

Fields in Waveguides – a Guide for Pedestrians

PETER TENENBAUM
DRAFT June 13, 2003

1 Introduction

The heart of a linear collider is the main linear accelerator, which uses high-power radio-frequency (RF) waves to impart energy to the beam. A quantitative understanding of how the linac works is essential to comprehend the capabilities and limitations of a large linac, and hence a linear collider.

This Note is intended to be a reasonably comprehensive guide to the mysteries of the multi-cell RF cavity (aka “accelerator structure”), and is intended for people who like to be able to see “all” (or at least most) of the ugly math that is typically left as an exercise to the reader in standard textbooks.

1.1 Maxwell’s Equations in MKSA Units

Any discussion of the applications of time-dependent electromagnetic fields must begin with Maxwell’s equations [1]:

$$\begin{aligned}\vec{\nabla} \cdot \vec{D} &= \rho, \\ \vec{\nabla} \cdot \vec{B} &= 0, \\ \vec{\nabla} \times \vec{H} &= \vec{J} + \frac{\partial \vec{D}}{\partial t}, \\ \vec{\nabla} \times \vec{E} &= -\frac{\partial \vec{B}}{\partial t}.\end{aligned}\tag{1}$$

In Equation 1, the four electromagnetic vectors \vec{E} , \vec{D} , \vec{B} , \vec{H} are all present, and MKSA units are assumed. The number of vector quantities can be reduced by replacing \vec{B} with $\mu\vec{H}$ and replacing \vec{D} with $\epsilon\vec{E}$.

1.2 The Wave Equation

Maxwell’s equation can be combined into a wave equation by making use of the vector identity:

$$\vec{\nabla} \times (\vec{\nabla} \times \vec{A}) \equiv \vec{\nabla}(\vec{\nabla} \cdot \vec{A}) - \nabla^2 \vec{A}.\tag{2}$$

Let us apply the identity above to the Maxwell’s electric field curl equation:

$$\begin{aligned}\vec{\nabla} \times (\vec{\nabla} \times \vec{E}) &= \vec{\nabla}(\vec{\nabla} \cdot \vec{E}) - \nabla^2 \vec{E} \\ &= -\vec{\nabla} \times \frac{\partial \vec{B}}{\partial t}.\end{aligned}\tag{3}$$

In a region of space free of charges, $\vec{\nabla} \cdot \vec{E} = 0$. Assuming that we are only interested in well-behaved analytic functions (ie, those for which we can reverse the order of differentiations with impunity), we can transform the magnetic field term in Equation 3 from the curl of a time derivative to the time derivative of a curl:

$$\nabla^2 \vec{E} = \frac{\partial}{\partial t} \vec{\nabla} \times \vec{B}.\tag{4}$$

If we replace \vec{B} with $\mu\vec{H}$, then the RHS of Equation 4 can be replaced with the Maxwell's magnetic curl equation:

$$\begin{aligned}\nabla^2\vec{E} &= \mu\frac{\partial}{\partial t}\vec{\nabla}\times\vec{H} \\ &= \mu\left[\frac{\partial^2\vec{D}}{\partial t^2}+\frac{\partial\vec{J}}{\partial t}\right].\end{aligned}\tag{5}$$

If we assume that the region of interest is also current-free, we can replace \vec{D} with $\epsilon\vec{E}$ to obtain:

$$\nabla^2\vec{E}-\mu\epsilon\frac{\partial^2\vec{E}}{\partial t^2}=0.\tag{6}$$

Equation 6 is a *wave equation* for the electric field. A similar process can be followed to obtain a wave equation for the magnetic field:

$$\nabla^2\vec{H}-\mu\epsilon\frac{\partial^2\vec{H}}{\partial t^2}=0.\tag{7}$$

1.3 Solution to the Wave Equation in Free Space

Equations 6 and 7 can (almost!) be solved by inspection: the solutions will be superpositions of travelling plane waves. Let us define the z axis to be parallel to the direction of propagation. The form of the solution will then be:

$$\begin{aligned}\vec{E} &= \vec{E}_0\exp[i(\omega t-kz)], \\ \vec{H} &= \vec{H}_0\exp[i(\omega t-kz)].\end{aligned}\tag{8}$$

If we consider solutions in free space (ie, no boundaries or boundary conditions), then since space is isotropic and homogeneous \vec{E}_0 and \vec{H}_0 must be constant over all time and all space.

