}\kappa\omega_0^2 V_1. \tag{44}$$

With some algebra one can show that

$$\Omega_0^2 = (1 - \kappa/2)\omega_0^2, \quad \Omega_\pi^2 = (1 + \kappa/2)\omega_0^2.$$

If it is helpful, one may think of the equivalent circuit of Fig. 15. One can check that the circuit parameters satisfy $\Omega_0^2 = 1/LC$ and $C/C' \approx \kappa/2$.

At this point, having examined the kinematics of modes, we have determined that there must be a figure κ, but we haven't determined what this circuit parameter is—how it depends on the cell geometry. To calculate cell-to-cell coupling from the geometry we apply Slater's Theorem. We suppose the cells are excited in the π-mode and sketch a cylindrical volume with radius equal to that of the coupling iris a, as indicated in Fig. 16. In this accelerating mode, electric field is large in the vicinity of the iris, and magnetic field is small. As a result, the iris experiences mechanical tension due to the pull of the electric field lines. This tension is transmitted by the field lines through the endcaps of our imaginary cylinder to the wall separating the cells. We may express the time-averaged balance of forces as

$$\int_{endcap} dS\, T^{zz} = -\int_{sidewall} dS\, T^{zr} = \int_{sidewall} dS\, T^{rr}.$$

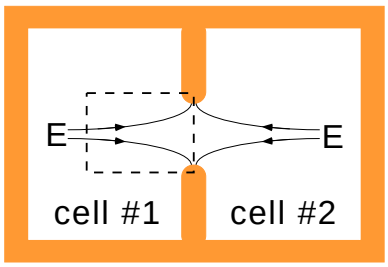

Figure 16. We apply Slater's theorem to calculate the coupling of two-cells.

The first equality expresses conservation of axial (z) momentum, and the second conservation of radial (r) momentum. Integrals over the cylinder endcap centered in the iris vanish since the axial field vanishes there. Thus the axial force may be estimated with an integral of $T^{zz} \approx \varepsilon_0 E^2/2$ over the left endcap, in terms of the peak electric field E,

$$\int_{sidewall} dS\, T^{rr} = \int_{endcap} dS\, T^{zz} \approx \pi\, a^2 \frac{1}{2}\varepsilon_0 E^2 \times \frac{1}{2}$$

where the last factor of $1/2$ results from time-averaging. If we permit the iris to relax adiabatically, deforming in radius by an amount $\delta a < 0$, the mode will do work, and the stored energy U, will be reduced

$$\delta U = \delta a \int_{sidewall} dS\, T^{rr} \times 2 \approx \pi\, a^2 \delta a \frac{1}{2}\varepsilon_0 E^2$$

where the factor of two in the first equality results from counting both the left and right halves of the iris. Closing the iris, the mode loses energy

$$\Delta U = -\frac{\pi}{6}\, a^3 \varepsilon_0 E^2.$$

Total stored energy is meanwhile $U \approx \varepsilon_0 E^2 V/8$, with V the total volume of the two cells. Adiabatic invariance of U/ω then implies that with the closing of the iris, the π-mode frequency shifts by an amount

$$\frac{\Delta\omega}{\omega} = -\frac{4\pi}{3}\frac{a^3}{V}.$$

Once the iris is closed we know that the frequency must be restored to the unperturbed, closed-cavity value, and this is just Ω_0. Thus we have computed the frequency separation of the 0 and π modes, $\Delta\omega \approx \Omega_0 - \Omega_\pi$. The coupling constant is then

$$\kappa \approx \frac{8\pi}{3}\frac{a^3}{V}. \tag{45}$$

Subsituting $V = 2\pi LR^2$ with L the length of a single cell, and R the cell radius one obtains $\kappa \approx 1.3a^3/LR^2$. A more precise calculation provides the coefficient 1.57 [1].

Appreciating the behavior of the modes let us characterize their properties for acceleration. The 0-mode is just the mode of a single cavity, and thus its $[R/Q]$ can be no better than the single-cell case. For the π-mode on the other hand we see that a particle accelerated in cell 1, and leaving this cell at about the time the field is reversing, arrives in cell 2 just as the field in that cell is ready to accelerate. Thus one expects the net voltage to be the sum of each cell voltage. Stored energy meanwhile is roughly the sum of the energies stored in each cell. With twice the voltage of a single cell, for twice the energy, we see that $[R/Q] = V^2/\omega U$ is twice the $[R/Q]$ for a single cell. As for wall Q, the magnetic field is small near the iris, thus for a given single-cell voltage, the dissipated Poynting flux in the walls is unchanged by the coupling. The wall area has increased by a factor of 2, so the dissipated power is twice as large. However, the stored energy is also larger by 2, so Q_w is unchanged. This implies that shunt impedance $R_s = Q_w[R/Q]$ is twice the single-cell case. Thus if our microwave power source was capable of providing a 2 MV "gap-voltage" with one cell, we should be able to arrange 3 MV with a two-cell structure.

2.12. MULTI-CELL STRUCTURES

This brush with the 2-cell structure encourages us to continue right on to the case of N-cells, expecting shunt-impedance larger by a factor of about N. Actually though it is simpler to proceed directly to the infinite structure. According to our analysis of the effect of a single-coupling iris, as summarized in Eqs. (43) and (44), we expect that cell voltages in the infinite structure will satisfy

$$\left(\frac{d^2}{dt^2} + \omega_0^2\right) V_n = \frac{1}{2}\kappa\omega_0^2 \left(V_{n-1} + V_{n+1}\right), \tag{46}$$

where n indexes the cells, and we continue to neglect Ohmic losses. Since we have made two iris cuts in each cell, our circuit parameter ω_0 differs from the 2-cell case, as it is shifted by twice as much relative to the unperturbed cell frequency Ω_0, $\omega_0^2 = \Omega_0^2/(1-\kappa)$.

To appreciate the behavior of this coupled system consider the case of a single-frequency excitation, imposed somehow on one particular cell. In steady-state, a general solution for cell voltages may be expressed as $V_n = \Re\tilde{V}_n e^{j\omega t}$, for some choice of cell phasors $\tilde{V}_n$. The phasors describe, among other things, the flow of energy along the waveguide. One could compute this from the Poynting flux integrated over the coupling hole,

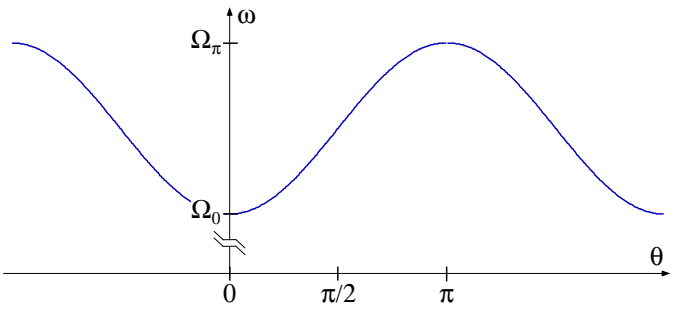

Figure 17. Brillouin curve for a periodic structure.

however, it's enough for us to recognize that conservation of energy requires $|\tilde{V}_{n+1}| = |\tilde{V}_n|$. Thus adjacent cells are related by a phase-factor, and from periodicity, this factor must be the same for any two successive cells. Let us denote this *cell-to-cell phase-advance* $\theta\,(\omega)$. To the right of the cell being excited we observe a rightward travelling wave,

$$\tilde{V}_{n+1} = e^{-j\theta}\tilde{V}_n. \tag{47}$$

Substituting this in Eq. (46) we find

$$\omega^2 = \omega_0^2\,(1 - \kappa\cos\theta)\,, \tag{48}$$

the *dispersion relation* for our periodic line. This is illustrated in Fig. 17. Given that the result is periodic in θ, and symmetric it is enough to plot the result on the interval $(0,\pi)$. The vertical scale in this sketch has a break to indicate that the relative separation of 0 and π modes may be quite small. This separation is significant as it is the full-width of the *passband*, the range in frequency, associated with the original cavity mode, over which the structure will permit microwave propagation.

To construct a finite periodic structure, it is enough to know that interior cells in the structure support forward and backward waves with the dispersion characteristics of the infinite structure. A general solution for cell voltages may consist of both left and right travelling waves, and then must take the form

$$\tilde{V}_n = \tilde{V}^+ e^{-jn\theta} + \tilde{V}^- e^{jn\theta}.$$

End cells determine the boundary conditions on forward and reverse amplitudes $\tilde{V}^{\pm}$.

For a standing-wave structure, one has left and right-going waves of equal amplitude, with relative phase determined by the end cells. As we saw in the case of the two-cell structure, maximum shunt-impedance for

a standing-wave structure occurs for π phase-advance per cell. Structure length is limited by mode density near π-mode, as mode separation should be greater than the natural $1/Q_w$ mode-width.

A travelling-wave structure employs a power feed at one end, and an output-waveguide at the other. End cells are matched, designed to insure that in steady-state there is no reflection on the input line, and no standing-wave in the structure. The rate at which energy flows into the coupling cell matches the rate at which energy flows through the structure and into the walls. Propagation through the structure is governed by the group velocity, as determined from the slope of the Brillouin curve, and is given by

$$\beta_g = \frac{V_g}{c} = \frac{L}{c}\frac{d\omega}{d\theta} = \frac{1}{2}\left(\frac{\omega_0 L}{c}\right)\frac{\kappa \sin\theta}{\sqrt{1-\kappa\cos\theta}}, \tag{49}$$

with L the cell period. Group velocity is maximum and dispersion is minimum in mid-band, and group velocity approaches zero at 0 and π phase-advance. In steady-state, interior cell-voltages are

$$V_n(t) = \Re \tilde{V}^+ e^{j\omega t - jn\theta},$$

and the net voltage experienced by a particle entering the structure at time t_0 is

$$V_{NL}(t_0) = \sum_n V_n\left(t_0 + \frac{nL}{c}\right) = \Re \tilde{V}^+ e^{j\omega t_0} \sum_n \exp jn\left(\frac{\omega L}{c} - \theta(\omega)\right).$$

Thus secular energy-gain requires synchronism between the particle, and the forward-wave in the structure. The transit angle must equal the geometric phase-advance per cell,

$$\frac{\omega L}{c} = \theta(\omega). \tag{50}$$

To appreciate the scalings for multi-cell structures we should include losses. For the standing-wave structure this is straightforward, amounting to an effective wall-Q for the operating mode. For the travelling-wave structure we could proceed by fashioning a multi-cell circuit model, as depicted in Fig. 18. However a simpler approach is possible. We view the structure as a transmission line, with input and output matched at the operating frequency. We suppose a steady-state power P_{in} is provided to the input, and compute the net no-load voltage.

The steady flow of power P through the structure at location z, at group velocity V_g, implies stored energy per unit length $u = P/V_g$. Due to losses P may diminish through the line. Picking a particular cell one may say that

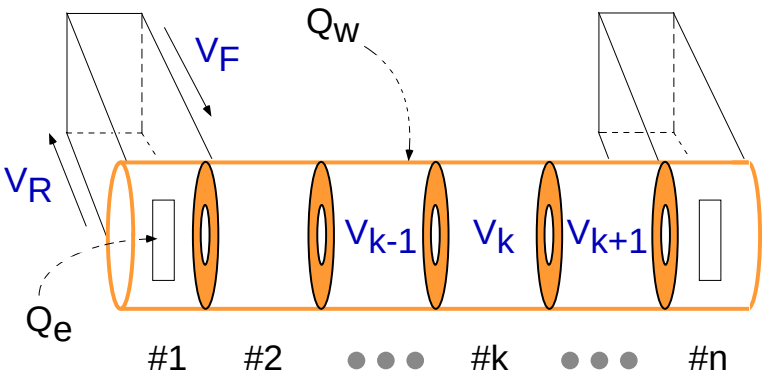

Figure 18. An N-cell travelling-wave accelerator structure may be modelled as N coupled resonators, or, more simply, as a transmission line.

if P_- is the power flowing into the cell, and P_+ the power flowing out, the difference must be due to power flowing into the walls,

$$P_- - P_+ \approx -L\frac{dP}{dz} = \frac{\omega}{Q_w}(uL),$$

where L is the cell-length, uL is the energy stored in the cell, and z is the displacement from the input cell. This may be expressed as

$$\frac{d}{dz}(V_g u) = -\frac{\omega u}{Q_w}. \tag{51}$$

This result determines the profile for stored energy in the structure, in terms of a group-velocity profile determined by the variation in beam-port radius, with initial condition set by $u(0) = P_{in}/V_g(0)$. Voltage may be determined from the shunt-impedance per unit length $r \equiv R/L$, with R the shunt-impedance associated with a single-cell. Identifying the local gradient $G \equiv V/L$, with V the local single-cell voltage, one finds

$$r = \frac{Q_w}{\omega u}G^2 = \frac{G^2}{-dP/dz}$$

To illustrate, consider a *constant-impedance* structure, one with constant iris-radius, and therefore constant V_g. One can show that

$$G(z) = \left(\frac{\omega u}{Q_w}r\right)^{1/2} = G(0)\exp(-\alpha z),$$

where the attenuation constant is

$$\alpha = \frac{\omega}{2Q_w V_g}.$$

The net voltage is then obtained by integration

$$V = \int_0^{L_s} G\,(z)dz = G\,(0)\,\frac{1-e^{-\alpha L_s}}{\alpha},$$

with $L_s = NL$ the structure length, and N the number of cells. In terms of the *attenuation parameter*,

$$\tau = \alpha\, L_s = \frac{\omega L_s}{2\,Q_w V_g}, \tag{52}$$

one may express this result as

$$V = (P_{in} R_{shunt})^{1/2}\,(1-e^{-\tau})\left(\frac{2}{\tau}\right)^{1/2}, \tag{53}$$

where $R_{shunt} = rL$, is larger by N than the shunt-impedance for a single-cell. One can show that maximum no-load voltage occurs for $\tau \approx 1.26$ and is given by $V \approx 0.9\ (P_{in}\,R_s)^{1/2}$. The attenuation parameter also determines the power to the load $P_{out} = P_{in}e^{-2\tau}$, and the fill-time, $T_f = L_s/V_g$, is just

$$T_f = \frac{2Q_w}{\omega}\tau. \tag{54}$$

For example, the Mark III accelerator employed constant impedance structures, operated at 2856 MHz, and $\pi/2$ phase-advance per cell [17]. Group velocity was $0.01c$ and shunt-impedance per unit length was $r \approx$47.3 MΩ/m. Structure length was 3.05 m, and fill-time was 1μs. With $Q_w \approx 1.0 \times 10^4$, attenuation parameter was $\tau \approx 0.90$. For their fixed-target experiments, the Mark III group developed a 20 MW klystron, and designed and built a linac powered by 21 such tubes. With one tube powering one structure, no-load voltage of 47.5 MV/tube could be achieved, and maximum beam energy was 1 GeV. Subsequent adventures, with a 20-GeV beam and a thousand or so *constant-gradient* structures can be found in [2].

2.13. SPACE-HARMONICS

It is not always adequate to have only the simple description of the multi-cell structure in terms of cell voltages as sampled by a relativistic beam. Where the sampling is done by a non-relativistic beam (in an injector), or by a bench-measurement device (a perturbing bead), one would appreciate a prescription for understanding the observations, and this means a more detailed picture of the fields in a multi-cell structure. This picture can be approached quite readily based on the analysis of the last section. We

consider a right-travelling wave, in steady-state, at angular frequency ω. Based on the analysis of cell voltages leading to Eq. (47), we may infer that the steady-state axial electric field phasor satisfies the phase-advance boundary condition

$$\tilde{E}_z\left(\vec{r}_\perp, z+L\right) = e^{-j\theta}\tilde{E}_z\left(\vec{r}_\perp, z\right). \tag{55}$$

Defining wavenumber $\beta_0 = \theta/L$, and

$$\tilde{e}_z\left(\vec{r}_\perp, z\right) = e^{-j\beta_0 z}\tilde{E}_z\left(\vec{r}_\perp, z\right),$$

we observe that $\tilde{e}_z\left(\vec{r}_\perp, z+L\right) = \tilde{e}_z\left(\vec{r}_\perp, z\right)$, $\tilde{e}_z$ is a periodic function of z. This implies that we may expand it in a Fourier series,

$$\tilde{e}_z\left(\vec{r}_\perp, z\right) = \sum_{n=-\infty}^{+\infty} e_n\left(\vec{r}_\perp\right) e^{j2\pi n\, z/L}.$$

This in turn provides a decomposition for the electric field

$$\tilde{E}_z\left(\vec{r}_\perp, z\right) = \sum_{n=-\infty}^{+\infty} e_n\left(\vec{r}_\perp\right) e^{j\beta_n z}, \tag{56}$$

where the wavenumber

$$\beta_n = \beta_0 + \frac{2\pi\, n}{L}. \tag{57}$$

This result is Floquet's theorem. Applying Maxwell's equations to the electric field, we find that the harmonic amplitudes satisfy

$$\left(\nabla_\perp^2 + \frac{\omega^2}{c^2} - \beta_n^2\right) e_n = 0.$$

For a structure with circular symmetry, for example

$$e_n\left(\vec{r}_\perp\right) = E_n I_0\left(r\sqrt{\beta_n^2 - \frac{\omega^2}{c^2}}\right), \tag{58}$$

with I_0 the modified Bessel function. In general higher multipole content may be present, when not excluded by symmetry. With this "space-harmonic" decomposition, the problem of describing the fields is reduced to a determination of the space-harmonic amplitudes E_n.

Meanwhile, a relativistic beam on a ballistic orbit in effect performs a spatial Fourier transform on the fields in the structure, selecting out the synchronous component at each frequency ω, corresponding to wavenumber

ω/c. For the powered structure, the voltage gain of a speed-of-light particle is determined by the $n = 0$ space-harmonic, as this satisfies $\beta_0 = \omega/c$ by design. This is why the model of the multi-cell structure in terms of single-cell voltages is adequate, from the descriptive point of view. The space-harmonic decomposition is helpful for design. In the early days, multi-cell structure design proceeded by analytic calculation [18] and bench-measurements on test stacks of a few cells [2]. Such calculations and measurements invoke the space-harmonic decomposition. In modern times one may in addition employ numerical design with prescribed phase-advance boundary conditions, as in Eq. (55).

To summarize, in pursuing the theory of the microwave accelerator we have followed the logic illustrated in Fig. 6. starting with a terminated interaction, and forming a cavity amenable to resonant excitation. Having arrived at the concept of shunt-impedance, we went on to find that shunt-impedance can be improved by use of multiple cells. One can achieve a high accelerating voltage, with a low peak power, limited only by losses, and the structure tune. These are the basic principles of the microwave linac as formulated by Ginzton, Hansen and Kennedy [19]. They leave a great deal of freedom to the structure designer as one can see in the literature [20, 21].

2.14. MICROWAVE CIRCUITS

To appreciate the practical aspects of a linac rf system, and associated instrumentation, it is helpful to take a look at a few common rf components, as depicted schematically in Fig. 19. The most conspicuous items in the figure are the lines, indicating waveguide and the circles, indicating connectors. We'll consider these first, and then look at common circuit elements.

Considerations in the choice of waveguide include: power-handling, multi-moding, attenuation, phase-stability, coefficient of thermal expansion, convenience and cost. Waveguide appears in two applications in an accelerator system. One is for transport of high-power, and the other is for monitoring of the beam and the rf system, "instrumentation". For instrumentation work, one typically employs coaxial cable or "coax", as this is cheapest to obtain and install. The accelerator itself is a high-power application, demanding low-attenuation, and resilience against "breakdown". For this, hollow rectangular waveguide is a common choice.

Attenuation includes skin-effect losses, dominant at low-frequency, and, in coax, bulk dielectric losses, dominant at high-frequency. Attenuation is accompanied by dispersion thanks to the Kramers-Kronig relations [22]. Cumulative phase-shift through a length of line depends on the guide length, and thus the temperature. Where phase between two outputs is

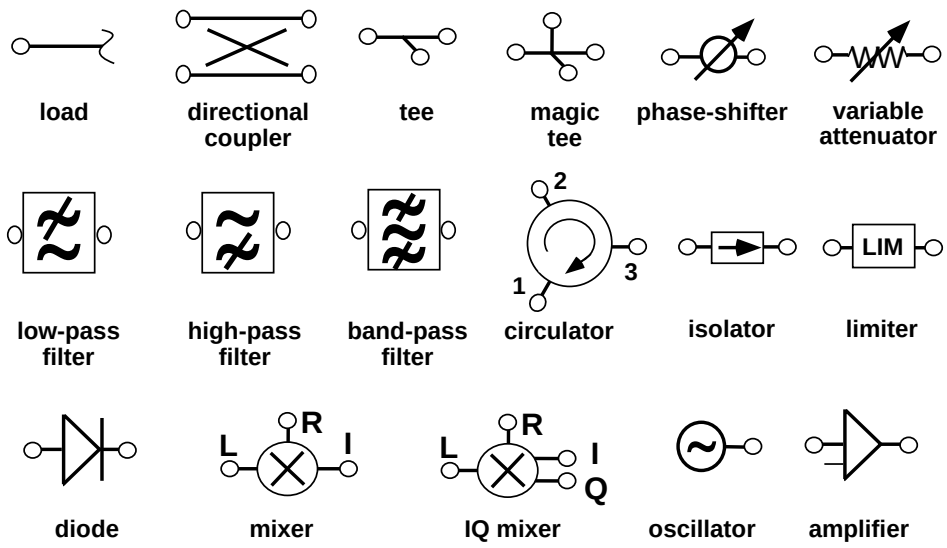

Figure 19. Components commonly found in microwave circuits.

the concern, one may require temperature-stabilized water-flow, a thermal blanket, and/or special low-expansion coefficient cable. For instrumentation work, if one can arrange that phase difference between two cable outputs is the relevant quantity, it may be enough that the cables are of roughly equal length, and traverse the same run. Phase is in general sensitive to mechanical flexure, and guide should not be unintentionally crushed, dented, bent, made to bear a load or stepped on. Intentional dimpling may be employed to adjust phase-length of rectangular guide, which for that matter is used to bear a load, commonly its own weight. Where precise phase relationships are required within an instrumentation circuit, semi-rigid cable may be employed, and the circuit enclosed in a box to preserve mechanical integrity, as well as to provide shielding against electromagnetic interference and pulsed noise—"technical noise".