Applying Equation 6 to the suggested solution in Equation 8, we find:

$$\begin{aligned}\nabla^2\vec{E} &= \left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)\vec{E}_0\exp[i(\omega t-kz)] \\ &= \mu\epsilon\frac{\partial^2}{\partial t^2}\vec{E}_0\exp[i(\omega t-kz)].\end{aligned}\tag{9}$$

Because \vec{E}_0 is constant over all time and space, both the laplacian and the time derivative operate only upon the complex exponential. After appropriate cancellation of the constant and the exponential itself, what remains is:

$$-k^2=-\mu\epsilon\omega^2.\tag{10}$$

Equation 10 relates the wave number, $k\equiv 2\pi/\lambda$, to the angular frequency, $\omega\equiv 2\pi\nu$, for any wave which can propagate in free space. In particular, Equation 10 shows that the phase velocity of any such wave, ω/k , will be $1/\sqrt{\mu\epsilon}$, and it can also be easily shown that the group velocity, $\partial\omega/\partial k$, will also be $1/\sqrt{\mu\epsilon}$. In vacuum, it is well known that $1/\sqrt{\mu_0\epsilon_0}=c$. So: the solution is a wave which propagates in the z direction at the speed of light.

A problem appears when Equation 8 is subject to the other constraints of Maxwell's equations. Consider for example the electric divergence equation, which requires that in the absence of electric

charges the divergence of the electric field must vanish. Since the solution in Equation 8 varies only in z , the divergence equation reduces to:

$$\vec{\nabla} \cdot \vec{E} = \frac{\partial E_z}{\partial z} = -ikE_{0,z} \exp[i(\omega t - kz)] = 0. \quad (11)$$

Equation 11 implies that either the longitudinal component of \vec{E} vanishes, or else the momentum vector k vanishes. The latter case corresponds to electrostatic acceleration, which for engineering reasons is unacceptable for final energies of more than a few MeV. The former case corresponds to a purely transverse electric wave, which will accelerate charged particles normal to the direction of wave propagation. To see why this is unacceptable for accelerating anything, consider a particle which is already ultra-relativistic and therefore moving at a speed close to c ; this particle interacts with a wave which propagates in z and has its electric field oriented along x ; the interaction begins at $t = x = z = 0$. At this time, the electric field is $E_{0,x}$, and in a time dt the particle's energy gain is given by the product of the electric field and the distance over which the field is applied, or $\Delta U = E_{0,x}cdt$; the particle is accelerated in the $+x$ direction. One half-period later, the sign of the electric field is reversed and the change in kinetic energy is $\Delta U = -E_{0,x}cdt$. Thus, we see that the particle is alternately accelerated and decelerated in the x direction, and no net energy increase is possible.

If a solution of the form shown above – a travelling-wave solution – is to be used for accelerating particles, it will be necessary to arrange for the electric field parallel to the direction of travel to be nonzero. Equation 11 shows that the problem with the free-space solution is that the field parallel to the direction of travel must be zero to satisfy the electric divergence equation. This constraint can be relaxed by permitting \vec{E}_0 and \vec{H}_0 to be functions of the transverse coordinates. By doing this, the derivatives $\partial E_x/\partial x$ and $\partial E_y/\partial y$ will be nonzero and can be used to balance a nonzero value of $\partial E_z/\partial z$. Arranging for transverse variation in \vec{E}_0 and \vec{H}_0 , in turn, requires that the transverse symmetry of free space be broken by some form of boundary conditions. As a trial, let us consider a conducting circular pipe of inner radius b oriented along the z axis, such that the center of the pipe corresponds to $x = y = 0$. Because the pipe exhibits cylindrical symmetry, we will use cylindrical coordinates (r, θ, z) to explore this solution. Such a pipe is usually referred to as a *waveguide*.

1.4 Solution to the Wave Equation in a Circular Waveguide

To reiterate: we seek a solution to Maxwell's equations which is of the form:

$$\begin{aligned} \vec{E} &= \vec{E}_0 \exp[i(\omega t - kz)], \\ \vec{H} &= \vec{H}_0 \exp[i(\omega t - kz)], \end{aligned} \quad (12)$$

where \vec{E}_0 and \vec{H}_0 are functions of transverse coordinates r and θ , but not of z or t . By limiting ourselves to solutions of this form, we can make a few alterations in the way that z and t partial derivatives are presented:

$$\begin{aligned} \frac{\partial}{\partial z} &= -ik, \quad \frac{\partial^2}{\partial z^2} = -k^2, \\ \frac{\partial}{\partial t} &= i\omega, \quad \frac{\partial^2}{\partial t^2} = -\omega^2. \end{aligned} \quad (13)$$

1.4.1 Boundary Conditions

The solutions must also obey the boundary conditions of a conducting pipe at $r = b$. The boundary conditions for electric and magnetic fields are derived in [2], and will only be qualitatively reviewed here¹: at any boundary between media, the *normal* component of \vec{B} and the *tangential* component of \vec{E} are continuous across the boundary. The normal electric field across the boundary must obey the relations:

$$\begin{aligned}\epsilon_1 E_{1n} - \epsilon_2 E_{2n} &= \Sigma, \\ \sigma_1 E_{1n} - \sigma_2 E_{2n} &= i\omega\Sigma,\end{aligned}\tag{14}$$

where ϵ_1 , ϵ_2 are the permeabilities of the two regions, σ_1 , σ_2 are the conductivities of the two regions (with units of inverse ohms/meter in MKS units), and Σ is the surface charge density at the boundary. If the conductivity of region 1 is zero (vacuum), and that of region 2 is infinite (conductor), then Equation 14 shows that E_{2n} must be zero and E_{1n} must be Σ/ϵ_1 .

Within the conducting material, the magnetic curl equation can be written as:

$$\vec{E}_2 = \frac{1}{\sigma_2 + i\epsilon_2\omega} \vec{\nabla} \times \vec{H}_2,\tag{15}$$

where we have used Ohm's law, $\vec{J} = \sigma\vec{E}$, to replace the current with the electric field. In the limit of infinite conductivity, the electric field within the conductor must be identically zero. Since the tangential component of \vec{E} is continuous across the boundary, and that component is zero within the conductor, it follows that the tangential component of \vec{E} must vanish at the waveguide boundary. Similarly, the electric curl equation can be used to show that \vec{H} vanishes within the conductor, and therefore the normal component of \vec{B} (and hence \vec{H}) is zero at the boundary. In the case of a perfectly-conducting evacuated waveguide, then: the electric and magnetic field both vanish completely within the conductor; the tangential electric field and the normal magnetic field must go to zero at the boundary between vacuum and conductor.

A further boundary condition on the longitudinal magnetic field can also be deduced by considering the form of the curl operator in cylindrical coordinates:

$$\vec{\nabla} \times \vec{A} = \left(\frac{1}{r} \frac{\partial A_z}{\partial \theta} - \frac{\partial A_\theta}{\partial z}, \frac{\partial A_r}{\partial z} - \frac{\partial A_z}{\partial r}, \frac{1}{r} \frac{\partial(rA_\theta)}{\partial r} - \frac{1}{r} \frac{\partial A_r}{\partial \theta} \right).\tag{16}$$

Since $E_\theta \rightarrow 0$ at the boundary, and $\partial\vec{E}/\partial t = i\omega\vec{E}$, it follows that $\partial E_\theta/\partial t \rightarrow 0$ at the boundary, which in turn implies that $(\vec{\nabla} \times \vec{H})_\theta \rightarrow 0$. Equation 16 shows that $(\vec{\nabla} \times \vec{H})_\theta = \partial H_r/\partial z - \partial H_z/\partial r$. Since $\partial H_r/\partial z = -ikH_r$, and $H_r \rightarrow 0$, $\partial H_r/\partial z \rightarrow 0$ on the boundary. Since $\partial H_r/\partial z - \partial H_z/\partial r \rightarrow 0$ and $\partial H_r/\partial z \rightarrow 0$, it follows that $\partial H_z/\partial r \rightarrow 0$ on the boundary.

In summary, the presence of the waveguide requires that E_θ , E_z , H_r , and $\partial H_z/\partial r$ vanish at $r = b$.