Rectangular Guide

A selection of common rectangular waveguide dimensions is indicated in Table 3. Due to the proliferation of "standards" for band nomenclature, there really isn't a single standard band designation as such, and the nomenclature of Table 3 may be considered vernacular. More detailed tables are available from waveguide manufacturers, including tolerances on guide dimension, and an assortment of standard flanges [23]. Accelerators have been built and operated in all of the bands listed in Table 3.

The mode of rectangular guide with the lowest cut-off frequency is

TABLE 3. Common rectangular waveguide ("WR") interior dimensions, recommended operating band, and cut-off frequency. Also indicated is the larger band designation into which the waveguide may be classed.

Guide	Size (inch)	Rec. (GHz)	f_c (GHz)	Band	(GHz)
WR650	6.500×3.250	1.12 - 1.70	0.91	L	(1.0 - 2.0)
WR284	2.840×1.340	2.60 - 3.95	2.08	S	(2.0 - 4.0)
WR187	1.872×0.872	3.95 - 5.85	3.15	C	(4.0 - 8.0)
WR90	0.900×0.400	8.20 - 12.40	6.56	X	(8.0 - 12.0)
WR62	0.622×0.311	12.40 - 18.00	9.49	Ku	(12.0 - 18.0)
WR42	0.420×0.170	18.00 - 26.50	14.05	K	(18.0 - 27.0)
WR28	0.280×0.140	26.50 - 40.00	21.08	Ka	(27.0 - 40.0)

the TE_{10} mode. This mode has electric field oriented parallel to the b-dimension, as in Fig. 4, peaking in the guide center, and dropping monotonically to zero at the edges of the guide, so as to satisfy conducting boundary conditions. The TE_{10} mode dispersion relation is, neglecting losses,

$$\left(\frac{\omega}{c}\right)^2 = \left(\frac{\pi}{a}\right)^2 + \beta^2, \tag{59}$$

relating angular frequency ω to wavenumber β and guide long dimension a, as in Fig. 4. This is to say that the spatial variation of a forward wave in the guide is characterized by guide wavelength $2\pi/\beta$ different from the free-space wavelength $2\pi c/\omega$. The dispersion relation, Eq. (59) looks different from that of free-space, and the reason for this is that the mode is a superposition of *two* free-space waves, each propagating at an angle to the guide-axis, and phased in such a way as to satisfy conducting boundary conditions. Cut-off occurs at frequency $f_c = c/2a$ and corresponds to a propagation angle of 90 °, i.e., no propagation. The lower frequency listed in column 3 of Table 3, $1.25 \times f_c$, is recommended to insure that the group velocity, $d\omega/d\beta$, is not too low. The upper frequency in column 3, $1.90 \times f_c$ is just below the cut-off for the next mode(s) in the waveguide.

Waveguide is not lossless, of course, and finite surface resistance damps wave amplitude, amounting to attenuation per unit length given by [24]

$$\alpha = \frac{R_s}{Z_0}\left[\frac{\omega^2}{c^2}\frac{1}{b} + \frac{2\pi^2}{a^3}\right]\frac{c}{\omega}\frac{1}{\beta}. \tag{60}$$

At the lower end of the operating band, the wave has lower group velocity and spends more time in a given length of guide. Thus attenuation becomes large near cut-off. At higher frequencies, surface resistance is increasing, and

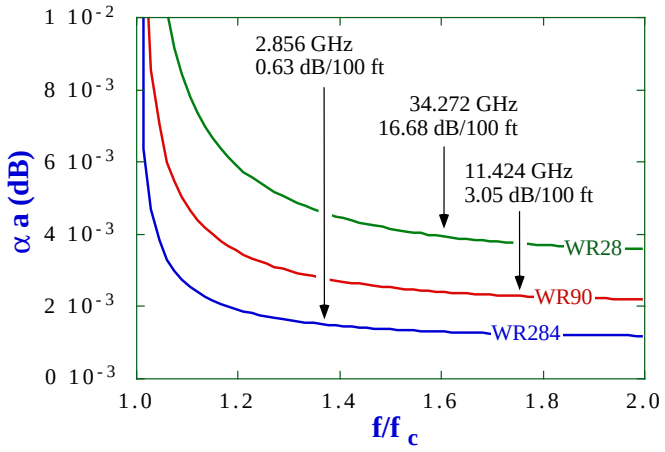

Figure 20. Attenuation α normalized to waveguide width a, versus frequency for three sizes of waveguide.

so attenuation ultimately must increase. Minimum attenuation occurs at $2.3 - 2.4 \times f_c$, outside the single-mode operating band. In Fig. 20 one can see Eq. (60) illustrated for three example waveguide sizes. As a point of reference, OFE copper WR284 has attenuation of 0.629 dB/100 ft at 2856 MHz.

As a practical matter, waveguide insertion loss must be accounted for in the layout of the accelerator microwave network. As an example, the "thick-wall" WR284 network for two successive tube stations in the Two-Mile Accelerator is sketched in Fig. 21. Waveguide arm phase-lengths range from 118 to 139 λ_g, where the guide wavelength $\lambda_g = 2\pi/\beta \approx 15.3$ cm. Over these 60-ft lengths, one expects attenuation in the range of 0.37-0.43 dB. The measured results lie in the range 0.41-0.68 dB, including coupler and flange insertion loss [2]. This means that 10–15% of the power is lost before it reaches the accelerator. Theoretically one could reduce the attenuation loss by a factor of about 2.3, by employing nearly-overmoded custom-made guide. However, the cost-savings for this accelerator wouldn't have been worth it. Even in modern-times, with a utility savings accruing at 3 cents/kW-hr it would take years to "pay" for the cost of the mode-converters from the custom-guide to standard guide. For future machines, however, where rf power consumption may approach several hundred megawatts, waveguide loss is taken quite seriously. Much interest attaches to the design of overmoded or "quasi-optical", low-loss, high-power waveguide components [25].

Fig. 21 also indicates the waveguide coupler layout employed. This layout can be understood in part from the discussion of Sec. 2.8, and beam

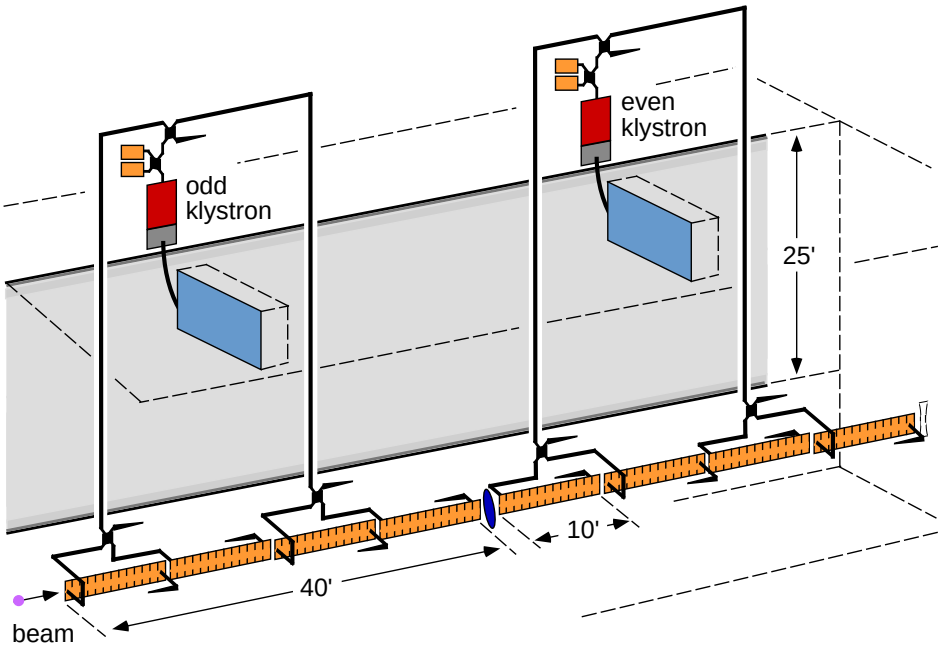

Figure 21. Waveguide layout for a standard pair of klystrons on the Two-Mile Accelerator. This configuration "babaabab" repeats along the linac, as discussed in connection with coupler asymmetry.

deflection due to phase asymmetry. This effect was reduced by offset of the coupling cells, and waveguide feed layout designed to provide partial cancellations of rf kicks. Kick cancellation is not local, however, but is completed only over the 80-ft length spanned by eight structures, and then only approximately.

One may also find, for high-power systems, pressurized waveguide; these are intended to function with electronegative gas, such as SF_6, to inhibit breakdown. These are common in continuous-wave (CW) systems, but are found on pulsed systems as well.

Connectors and Cable

For instrumentation work, and bench measurements one may rely on an assortment of cable and connectors. Commonly used coaxial connectors include the General Radio Corp. (GRC) connector, Navy-type (N) connector, sub-miniature A (SMA), and Amphenol Precision Connectors (APC), 3.5-mm and 7.0-mm. The variety of connectors reflects the frequency range, and some history. The GRC connectors date back to near the Paleolithic Era, but are still to be found in legacy systems, for example, all Two-Mile Accelerators. APC connectors are intended to be used where repeated con-

nection and disconnection will be made, as in bench measurements and calibrations. Type N was the predecessor for the APC-7. SMA is common, and particularly appropriate where a relatively static connection is to be made. It looks a lot like the APC-3.5, except that it contains dielectric inner liner, and is therefore more subject to wear. BNC is really a different category of connector, perfectly adequate for low-frequency signals, but not recommended for microwave work. Other commonly used connectors include "instrument flange" fittings for rectangular waveguide, and adaptors from waveguide instrument flange to standard coax, and *precision* adaptors. For accelerator work one has in addition vacuum-waveguide flanges that mate to knife-edge gaskets to form a crush-seal. These and associated adaptors to instrument flange are not commercial standards.

While folks often refer to cable in terms of the connector they are accustomed to seeing on it, the cable itself may have a name, here are a few.

- *RG-214* is often referred to as "N-type cable" and sees routine use for microwave signals in an S-Band linac system. It has a black polyvinylchloride jacket with 0.425" outer diameter (OD), and a polyethylene dielectric. Attenuation constant is 18 dB/100 ft at 3 GHz and 27 dB/100 ft at 5 GHz.
- $Heliax^{TM}$ is a brand name of Andrews Corporation, for coax with a dielectric that is mostly air, with outer conductor supported on a helical foam ribbon. As a result of the diminished volume of lossy-dielectric, attenuation per unit length is low. It can be delivered, already cut to length and connectorized, in a variety of sizes. It is excellent for high-frequency beamline instrumentation, particularly where low signal level and noise are concerns.
- *RG-223/U* is often referred to as "BPM cable" as it is commonly used to transmit signals from beam position monitor (BPM) striplines. It shouldn't be used for microwave work, unless attenuation and dispersion are desired. Attenuation constant is 3.2 dB/100 ft at 100 MHz, 16.5 dB/100 ft at 1 GHz and 46.0 dB/100 ft at 5 GHz.
- *RG-58C* is often referred to as "BNC cable" and isn't to be used for microwave frequencies. RG-58C has an attenuation constant of 4.9 dB/100 ft at 100MHz.

RG-214, RG-58C and RG-223/U have a group velocity of 0.66c. In addition to these, there is a wealth of specialty cable available commercially, designed to provide either low loss, low thermal expansion, insensitivity to flexure, or mechanical ruggedness.

Having employed low attenuation cable to extract a marvelous beam-induced or other signal, one will probably need an *attenuator*. These can be fixed, or variable, for example, mechanically variable, by knob or screw

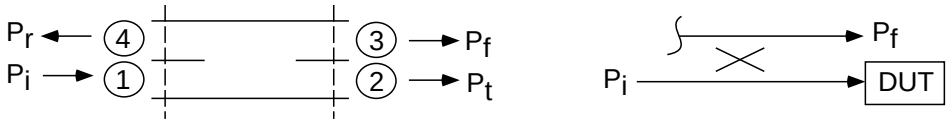

Figure 22. A directional coupler permits one to monitor a circuit. The circuit sketch on the right illustrates a typical use, to monitor forward power to a device-under-test (DUT) attached to port 2, with a matched load on the reverse arm, port 4.

adjustment. Beams are pretty variable, so variable attenuators are good to have. Like most microwave components, attenuators come with a rating for VSWR and insertion loss over a specified bandwidth.

To understand a circuit it often helps to disconnect a new device and replace it with a *matched load* . Matched loads come in all shapes and sizes, with a specification for VSWR and bandwidth. A cavity with $\beta = 1$ is an example of a matched load. Generally however one prefers a match over a broad band, and this implies use of high-loss material, equivalent to very low Q_w in the parlance of cavities. S-Band structure output loads may employ a Kanthal-loaded tapered geometry, for example.

Directional Coupler

To monitor the waves in a circuit it is helpful to tap off a small bit of the forward and reverse waveforms, while not interfering appreciably with the circuit. One needs a *directional coupler* as depicted in Fig. 22.

The function of many basic components can be understood from their description in terms of an S-matrix, and the directional coupler is a good example of this. The S-matrix for an N-port device is an $N \times N$ matrix defined with respect to reference planes on the connecting waveguide, and a common normalization for waveguide impedance. With phasor $\tilde{V}_k^+$ incident on port k, and all other ports terminated in a matched load, the outgoing phasor on port l is $S_{lk}\tilde{V}_k^+$. For elements that are free of sources the S-matrix is symmetric and unitary, and this is a fair approximation even for copper boundaries with their finite resistivity, provided the microwaves don't spend too much time in the device, i.e., for a broadband device.

In this context, an ideal directional coupler is a four-port element that is symmetric upon reflection, is lossless and is perfectly matched. These conditions, together with unitarity restrict the S-matrix to the form

$$S = \begin{pmatrix} 0 & \alpha & j\sqrt{1-\alpha^2} & 0 \\ \alpha & 0 & 0 & j\sqrt{1-\alpha^2} \\ j\sqrt{1-\alpha^2} & 0 & 0 & \alpha \\ 0 & j\sqrt{1-\alpha^2} & \alpha & 0 \end{pmatrix}, \qquad (61)$$

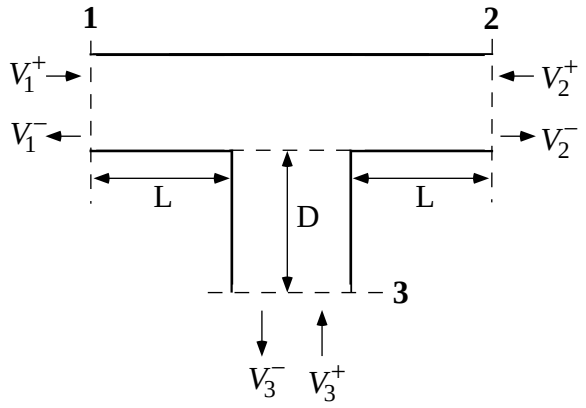

Figure 23. A tee is a three-port element with symmetry between two arms.

for some parameter $\alpha \leq 1$, and some choice of reference planes. A real directional coupler is described by coupling C and directivity D, defined with respect to the quantities indicated in the sketch: power incident on port 1, P_i, transmitted power to port 2, P_t, forward power on port 3, P_f, and reverse power, P_r, on port 4. The coupling and directivity are, in units of decibels (dB),

$$C = 10\log_{10}\left(\frac{P_i}{P_f}\right), \quad D = 10\log_{10}\left(\frac{P_f}{P_r}\right). \tag{62}$$

The isolation is defined according to $I = 10\log_{10}(P_i/P_r)$, so that $C+D = I$. The 3-dB coupler is rather special in that, with a matched load on one arm, it can distribute power equally to the two output arms. Couplers come in high-power, vacuum waveguide and low-power, instrument-connectorized versions.

Tee

The simplest-looking element is the *tee* as illustrated in Fig. 23. Normalizing all ports to a common impedance, one may apply unitarity to determine the form the S-Matrix must take. After a choice of reference planes (L and D in the figure), one can show that the S-Matrix may be parameterized in terms of two real variables ϕ and θ, such that [25]

$$S = \begin{pmatrix} \frac{1}{2}\left(e^{j\phi} - \cos\theta\right) & \frac{1}{2}\left(-e^{j\phi} - \cos\theta\right) & \frac{\sin\theta}{\sqrt{2}} \\ \frac{1}{2}\left(-e^{j\phi} - \cos\theta\right) & \frac{1}{2}\left(e^{j\phi} - \cos\theta\right) & \frac{\sin\theta}{\sqrt{2}} \\ \frac{\sin\theta}{\sqrt{2}} & \frac{\sin\theta}{\sqrt{2}} & \cos\theta \end{pmatrix}. \tag{63}$$

The ideal lossless tee cannot be matched ($|S_{11}| \neq 0$), except in the degenerate case, $\sin\theta = 0$.

The tee forms the basis for many useful devices. With a shorting plane on arm 3, and a signal incident on port 1, the signal transmitted through from port 1 to port 2 is a superposition of a wave transmitted directly to port 2, and one that travels down arm 3, and, after multiple reflections in arm 3, returns to be split between arms 1 and 2. The impedance seen looking into arm 1 then depends on the position of the shorting plane on arm 3. A "stub-tuner" consists of one or more such tees, with moveable shorting planes, and can be quite helpful in providing a match when arm 2 is connected to a DUT. A tee with moveable short can also function as a reflective switch. Considering the frequency dependence of the constructive or destructive intereference, one recognizes that this device may also function as a tunable band-pass or a band-reject ("notch") filter.

Speaking of which, we have it seems already touched on the basic elements needed to fashion filters. Hollow waveguide is a high-pass filter, due to the phenomenon of cut-off. A tee can function as a band-pass filter. A resonant cavity can serve as a narrow-band filter. More generally an assortment of tee's and cavities ("poles") can be employed to fashion a two-port element with a custom-made S_{21}. One can find a wide assortment of commercial tubular, cavity and waveguide microwave filters, fixed filters, mechanically-tuned, or voltage-controlled.

Magic Tee

A hybrid tee is a four-port device as illustrated in Fig. 24. In the ideal case, it is lossless and the symmetry seen in the figure precludes coupling of ports 1 and 4. A *magic tee* is a hybrid tee that is matched on all ports; in practice, matching relies on obstacles situated in the junction. Constraints of symmetry and unitarity reduce the possible forms of the S-matrix to just one, after a choice of reference planes,

$$S = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & -1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \end{pmatrix}. \tag{64}$$

From this result, or just from the symmetry evident in Fig. 24, one can show that for two incident voltages, on ports 2 and 3, port 1 provides a an output proportional to the sum, and port 4 an output proportional to the difference. In conjunction with a phase-shifter and a variable attenuator, the magic tee can be employed to balance two large signals, permitting easy inspection of a small residual.

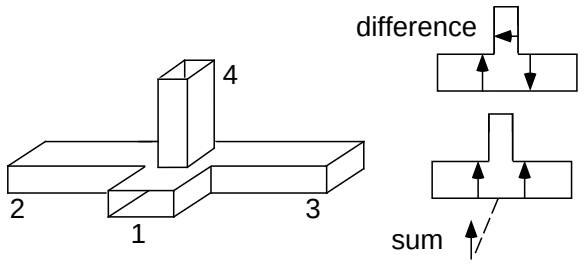

Figure 24. A magic tee is a symmetric four-port device matched on all ports. With signals incident on through arms 2 and 3, the H-plane arm, port 1, provides a sum signal, and the E-plane arm, port 4, provides a difference signal.

Circulator, Isolator

An ideal *circulator* is a three-port element described by

$$S = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Inspecting this S-matrix one can see that the device is matched on all ports, and a signal incident on one arm is circulated to only one adjacent arm. Based on our analysis of the tee, one might think that such a device would be impossible to construct; however, our analysis of the tee assumed a reciprocal device, one with symmetric permeability and permittivity tensors. A circulator meanwhile incorporates μ-material and a static magnetic field, and is non-reciprocal, by design. A real circulator has noticeable insertion loss. With one arm terminated in a matched load, a circulator becomes a two-port device, then called an *isolator*. These devices are almost essential for isolating a source from a load, to avoid pulling of the source frequency or amplitude, and in isolating one part of a circuit from another, to avoid an inadvertent standing-wave and resulting modification of the circuit operating point. Mention of operating point hints that some circuits are non-linear. Let us consider the basic non-linear elements.

Crystal Detector

To monitor the time-variation of a low-frequency voltage one may use an oscilloscope. This method doesn't work well with microwaves, however, since typical scope bandwidths are lower than 1 GHz, and typical microwave frequencies are above 1 GHz. To monitor amplitude one needs a microwave diode, a *crystal detector*. A diode is a non-linear element, with ideal characteristic $i = i_s(e^{eV/k_BT} - 1)$, where $k_B \approx 1\text{eV}/(38.7 \times 300\,\text{K})$ is Boltzmann's constant, T is temperature, and i_s is the saturation current. A microwave diode is a diode encapsulated in a matched waveguide mount (coaxial or

other). The problem of designing such a mount was once a major research problem [26]; nowadays there is a variety of commercial detectors to choose from, depending on one's needs for dynamic range, bandwidth, sensitivity, and output polarity. Microwave diodes are ubiquitous elements, appearing in all microwave detection circuits. In addition to power-detectors, there are also special-purpose diodes, including step-recovery diodes for generating harmonics, as in a frequency-multiplier, and PIN diodes for fast switching or phase-shifting.

The first characteristic to know about a diode is that it can be destroyed by too high an input level. If routine operation will involve extreme signal ranges, one needs also attenuation and a *limiter*. A limiter is a two-port element that changes state ("breaks down") above a certain power level, absorbing or reflecting the incident signal. A limiter is essential where expensive test equipment is employed with beam-induced signals. At high-power limiters may generate un-wanted harmonics, and this is not usually desirable; they are intended to function like a circuit-breaker.