1.4.2 Longitudinal Components

We can now solve Equation 6 for the case of the longitudinal components of \vec{E} and \vec{H} . Let us begin with the electric field, and recall that we seek a solution of the form $E_z = E_{0,z}(r, \theta) \exp[i(\omega t - kz)]$. We can rewrite the z component of the wave equation thus:

$$\nabla_\perp^2 E_z - k^2 E_z + \mu\epsilon\omega^2 E_z = 0,\tag{17}$$

¹unless I find some time do write it up here.

where we have defined the transverse component of the laplacian, $\nabla_{\perp}^2 \equiv \nabla^2 - \partial^2/\partial z^2$, and replaced the longitudinal and time derivatives as shown in Equation 13. We can further simplify Equation 17 by defining $k_c^2 \equiv \mu\epsilon\omega^2 - k^2$, and cancelling the common factor of $\exp[i(\omega t - kz)]$ from all terms:

$$\nabla_{\perp}^2 E_{0,z} + k_c^2 E_{0,z} = 0. \quad (18)$$

A solution for Equation 18 can be sought by separation of variables: define $E_{0,z}(r, \theta) \equiv R(r)\Theta(\theta)$, and expand the ∇_{\perp}^2 operator (available in any decent textbook on PDE's or electrodynamics):

$$\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial}{\partial r} R\Theta \right) + \frac{1}{r^2} \left(\frac{\partial^2}{\partial \theta^2} R\Theta \right) + k_c^2 R\Theta = 0. \quad (19)$$

The function R can be pulled out of the θ -derivative, and similarly the function Θ can be pulled out of the r -derivative:

$$\frac{\Theta}{r} \frac{\partial}{\partial r} \left(r \frac{\partial}{\partial r} R \right) + \frac{R}{r^2} \left(\frac{\partial^2}{\partial \theta^2} \Theta \right) + k_c^2 R\Theta = 0. \quad (20)$$

Dividing by $R\Theta$, multiplying by r^2 , and rearranging terms yields:

$$\frac{r}{R} \frac{\partial}{\partial r} \left(r \frac{\partial R}{\partial r} \right) + k_c^2 r^2 + \frac{1}{\Theta} \left(\frac{\partial^2 \Theta}{\partial \theta^2} \right) = 0. \quad (21)$$

Equation 21 shows that the terms with θ dependence and the terms with r dependence have been completely separated. We can advance the solution of Equation 21 by requiring that the two components be equal and opposite, defining separation constant \mathcal{Q} :

$$-\frac{1}{\Theta} \left(\frac{\partial^2 \Theta}{\partial \theta^2} \right) = \frac{r}{R} \frac{\partial}{\partial r} \left(r \frac{\partial R}{\partial r} \right) + k_c^2 r^2 = \mathcal{Q}. \quad (22)$$

The two components of Equation 22 can be solved separately. The first component, which defines $\Theta(\theta)$, can be solved almost by inspection: $\Theta = \cos(n\theta + \theta_n)$, where $\mathcal{Q} = n^2$ and periodicity – the requirement that $\Theta(\theta) = \Theta(\theta + 2\pi)$ – constrains n to be an integer. This constraint can be inserted back into the other component of Equation 18:

$$= \frac{r}{R} \frac{\partial}{\partial r} \left(r \frac{\partial R}{\partial r} \right) + k_c^2 r^2 = n^2. \quad (23)$$

Multiplying through by R and expanding the derivative yields:

$$r^2 R'' + rR' + R(k_c^2 r^2 - n^2) = 0. \quad (24)$$

Equation 24 is *Bessel's equation of order n* ; the solution is a linear combination of *Bessel Functions*

$$R = a_n J_n(k_c r) + d_n Y_n(k_c r). \quad (25)$$

Since $Y_n(0)$ diverges, we can limit ourselves to solutions for which $d_n \equiv 0$, and write down a solution for $E_{0,z}$ in terms of a series in n :

$$E_{0,z} = \sum_{n=0}^{\infty} a_n J_n(k_c r) \cos(n\theta + \theta_n). \quad (26)$$

A similar solution can be derived for $H_{0,z}$.

1.4.3 Applying Boundary Conditions

With a general solution for $E_{0,z}$ and $H_{0,z}$ in hand, we can now apply the necessary boundary conditions. The electric field boundary condition requires that $E_{0,z}$ vanish at $r = b$. This in turn implies that $k_c b = z_{np}$, where z_{np} is the p th zero of J_n . The solution for $E_{0,z}$ is therefore more readily expressed as a double sum:

$$E_{0,z} = \sum_{p=1}^{\infty} \sum_{n=0}^{\infty} a_{np} J_n(k_{c,np} r) \cos(n\theta + \theta_{np}). \quad (27)$$

The constraint on k_c values has an interesting implication on the waves which flow in the waveguide. We can use this constraint in an expansion of k_c :

$$k_{c,np} = \frac{z_{np}}{b} = \sqrt{\mu\epsilon\omega^2 - k^2}. \quad (28)$$

Now consider the case in which $k^2 = 0$, corresponding to infinite wavelength. Equation 28 implies that the wave with infinite wavelength must have a nonzero frequency. We define this frequency to be the *cutoff frequency* of the waveguide:

$$\omega_c = \frac{1}{\sqrt{\mu\epsilon}} \frac{z_{np}}{b}. \quad (29)$$

This allows us to solve Equation 28 for k in terms of ω and ω_c :

$$k^2 = \mu\epsilon (\omega^2 - \omega_c^2). \quad (30)$$

If a wave with a frequency greater than ω_c is introduced into our waveguide, Equation 30 tells us that k will be real-valued, as we require. If, on the other hand, $\omega < \omega_c$, then k will be imaginary. Substituting an imaginary value of k into Equation 12, we find that the solution no longer has the form of an oscillation in z , but rather an exponential decay or growth in z . If we sensibly reject the exponential-growth solution, the implication is that a wave with a frequency below the cutoff frequency decays exponentially in a waveguide. Such a wave is called an *evanescent* wave.

In summary, breaking the transverse symmetry of an electromagnetic plane wave via a waveguide permits a solution in which the longitudinal electric and magnetic fields do not vanish, but at the expense of forbidding waves with excessively low frequencies from propagating in the guide.

Additional complications can be observed if Equation 30 is used to compute the phase and group velocity of a wave in a waveguide. The group velocity, $d\omega/dk$, is given by:

$$v_{\text{gr}} = \frac{d\omega}{dk} = \frac{1}{\sqrt{\mu\epsilon}} \frac{\sqrt{\omega^2 - \omega_c^2}}{\omega}. \quad (31)$$

Since $\sqrt{\omega^2 - \omega_c^2} < \omega$ for frequencies above cutoff, the group velocity is less than the speed of light, and is a function of frequency; asymptotically, as $\omega \rightarrow \infty$, the group velocity approaches the speed of light. The phase velocity, ω/k , is given by:

$$v_{\text{ph}} = \frac{\omega}{k} = \sqrt{\frac{1}{\mu\epsilon} + \frac{\omega_c^2}{k^2}}. \quad (32)$$

The phase velocity, like the group velocity, is a function of frequency. Unlike the group velocity, the phase velocity is greater than the speed of light; indeed, it can be shown that the product of the phase velocity and the group velocity for any such wave is c^2 . This means that the electromagnetic

wave in a regular waveguide is unacceptable for use in accelerating particles despite its longitudinal electric field. This is because the particles in question will have velocities below that of light. As the wave and the particle travel down the waveguide, the accelerating phase of the wave will overtake the particle and the decelerating phase will catch up with the particle; as with the transverse wave described previously, over one oscillation this wave will provide equal acceleration and deceleration, for a net acceleration of zero.

1.4.4 TE and TM Modes

Although electromagnetic waves in a regular waveguide are not suitable for acceleration, they have many other useful characteristics (for example, waveguides can be used to transport waves from a source to a more suitable accelerating structure). Furthermore, we will find that the waves which are ultimately suitable for acceleration share many characteristics with those described above. For this reason, we will complete our study of electromagnetic waves in regular circular waveguide.

By analogy with the longitudinal electric field solution, we can derive a solution to the longitudinal magnetic field, $H_{0,z}$:

$$H_{0,z} = \sum_{v=1}^{\infty} \sum_{u=0}^{\infty} f_{uv} J_u(k_{c,uv}r) \cos(u\theta + \theta_{uv}). \quad (33)$$

Because its boundary conditions are different, the cutoff wave number $k_{c,uv}$ is different from the electric field cutoff. In particular, since it is the radial derivative of the field which vanishes at the boundary, we find for the magnetic field that $k_{uv}b = y_{uv}$, where y_{uv} is the v th zero of J'_u , and $J'_u(\mathcal{R}) \equiv dJ_u(\mathcal{R})/d\mathcal{R}$.

In general, the zeros of J_n are different from the zeros of J'_u . This means that the longitudinal electric and magnetic fields have different cutoff wave numbers, and for a given frequency they will have different phase and group velocities. This leads to the general statement that for a given frequency ω and wave cutoff number k_c , *a single wave cannot have both a longitudinal electric field and a longitudinal magnetic field*. A field with no longitudinal magnetic field is called a “transverse magnetic,” or TM mode; usually the modes are referred to as TM_{np} , where n and p are defined as above. Physically, n can be interpreted to tell the number of nulls in the field pattern as one goes around the azimuth (actually, the number of nulls is $2n$), while p tells the number of nulls encountered radially between $r = 0$ and $r = b$. Similarly, a field with no longitudinal electric field is called a “transverse electric,” or TE_{uv} mode.