It is instructive to consider a simple model of the ideal microwave diode. Let us apply a single-frequency voltage waveform to the ideal diode characteristic. We express $eV/k_BT = A\cos\phi$ with $\phi = \omega\, t + \phi_0$, t is time, ω is the angular frequency of the signal, and ϕ_0 is the phase of the signal. Using the identity,

$$e^{A\cos\phi} = \sum_{k=-\infty}^{+\infty} I_k\,(A)\, e^{jk\phi}, \tag{65}$$

with I_k the modified Bessel function, we may express the current through the diode as

$$i = \sum_{k=-\infty}^{+\infty} i_k\,(A)\, e^{jk\phi},$$

for some coefficients i_k. The "DC" component is

$$i_0 = i_s\,(I_0\,(A) - 1) \approx \frac{1}{4} i_s A^2,$$

and in the last equality we employ the small-argument expansion for I_0. Combined with a low-pass filter characteristic (integral to the crystal mount or in the output circuit) crystals evidently provide a "square-law" response, for small signal amplitude, with current output proportional to incident microwave power. Where the crystal is being employed as the primary power-monitor, it could be labelled, calibrated, and re-calibrated occasionally. For best results, the calibration waveform should mimic as closely as possible the waveform to be monitored, because a crystal mount will not transmit all frequencies equally. Crystals may be employed beyond the range of

square-law response; this requires a polynomial fit based on the calibration data covering the full range of power-levels to be employed.

As an example we could consider monitoring an S-Band klystron-amplifier output. A typical setup would include a 50 or so dB directional coupler, a low-pass filter to block out harmonics generated by the klystron, attenuators, and a crystal mount, with a replaceable 1N21B cartridge. Where high-precision is required, an occasional check of calibrations might employ calorimetry, inferring average power based on temperature variation of cooling water to a load. In addition, there are several instruments that can greatly simplify circuit tests, as well as instrumentation problems themselves. These include the power-meter, frequency-counter, spectrum analyzer, and the scalar or vector (phasor) network analyzer. Items such as these come with documentation roughly proportional to cost, and thus the manufacturers' literature provides a wealth of additional information.

Mixer

Where phase-information is of interest it is common to employ a *mixer*, a three-port device, accepting two high-frequency inputs, and providing output at the difference frequency. The simplest concept for a mixer consists of a tee to sum two signals, labelled R and L, and a diode to detect the output. The voltage across the diode takes the form $eV/k_BT = A_R \cos\phi_R + A_L \cos\phi_L$, where $\phi_R = \omega_R t + \phi_{0R}$, and $\phi_L = \omega_L t + \phi_{0L}$. Using the identity, Eq. (65), twice, we may express the diode current as

$$i = i_s \sum_{m,n=-\infty}^{+\infty} I_m(A_R)\, I_n(A_L)\, e^{j(m\phi_R + n\phi_L)} - i_s.$$

Thus one may expect the output to include all frequencies of the form $m\omega_R + n\omega_L$. This isn't usually satisfactory, since one is interested in the component at the intermediate frequency ("IF") $\omega_{IF} = \omega_R - \omega_L << \omega_R, \omega_L$. To remove the "DC" offset

$$i_0 \approx i_s \left(I_0(A_R)\, I_0(A_L) - 1\right)$$

the mixer may incorporate two diodes, exposed to opposite polarity versions of the R signal, with outputs subtracted. This can be accomplished for example with a magic tee, the L-signal incident on the H-plane arm, the R-signal on the E-plane arm, and diode mounts on the thru-arm ports, with balanced outputs. The lowest frequency component remaining is then

$$i_{IF} = 2 i_s I_1(A_R)\, I_1(A_L) \cos(\phi_R - \phi_L)$$

In phasor-form we may write this as $i_{IF} = \Re \tilde{i}_{IF} e^{j\omega_{IF} t}$, where

$$\tilde{i}_{IF} = 2 i_s I_1(A_R)\, I_1(A_L)\, e^{j(\phi_{Ro} - \phi_{Lo})}.$$

This last result is quite helpful, for conceptual purposes. Ideal operation of a mixer invokes a CW signal of known, constant amplitude and phase for the L port. If one thinks of the R input in simple phasor terms, as $V_R = \Re\tilde{V}_R e^{j\omega_R t}$, with phase referred to the L signal, then the IF voltage produced by the circuit $V_{IF} = \Re\tilde{V}_{IF} e^{j\omega_{IF} t}$ is simply a phasor $\tilde{V}_{IF} \approx k\tilde{V}_R$ providing an approximately linear analog of the input signal, referred in phase to the L port, with IF carrier frequency. The last equality is written with the approximation $A_R << 1$, in mind, and the constant k depends on the mixer specification, and the L-amplitude. This conversion, as well as features of the unwanted higher-frequency terms will be described in plots provided by the manufacturer for the particular mixer.

Mixers come in a variety of forms more sophisticated than a simple balanced mixer. The "IQ" or "dual-output" mixer is a four-port device, providing two IF outputs representing the in-phase and quadrature components of the R-signal referred to the L-signal. One output provides $V_{IF-I} = \Re\tilde{V}_{IF} e^{j\omega_{IF} t}$, and the other $V_{IF-Q} = \Im\tilde{V}_{IF} e^{j\omega_{IF} t}$. In practical terms, mixers permit one to take a high-frequency signal down below 1 GHz, where filters, amplifiers and other components are cheap, at the expense of some additional caution in designing the circuit. Related to the mixer concept is the "modulator", where a pulsed bias voltage is applied to shift the phase or amplitude of a high-frequency signal, to produce a modulated high-frequency signal.

Oscillators and Amplifiers

Just a few words about oscillators and amplifiers, as there are libraries on the subject. When using old, cast-off, or "don't-know-where-it's-been" sources or amps, one should keep in mind that front-panel displays are for monitoring not measuring. A frequency-counter and power-meter are helpful to check what an instrument is actually putting out. Since diodes have a temperature-dependent characteristic, most solid state devices accept temperature as a "control knob". So it is good practice to exert some control over temperature. Silicon is a good particle detector, so it is preferable not to situate solid-state components in an accelerator vault.

We could also say a word about "noise". For a system in equilibrium at temperature T, thermodynamics tells us to expect, in a bandwidth Δf noise-power $P_n = k_B T \Delta f$. Most systems aren't yet aware of this equilibrium. Accelerator vaults and associated areas are flooded with all manner of signals, ranging from ionizing radiation, to rf frequencies leaking out of unterminated couplers, to pulsed noise from the modulator circuits, and ac noise at the power-line frequency and harmonics. Diurnal variations round out the low-frequency end of the "noise" spectrum.

Amplifiers also don't do as well as the black-body, and so usually come

with a specification for "noise figure" or NF. Given power gain G for the amplifier, one may determine the noise power output by the device P_N, above that expected at the operating temperature T, according to

$$NF = 1 + \frac{P_N}{G k_B T \Delta f}.$$

Operationally, in terms of loss in the ratio of signal-to-noise,

$$NF = \frac{(S/N)_{input}}{(S/N)_{output}}.$$

By convention NF is specified at 290 K (17°C), usually in decibels, as $10 \log_{10} NF$, or in terms of equivalent noise temperature $T_{eq} = (NF - 1)T$.

With microwave circuit components in-hand, we can turn to instrumentation; but first let us examine what there is to instrument.

3. Physics of Beams

Thus far we have discussed the microwave accelerator in detail, and left the beam pictured as a little point charge. In this section, we take a closer look at the beam, first some formal-looking details and then the more colloquial aspects of beam physics. The theoretical development of the subject of beams is extensive, but what is actually known about beams is limited by the instrumentation available.

3.1. FORMAL-WEAR FOR BEAMS

For purposes here a beam is a collection of electrons of comparable momenta, moving at about the speed of light in the lab-frame. One may be concerned that this definition excludes discussion of low-voltage beams, but even an 80 keV electron is moving at $0.5c$. And typical accelerating fields are on the order of several MeV/m, so we omit at most about an inch of linac, and actually, not even that.

To describe our beam, we might at first be inclined to pull out quantum electrodynamics and set to work. Then we realize as we observe collisions, that we need to pull out the whole Standard Model. At this point we realize that we really can't describe beams completely until we have figured out the Theory of Everything. Fortunately such a complete description of beams isn't needed. Accelerator phenomena take place at low center of mass energy where electrodynamics rules. If we trace the evolution of a beam from emission through collisions we find that it's behavior is well-described by classical mechanics and electrodynamics, with only a few exceptions. For the present work, we will limit ourselves to the linac, a classical venue.

Kinetic Theory

Adopting a classical description then, we may associate with each electron a position $\vec{R}$ and a velocity $\vec{V} = d\vec{R}/dt$, and momentum $\vec{P}$. Let us suppose our beam contains N electrons. The beam is completely specified given the positions and momenta $\left\{\vec{R}_i(t), \vec{P}_i(t)\right\}$, with $i = 1, ..., N$ at some instant of time t.

In principle, to describe the evolution of this system, one need merely determine the fields at this time, including fields due to the beam, and then solve $6N$ ordinary differential equations coupled to Maxwell's equations. This is really too much work though, and all to produce too much information. Such a solution would include the Coulomb field of each and every electron, and detailed information on electron-electron collisions mediated by these fields. And electrons don't even collide too often, so one is storing quite a lot of information in RAM, in order to accurately model something

that is really a higher-order effect. There is a simpler, approximate method of describing such a beam.

We start from the *Klimontovich* distribution,

$$F(\vec{r},\vec{p},t)=\sum_{k=1}^{N}\delta^3\left[\vec{r}-\vec{R}_k(t)\right]\delta^3\left[\vec{p}-\vec{P}_k(t)\right]$$

the density in the six-dimensional space with coordinates $(\vec{r},\vec{p})$. We observe by differentiation, and use of the Lorentz force law, that

$$\frac{\partial F}{\partial t}+\vec{v}\bullet\vec{\nabla}_{\vec{r}}F-\frac{e}{m}\left(\vec{E}+\vec{v}\times\vec{B}\right)\bullet\vec{\nabla}_{\vec{p}}F=0,$$

where $\vec{v}$ is velocity. We emphasize that this is a rather odd equation, not because of its form, a continuity equation, but because F is a spiky thing, and $\vec{E},\vec{B}$ also behave badly in the vicinity of each electron. A clearer picture of the system may be formed by selecting a length-scale that distinguishes between the microscopic and macroscopic features. We average on this microscopic length-scale,

$$f=\langle F\rangle,\quad \vec{e}=\left\langle\vec{E}\right\rangle,\quad \vec{b}=\left\langle\vec{B}\right\rangle,$$

permitting us to express,

$$F=f+\delta f,\quad \vec{E}=\vec{e}+\delta\vec{e},\quad \vec{B}=\vec{b}+\delta\vec{b}.$$

The smoothed distribution satisfies,

$$\frac{\partial f}{\partial t}+\vec{v}\bullet\vec{\nabla}_{\vec{r}}f-\frac{e}{m}\left(\vec{e}+\vec{v}\times\vec{b}\right)\bullet\vec{\nabla}_{\vec{p}}f=\frac{e}{m}\left\langle\left(\delta\vec{e}+\vec{v}\times\delta\vec{b}\right)\bullet\vec{\nabla}_{\vec{p}}\,\delta f\right\rangle.$$

In the absence of fluctuations in the actual density, the right-side vanishes. Such fluctuations disappear however only in the formal limit, $e\to 0, m\to 0$, e/m constant. In this limit, all cross-sections vanish, there are no single-particle emissions or collisions as such. With finite e, fluctuations represent a correction, but a small correction, and not a dominant effect. Thus we do well to describe the distribution in iterative fashion. At lowest-order, neglecting two-particle correlations, we have the *Vlasov* equation,

$$\frac{\partial f}{\partial t}+\vec{v}\bullet\vec{\nabla}_{\vec{r}}f-\frac{e}{m}\left(\vec{e}+\vec{v}\times\vec{b}\right)\bullet\vec{\nabla}_{\vec{p}}f=0. \tag{66}$$

At this order too, we have Maxwell's Equations (for the smoothed fields) determined by moments of f,

$$\rho(\vec{r},t)=-e\int d^3\vec{p}\,f(\vec{r},\vec{p},t),$$

$$\vec{J}(\vec{r},t) = -e \int d^3\vec{p}\,\vec{v}\,f(\vec{r},\vec{p},t)\,.$$

One can check that

$$\frac{\partial \rho}{\partial t} + \vec{\nabla} \bullet \vec{J} = 0.$$

To simplify our notation we will denote the smoothed fields by $\vec{E}$, $\vec{B}$.

The Vlasov equation is a continuity equation in phase-space. It states that the convective derivative of f along a physical orbit vanishes. If one moves with an electron in phase-space, one finds that the local density of electrons, f, is constant. Said differently, trajectories in phase-space don't cross.

Let us emphasize a few points to put aside some common misconceptions. Beam-induced fields are not omitted in this treatment, merely microscopic fluctuations in these fields. Thus fields of electrostatic character ("space-charge" fields) fit perfectly well in this description. But let us not call them "Coulomb fields", since we have averaged out the fluctuations that this phrase suggests, and neglected the collisions that these fields mediate. Collisions of course would exert a force specific to the scattering particles and cause phase-space trajectories to cross. In addition to intra-beam scattering, we have also omitted other collisions, for example, between an electron and an ion resident in the beamline. Radiation is not omitted in this treatment, merely fluctuations in the radiation fields, radiation seen by one electron, but not its neighbor. But let us not say that "incoherent radiation" is omitted, since an externally-produced wave with poor coherence fits easily in this description. Finally we emphasize, as a caveat, that important exceptions to this picture appear in beam-physics problems other than the electron-linac. These are fluctuation-dissipation problems, associated with scattering, radiation, or cooling. In these problems, corrections to the right-side of the Vlasov equation may often be approximated as in the Fokker-Planck equation.

Vlasov Equation

To fully enjoy the Maxwell-Vlasov equations we should consider a practical computational interpretation. Suppose that we start at $t = 0$, and solve these equations for the distribution and the fields, taking note of the fields at every step in time. Then let us return to $t = 0$, and generate a finite collection of "test-particles", i.e., coordinates $(\vec{r},\vec{p})$. We'll suppose this initialization or "loading" in phase-space represents a sampling of the original initial Vlasov distribution, when smoothed on some appropriately small length scale ("binned", "allocated"). We then "track" these individual test particles by integrating the equations of motion in the fields already determined. As far as the particles are concerned, these are prescribed external

fields. Along the way, we compute charge and current densities associated with the test particles, averaged on the microscopic length-scale. When we are finished tracking, we should find that we have computed the evolved Vlasov distribution, with precision limited by the statistics of our original sampling. Actually we realize that we could return to $t = 0$ again, and track the particles while computing the fields from the "macroparticle" distribution itself. We should arrive at the same result, if we have adequately modelled the charge and current densities with our sampling.

This is the basis for "particle-in-cell" simulations of beams. It is also the basis for a Hamiltonian description of beams, in which we picture individual particles with motion described by a single Hamiltonian, H. Particles do not interact with each other directly, they simply move in a Hamiltonian prescribed by the electromagnetic fields. The Vlasov equation may then be expressed simply in terms of the Poisson bracket in canonical coordinates $(\vec{q}, \vec{p})$,

$$[f, H] = \frac{\partial H}{\partial \vec{p}} \bullet \frac{\partial f}{\partial \vec{q}} - \frac{\partial H}{\partial \vec{q}} \bullet \frac{\partial f}{\partial \vec{p}}, \tag{67}$$

as

$$\frac{df}{dt} = [f, H] + \frac{\partial f}{\partial t} = 0. \tag{68}$$

Viewing each particle as an independent Hamiltonian system, and the beam as a statistical ensemble of such non-interacting systems, the Vlasov equation amounts to a statement of *Liouville's theorem*.

3.2. BEAMS AT WORK

Time-Structure of Beams

In describing the beam and the accelerator it is helpful to distinguish between the *main-linac*, the region of linac where $V \approx c$, and the *injector*, where $V < c$. Our work on the accelerator structure dictates some basic features we must require of our beam distribution in the main-linac. In this section we describe these requirements.

At high-gradient it is not economically (nor thermally) feasible to operate the linac continuously, thus it is *pulsed*. Machine pulses repeat at some frequency f_{rep} with 10–360 Hz a typical range. The length of the rf pulse should be enough to fill the structure, at least about 1 μs at S-Band. The maximum length is constrained by the klystron modulator circuit, with 1–10 μs typical. Within an rf pulse, beam electrons should reside near the accelerating peak of the voltage. In this way one arrives at the picture of Fig. 25, depicting the time-structure of the beam, as dictated by the accelerating mechanism.

Some freedom still remains within this picture. Depending on the capabilities of the injector, one may operate with every bucket filled, where then

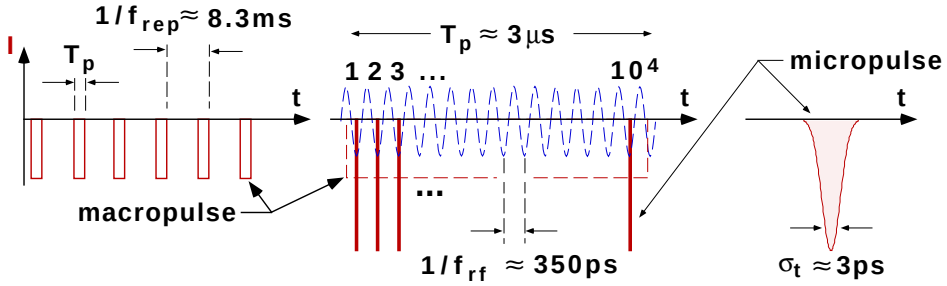

Figure 25. Typical time-structure of a beam in a normal conducting rf linac, viewed on three different time-scales. For illustration here pulse repetition frequency $f_{rep} = 120$ Hz, macropulse length $T_p \approx 3$ μs, with every bucket filled, so that micropulses or "bunches" are spaced at the rf period $1/f_{rf} \approx 350$ ps.

bunches are spaced at the rf period $1/f_{rf}$, one may skip buckets, or one may fill just one bucket. The beam pulse duration T_p may be the full length of the rf pulse, or shorter, 50 ns–3 μs being typical. Intrapulse current is set by the requirements for experimentation, and limited by collective effects, longitudinal and transverse wakefields, beam-loading and beam break-up. For definiteness we often picture a single bunch as a Gaussian current profile,

$$I_b(t) = \frac{Q_b}{\sqrt{2\pi}\sigma_t} \exp\left(-\frac{t^2}{2\sigma_t^2}\right), \tag{69}$$

as a function of time t, corresponding to bunch charge Q_b and bunch length $c\sigma_t$. Real bunch profiles exhibit a variety of shapes, and may contain fine structure. In general one should be alert to the fact that there are typically three currents that may be quoted in connection with a beam. The peak current within a bunch is quite high, approaching 10^3 A. For a bunch-train, the average current, within the pulse, is just $Q_b f_{rf}$, when every bucket is filled, and this may be on the order of 10^0 A. Finally the average current is just the total charge per pulse multiplied by the machine repetition frequency, and this may be on the order of 10^{-5} A.

In the limit of low-current, where collective phenomena may be neglected, the function of the linac rf system is quite simple. Let us consider an electron entering the main-linac at time t. We suppose there are N klystrons powering the linac, and each klystron may be powering one or several structures. Consider the net voltage witnessed by the relativistic electron, absent beam-loading ("no-load" voltage),

$$V = \sum_{n=1}^{N} V_n \cos(\omega t + \varphi_n),$$

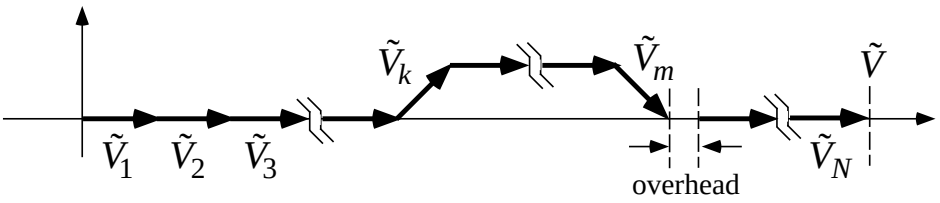

Figure 26. Net no-load voltage in a linac is a sum of phasors from each tube. Phase-closure need not be achieved locally. It does require "overhead" in achievable voltage.

where each klystron is responsible for a voltage-phasor $\tilde{V}_n = V_n e^{j\varphi_n}$ established in the structure(s) it powers. This may be expressed more simply as $V = \Re \tilde{V} e^{j\omega t}$, where

$$\tilde{V} = \sum_{n=1}^{N} \tilde{V}_n .$$

Thus we may express $V = |\tilde{V}| \cos(\omega t + \varphi)$, with $\tilde{V} = |\tilde{V}| e^{j\varphi}$. Evidently, in the absence of beam-induced voltage, and considering only the longitudinal phase-space, the linac is simply a sinusoidal kick. It is quantified by a single-phasor, the no-load voltage of the linac. Considering then an electron bunch, one can see that if one tube is improperly phased, then the center energy and the energy spread deviate from the design after propagating through the structure(s) powered by this tube. However, since the result at the end of the linac is determined simply by a sum of phasors, a later tube may be re-phased to correct the error and achieve "phase-closure", as illustrated in Fig. 26.

Notice that the effect of the phase-error is to reduce the net length of the no-load voltage phasor. This implies that phase-closure requires "overhead" in maximum voltage available to the linac from the tube complement. This may be achieved for example with spare tubes on "stand-by", operating at the machine repetition rate, triggered after the beam has passed, but available to be put on the beam on short notice.

In the meantime, phase-closure is no guarantee that the beam hasn't noticed the mis-phased tube, for the transverse motion depends on energy, and the consequence of mis-phasing is a change to the energy profile along the linac. In fact, to understand the beam dynamics in detail, it is important to appreciate the *local* variation in beam energy as the simple picture of "linac as sine-wave" does not provide sufficient detail for tracking particles through the machine. An electron's local energy variation (absent

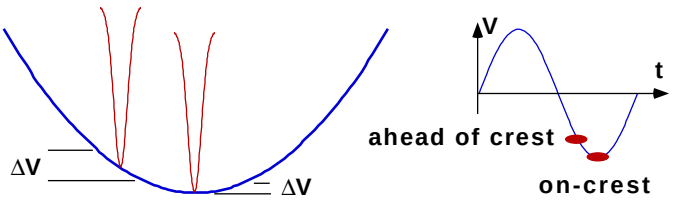

Figure 27. Energy spread of the beam is determined in the first approximation by bunch length and phasing.

wakefields) may be described by

$$\frac{d}{ds}mc^2\gamma = eG\cos(\psi+\phi), \tag{70}$$

where the voltage produced by the local structure has been quantified in terms of local gradient G, and phase ϕ. The electron phase relative to a properly-phased ($\phi = 0$) accelerating wave is $\psi = \beta\, s - \omega\, t$, and varies according to

$$\frac{d\psi}{ds} = \beta - \frac{\omega}{V}. \tag{71}$$

With $V \approx c$ and phase velocity of the accelerating mode $V_\phi = \omega/\beta \approx c$, the phase ψ is constant. The electron is "frozen" longitudinally, unless it is bumped by a mis-phased tube.

This model is helpful in understanding the structure required of a micropulse, the collection of charge residing within one bucket, a single bunch. The longitudinal structure of a single bunch is constrained in the first approximation simply by the sinusoidal shape of the accelerating waveform. The problem is sketched in Fig. 27. Let us suppose for illustration that the initial distribution for the bunch takes the form of Eq. (69), and let us express this in terms of phase,

$$\frac{dQ_b}{d\psi} = \frac{Q_b}{\sqrt{2\pi}\sigma_\psi}\exp\left\{-\frac{(\psi-\bar{\psi})^2}{2\sigma_\psi^2}\right\}.$$

Using this distribution and Eq. (70), and performing integrals one can check the simple rule of thumb relating root-mean-square (rms) phase-width and asymptotic rms energy spread

$$\sigma_\psi\,(^\circ RF) \approx 6.8\sqrt{\sigma_\gamma\,(\%)}. \tag{72}$$

This is just the first approximation to the beam energy distribution, valid at low charge. The actual voltage witnessed by beam particles is a

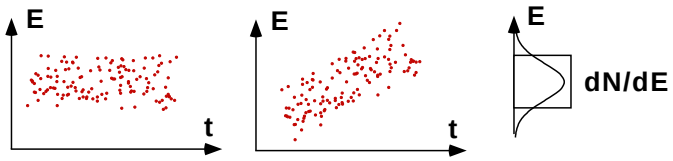

Figure 28. Illustrating projection of the phase-space distribution

superposition of the applied (or "no-load") voltage, and an effective voltage due to collective radiation, the longitudinal wakefield. The longitudinal structure determines the frequency components present in the beam spectrum, and thus the character of coherent radiation in the linac. Where single-bunch charge is high and bunch length is short, the wakefield results in a significant loss in mean gradient, of order a few percent. At the same time, due to wakefields, phasing ahead of crest can actually *reduce* energy spread, below that of Eq. (72). Long and short-range wakefields are important in determining the energy distribution of the beam, be it a single bunch, or a train of bunches. Energy spread is not all bad, by the way, provided it is erased before the high-energy experiment can see it. Temporary use of energy-spread, along some region of the linac, may help to inhibit beam break-up ("BNS damping").

In connection with energy-spread, transverse oscillations and the like, it is important to distinguish between the distribution in 6-dimensional phase-space, to which Liouville's theorem applies, and *projected* distributions. This is illustrated in Fig. 28, showing a sampling of two distributions, in energy and time. Such a sketch invites one to distinguish between *projected* distributions and the full distribution. As one can see, a projection of the distribution onto the energy-axis (histogram by energy) shows merely energy-spread. In the full distribution, energy-spread may in fact be correlated or uncorrelated with another phase-space dimension. The two samples seen in Fig. 28 have the same "microscopic" density in 2-dimensional phase-space, but this fact is lost in projection. We could add that a distribution in E, t is itself a projection, so the actual phase-space density is not made clear in this figure, either.

This matter of microscopic density in the full-phase space is sometimes referred to in terms of "correlated" and "uncorrelated" beam attributes, energy-spread being just one example. Liouville's theorem tells us that we cannot increase the density in 6-dimensional phase-space. Uncorrelated energy-spread cannot be removed then, in the absence of fluctuation-dissipation effects. However, where a correlation exists, it may be exploited. The phase-closure problem illustrates this practical difference. If uncorrelated energy spread were present on injection, it could not be reduced by

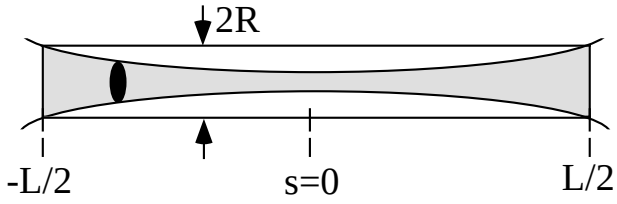

Figure 29. We consider the problem of propagating a beam through a pipe.

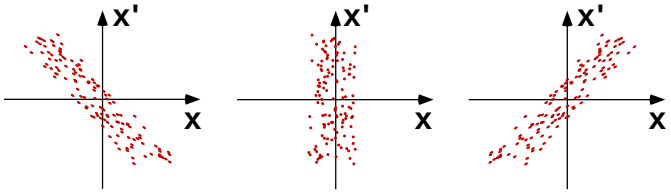

Figure 30. Beam motion in a drift corresponds to shear in trace-space.

re-phasing a tube.

In this section we have glimpsed the requirements on the time-structure of the beam. Let us consider next requirements on the distribution projected in the transverse variables.

Transverse Phase-Space

High rate for an experiment imposes requirements on the transverse phase-space permitted to the beam. However, before considering such things, there is a logical prerequisite: What is required to get the beam through the linac? Let us illustrate the first and most basic ingredient, *emittance*. We pose the problem illustrated in Fig. 29 of transmitting a beam through a pipe.

We expect at first to pull out the Lorentz force law. However, let us suppose there are no externally applied forces. Let us neglect also forces generated by the beam itself. For a relativistic beam, electrostatic and magnetostatic forces cancel to within a factor of $1/\gamma^2$. Radiative forces, wakefields, are present but we may neglect them if charge is sufficiently low. We consider then simply a relativistic, low-intensity, ballistic beam. What determines the behavior of such a beam in a drift?

Since motion is ballistic, particle trajectories are straight lines determined by initial position x_0 and initial angle x_0',

$$x = x_0 + x_0' s, \quad x' = x_0',$$

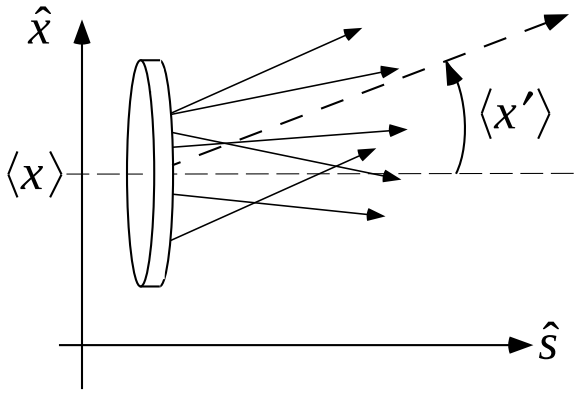

Figure 31. We consider a "slice" of beam.

where s is the displacement down the pipe, and the angle of the trajectory is $x' = dx/ds = p_x/p_s$. One can predict the outcome of the experiment pictured in Fig. 29 with some simple sketches as seen in Fig. 30. There we see a sampling of the beam distribution, projected onto the $x - x'$ plane. This is sometimes referred to as a "trace-space" depiction of the beam, since with it, one may trace out the evolution of the beam. A particle's angle is constant, and the transverse position coordinate varies linearly with displacement s. Asymptotically one can see that the size of the beam will diverge with s, and the beam will scrape the pipe if the drift length L is large enough. Analyzing the geometry of this bundle of rays, we may quantify this divergence.

Let us define the *beam position* in the horizontal plane as $x_b = \langle x \rangle$. The brackets $\langle ... \rangle$ indicate an average over some subset of the beam distribution. Oftentimes this subset will refer to a single bunch. However, what subset we choose depends on the application, and given the variety in the time structure of the beam, we have many choices. In practice one may consult what is measurable, but for now we will refer to this subset simply as a "beam-slice" or just "the beam" for short, with a picture as in Fig. 31. The beam position behaves as a single-particle travelling at angle $x'_b = \langle x' \rangle$. Steered properly one could imagine a single particle making it through a great length of pipe. However, unlike a single particle, the beam has a finite size, one that eventually grows with s. Let's see this. Second moments of the beam are

$$\sigma_x^2 = \left\langle x^2 \right\rangle - \langle x \rangle^2 , \quad \sigma_{x'}^2 = \left\langle x'^2 \right\rangle - \langle x' \rangle^2 , \quad \sigma_{xx'} = \langle x x' \rangle - \langle x' \rangle \langle x \rangle .$$

In our drift, the second moments vary according to

$$\left\langle x^2 \right\rangle = \left\langle x_0^2 \right\rangle + 2\left\langle x_0 x_0' \right\rangle s + \left\langle x_0'^2 \right\rangle s^2, \quad \left\langle x\, x' \right\rangle = \left\langle x_0\, x_0' \right\rangle + \left\langle x_0'^2 \right\rangle s$$

Completing the square in the expression for σ_x^2 one can show that

$$\sigma_x^2(s) = \frac{\varepsilon^2}{\sigma_{x'}^2} + \sigma_{x'}^2 (s - s_*)^2$$

where

$$\varepsilon^2 = \sigma_x^2 \sigma_{x'}^2 - \sigma_{xx'}^2. \tag{73}$$

The quantity ε is referred to as the root-mean-square *emittance*. One can show that it is a constant in such a drift. The beam forms a waist at $s_* = -\sigma_{xx'0}/\sigma_{x'}^2$, and the spot-size there is $\sigma_{x*} = \varepsilon/\sigma_{x'}$. This expression shows that to make a small spot, one must arrange a high angle of convergence $\sigma_{x'}$. Equivalently one may write

$$\sigma_x^2(s) = \sigma_{x*}^2 + \frac{\varepsilon^2}{\sigma_{x*}^2}(s - s_*)^2 = \sigma_{x*}^2 \left\{ 1 + \frac{(s - s_*)^2}{\beta_*^2} \right\}, \tag{74}$$

where $\beta_* = \sigma_{x*}^2/\varepsilon$. This form emphasizes that a small-spot implies rapid divergence, on a length-scale varying quadratically with the minimum spot-size.

This divergent behavior ("hourglass") in a drift is similar to the diffraction of light, also a bundle of rays. Diffraction of coherent light causes the intensity-weighted beam-size to double after a waist in a Rayleigh length $L_R = 4\pi\sigma_x^2/\lambda$, where the rms width may be expressed in terms of the laser waist w as $\sigma_x = w/2$, and w determines the field amplitude variation $\propto exp\,(-r^2/w^2)$. Roughly speaking then, a beam of electrons behaves like a beam of photons, with wavelength $\lambda \approx 4\pi\,\varepsilon$. This might seem odd given that diffraction in the case of photons is sometimes described as a characteristic feature of their wave-nature. In fact, we have seen, one doesn't need waves to see diffraction. The wave nature of photons lies instead in the statement that the photon beam has emittance no less than $\lambda/4\pi$. Similarly, one may infer that the emittance of an electron beam can be no less than $\lambda_D/4\pi$, where λ_D is the de Broglie wavelength. Beams today are orders of magnitude away from this quantum-limited emittance, leaving us plenty of room for advances in beam-physics.

We may translate this analysis of the beam drifting through the pipe, into a statement about the beamline consisting of the pipe and the drift, and what emittance it can "accept". Given that a beam with too large an emittance will be scraped by the pipe, let us determine how to maximize transmission. We suppose that by adjustment of the incoming beam moments, we may place the waist at the pipe center. We adjust the waist σ_{x*}

so as to minimize the beam size at the pipe exit, and thereby minimize scraping. One can show then that maximum beam-size is $\sigma^2 = \varepsilon L$, with L the pipe length. So, for example, if our beam is round, of uniform density with radius $r_b(s)$, then $\sigma = r_b/2$, and scraping occurs if $r_b > R$. Stated differently, scraping occurs if $\varepsilon > A$, where $A = R^2/4L$, with R the pipe radius and L the pipe length L. This is the *acceptance* of a circular pipe for a uniform beam, with waist centered in the pipe.

This notion of acceptance is a useful figure of merit, and may be generalized to more elaborate beamlines. Evidently specification of the beamline for this purpose means a description of the physical aperture as a function of s, and the beam "optics". While in principle one may configure an aperture and optics to "select" an emittance, this is not common on microwave linacs. Intentional scraping appears only in special apertures, "collimators". These may be employed, for example, to define energy or to remove beam "halo". Halo refers to a small population of electrons separated by several σ from the beam "core", and may be produced by scattering upstream. Other than the aperture of the collimator jaws, the smallest aperture on a microwave linac is typically the cell-to-cell coupling iris of the accelerating structure, about $1''$ in diameter at S-Band. Good transmission through this narrow constriction requires magnetic focusing.

Appreciating the significance of emittance, one might like to generalize the notion, to devise a quantity that is invariant and descriptive of limitations on focusing the beam. However, while our beam may be described by some Hamiltonian, there are quite a few we could cook up, and they may generate different invariants. Nevertheless, in the vicinity of the target, or the beam-collision point, the beam still passes through a drift, and its optical behavior still refers to the emittance already defined. The *normalized emittance*, ε_n,

$$\varepsilon_n^2 = (\sigma_x^2 \sigma_{p_x}^2 - \sigma_{x-p_x}^2)/m^2c^2. \tag{75}$$

is a useful figure of merit, as one can show that this quantity is a constant, in the presence of acceleration that is uniform throughout the beam-slice. This one can understand since x and p_x are unchanged by a Lorentz boost. Normalized emittance is also constant for a linear restoring force in x. For a high-energy beam then, and given ε_n, one has $\varepsilon \approx \varepsilon_n/\gamma$ and this is useful for simple calculations. This is to say that if one knows the normalized emittance at the 1 GeV point in the linac, one may hope that the same figure applies at the 45 GeV point, so that the emittance is smaller by $\times 45$. In practice, one will find that normalized emittance grows through the linac. To appreciate how this occurs, one must contemplate a number of features of the beam dynamics: the choice of beam-slice, and the non-linear and energy-dependent ("chromatic") character of the forces at work. While Liouville may assure us that microscopic density in phase-space is constant,

high-energy experiments represent projections of phase-space. As we saw in Fig. 28, Liouville's theorem does not prevent a projected distribution from becoming diluted.

Before we get too far out in phase-space describing beams, let's take a look at what is observable in practice.

3.3. INSTRUMENTATION

As we have seen a beam may be described in theory by a distribution in six-variables, 3 position variables, and 3 momentum variables — a distribution in phase-space. The task of the machine physicist is to produce beams with the phase-space desired by the high-energy experimentalist. This implies a requirement to monitor the relevant attributes of the phase-space distribution.

In fact, many attributes are not monitored, due to technical difficulties, and the problem of beamline instrumentation is a very active area of research. It is quite feasible to monitor intensity and first moments on a pulse-to-pulse basis, transverse position, phase, and energy. Monitoring of second-moments from shot to shot is more of a challenge. Quantities typically available over several machine pulses include transverse size (rms), transverse asymmetry (skew), and energy spread.

To diagnose attributes of a beam-distribution, we must cause the beam to couple to an external circuit. This coupling may be electrostatic, inductive or electromagnetic, and we may classify instruments accordingly. Inductive pickups may be coreless (fast) or employ a high-μ core. The workhorses of beamline instrumentation are typically low-frequency electrostatic or inductive pickups. "Low-frequency" refers to the frequencies actually employed in the output circuit; all pickups exhibit parasitic high-frequency coupling impedance (wakefields) to some degree. Within the category of radiative diagnostics there are coherent and incoherent radiators, including microwave cavities, gaps, and other structures, as well as sources of synchrotron, Cherenkov, transition or other radiation. At frequencies much lower than $1/\sigma_t$ one might employ a microwave detection circuit, and at much higher frequencies a camera, or streak-camera. This classification is not all-inclusive, as there are diverse ways to employ foils, secondary emission, kickers and screens.

Let us illustrate some common instruments, and for definiteness we may think of the beam in terms of the picture of Fig. 25. At 120 Hz, one beam pulse arrives every 8.3 ms. A beam pulse may consist simply of one bunch of charge with a length of 1 mm, and transverse size of 100 μm - 1 mm, or a pulse may consist of a train of such bunches, a train lasting from 100 ns to 3 μs.

Ion-Chamber

The most basic instrument for an accelerator is one that registers *beam loss*. Without such a monitor, symptoms of a mis-steered beam may still be visible, appearing on other instrumentation, that we will discuss, toroids and beam-position monitors. Without instrumentation of some kind, deposition of MW/cm^2 in a beamline component might appear first on a temperature monitor, and perhaps not long after that, a vacuum pressure readout (e.g., an ion pump current). Protection of the machine favors a dedicated diagnostic of beam-loss, and an ion-chamber [27] is the most common instrument for this purpose. Slightly more elaborate is a length of gas-filled coaxial line, as proposed by Panofsky, with a high-voltage suitable for operation as a proportional counter. Stretched the length of the linac such a "Panofsky long ion-chamber" (PLIC), can serve as a monitor of *localized* loss, with resolution on the order of a meter. In addition to machine protection (shut-off or rate-limit), a PLIC can assist in beam-steering. At the SLC there are eight PLIC systems, implemented with 250 V applied to 1/2" Heliax^{TM} (Andrews Corp.), loaded with 95% Ar gas, with Freon and CO_2 [28]. The primary concern in the design of such a system is to avoid shadowing of the cable by beamline components, and this may dictate pairs of cables, or threading of cable through magnet bores where feasible.

Toroid

One step up from a beam-loss monitor is a current monitor, most commonly for microwave linacs, a toroid. As seen in Fig. 32, a simple current monitor can be fashioned from a toroidal ring wrapped with a coil of N turns, coupled out through cable of characteristic impedance R. The azimuthal magnetic field produced by the beam current I_b and the toroid circuit current i is, from Ampere's Law,

$$H = \frac{(I_b - iN)}{2\pi\, r}$$

with r the mean radius of the toroid. The magnetic flux through the toroid circuit is then

$$\phi = \mu\, AN \frac{(I_b - iN)}{2\pi\, r},$$

with A the cross-section of the toroid. Faraday's law then determines the voltage developed across the terminals $V = d\phi/dt$. With output terminated in impedance R, the circuit current satisfies

$$\frac{di}{dt} + \frac{R}{L} i = \frac{1}{N} \frac{dI_b}{dt},$$

where the inductance is $L = \mu\, AN^2/2\pi\, r$. If L/R is large compared to the beam-pulse length then $V \approx I_b R/N$. In this limit the circuit is equivalent to

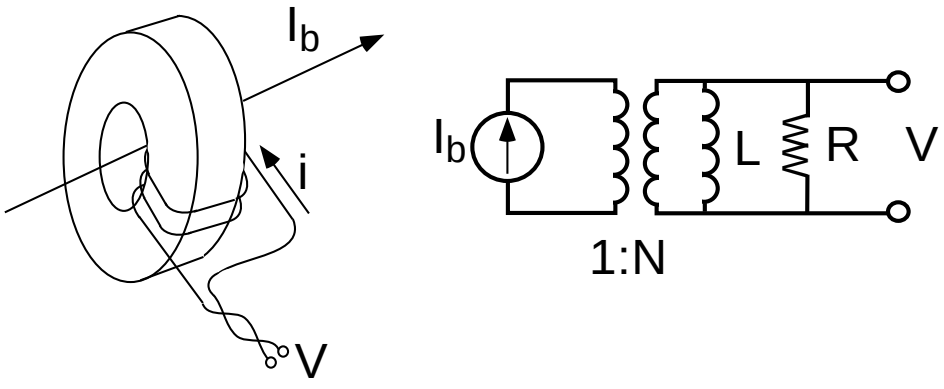

Figure 32. A toroid current monitor and equivalent circuit.

a current-source providing $i \approx I/N$ to the cable. Example implementations may be found in the literature [29, 30].

Screens and Radiative Diagnostics

Screens or "paddles" are among the simplest and most common instruments employed on a linac, usually in combination with a camera. A screen is generically a piece of metal or ceramic with a phosphorescent coating, and a graticule to register scale in the tranverse dimensions x, y. The light from the screen is imaged and employed to infer the transverse beam distribution. As a monitor of position the screen aids in steering, and as a check of beam energy, when employed after a bend magnet. Placing a screen at an image point for a target permits one to tune up the beam on the screen first, before delivery to the target. The drawbacks to screens are that they are destructive to the beam and are eventually destroyed themselves by an intense beam. Operational issues include burn-spots on the screen and saturation of the camera. In addition, beam-jitter on the screen affects the inference of beam-moments. Examples of different screen, graticule and camera configurations can be seen in [31, 32, 33]. Analysis of beam profile data from a screen may be found in [34].

Other radiative diagnostics include transition radiation (e.g., from the front or rear of a screen), synchrotron radiation from a bending magnet or wiggler magnet, and Cherenkov radiation. Depending on the application, diagnosed spectra may extend from the infrared through the ultra-violet and X-ray. Time-resolution, if needed, can be aided by filtering the light. Detection may be through a scintillator and photo-multiplier tube in the case of ionizing radation, or a photo-diode for wavelengths up to 20 μm. A more elaborate system may employ a gated CCD [35] or streak-camera

[36, 37]. The latter permits resolution of features on the scale of a single 3 ps bunch. Time resolution is not always important though. An early application of synchrotron and Cherenkov radiation for monitoring of a beam spot in the mm-cm range can be seen in [38]. At the other extreme one finds modern work employing Compton-scattering of laser-light to measure a 100-nm beam spot [39].

Beam-Position Monitor

There are several different kinds of BPM in common use, button electrode, stripline, and resonant cavity. Let us consider the stripline as it is fairly common; the geometry is seen in Fig. 33.

A transient current (a beam) propagating through a conducting pipe is accompanied by equal and opposite current flowing through the pipe-wall. The distribution of current around the pipe depends however on the beam position relative to the pipe-axis. The stripline BPM cuts the pipe wall and intercepts some of this wall current. Considering one stripline, as seen in Fig. 33, there is a net current $-i$ incident on the upstream end of the stripline. For example, for a centered beam, stripline width w and pipe radius r, this current is just $i = I_b(w/2\pi r)$. As the beam arrives at the upstream end of the BPM, current is induced on the coaxial output, and the interior stripline. If the characteristic impedance of these lines is Z, then the voltage induced on each is $V = Zi/2$, since they appear in parallel. In this way two waves each with current $i/2$ are launched, one up the coax, the other down the interior stripline. As the beam reaches the downstream end of the BPM it induces an opposite polarity pulse on the output coaxial line corresponding to current $-i/2$. If the phase-velocity of the wave on the stripline matches the beam velocity, this signal cancels the forward TEM wave on the downstream coaxial output. A current $-i/2$ is then launched backward on the interior stripline. Thus the voltage appearing on the upstream coaxial output is bipolar in character,

$$V_U(t) = \frac{1}{2} Z \left[i(t) - i\left(t - \frac{2l}{V}\right) \right],$$

with l the stripline length. A matched stripline of this type is sometimes referred to as a directional coupler, since $V_D \approx 0$. This implies that such a device may be employed to monitor two counter-propagating beams in a common pipe.

To determine the position dependence of the wall current $-i$, we may employ the method of images [22], and expand the result as an infinite series

$$i = \frac{w}{2\pi r} I_b \left[1 + 2 \sum_{n=1}^{\infty} \left(\frac{\rho}{r}\right)^n \cos n(\phi - \theta) \right].$$

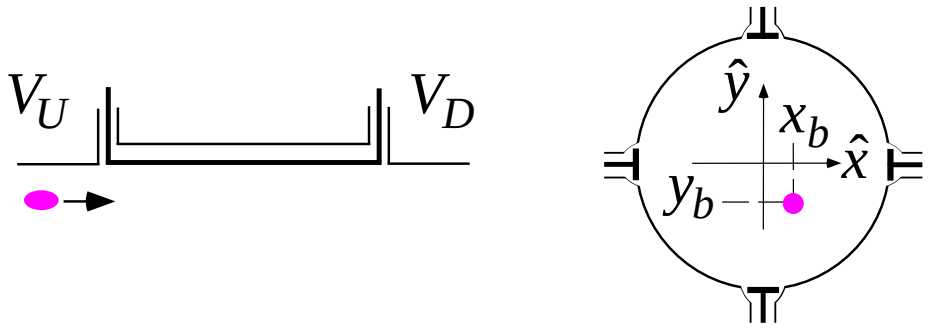

Figure 33. A stripline BPM, side-view of one stripline, and assembly end-view.

where ϕ specifies the position of the stripline, r is the pipe-radius, and the beam-position is given by $x = \rho\cos\theta$, $y = \rho\sin\theta$. The BPM illustrated in Fig. 33 shows four striplines at angles $\phi = 0$, $\pi/2$, π, and $3\pi/2$. We can illustrate the problem of position monitoring more simply however with two striplines at angles $\phi = 0$ and $\phi = \pi$. In this case we have wall currents

$$i_{\phi=0} \approx \frac{w}{2\pi\, r} I_b \left[1 + 2\frac{x}{r} + 2\frac{x^2 - y^2}{r^2} + 2\frac{x^3 - 3xy^2}{r^3} + \ldots\right],$$

$$i_{\phi=\pi} \approx \frac{w}{2\pi\, r} I_b \left[1 - 2\frac{x}{r} + 2\frac{x^2 - y^2}{r^2} - 2\frac{x^3 - 3xy^2}{r^3} + \ldots\right].$$

In processing the signals from the coaxial outputs, one may employ tees to form sum and difference signals

$$\Sigma i = i_{\phi=0} + i_{\phi=\pi}, \Delta i = i_{\phi=0} - i_{\phi=\pi} \tag{76}$$

The sum provides a normalization signal with which to extract position x from the difference. Evidently BPM's exhibit intrinsic non-linearity.

The electronics processing the stripline readouts will perform a low-pass ($\approx$50 MHz) filter on each channel, and add and subtract by means of tees. Following this will be a combination of limiters, amplifiers and digitization. The processor circuit will also exhibit non-linearity, as well as scale and offset errors. One may find too that BPM processor outputs are multiplexed ("MUX'd") for readout from the control system, and this can on occasion cause confusion, since orbits inferred from MUX'd BPM readouts may correspond to different machine pulses, and not a single physical orbit. Further discussion of BPM electronics may be found in [40, 41].

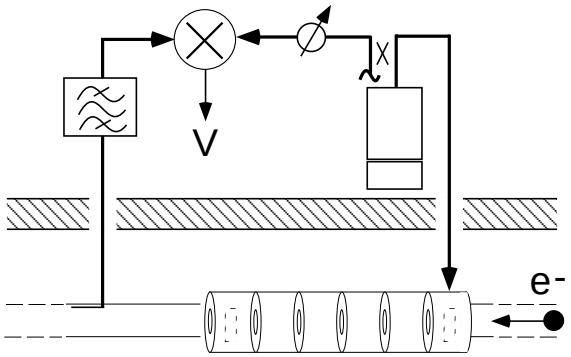

Figure 34. A simple phase-monitor comparing the filtered signal taken from a stripline within the accelerator vault to a signal sampled from the klystron output waveform.

Beam-Phase Monitor

In a microwave linac one typically has a number of phases to control and this implies, to monitor. The injector system defines a reference phase and one wishes to control all tube phases relative to the injector, in principle. Maintenance of accurate phase-relationships in a multi-tube linac requires a single clock, a "master oscillator" from which all rf signals are derived. In addition, one would like this drive signal to be distributed to all tubes in a phase-stable fashion. Meanwhile, to set the phase of a tube, one requires a diagnostic of the tube phase relative to the beam. So for example, a beam traversing an unpowered structure will induce a microwave signal on a forward coupler placed on the structure output to the load. If the tube is phased to put the beam on the accelerating crest, then the beam-induced voltage as seen on the output will be π out of phase with the component due to the tube. This is the basis for the "induction" phasing technique employed for the original Two-Mile linac rf system [2]. Beam-phase relative to a tube can also be assessed with the help of a BPM placed downstream, after a bend magnet. In this case adjustment of an in-line phase-shifter at the tube input will produce a sinusoidal variation of beam-position. On-crest acceleration corresponds to the phase setting where the BPM reading is at an extremum. Thus there are a number of ways to get at the phase-information one requires.

To appreciate the systematic issues associated with a phase-monitor, consider the system seen in Fig. 34, comparing local beam-phase to klystron output phase. This could provide a pulse-to-pulse beam-phase reference, not an absolute figure, but one whose phase offset could be calibrated out by comparison with one of the other techniques. Cable lengths may vary due to temperature drifts associated with the diurnal cycle, or thermal equilibration of the accelerator vault. In this case the phase-monitor output will

drift. In the SLC, the full range of daily temperature variation is 20°C upstairs in the klystron gallery and 1°C downstairs in the tunnel. To monitor for phase-drift one may employ resistive temperature detectors, and to control it one may employ phase-stabilized cable, self-regulating heater tape, and still more elaborate measures [42]. A more conventional solution is to provide for temperature control of electronics buildings. However, even in this case it is possible to "detect" the air-conditioner cycle with the help of the electronics in one's processor. If one actually employs a single stripline signal then the output is weakly dependent on position and this will appear in the phase signal. This can be corrected by combining the stripline outputs. The same principles can be employed to monitor the pulse-to-pulse timing jitter of two bunches [43].

Bunch-Length Monitor

The simplest monitor of bunch length employs a gap in the beam-tube to permit the beam to radiate coherently, as seen in Fig. 35, with gap length L and pipe radius R. The power spectrum radiated will depend on the geometry, and the beam spectrum. For example, for a Gaussian beam, the Fourier transform is

$$\tilde{I}_b(\omega) = \int_{-\infty}^{\infty} \frac{d\omega}{\sqrt{2\pi}} e^{-j\omega t} I_b(t) = \frac{Q_b}{\sqrt{2\pi}} \exp\left(-\tfrac{1}{2}\omega^2\sigma_t^2\right), \qquad (77)$$

Roughly half of the incident spectrum lies at wavelengths below $\lambda \approx 4\pi\sigma_z$, and this determines the frequency range of interest for the detection circuit. For a 1 mm bunch length, one is interested in frequencies in K-band. A crystal detector, looking into a filter blocking out low-frequencies, provides a monitor of power radiated. The crystal should be positioned outside the vault, well away from any pulsed-noise or ionizing radiation sources. One can then extract a signal to monitor bunch length, after normalization by a beam-intensity signal derived from the low-frequency part of the spectrum or a toroid.

In addition to frequency, we require some appreciation of the signal level. In the first approximation, this system is just a gap with capacitance of order $C \approx \varepsilon_0 \pi R^2/L$ and loss-factor $k_l \approx 1/C$. However, for this application one is interested to grasp the frequency dependence of the coupling. In practice one will filter out the low frequency signal to permit clean detection of the high-frequency components. For typical parameters $R, L >> \sigma_z$, and we are interested in frequencies for which the geometry is no longer a lumped circuit, but an obstacle that serves to diffract the beam self-fields out of the pipe. We may estimate the power radiated by considering the beam self-fields in the frequency domain $E_r = Z_0 I/2\pi r$, $H_\phi = I/2\pi r$, and

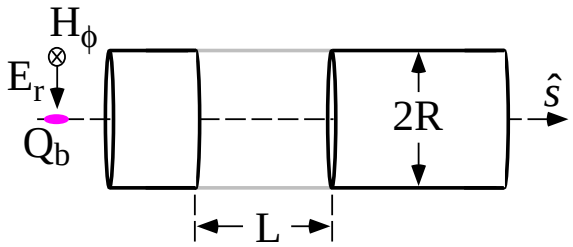

Figure 35. A gap-based monitor makes use of the radiation of a beam through an interruption in the return current-path in the beam-pipe, bridged by a dielectric gap.

the Poynting flux near the pipe wall, $S = E_r H_\phi = Z_0 (I/2\pi R)^2$. We need only compute the fraction of this flux that leaves the pipe, to gauge the signal level to be expected. We treat the dielectric liner as transparent, for simplicity. If we were to compute in detail the fields at the far side of the gap, we would find a superposition of waves with transverse wavenumber $k_\perp$ and longitudinal wavenumber $k_{||}$, satisfying $k_{||}^2 + k_\perp^2 = (\omega/c)^2$. To represent fields diffracting from a radius $r = R - \delta r$, we expect to require transverse wavenumbers $k_\perp \approx \pi/\delta r$. In the high-frequency limit ($\omega R/c >> 1$), all component plane-waves propagate at small angles and the deviation in $k_{||}$ from the unscattered value ω/c is small $\delta k_{||} << k_{||}$. Meanwhile, waves add constructively at the far side of the gap, when they are in phase $\delta\varphi \approx k_\perp \delta r + \delta k_{||} L \approx 0$. Combining these estimates one arrives at a figure for the annulus that leaves the pipe,

$$\delta r \leq \left(\frac{L}{4\pi\, k_{||}} \right)^{1/2}$$

This determines the area intercepted due to diffraction, $A \approx 2\pi R \delta r$. The radiated spectrum is then

$$\frac{dU}{d\omega} \approx 2SA \approx \left(\frac{Z_0}{\pi\, R} \right) \left(\frac{L}{4\pi\, \omega/c} \right)^{1/2} |\tilde{I}_b|^2,$$

and the factor of two accounts for positive and negative frequencies. Considering a flat high-pass filter on the interval $[\omega_c, \infty]$, we integrate to obtain

$$k_{eff}\,(\omega_c) = \frac{1}{Q_b^2} \int_{\omega_c}^{\omega} d\omega \frac{dU}{d\omega} = \frac{\Gamma(1/4)}{8\,\pi^{5/2}} \frac{Z_0 c}{R} \left(\frac{L}{c\sigma_t} \right)^{1/2} f\,(\omega_c \sigma_t)\,. \qquad (78)$$

The function

$$f\,(u) = \frac{1}{\Gamma(1/4)} \int_{u^2}^{\infty} dx\; x^{-3/4} e^{-x},$$

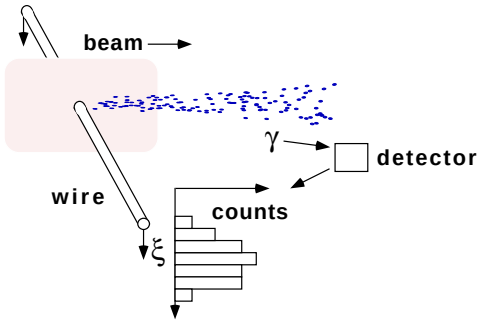

Figure 36. Use of a wire target to infer beam-size, for example, by detection of bremsstrahlung photons with a scintillator, PMT and GADC.

and $f(0) = 1$, with $\Gamma(1/4) \approx 3.6$. For results more precise than this one may employ a numerical field-solver. As an example, for $L \approx 1$ cm, $R \approx 1.27$ cm, and $c\sigma_t \approx 1$ mm, we have $k_{eff}(0) \approx 0.7$ V/pC. Thus a 1 nC bunch deposits about 0.7 μJ, in about 33 ps, or 20 kW peak. Even after dispersion in the connecting guide the peak power will be quite high, and this must be accounted for in the filtering. One may guess that from the point of view of nearby beamline instrumentation in the vault, a gap is also an excellent "noise" generator, unless it is enclosed in a conducting box. Further adventures with bunch length monitoring, and gap-pickups can be found in [43].

While we have computed the energy radiated out of the pipe, in fact, half of the energy is radiated *into* the pipe. Associated with the gap there is a high-frequency wakefield that may interact with particles that follow [7]. This confirms that *any* discontinuity in the beam-pipe will interact with the beam, as claimed in Sec. 2.9. In fact, this *diffraction-model* impedance, evaluated for the accelerating structure iris, is the dominant contributor to the broadband impedance of a linac.

Wire-Scanner

A direct approach to extracting beam-profile information is to move a wire target through the beam path, over the course of several beam pulses. A proportional signal can be derived either from secondary electron emission or bremsstrahlung [44]. In the latter case, at each wire position, the beam produces bremsstrahlung radiation that may be detected with the help of a scintillator and photo-multiplier tube (PMT), and acquired with a gated analog to digital converter (GADC). The scheme is depicted in Fig. 36. The principle of operation is that a histogram of counts versus wire position provides a root-mean-square (rms) width dependent on the beam size. Adopting coordinates u, v in the transverse plane, as seen in Fig. 37, let us

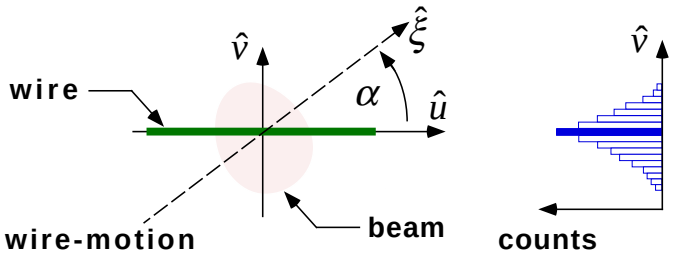

Figure 37. We consider orthogonal coordinates u, v in the transverse plane, with a wire oriented in the u-direction, and travelling along an axis ξ oriented at an angle α to the wire. The beam travels into the page.

denote by $N_e(v)$ the beam distribution in the transverse plane, integrated over a time-scale corresponding to the gate-width for the GADC, and projected onto the v-axis. The number of γ's generated by an infinitely thin wire is proportional to $N_e(v)$. Thus by moving such a wire in ξ, with small steps $\delta\xi$, inferring the step in v, $\delta v = \sin(\alpha)\delta\xi$, and determining $N_\gamma(v)$, one could in principle compute $N_e(v)$ up to an overall scale factor depending on the geometry traversed by the γ's and the detection system. This distribution $N_\gamma(v)$ will have the rms width σ_v of the actual beam distribution in v. The second moment of the inferred distribution referred to a coordinate origin v_w is then $\langle v^2\rangle = \langle (v_b - v_w)^2\rangle + \sigma_v^2$, where the brackets indicate an average over all wire positions weighted by N_γ. A wire of finite dimension may be thought of as a collection of infinitesimal wires each with a different position v_w with respect to the wire-centroid. The inferred distribution $N_\gamma(v)$ in this case is just the sum of the individual distributions from each wire and the rms width Σ_v may be determined from

$$\Sigma_v^2 = \left\langle v^2 \right\rangle - \langle v\rangle^2 = \left\langle (v_b - v_w)^2 \right\rangle + \sigma_v^2 = \sigma_v^2 + \sigma_w^2,$$

where the brackets again refer to an average over all (infinitesimal) wire positions weighted by N_γ. For example, $\sigma_w^2 = D^2/16$ for a wire with circular cross-section of diameter D. Thus a 50-μm diameter wire contributes 12.5-μm in quadrature.

These figures refer to the coordinate v normal to the wire-orientation. Typically several wires are employed and it is convenient to refer moments to a common coordinate system x, y. If the u, v axes are obtained from the x, y axes with a rotation by θ_w, so that $v = -x\sin(\theta_w) + y\cos(\theta_w)$ then

$$\sigma_v^2 = \sigma_x^2 \sin^2\theta_w - \sigma_{xy}\sin 2\theta_w + \sigma_y^2\cos^2\theta_w.$$

Evidently one may infer three second-moments with three wires, provided the wire orientations θ_w are not degenerate, for example, $0°, 45°, 90°$. Fitting of inferred distributions can be quite elaborate in that asymmetric

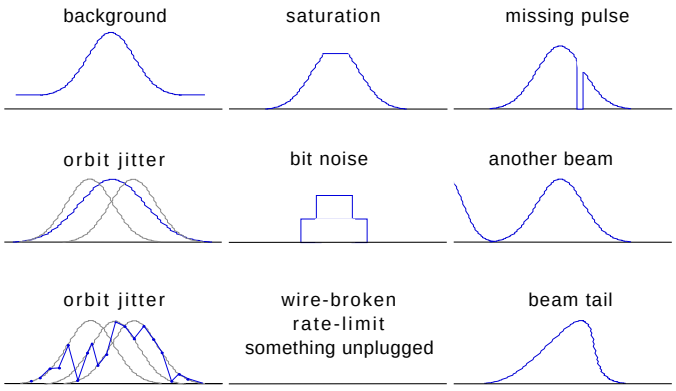

Figure 38. Some of the features one may see with a wire-scan.

profiles reveal beam dynamics significant to the process of machine tune-up.

Systematic issues are illustrated in Fig. 38 and include background from upstream sources of ionizing radiation flux (e.g. beam-scraping and linac "dark current"), wire-breakage, wire-vibration, and beam orbit motion from pulse to pulse. The latter can be corrected with the help of the BPM system and the control software. The last sketch actually may be showing something related to beam dynamics, a beam profile asymmetry or "tail".

Due to problems with breakage, wire-scanners are an ongoing area for research. In the earliest wire-scanners, tungsten wire (gold-plated to permit soldering) was chosen for its high $Z = 74$. More recently, carbon wire has been employed, with detection of electrons at 90° to the beam-axis, rather than small-angle bremsstrahlung photons. The difficulty of maintaining the small diameter wire needed to resolve micron-beam dimensions has motivated development of a "laser-wire" system based on Compton scattering rather than bremsstrahlung.

3.4. MAGNETIC MULTIPOLES

Appreciating what features of the beam are observable, let us consider next the matter of getting a relativistic beam through the main linac. We found in Sec. 3.2 that even a relativistic beam will diverge due to emittance and for this reason some beam-confining forces are necessary. We have already figured out that magnetic forces are preferable to the electric variety, in the case of a relativistic beam. What manner of magnetic forces may we apply?

Following the first rule of beamline design, we put the magnets on the beamline, not *in* the beam. In this case, the applied magnetic induction satisfies $\vec{\nabla} \times \vec{B} = 0$ and $\vec{\nabla} \bullet \vec{B} = 0$, within the beam-path. Thus we have a

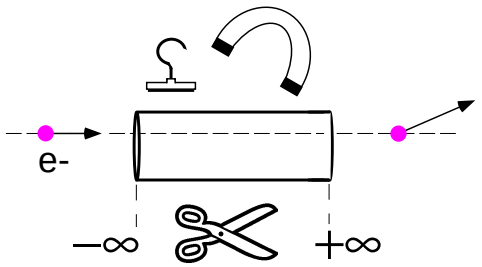

Figure 39. Deflection by *any* "thin magnet" may be decomposed into multipoles.

choice, we may express $\vec{B}$ as a curl or a gradient; we choose gradient,

$$\vec{B} = \vec{\nabla}\varphi. \tag{79}$$

The solenoidal condition then implies that the magnetic potential φ satisfies Laplace's equation, $\nabla^2\varphi = 0$. Boundary conditions then determine the character of the magnetic potential, and they are shaped by the magnetic materials and coil currents outside the beam-pipe, or at least outside the beam-path.

Laplace's equation is a strong constraint on the form the magnetic potential can take. To see this, consider a magnet of some kind as depicted in Fig. 39; let us compute the net kick to an incident electron, with the approximation that the electron orbit is ballistic through the magnet (a "thin" magnet). We integrate to compute the impulse received by the electron,

$$\Delta\vec{p} = -e\int_{-\infty}^{+\infty} dt\ \vec{V}\times\vec{B},$$

or

$$\Delta p_x = -e\int_{-\infty}^{+\infty} dt\ (V_yB_s - V_sB_y) = -e\int_{-\infty}^{+\infty} ds\ \left(\frac{V_y}{V_s}\frac{\partial\varphi}{\partial s} - \frac{\partial\varphi}{\partial y}\right).$$

Since we are calculating the impulse to first order in the fields, V_y may be taken to be constant and this term integrates to 0. Proceeding in a similar fashion with Δp_y, we find that

$$\Delta\vec{p} = -e\hat{s}\times\vec{\nabla}_\perp\psi, \tag{80}$$

where,

$$\vec{\nabla}_\perp = \hat{x}\frac{\partial}{\partial x} + \hat{y}\frac{\partial}{\partial y} = \hat{r}\frac{\partial}{\partial r} + \hat{\phi}\frac{1}{r}\frac{\partial}{\partial\phi},$$

$$\psi\left(\vec{r}_{\perp}\right)=\int_{-\infty}^{+\infty} ds\ \varphi\left(s,\vec{r}_{\perp}\right). \tag{81}$$

Thus the kinds of kicks we may impart to a paraxial beam at first order in the applied magnetic field are described by the function ψ. Notice that,

$$\nabla_{\perp}^{2}\psi\left(\vec{r}_{\perp}\right)=\int_{-\infty}^{+\infty} ds\ \nabla_{\perp}^{2}\varphi\left(s,\vec{r}_{\perp}\right)=\int_{-\infty}^{+\infty} ds\ \left(\nabla^{2}-\frac{\partial^{2}}{\partial s^{2}}\right)\varphi\left(s,\vec{r}_{\perp}\right)$$

$$=\int_{-\infty}^{+\infty} ds\ \left(-\frac{\partial^{2}}{\partial s^{2}}\right)\varphi\left(s,\vec{r}_{\perp}\right)=-B_{s}\left(s,\vec{r}_{\perp}\right)|_{-\infty}^{+\infty}=0.$$

In the region of the beam-pipe then we may express ψ as

$$\psi\left(\vec{r}_{\perp}\right)=\sum_{m=0}^{\infty} r^{m}\left\{a_{m}\cos\left(m\phi\right)+b_{m}\sin\left(m\phi\right)\right\}, \tag{82}$$

or, in Cartesian coordinates,

$$\psi=a_{0}+a_{1}x+b_{1}y+a_{2}(x^{2}-y^{2})+2b_{2}xy+a_{3}(x^{3}-3xy^{2})+b_{3}(3x^{2}y-y^{3})+\ldots.$$

These terms come with nomenclature as indicated in Table 4. Together

TABLE 4. Magnetic multipoles associated with a thin magnet, and the corresponding first-order kicks, per unit charge.

Term	Δp_x	Δp_y	Nomenclature
a_1	0	ea_1	vertical bend
b_1	$-b_1$	0	horizontal bend
a_2	$2a_2y$	$2a_2x$	skew quadrupole
b_2	$-2b_2x$	$2b_2y$	normal quadrupole
a_3	$6a_3xy$	$3a_3(x^2-y^2)$	skew sextupole
b_3	$-3b_3(x^2-y^2)$	$6b_3xy$	normal sextupole

with this mathematical description in terms of multipoles, one may construct a picture for each multipole, based on the corresponding magnetic equipotentials.

There is more variety to magnets than is illustrated by the simple kicks given here. A sufficiently thick or strong magnet can appreciably perturb the beam orbit within the magnet itself. In the case of linacs however one

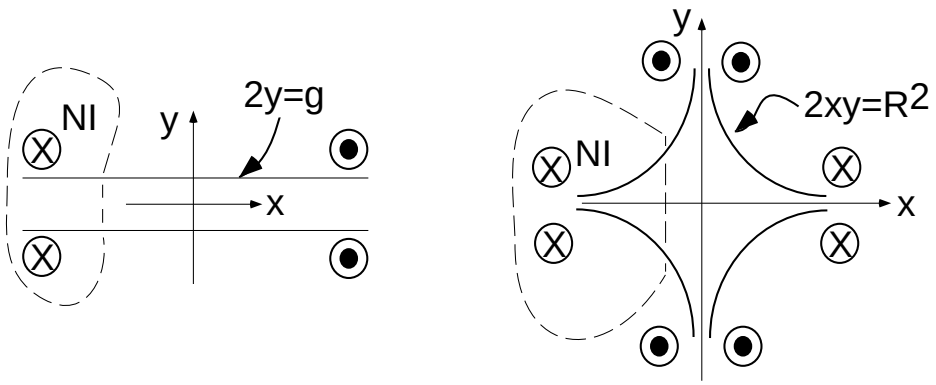

Figure 40. The horizontal bend, and the normal quad are the basic building blocks for optical systems.

is interested simply to transport the beam from the injector to the application (target, collision point) controlling the beam-orbit and the beam-size. Typically one prefers that x and y be uncoupled. For simplicity one prefers that the orbit should lie in a plane. With these assumptions, one is interested primarily in the horizontal bend and the normal quadrupole as in Fig. 40. These two kinds of magnets are the starting point for the subject of beam-optics.

Applying Ampere's Law to the geometries shown one may relate the kick seen by the beam to the coil currents. For the horizontal bend magnet we have $\vec{B} = B_p \hat{y}$, and the pole-tip field $B_p = \mu_0(2NI/g)$, in terms of the gap g, and the product of coil current I, and the number of turns N.

A normal quadrupole field corresponds to $\vec{B} = \vec{\nabla}\{kxy\}$ where $k = \partial B_x/\partial y = \partial B_y/\partial x$ is the quadrupole gradient. This field is defined by an equipotential surface $xy = R^2/2$, a hyperbola, with R the radial distance from the center-axis to the pole-tip. The field at the pole-tip is then $B_p = kR$. We suppose that the magnetic flux is driven by coils carrying current NI placed with the symmetry indicated in Fig. 40. We may compute k in terms of the current and the geometry, using Ampere's Law,

$$-2NI = \oint \vec{H} \bullet d\vec{l} = \int\limits_{pole} \frac{\vec{B}}{\mu} \bullet d\vec{l} + \int\limits_{gap} \frac{\vec{B}}{\mu_0} \bullet d\vec{l}.$$

Since $\mu >> \mu_0$, we need only the integral in the gap and this is

$$\int\limits_{gap} \frac{\vec{B}}{\mu_0} \bullet d\vec{l} = \frac{1}{\mu_0} \int\limits_{-R/\sqrt{2}}^{R/\sqrt{2}} B_y dy = \frac{kR^2}{\mu_0}.$$

Thus we find $k = \mu_0(2NI/R^2)$, expressing quadrupole gradient in terms of the coil current. The validity of this analysis requires that the pole tip field be small enough that the permeability is not driven to saturation. Beyond saturation, the differential response of the material to any further increase in coil current, is that of vacuum; additional magnetic flux ceases to be confined by the magnet material, and the on-axis gradient increases less rapidly with coil current. Moreover the $m = 2$ symmetry is no longer enforced by the pole-tip geometry and aberrations appear. For iron this implies a maximum field of 2 T, and for quadrupoles, a maximum pole-tip field of 1 T. These numbers are important constraints on our ability to confine beams. In practice the "strength" of the lattice is the ultimate limit on tolerable single-bunch charge, due to the effect of wakefields. Further reading on "iron-dominated" magnets may be found in [45].

3.5. MOTION IN A PLANE

To formulate linac optics it would be simplest, and nearly correct to consider only quadrupoles. However, an offset quad introduces a dipole component. In addition, dipoles are present in the form of *correctors*, used to compensate offset quads and to steer the beam. Dipoles are also useful in that they can be employed for energy analysis and longitudinal bunch compression. For these reasons it is useful to formulate a description of the motion incorporating dipoles, and not just quadrupoles. Higher multipoles, particularly sextupoles, are also handy; however, they may be treated as a higher-order correction to the motion governed by dipoles and quads. We assume bend magnets are separated from accelerating structures, so that we may take the electromagnetic fields to be zero in the following analysis.

We define a reference orbit in the machine corresponding to the motion of a particular particle, with position $\vec{r}_0$, velocity $\vec{v}_0$ and momentum $\vec{p}_0$, varying in time. We suppose this motion lies in a plane, and designate by $\hat{y}$ the unit vector normal to the plane. We denote the local tangent to the orbit $\hat{s} = \vec{v}_0/|\vec{v}_0|$, and we define $\hat{x} = \hat{y} \times \hat{s}$. These unit vectors form the basis for a local right-handed coordinate system. We may parameterize the orbit with variable s such that $ds/dt = |\vec{v}_0|$; s is arc-length traversed by the reference particle. Geometry dictates that there is a quantity κ, such that

$$\frac{d\hat{s}}{ds} = -\kappa\hat{x}, \quad \frac{d\hat{x}}{ds} = \kappa\hat{s}, \quad \frac{d\hat{y}}{ds} = 0.$$

κ is the local curvature of the orbit. We may view a length ds of the reference trajectory as a section of circular arc with radius $\rho = 1/\kappa$. Consulting the Lorentz force law,

$$\frac{d\vec{p}_0}{dt} = |\vec{v}_0|\,\frac{d\vec{p}_0}{ds} = |\vec{v}_0|\,\frac{d}{ds}\,|\vec{p}_0|\,\hat{v}_0 = |\vec{v}_0|\,|\vec{p}_0|\,(-\kappa\hat{x}) = -e\vec{v}_0 \times \vec{B}\,(\vec{r}_0)\,,$$

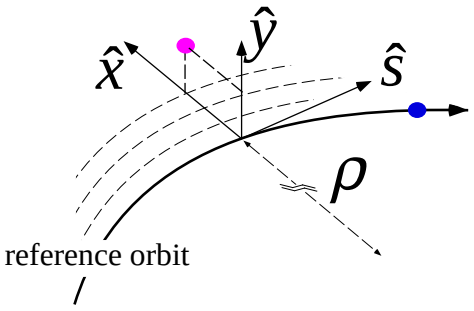

Figure 41. We describe particle motion relative to a reference orbit with local curvature κ.

we see that the assumption of motion in a plane requires $B_x = 0$ on the reference orbit. The curvature is determined by B_y,

$$\kappa = \frac{1}{\rho} = \frac{-eB_y\left(\vec{r}_0\right)}{|\vec{p}_0|}. \tag{83}$$

In practical units, $3.34 p_0(\mathrm{GeV/c}) = B(\mathrm{T})\rho(\mathrm{m})$.

Having determined the reference orbit, we may locate orbits of nearby particles relative to it, parameterizing their motion by s as well. We consider one such particle, finding it at local coordinates $\vec{r}_\perp = x\hat{x} + y\hat{y}$, as depicted in Fig. 41. We describe this particle's motion to linear order in deviations from the reference orbit. Taking into account the rotation of the horizontal axis due to curvature, we express the rate of change of position with reference arc-length s as

$$\frac{d\vec{r}}{ds} = x'\hat{x} + y'\hat{y} + (1+\kappa x)\,\hat{s}.$$

Thus the rate at which length is traversed by the particle is

$$\frac{dS}{ds} = \sqrt{(1+\kappa x)^2 + x'^2 + y'^2} \approx (1+\kappa x)\,.$$

This states that at linear order the particle orbit is tangent to the reference trajectory, and has the appearance of a concentric circular arc, with bending radius $\rho_1 = \rho + x$.

The Lorentz force law then takes the form

$$\frac{d}{ds} p_1 \frac{dx}{ds} \approx \frac{p_1}{\rho_1} + e\left(B_y - y' B_s\right),$$

$$\frac{d}{ds} p_1 \frac{dy}{ds} \approx -e\left(B_x - x' B_s\right),$$

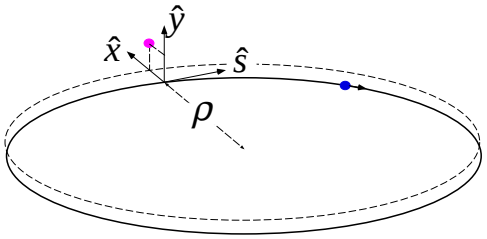

Figure 42. Two circles in a plane intersecting twice. One orbit is "focused" about the other.

and we permit a momentum deviation $p_1 = p_0 (1 + \delta)$.

With the assumption of motion in a plane , we have set $a_1 = 0$, in the notation of Eq. (82). We will make the additional assumption that $a_m = 0$ for all m, considering such skew terms to be small corrections to paraxial motion about the design orbit. This amounts to the assumption of *mid-plane symmetry*, $\varphi(x, y, s) = -\varphi(x, -y, s)$. In this case, on the reference trajectory, $\partial B_y/\partial y = B_s = \partial B_x/\partial x = 0$. Since $\partial B_x/\partial y = \partial B_y/\partial x$, from Ampere's Law, the local magnetic field is characterized by just two quantities, the dipole component B_y and the quadrupole gradient $\partial B_y/\partial x$ evaluated on the reference orbit. Paraxial fields are then

$$B_y(\vec{r}_0 + \vec{r}_\perp) \approx B_y(\vec{r}_0) + \frac{\partial B_y}{\partial x} x, \quad B_x(\vec{r}_0 + \vec{r}_\perp) \approx \frac{\partial B_y}{\partial x} y$$

The equations of motion reduce to

$$x'' + (1 - n)\kappa^2 x = \kappa\delta, \tag{84}$$

$$y'' + n\kappa^2 y = 0, \tag{85}$$

where the *field-index*

$$n = -\frac{1}{\kappa B_y}\frac{\partial B_y}{\partial x},$$

is evaluated on the reference orbit.

The result for zero gradient, $n = 0$, may be understood with reference to Fig. 42, corresponding to constant curvature, i.e., motion in a uniform magnetic field. Where two circles are displaced from each other, one will appear, to be "focused" about the other. If the radii (momenta) are different, there is in addition a constant offset. Meanwhile, there is no way for particles to distinguish the reference plane from another displaced in y. Thus y perturbations are not restored to the reference trajectory.

A modest gradient, $0 < n < 1$, evidently provides focusing in both planes. Excursions in y are met with a horizontal field deflecting particles

toward the axis. Excursions in x see a diminished vertical field, but are still deflected toward the reference orbit, provided the gradient is not too large.

Less evident is the possibility to confine a beam *without* a dipole field, employing only a quadrupole gradient alternating in sign. This concept of "alternating-gradient focusing" was first recognized (and patented) by Christofilos in 1950 [46]. Omission of the dipole field implies in practice a reduction in magnetic field energy, a smaller magnet, and a more compact, high-energy, "strong-focusing" machine. Historically the development of these concepts took place in the context of circular machines. For the linac lattice, we are interested exclusively in alternating gradient focusing.

A general solution for Eq. (84) may be expressed as a superposition of a particular solution of the inhomogeneous equation, and a solution of the associated homogeneous equation. Thus we may express

$$x = x_\beta + D\delta, \quad y = y_\beta. \tag{86}$$

The terms x_β and y_β are referred to as *betatron oscillations*, since they were first investigated for an accelerator of the same name. The term D is referred to as *dispersion* and quantifies the separation of off-momentum rays due to a bend. The betatron oscillations satisfy

$$\frac{d^2 x_\beta}{ds^2} + \hat{K}_x x_\beta = 0, \tag{87}$$

$$\frac{d^2 y_\beta}{ds^2} + \hat{K}_y y_\beta = 0, \tag{88}$$

where the focusing or "K-model" is specified by the reference momentum and the arrangement of dipoles and quadrupoles,

$$\hat{K}_x = -\frac{e}{p_0}\frac{\partial B_y}{\partial x}, \quad \hat{K}_y = \frac{1}{\rho^2} - \hat{K}_x. \tag{89}$$

with ρ the local radius of curvature. The dispersion satisfies

$$\frac{d^2 D}{ds^2} + \hat{K}_x D = \frac{1}{\rho}, \tag{90}$$

Dispersion may be computed directly by fashioning a Green's function from two independent solutions of Eq. (87), and convoluting it with the bending radius. That is to say, after a bend magnet, dispersion behaves as a betatron oscillation would, given the same initial conditions. Thus an appreciation of linear optics requires attention to the betatron motion. This motion is no more involved than what one finds in a normal quadrupole lattice, the problem of primary interest to us, in any case, for the linac.

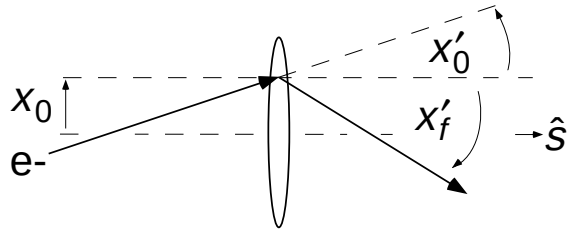

Figure 43. A quadrupole is a lens.

3.6. LINEAR OPTICS

Let us examine beam motion with normal quads. Consider an electron travelling down the beam pipe and encountering a quadrupole field. The motion is depicted in Fig. 43. By convention we refer to a quad that deflects electrons toward the x-axis as a focusing or "F" quad. Notice then that an F quad is defocusing for electrons in the y-plane, and focusing for positrons in the y-plane. Just to the left of the quad the electron horizontal position is x_0, and its angle with respect to the beamline axis is x_0'. Just to the right of the quad its position is still x_0, since this is a thin quad, but its angle has changed to x_f'. If we know the normalized integrated field strength for this quad, at its present current setting, we may compute the kick $\Delta x' = -x_0/f$, where the length f is determined by the integrated field gradient, and the electron momentum,

$$\frac{1}{f} = -\frac{e}{p}\int_{-\infty}^{+\infty} ds\ \frac{\partial B_y}{\partial x}. \tag{91}$$

The motion as a whole is mapped through the element according to,

$$x_f' = x_0' - \frac{x_0}{f}, \quad x_f = x_0$$

The quantity f is the focal length for this quad, at this momentum. Transport through the quad may be expressed in matrix form as

$$\begin{pmatrix} x \\ x' \end{pmatrix}_f = R \begin{pmatrix} x \\ x' \end{pmatrix}_0, \tag{92}$$

where

$$R = \begin{pmatrix} 1 & 0 \\ -1/f & 1 \end{pmatrix}$$

Similarly one may describe a drift of length L by the map,

$$R = \begin{pmatrix} 1 & L \\ 0 & 1 \end{pmatrix}. \tag{93}$$

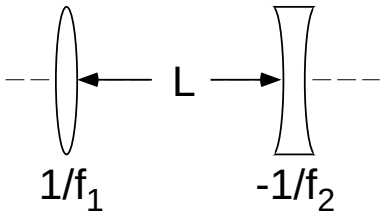

Figure 44. A quad doublet can provide focusing in both planes.

Such "R-matrices" are handy in that one can concatenate them to determine the overall effect of a series of elements. Moreover one can show that any such matrix may be decomposed into an equivalent drift-quad-drift,

$$R = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix} = \begin{pmatrix} 1 & q \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1/f_* & 1 \end{pmatrix} \begin{pmatrix} 1 & p \\ 0 & 1 \end{pmatrix}. \tag{94}$$

Thus any combination of quads and drifts may be characterized by the location of the 1st principal plane forward from the element entrance, p, the effective focal length f_*, and the location of the 2nd principal plane, backward from the element exit, q. For example, it is straightforward to show that the beamline depicted in Fig. 44, consisting of a quad with focal length f_1, followed by a drift of length L, followed by a quad of focal length $-f_2$ may provide *positive* f_*—focusing—in both planes. For this reason such a "doublet" is a handy building block for more elaborate optical systems. The simplest linac lattice is just a periodic arrangement of such doublets, a "FODO" lattice. The principal of such alternating-gradient focusing elements is that the beam envelope may be large in a focusing quad, and small in a defocusing quad, so that the average radial "pressure" provided by the magnetic fields is positive.

In general one is interested to know the beam-behavior throughout the lattice. This would seem to imply knowledge of the map between any two points. However, the most general description of transport is simpler than that. In an arbitrary lattice of quadrupoles and drifts, particle motion satisfies

$$\frac{d}{ds}\gamma\frac{dx}{ds} + K(s)\,x = 0, \tag{95}$$

where quad locations and settings determine $K \approx \gamma\hat{K}_x$, and we suppose $V \approx c$. For a prescribed energy profile $\gamma(s)$ there are two independent solutions to this homogeneous second order ordinary differential equation. Selecting an initial reference location, s_0 on the beamline, we may define a "cosine-like" solution $C(s, s_0)$ with initial conditions $C = 1$ and $C' = 0$, and a "sine-like" solution $S(s, s_0)$ with initial conditions $S = 0$ and $S' = 1$. In

terms of these functions, an arbitrary "betatron" orbit may be represented as $x = x(s_0)C + x'(s_0)S$. The R-matrix for transport from s_0 to s is then

$$R(s_0 \to s) = \begin{pmatrix} C(s,s_0) & S(s,s_0) \\ C'(s,s_0) & S'(s,s_0) \end{pmatrix}. \tag{96}$$

Evidently, the C and S functions provide a complete description of the transport properties of the lattice. They also determine the Green's function with which one may incorporate perturbations to Eq. (95), due to correctors, quad-misalignments, rf kicks as in Eq. (36), and wakefields as in Eq. (42).

A complementary approach employs the "machine functions", or "Twiss parameters". We consider Eq. (95) in the absence of acceleration and look for a solution of the form

$$x = \Re\left(A\,(s)\,e^{j\phi(s)}\right),$$

with A and ϕ real functions, to be determined. Plugging this into the equation of motion we find

$$A'' + \hat{K}A - (\phi')^2 A = 0, \quad (\phi' A^2)' = 0,$$

where the prime denotes the derivative with respect to s. One can check that these equations are solved by

$$A\,(s) = c_1\sqrt{\beta\,(s)}, \quad \phi\,(s) = c_2 + \psi\,(s)$$

for any two constants c_1 and c_2, provided the functions β and ψ satisfy

$$\frac{d^2}{ds^2}\beta^{1/2} + \hat{K}\beta^{1/2} = \frac{1}{\beta^{3/2}}, \tag{97}$$

$$\psi\,(s) = \int_0^s ds' \frac{1}{\beta\,(s')}. \tag{98}$$

The most general solution of the equations of motion may then be expressed as

$$x = \Re\left\{\chi\sqrt{\beta\,(s)}e^{j\psi(s)}\right\},$$

where the phasor $\chi = c_1 \exp(jc_2)$ is a constant. This constant may be expressed, after a bit of algebra as

$$\chi e^{j\psi} = \frac{x}{\sqrt{\beta}}(1 - j\alpha) - jx'\sqrt{\beta},$$

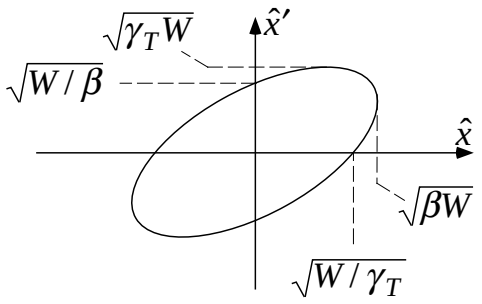

Figure 45. The Courant-Synder invariant defines an ellipse parameterized by β, α and γ_T.

where we introduce

$$\alpha = -\beta'/2. \tag{99}$$

One can go on to show that $\tan c_2 = -\alpha_0 - \beta_0 x_0'/x_0$, and $c_1^2 = W$, where

$$W = |\chi|^2 = \left(\frac{\alpha}{\sqrt{\beta}}x + \sqrt{\beta}x'\right)^2 + \frac{x^2}{\beta} = \gamma_T x^2 + 2\alpha x x' + \beta x'^2, \tag{100}$$

and we abbreviate

$$\gamma_T = \frac{1+\alpha^2}{\beta}. \tag{101}$$

The quantity W is an invariant of the motion, as noted by Courant and Snyder. This all implies that the constants c_1 and c_2 determine the orbit, and are determined by the initial conditions. The optical functions (or Twiss parameters) α and β describe on the other hand, the lattice, independent of a particular particle's initial conditions. The form of the Courant-Synder invariant indicates that a particle orbit in trace-space may be depicted as motion on an s-dependent ellipse as seen in Fig. 45. The optical functions specify the rotation and deformation of such ellipses as a function of s. Particle motion is specified by the action variable, labelling the particular ellipse the particle is on, and an angle variable labelling the location of the particle on the ellipse circumference. One then may picture the beam as a whole in terms of a collection of concentric ellipses, each ellipse with particles distributed around its circumference.

Since lattice transport is, we find, governed by α and β, we should be able to express the cosine-like and sine-like functions in terms of them. In fact,

$$C(s,0) = \sqrt{\frac{\beta(s)}{\beta(0)}}\left(\cos\psi + \alpha(0)\sin\psi\right), \tag{102}$$

$$S\left(s,0\right)=\sqrt{\beta\left(s\right)\beta\left(0\right)}\sin\psi, \tag{103}$$

as one may check by confirming that these expressions satisfy the equation of motion, and the appropriate initial conditions for the C and S functions. With this we may express the R-matrix as

$$R\left(s,0\right)=\begin{pmatrix}\sqrt{\frac{\beta(s)}{\beta(0)}}\left(\cos\psi+\alpha_0\sin\psi\right) & \sqrt{\beta\left(s\right)\beta\left(0\right)}\sin\psi\\ \frac{\left(\alpha_0-\alpha\right)\cos\psi-\left(1+\alpha_0\alpha\right)\sin\psi}{\sqrt{\beta(s)\beta(0)}} & \sqrt{\frac{\beta(0)}{\beta(s)}}\left(\cos\psi-\alpha\sin\psi\right)\end{pmatrix}. \tag{104}$$

One can show also that optical functions themselves may be "transported" according to

$$\begin{pmatrix}\beta\\ \alpha\\ \gamma_T\end{pmatrix}_s=\begin{pmatrix}C^2 & -2SC & S^2\\ -CC & S'C+SC' & -SS'\\ C'^2 & -2S'C' & S'^2\end{pmatrix}\begin{pmatrix}\beta\\ \alpha\\ \gamma_T\end{pmatrix}_0, \tag{105}$$

and for that matter, the beam second moments may be mapped using

$$\begin{pmatrix}\sigma_x^2\\ \sigma_{xx'}\\ \sigma_{x'}^2\end{pmatrix}_s=\begin{pmatrix}C^2 & 2SC & S^2\\ CC & S'C+SC' & SS'\\ C'^2 & 2S'C' & S'^2\end{pmatrix}\begin{pmatrix}\sigma_x^2\\ \sigma_{xx'}\\ \sigma_{x'}^2\end{pmatrix}_0. \tag{106}$$

While the C and S functions are unique, the β function is not. This might seem like a mere mathematical curiosity, in fact it is a physics problem, the problem of *matching*. For example, in a FODO lattice with quad separation L, and quad focal length f, alternating in sign, there is a periodic solution for the beta function with maxima and minima given by,

$$\beta_\pm=2f\left(\frac{1\pm\frac{L}{2f}}{1\mp\frac{L}{2f}}\right)^{1/2}. \tag{107}$$

Phase advance μ per period (length $2L$) is given by $\sin(\mu/2)=L/2f$. One may check this using Eqs. (97) and (98). The existence of this solution, however, does not imply that the beam will be able to find it by itself. To see this, let us suppose that at $s=0$ a beam is incident on our lattice. We follow the beam through the lattice and define at each point s

$$\hat{\beta}=\frac{\sigma_x^2}{\varepsilon},\quad \hat{\gamma}_T=\frac{\sigma_{x'}^2}{\varepsilon},\quad \hat{\alpha}=-\frac{\sigma_{xx'}}{\varepsilon}.$$

One can check that $\hat{\gamma}_T\hat{\beta}-\hat{\alpha}^2=1$. One can show that $\hat{\beta}$, $\hat{\alpha}$ are also solutions for the optical functions, by comparing Eqs. (105) and (106). Thus the beam will also define a choice of β-function, one that depends on the

optics prior to injection. This suggests two questions. What initial conditions should be employed to insure that $\hat{\beta}(s)$ matches the intended $\beta(s)$, the design ellipse? And, is there a particular $\beta(s)$ that is better than others? We would like to design a linac lattice and match into it. The problem is not unlike that of an old television set, also a kind of accelerator. If the picture is fuzzy, one will need to turn some knobs ("matching"), and such tuning is predicated on the assumption that the machine has been designed to provide for a good picture within the range of knob settings ("lattice design"). First we describe matching.

3.7. MATCHING

The beam optical functions $\hat{\beta}$, $\hat{\alpha}$ generate ellipses labelled by the invariant,

$$\hat{W} = \hat{\gamma}_T x^2 + 2\hat{\alpha} x x' + \hat{\beta} x'^2.$$

This is a single-particle quantity, one that depends, through $\hat{\beta}$ and $\hat{\alpha}$, on the initial conditions at injection into the linac, and the linac lattice. After averaging over the beam, we find that

$$\left\langle \hat{W} \right\rangle = \hat{\gamma}_T \sigma_x^2 + 2\hat{\alpha}\sigma_{xx'} + \hat{\beta}\sigma_{x'}^2 = 2\varepsilon.$$

Thus the beam-based Courant-Snyder invariant is determined by the emittance. Let us compare this result to the lattice Courant-Synder invariant, averaged over the beam. With some algebra we see that

$$\langle W \rangle = \gamma_T \sigma_x^2 + 2\alpha\sigma_{xx'} + \beta\sigma_{x'}^2$$

$$= \gamma_T \left(\varepsilon\hat{\beta}\right) + 2\alpha\left(-\varepsilon\alpha\right) + \beta\left(\varepsilon\gamma_T\right) = B_M \left\langle \hat{W} \right\rangle$$

where

$$B_M = \frac{\langle W \rangle}{\left\langle \hat{W} \right\rangle} = \frac{\langle W \rangle}{2\varepsilon} = \frac{1}{2}\left(1+\alpha^2\right)\frac{\hat{\beta}}{\beta} + \frac{1}{2}\left(1+\hat{\alpha}^2\right)\frac{\beta}{\hat{\beta}} - \alpha\hat{\alpha}, \qquad (108)$$

is the "beam magnification" factor. Minimum $B_M = 1$ corresponds to $\hat{\beta} = \beta$, $\hat{\alpha} = \alpha$, and a matched beam.

In practice the problem of matching, of arranging $B_M = 1$, is tied to the problem of emittance inference. Operationally it falls under the heading of "tune-up", preparation of the machine for experiment. There are several techniques employed for emittance inference, and the most common manage to rotate the beam in phase-space for display in coordinate space as seen in Fig. 46. This rotation may be achieved with a drift, by changing a quadrupole setting ("quad-scan"), or by some combination of these.

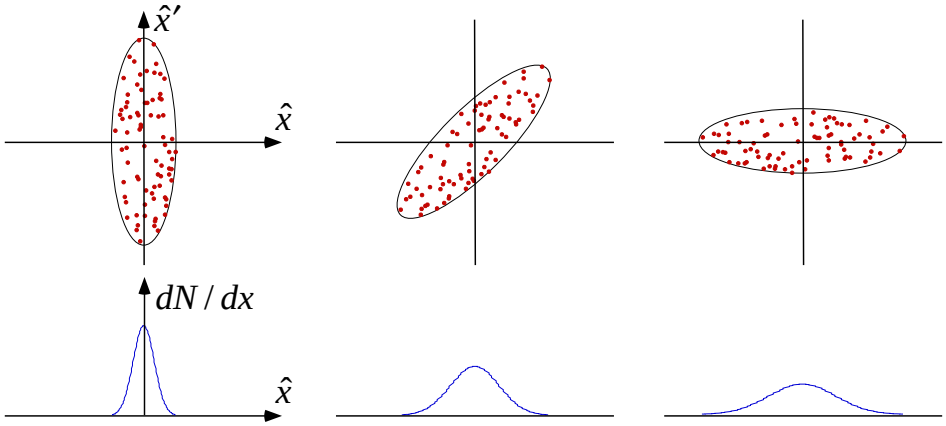

Figure 46. Rotation in phase-space and measurement of beam-size permit inference of emittance.

Measurement of size at three non-degenerate points in a drift is enough to determine ε in principle, assuming no x-y correlations. One can see an example of this technique in [47]. Measurement of size at additional points, and in x and y, should be anticipated as it permits an assessment of errors.

The quad-scan technique is illustrated in Fig. 47. The set-up consists of a thin focusing quad with focal length $1/k$, a drift of length L and a profile monitor (screen or wire-scanner). We are able to scan the quad strength, and measure beam-size at each quad setting, and we ask what this may tell us about emittance, and the beam optical functions. While the paraphernalia of linear optics may be brought to bear on this problem, let us pursue the more elementary approach employed to derive Eq. (74). In terms of beam-moments just after the quad (at $s = 0^+$), the beam-size at the screen may be expressed as

$$\sigma_x^2\left(k, L\right) = \sigma_x^2\left(0^+\right) + 2\sigma_{xx'}\left(0^+\right) L + \sigma_{x'}^2\left(0^+\right) L^2.$$

Beam moments just after the quad may be related to those just before the quad ($s = 0^-$) according to

$$\sigma_x^2\left(0^+\right) = \sigma_x^2\left(0^-\right)$$

$$\sigma_{xx'}\left(0^+\right) = \sigma_{xx'}\left(0^-\right) - k\sigma_x^2\left(0^-\right)$$

$$\sigma_{x'}^2\left(0^+\right) = \sigma_{x'}^2\left(0^-\right) - 2k\sigma_{xx'}\left(0^-\right) + k^2\sigma_x^2\left(0^-\right)$$

Putting these results together one can see then that beam-size at the screen is a quadratic function of quad-strength, with a minimum. Without loss of generality, we may refer quad-strength to the value for minimum spot, so

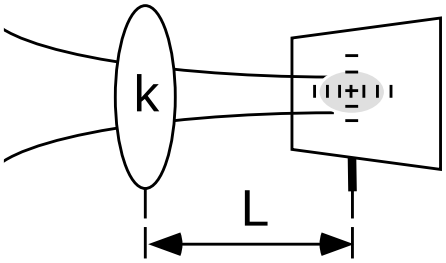

Figure 47. Emittance can be measured, and a beam matched, with the help of a quad-scan.

that $k = 0$ refers to this minimum. This is equivalent to the assumption $\sigma_{xx'}(0^-) = -\sigma_x^2(0^-)/L$. After some algebra one can show that

$$\sigma_x^2(k,L) = L^2\left\{\frac{\varepsilon^2}{\sigma_x^2(0^-)} + k^2\sigma_x^2(0^-)\right\} = \sigma_x^2(0,L) + \frac{k^2L^4\varepsilon^2}{\sigma_x^2(0,L)}.$$

Thus the minimum size in the quad-scan, combined with the curvature of the parabola permit one to infer the emittance. With quad set for the minimum size ($k = 0$), one can determine

$$\hat{\beta} = \frac{\sigma_x^2(0,L)}{\varepsilon}, \quad \hat{\alpha} = -\frac{\hat{\beta}}{L},$$

at the screen location.

We have mentioned a couple of techniques for emittance measurement and matching; however, it is not clear yet why matching might be desirable. For example, for given values for the beam $\hat{\alpha}$ and $\hat{\beta}$ at injection, one could simply define the lattice α and β such that the beam is "matched". Thus the implication of "matching" is really that some choices for α and β are better than others. What are "good" choices of optical functions to be matched to? How to design a linac lattice?

3.8. CHROMATICITY

At the crudest level a lattice should insure a beam-size small enough to avoid current-loss. The more subtle issue in lattice design may be found in the circumstance that beams are not mono-energetic, they include a distribution in momenta, and particles of different momentum are deflected by different angles in a magnetic field. For example, a beam with zero emittance and zero momentum spread may be focused to a point in x by a lattice consisting of a quadrupole and a drift. With momentum spread the spot is blurred. In view of our discussion of the optics of a drift, in Sec.

3.2, we may infer that the beam has developed an rms emittance due to the lattice. The density in phase-space is unchanged, but in projection it is diluted. A "good" lattice should minimize projected emittance despite the momentum-dependence of the optics, "chromaticity".

Recalling $\delta = (p - p_0)/p_0$, the fractional deviation of a particle momentum p from the design momentum p_0, the chromatic properties of the lattice may be quantified in terms of $\beta_\delta = \partial\beta/\partial\delta$. Associated with this "beta-function chromaticity" one has also "phase-advance chromaticity" corresponding to $\psi_\delta = \partial\psi/\partial\delta$. Due to ψ_δ particles with different momenta advance through betatron oscillations at a different rate. Due to the beam collective behavior (wakefields), some amount of phase-advance chromaticity is quite helpful as it inhibits the beam from acting in a coherent manner in driving destructive microwave modes of the linac. The problem for lattice design is to maintain a narrowly bounded β-function for off-momentum particles, i.e., small β_δ.

Thus to control chromatic emittance growth one prefers that the β-function be well-behaved for momenta represented in the beam distribution. Taking the FODO lattice as an example, there is a distinguished choice for the β-function, namely the periodic solution. For the periodic lattice, there is a nearby solution for the β-function for off-momentum particles that is also periodic, and the net dependence of β on momentum at any point is not more than is accumulated over a single period. In the absence of a distinguished symmetry, lattice designers typically proceed by attempting to maintain quasi-periodic machine functions. Beyond the linac, to further diminish chromatic effects, one may employ *chromatic correction* derived from the interaction of dispersion with sextupole fields [49].

3.9. LATTICE DIAGNOSTICS

Having implemented an ingenious lattice design, and carefully matched to it, the next task in commissioning a new lattice is to figure out why the beam won't go through the pipe. One needs methods of diagnosing errors in the lattice. An excellent, broad review of the subject of lattice and other diagnostics can be found in a recent article by Zimmermann [48]. Uncertainties in this problem are of two types: model errors and BPM errors, and we discuss each in turn.

The description thus far, of the magnet system and its effect on the beam, has considered the ideal case, where magnets perform as designed. In fact a "lattice" in practice means a lattice with errors, and they can be sporadic, recurrent and time-consuming to identify as such, and locate. The most common errors are (a) electrical - backwards wiring of a magnet, (b) mechanical - positioning of a magnet (rotated, tilted), (c) sporadic opera-

tional failure - turn-to-turn short in a magnet coil. Item (a) is a particular concern during turn-on after work has been done on the beamline. Item (b) is always present to some degree. Item (c) can occur at any time. As a matter of routine, one may expect to account for hysteresis, the circumstance that a given coil current setting does not correspond uniquely to a particular magnetic flux. Rather the flux depends on the history of the magnet since it was last measured. In practice, one may adopt a procedure for "standardizing" magnets by cycling of the coil currents.

Typically magnets are carefully designed to reduce spurious multipole components, but what components are present, measured and tabulated we needn't consider an error, as they may be incorporated in our model, after generalizing the linear map R, to include higher-order terms, the "T" and "U" coefficients [49].

Mechanical errors are always present to some degree and it is up to the designer to determine, from their analysis of beam dynamics, to what precision mechanical surveying and alignment are required. In practice, beam-based orbit analysis is more accurate in locating quad and BPM centroids than mechanical alignment is capable of achieving or correcting [50]. As to sporadic operational failures, these can be time-consuming in a large accelerator complex. They occur because coil insulation is not impervious to radiation, because the earth moves, because electrical circuits do not function properly when wet, etc. To locate the problem one must examine the beam dynamics, compare observed orbits to those predicted from the model, and by means of sleuth-work narrow down the "bad" region in the lattice to one or a few elements. At that point one can stop the beam, enter the vault and employ a compass to look for wrong-polarity or visually inspect for misplaced magnetic materials. Failing this one may examine the voltage drop across the coils to locate the short, with due safety precautions. If all else fails, one may disassemble the questionable magnets one by one. Accurate lattice diagnosis from "upstairs" is appreciated.

To identify "good" and "bad" regions in a lattice, one analyzes orbit data, looking for discrepancies. To illustrate in the simplest fashion, suppose one observes a betatron oscillation on a collection of beam position monitors numbered $n = 1, 2, ..., N$. Accepting a model for the lattice, embodied in cosine-like and sine-like functions, one chooses launch variables (x_0, x_0') as fit parameters in such a way as to minimize the error in the fit

$$\Delta^2 = \frac{1}{N}\sum_{n=1}^{N}\left\{x_{BPM}\left(n\right) - x_0 C\left(s_n\right) - x_0' S\left(s_n\right)\right\}^2$$

One can show that this is accomplished with the choice,

$$\begin{pmatrix} x_0 \\ x_0' \end{pmatrix} = \frac{1}{\overline{C^2}\,\overline{S^2} - (\overline{SC})^2}\begin{pmatrix} \overline{S^2} & -\overline{SC} \\ -\overline{SC} & \overline{C^2} \end{pmatrix}\begin{pmatrix} \overline{Cx} \\ \overline{Sx} \end{pmatrix},$$

where the vertical bar represents an average over the BPM positions. In practice, pursuing such a procedure, one quickly develops an interest in BPM errors. These include scale, offset, linearity, rotation, cables mislabeled, etc. Thus for model-checking one must go to greater lengths. The weighting may be improved by incorporating known BPM errors and the expected β function. Averaging may be employed to beat down BPM errors. Errors in energy profile lead to errors in inferred quad focal length; thus δ is an important variable and one may want to include dispersion in the orbit-fit. For that matter, to allow for x-y coupling a 5x5 R-matrix may be in order, with fitting for $(x_0, x_0', y_0, y_0', \delta)$. Wakefields too affect the beam orbit, in a current-dependent fashion, with current varying from pulse to pulse. A single bunch passing through a misaligned structure will appear to kick itself. Multiple bunches talk to each other thanks to undamped dipole modes. Each structure provides a kick due to coupler asymmetry. The energy distribution is not merely a function of all klystron phases and amplitudes, but is coupled to bunch-length through the longitudinal wakefield. As the number of variables proliferates, and one realizes that most of them are hidden, the dynamical relations become obscure and one may call on singular-value decomposition to analyze large collections of variables at once. This effort is aided by new and better software and new beamline instruments.

Having produced our high-current, low energy-spread, possibly low-emittance and rock-steady beam at high-energy let us consider how to arrange collisions. There are two kinds of schemes, fixed-target and collider. We consider each in turn, as they are implemented in the Two-Mile Accelerator complex seen in Fig. 48.

3.10. FIXED-TARGET

A generic fixed-target experiment is illustrated in Fig. 49, indicating a polarized electron gun, a diminutive linac, a significant-looking target, and a calorimeter. Not shown are magnets and collimation arranged to perform energy analysis and to remove backgrounds prior to the calorimeter. The energy available for collisions in the center-of-momentum (c.o.m) frame is $\sqrt{s}$, with

$$s = (E_1 + E_2)^2 - c^2 \left(\vec{p}_1 + \vec{p}_2\right)^2 , \tag{109}$$

where E_1 is the beam energy, E_2 is the energy of the target particle and $\vec{p}_i$ are the corresponding initial momenta. For a stationary target particle and a relativistic beam, $s \approx 2E_1E_2$, and c.o.m. energy varies as the square root of the incident beam energy.

To determine the event-rate we suppose the target is larger than the beam and uniform in the region of the beam, presenting a density n_t of

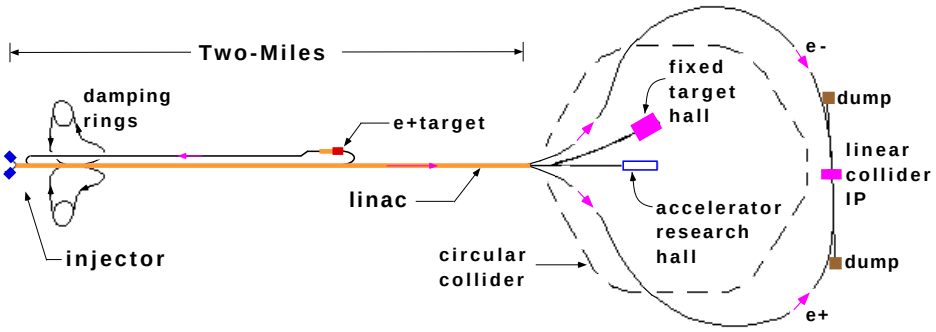

Figure 48. Layout of the Two-Mile Accelerator complex, indicating fixed-target and colliding beam configurations.

scattering centers, distributed over a length l. If there are N_b beam electrons incident per pulse, there are $N_b n_t l \sigma$ events or counts per pulse for a process characterized by cross-section σ for scattering into the detector. With the experiment run at machine repetition frequency f_{rep} we may express the rate as $L\sigma$, where the *luminosity* is

$$L = (N_b f_{rep})\,(n_t l)\,, \tag{110}$$

a product of two factors, one from the accelerator, and one from the target. The total number of events over the course of the run is then $N = TL\sigma$, where T is the time for which both the accelerator, the detector and all sub-systems are "up" and operating together. Luminosity, c.o.m. energy, and polarization circumscribe the capability of an accelerator for fixed-target experimentation. Beamline instrumentation may also be critical. We consider an example.

The E-158 experiment [13] proposes to scatter a polarized electron beam on the atomic electrons of an unpolarized liquid hydrogen target and to measure the asymmetry, $A_{LR} = (\sigma_R - \sigma_L)/(\sigma_R + \sigma_L)$, between the cross-sections for Moller scattering of right and left-handed electrons. The motivation for this is that the cross-section may be calculated precisely in the Standard Model, and thus a precision measurement may be employed to accurately discern *deviations* from the Standard Model. With the scheme depicted in Fig. 49, the left-right asymmetry may be inferred from the difference over many pulses between calorimeter readouts for a left-circularly polarized beam and a right-circularly polarized beam. One merely needs to count, precisely.

The setup consists of a Two-Mile linac, a transport line to the target (the "A-Line"), a liquid hydrogen target 1.5 m in length, and $3''$ in diameter, magnets, and a calorimeter. With each pulse, a 5-kJ "blowtorch" of ejecta

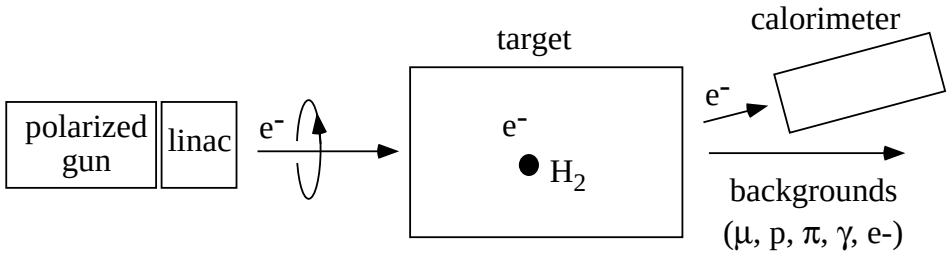

Figure 49. Scheme of a fixed-target experiment.

and disrupted beam will emerge from the target, and pass through a series of magnets that serve to bring scattered electrons with momenta from 12-24 GeV/c to the calorimeter, 60 m downstream. Most of the 500-kW average beam power will pass through the center hole of the annulus formed by the detector, and will dissipate itself in a beam dump. For this experiment $E_2 \approx$0.511 MeV, so that $\sqrt{s} \approx \sqrt{E_1}$, in units of MeV, and for $E_1 \approx 40 - 50$ GeV, $\sqrt{s} \approx$200 MeV. Luminosity is estimated from the target length $l \approx$ 1.5 m, and the number density of electrons in liquid hydrogen, $n_t \approx 4 \times 10^{22}$ cm^{-3}, so that $n_t l \approx 6 \times 10^{24}$ cm^{-2}. At 120 Hz, with $N_b \approx 1 \times 10^{11}$ electrons per pulse, $L \approx 8 \times 10^{37}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. To appreciate this number one must consult the statistical requirements for the experiment. How many Moller-scattered electrons are needed for an interesting experiment?

The theory of electroweak interactions predicts that $A_{LR} \approx 2 \times 10^{-7}$, and the goal of the experiment is to test this prediction at the 10%-level, requiring an uncertainty $\delta A_{LR} \approx 1 \times 10^{-8}$, or less. For a sample of size N, one expects statistical fluctuations with an rms $1/N^{1/2}$. Thus one should expect to make use of a sample at least as large as $N \approx 1/\left(\delta A_{LR}\right)^2 \approx 1 \times 10^{16}$. Meanwhile, the Moller differential cross-section integrated over the angular acceptance of the detector gives a cross-section $\sigma \approx 14\,\mu$barn for a 50 GeV beam, where 1 barn=10^{-24}cm^2. The total number of Moller-scattered electrons expected in up-time $T \approx 1 \times 10^7\,s$ (4 months) is then $N = TL\sigma \approx 10^{16}$, with $N_b \approx 10^{11}$. A more detailed accounting of the statistics, and allowance for 50% down-time, puts the requirement on the pulse intensity at $N_b \approx 4 - 6 \times 10^{11}$.

Notice that none of these considerations consulted emittance. The target is big, and the beam may as well be big too. The high-charge requirement does strain the linac energy budget, and requires creative techniques for beam-loading compensation, to maintain a narrow energy spectrum. Perhaps most challenging though is the high-precision required. If we imagine repeating this experiment many times, our confidence in results inferred on any one experiment would depend on the scatter in A_{LR}. If one requires

this scatter to be as low as 10% of A_{LR}, then with 10^9 machine pulses, one must bound single-pulse fluctuations at the level of 10^{-4} or so. So, to this level of precision, one must be able to bound or identify and tag any non-electroweak effects that could result in fluctuations in the combined probability of Moller electron production and transmission to the detector. Spurious effects could arise in principle, for example, if the laser on the gun jitters transversely in a manner correlated with laser polarization. In this case, the beam on the target might jitter, and so too would the scattered electrons. For this reason the BPM cavity seen in Fig. 12 is essential to the experiment.

From this example we see that for a precision fixed-target experiment, beamline instrumentation is crucial. Some practical limits are set by achievable no-load voltage with beam-loading compensation.

3.11. COLLIDING-BEAMS

To reach the highest energies it is desirable to collide a beam with another beam. For beams of equal energy E, the c.o.m. energy is then simply $\sqrt{s} = 2E$. The drawback in this concept is that luminosity is much lower than in a fixed target experiment. The physics goals for the Two-Mile Accelerator operated in this "SLC/SLD" mode are described in [51].

To determine luminosity, consider two beams with number densities n_1, n_2, interacting with cross-section σ to produce events at instantaneous rate per unit volume $dn/dt = 2\sigma c n_1 n_2$. We model each bunch as a tri-Gaussian,

$$n_k = \frac{N_k}{(2\pi)^{3/2}\,\sigma_{xk}\sigma_{yk}\sigma_{zk}} \exp\left\{-\frac{(x-x_k)^2}{2\sigma_{xk}^2} - \frac{(y-y_k)^2}{2\sigma_{yk}^2} - \frac{(s-V_k t)^2}{2\sigma_z^2}\right\},$$

with $k = 1, 2$ and $V_1 = c, V_2 = -c$. We suppose bunches collide at a rate f_{rep} and arrive at the average rate at which events are produced

$$\dot{N} = 2 f_{rep}\sigma c \int d^3\vec{r}\; dt\, n_1 n_2 = L\sigma.$$

After some integrations, luminosity may be expressed as

$$L = \frac{N_1 N_2 f_{rep}}{2\pi\Sigma_x\Sigma_y} \exp\left\{-\frac{(x_1-x_2)^2}{2\Sigma_x^2} - \frac{(y_1-y_2)^2}{2\Sigma_y^2}\right\}, \tag{111}$$

where

$$\Sigma_x^2 = \sigma_{x1}^2 + \sigma_{x2}^2, \quad \Sigma_y^2 = \sigma_{y1}^2 + \sigma_{y2}^2. \tag{112}$$

For illustration, values typical of the last SLC run are seen in Table 5 [52]. The rms beam sizes, omitting chromatic contributions, correspond to

TABLE 5. IP parameters SLC/SLD'97 [52].

parameter	e-x	e-y	e+x	e+y
ε_n (10^{-5} m-rad)	5.3	1.3	5.1	0.9
β_* (mm)	2.9	1.7	2.2	1.4
θ_* (μrad)	439	269	489	249
$\sqrt{\varepsilon\beta_*}$ (μm)	1.27	0.46	1.08	0.35

$\Sigma_x \approx 1.67\mu\text{m}, \Sigma_y \approx 0.58\mu\text{m}$, or $\Sigma_x \times \Sigma_y \approx 0.97\mu\text{m}^2$. For example, at 120 Hz, with $N \approx 3 \times 10^{10}$, this gives $L \approx 1.1 \times 10^{30}\text{cm}^{-2}\text{s}^{-1}$. On the Z-resonance, with a cross-section of 30 nb, this comes to 200 Z's per hour.

Equation (111) represents just the first approximation to the luminosity. In fact, chromatic effects are important and chromatic correction in the final focus is essential [49]. In addition, when such fine, intense beams collide it is important to account for the effect of beam-fields. Two oppositely-charged counter-propagating beams pinch each other, and this may result in luminosity enhancement, if they have enough time to constrict. Thus the bunch length should be long as long as is consistent with the depth of field, about 1 mm for the parameters of Table 5. This pinch effect accounted for a factor of two in luminosity during the last run. As particle trajectories are deflected by the oncoming beam, in the course of pinching, they will radiate synchrotron photons, or "beamstrahlung", and this is useful as a diagnostic [53]. Luminosity is also affected by slow drifts in machine optical parameters, and pulse-to-pulse orbit jitter, due for example to wakefield amplification of "noise" at the machine front-end. One may appreciate the tolerances on such effects based on the optics of a drift. Waist motion in s at the level of 1/3 of β_* provides a 10% reduction in luminosity. Typical orbit jitter is about 1/3 of the rms beam size, corresponding to a 20% reduction in luminosity. One can intuit that high luminosity requires a large constellation of systems to be functioning as one. Integrated luminosity requires exceptional attention to maintenance, and choreography by a team of expert operators, aided by beamline instruments, software and feedback [54].

4. Epilogue

We have covered a fair bit here, but there's more. A thorough discussion of the microwave aspects of colliders may be found in [55]. The subject of microwave linacs is also now available in textbooks [56, 57]. Beam optics is introduced in numerous texts [58, 59]. Features of collective beam dynamics are introduced in [7]. More specialized topics are easily located in the curricula of the US Particle Accelerator School, the CERN Schools, and the Joint Accelerator School. In between schools, there are many occasions to hear of the latest exploits, at the US Particle Accelerator Conference (PAC), the European PAC, the Asian PAC, and numerous other conferences. If one should tire of schools and conferences, there are many linacs in operation around the world where exciting work is taking place, for collider research, and for fixed-target experiments. And there is one linear collider.

Acknowledgements

These notes were prepared for the proceedings of the NATO Advanced Study Institute on Techniques and Concepts of High-Energy Physics, held in St. Croix, US Virgin Islands, June '98. This is a notoriously tough school, and it was only due to the invitation of Prof. Tom Ferbel that I was able to summon the courage to attend. These notes are the result and comments are welcome; please send them to *whittum@slac.stanford.edu*, unless I have moved to St. Croix.

Thanks to Perry Wilson and Mel Month for their encouragement. This work has benefited from collaboration and discussions with Marc Hill, Yury Kolomensky, Eddie Lin, Dennis Palmer, David Pritzkau, Mike Seidel, Bill Spence and Frank Zimmermann. Work by Jennifer Burney and helpful discussions with Clive Field were instructive. I am indebted to David Pritzkau and Frank Zimmermann for helpful comments on the manuscript. Thanks to Angie Seymour for her support.

Participation in this school was supported in part by U.S. Department of Energy, Contract DE-AC03-76SF00515.

References

1. D.H. Whittum, "Introduction to Electrodynamics for Microwave Linear Accelerators", *Proceedings of the US-CERN-Japan School on RF Engineering for Particle Accelerators* , (KEK, Tsukuba, to be published), SLAC-PUB-7802.
2. R.B. Neal, *The Stanford Two-Mile Accelerator* (W. A. Benjamin, New York, 1968).
3. J.T. Seeman, "The Stanford Linear Collider", *Ann. Rev. Nucl. Part. Sci.* 41 (1991) pp. 389-428.
4. W.W. Hansen "A Type of Electrical Resonator", *J. Appl. Phys.* 9 (1938) pp. 654-663.
5. J.C. Slater, *Microwave Electronics* (D. Van Nostrand, Boston, 1950).
6. W.K.H. Panofsky and W.A. Wenzel, *Rev. Sci. Instrum.* 27 (1956) p. 967.
7. A.W. Chao, *Physics of Collective Beam Instabilities in High Energy Accelerators* (Wiley, New York, 1993).
8. J.C. Nygard and R.F. Post, "Recent advances in high power microwave electron accelerators for physics research", *Nucl. Instrum. Meth.* 11 (1961) pp. 126-135.
9. K.A. Thompson and R.D. Ruth, "Controlling transverse multibunch instabilities in linacs of high-energy colliders", *Phys. Rev. D* 41 (1990) pp. 964-977.
10. J.T. Seeman, *et al.*, "Alignment issues of the SLC linac accelerating structures", *Proceedings of the 1991 Particle Accelerator Conference* (IEEE, New York, 1991), pp. 2949-2951, SLAC-PUB-5439.
11. G.A. Loew, *et al.*, "Linac beam interactions and instabilities", *Proceedings of the VII International Conference on High Energy Accelerators*, (Academy of Sciences, Yerevan, 1970) pp. 229-252.
12. O. Altenmueller and P. Brunet, "Some RF Characteristics of the Beam Phase Reference Cavity" SLAC TN-64-51, Sept. 1964 (unpublished).
13. R. Carr, *et al.*,"A Precision Measurement of the Weak Mixing Angle in Moller Scattering", K.S. Kumar, coordinator, SLAC-Proposal-E-158, 1997 (unpublished).
14. W. Bruns, "GdfidL: A finite difference program for arbitrarily small perturbations in rectangular geometries", *IEEE Trans. Magn.* 32, No. 3 (1996).
15. W.R. Fowkes, *et al.*, "Reduced field TE_{01} X-band travelling-wave window", *Proceedings of the 1995 Particle Accelerator Conference* (IEEE, New York, 1995) pp. 1587-1589, SLAC-PUB-6777.
16. E. Nelson, "Cavity Design Programs", *Proceedings of the US-CERN-Japan School on RF Engineering for Particle Accelerators*, (KEK, Tsukuba, to be published).
17. M. Chodorow, E. L. Ginzton, W. W. Hansen, R. L. Kyhl, R. B. Neal, W. K. H. Panofsky and The Staff, "Stanford High-Energy Linear Electron Accelerator (Mark III)", *Rev. Sci. Instrum.* 26 (1955) pp. 134-204.
18. E.L. Chu, "The Theory of Linear Electron Accelerators", Microwave Laboratory Report No. 140, Stanford University, 1951 (unpublished).
19. E.L. Ginzton, W.W. Hansen, and W.R. Kennedy, "A Linear Electron Accelerator", *Rev. Sci. Instrum.* 19 (1948) pp. 89-108.
20. E.A. Knapp, *et al.*, "Accelerating Structure Research at Los Alamos", *Proceedings of the 1966 Linear Accelerator Conference* (LANL, Los Alamos, 1966) pp. 83-87, LA-3609.
21. R.H. Miller, "Comparison of standing-wave and travelling-wave structures", *1986 Linear Accelerator Conference Proceedings* (SLAC, Stanford, 1986) SLAC-R-303, pp. 200-205.
22. J.D. Jackson, *Classical Electrodynamics* (John Wiley & Sons, New York, 1975).
23. Penn Engineering Components, 12750 Raymer St., North Hollywood, CA, 91605. Tel: 818-503-1511, Fax: 818-764-0195.
24. R.E. Collin, *Foundations for Microwave Engineering* (McGraw-Hill, Singapore, 1966).
25. S.G. Tantawi, *et al.*, "Active High Power RF Pulse Compression Using Optically Switched Resonant Delay Lines", *Advanced Accelerator Concepts*, AIP Proc. 398

(AIP, New York, 1997) pp. 813-821.
26. F. Seitz, "Research on silicon and germanium in World War II", *Physics Today*, Jan. 1995, pp. 22-34.
27. W.R. Leo, *Techniques for Nuclear and Particle Physics Experiments* (Springer-Verlag, Berlin, 1994).
28. J. Rolfe, *et al.*, "Long ion chamber systems for the SLC" *Proceedings of the 1989 Particle Accelerator Conference* (IEEE, New York, 1989) pp. 1531-1533.
29. T. Kobayashi, *et al.*, "Development of a beam current monitor by using an amorphous magnetic core", *Nucl. Instrum. Meth.* B79 (1993) pp. 785-787.
30. C. Nantista and C. Adolphsen, "Beam current monitors in the NLCTA", *Proceedings of the 1997 Particle Accelerator Conference* (IEEE, New York, 1997) pp. 2256-2258, SLAC-PUB-7524.
31. C. Nantista, *et al.*, "Beam profile monitors in the NLCTA", *Proceedings of the 1997 Particle Accelerator Conference* (IEEE, New York, 1998) pp. 2186-2188, SLAC-PUB-7523.
32. S. Yencho and D.R. Walz, "A high-resolution phosphor screen beam profile monitor", *Proceedings of the 1985 Particle Accelerator Conference*, IEEE Trans. Nucl. Sci. NS-32 (1985) pp. 2009-2011.
33. M.C. Ross, *et al.*, "High resolution beam profile monitors in the SLC", *Proceedings of the 1985 Particle Accelerator Conference*, *IEEE Trans. Nucl. Sci.* NS-32 (1985) pp. 2003-2005.
34. J.T. Seeman, F.-J. Decker, I. Hsu, and C. Young, "Characterization and monitoring of transverse beam tails", *Proceedings of the 1991 Particle Accelerator Conference* (IEEE, New York, 1991) pp. 1734-1736.
35. M. Minty, *et al.*, "Using a fast-gated camera for measurements of transverse beam distributions and damping times", *Accelerator Instrumentation: Fourth Annual Workshop*, AIP Conf. Proc. 281 (AIP, New York, 1992) pp. 158-167, SLAC-PUB-5993.
36. J.C. Sheppard, *et al.*, "Real time bunch length measurements in the SLC linac", *Proceedings of the 1985 Particle Accelerator Conference*, *IEEE Trans. Nucl. Sci.* NS-32 (1985) pp. 2006-2008.
37. R.L. Holtzapple, *Longitudinal Dynamics at the Stanford Linear Collider*, Ph.D. Thesis, Stanford University, June 1996.
38. R.W. Coombes and D. Neet, "Beam monitors based on light observation for the beam switchyard of the Stanford Two-Mile Linear Accelerator", *IEEE Trans. Nucl. Sci.*, NS-14 (1967) pp. 1111-1115.
39. T. Shintake, *et al.*, "Design of laser-Compton spot size monitor", *Proceedings of the XVth International Conference on High-Energy Accelerators*, *Int. J. Mod. Phys. A* (Proc. Suppl.) 2A (1993), pp. 215-218.
40. S.R. Smith, "Beam position monitor engineering", *Proceedings of the 7th Beam Instrumentation Workshop*, AIP Conf. Proc. 390 (AIP, New York, 1996), SLAC-PUB-7244.
41. H. Hayano, J.-L. Pellegrin, S. Smith and S. Williams "High resolution BPM for the FFTB", *Nucl. Instrum. Meth.* A320 (1992) pp. 47-52.
42. D. McCormick and M. Ross, "Control of coaxial cable propagation delay for a beam phase monitor", *Accelerator Instrumentation: Fourth Annual Workshop*, AIP Conf. Proc. 281 (AIP, New York, 1992) pp. 256-263, SLAC-PUB-6297.
43. F. Zimmermann, *et al.*, "Bunch-length and beam-timing monitors in the SLC final focus", *Proceedings of the Advanced Accelerator Concepts Workshop* (AIP, to be published).
44. C. Field, D. McCormick, P. Raimondi, and M. Ross "Wire Breakage in SLC Wire Profile Monitors", *Proceedings of the Eight Beam Instrumentation Workshop*, (SLAC, Stanford, to be published) SLAC-PUB-7832.
45. G.E. Fischer, "Iron dominated magnets", *Physics of Particle Accelerators* AIP Conf. Proc. 153 (1987) 1120.

46. A.C. Melissinos, "Nicholas C. Christofilos: His Contributions to Physics", *Proceedings of the CERN Accelerator School: Advanced Accelerator Physics Course*, S. Turner, ed., (CERN, Geneva, 1995) pp. 1067-1081, CERN-95-06.
47. K.D. Jacobs, J.B. Flanz, and T. Russ, "Emittance measurement at the Bates Linac", *Proceedings of the 1989 Particle Accelerator Conference*, (IEEE, New York, 1989) pp. 1526-1528.
48. F. Zimmermann, "Measurement and Correction of Accelerator Optics", *Proceedings of the Joint US-CERN-Japan-Russia School on Beam Measurement*, Montreux, Switzerland, May 11-20, 1998 (CERN, Geneva, to be published), SLAC-PUB-7844.
49. K.L. Brown, "A conceptual design of final focus systems for linear colliders", *Frontiers of Particle Beams*, M. Month and S. Turner, eds. (Springer-Verlag, Berlin, 1988) pp. 481-494.
50. T.L. Lavine, *et al.*, "Beam-based alignment technique for the SLC linac", *Proceedings of the 1989 Particle Accelerator Conference*, (IEEE, New York, 1989) pp. 977-979, SLAC-PUB-4902.
51. The SLD Collaboration, represented by M.J. Fero "First results from SLD with polarized electron beam at SLAC", *The Third Family and the Physics of Flavor*, (SLAC, Stanford, 1992) pp. 341-358, SLAC-PUB-6027.
52. R. Assmann, *et al.*, "Accelerator physics highlights in the 1997/98 SLC run", *Proceedings of the 1st Asian Particle Accelerator Conference* (KEK, to be published).
53. E. Gero, *et al.*, "Beamstrahlung as an optics tuning tool at the SLC IP", Proc. 1989 Particle Accelerator Conference (IEEE, New York, 1989) pp. 1542-1544.
54. T. Himel, "Feedback: Theory and Accelerator Applications", *Ann. Rev. Nucl. Part. Sci.* 47 (1997) pp. 157-192, SLAC-PUB-7398.
55. P.B. Wilson, "High Energy Electron Linacs: Application to Storage Ring RF Systems and Linear Colliders", *Physics of High Energy Particle Accelerators*, AIP Conf. Proc. 87 (AIP, New York, 1982) pp. 452-563, SLAC-PUB-2884.
56. T. Wangler, *RF Linear Accelerators* (John Wiley & Sons, New York, 1998)
57. H. Padamsee, J. Knobloch, and T. Hays, *RF Superconductivity for Accelerators* (Wiley, New York, 1998).
58. H. Wiedemann, *Particle Accelerator Physics*(Springer-Verlag, Berlin, 1993).
59. D.C. Carey, *The Optics of Charged Particle Beams*, (Harwood, Chur, 1987)