1.4.5 Solving for the Fields of the TM_{np} Mode

For a TM_{np} mode, the longitudinal electric field is given by:

$$E_{0,z} = J_n(k_{c,np}r) \cos(n\theta), \quad (34)$$

where we have selected our coordinate system such that $\theta_{np} \equiv 0$. The longitudinal magnetic field, $H_{0,z}$, is known to be zero. It is now possible to solve for the remaining field components of the TM_{np} mode (specifically: $H_{0,r}$, $H_{0,\theta}$, $E_{0,r}$, $E_{0,\theta}$). Four equations are required to solve for these four unknowns.

The θ component of the magnetic curl equation Tells us that $\partial H_{0,r}/\partial z - \partial H_{0,z}/\partial r = \epsilon \partial E_{0,\theta}/\partial t$. Since $H_{0,z} \equiv 0$ and we can replace time and longitudinal derivatives via Equation 13, we can write:

$$kH_{0,r} = -\epsilon\omega E_{0,\theta}. \quad (35)$$

The r component of the magnetic curl equation tells us that $1/r\partial H_{0,z}/\partial\theta - \partial H_{0,\theta}/\partial z = \epsilon\partial E_{0,r}/\partial t$. By similar operations to those in the previous paragraph, we find:

$$kH_{0,\theta} = \epsilon\omega E_{0,r}. \quad (36)$$

The θ component of the electric curl equation tells us that $\partial E_{0,r}/\partial z - \partial E_{0,z}/\partial r = -\mu\partial H_{0,\theta}/\partial t$. We can replace the time and longitudinal derivatives, and can also replace $E_{0,r}$ with $H_{0,\theta}$ via Equation 36 to find:

$$\frac{\partial E_{0,z}}{\partial r} = \frac{ik_c^2}{\omega\epsilon} H_{0,\theta}. \quad (37)$$

The magnetic divergence equation tells us that $\vec{\nabla} \cdot \vec{H}_0 = 0$. Since $H_{0,z}$ is identically zero, we can write:

$$\frac{\partial H_{0,\theta}}{\partial\theta} = -\frac{\partial}{\partial r}(rH_{0,r}). \quad (38)$$

Equation 37 can be used immediately to obtain $H_{0,\theta}$, which in turn allows $H_{0,r}$ to be determined from Equation 38 and $E_{0,r}$ from Equation 36. Finally, Equation 35 relates $E_{0,\theta}$ to $H_{0,r}$. The solution for the fields in a TM_{np} mode are:

$$\begin{aligned} E_{0,z} &= J_n(k_{c,np}r) \cos(n\theta), \\ H_{0,z} &= 0, \\ E_{0,r} &= \frac{-ik}{k_{c,np}} J'_n(k_{c,np}r) \cos(n\theta), \\ E_{0,\theta} &= \frac{ikn}{k_{c,np}^2 r} J_n(k_{c,np}r) \sin(n\theta), \\ H_{0,r} &= \frac{-i\omega\epsilon n}{k_{c,np}^2} J_n(k_{c,np}r) \sin(n\theta), \\ H_{0,\theta} &= \frac{-i\omega\epsilon}{k_{c,np}} J'_n(k_{c,np}r) \cos(n\theta). \end{aligned} \quad (39)$$

A similar painful procedure can be used to derive the fields of a TE mode. We point out that for $n \neq 0$ there are two polarizations possible for each TE or TM mode; the second polarization can be obtained from Equation 39 by replacing sine with cosine and cosine with -sine.

Figure 1 shows the field patterns associated with several of the lower TE and TM modes. Figure 2 shows the relationship between cutoff frequencies for various modes in a circular waveguide. Note that the TE_{11} mode has the lowest cutoff; frequencies which are so low that they can only propagate in the TE_{11} mode are called *single-moded* for this reason.

2 Single-Celled Accelerating Cavities

In 1.3, we saw that in free space waves of any frequency can propagate; that both the phase velocity and the group velocity of such waves are the speed of light; and that such waves have a purely transverse polarization, so that they are not usable for acceleration of charged particles. In 1.4 we saw that in a regular cylindrical waveguide only waves above the cutoff frequency can propagate; that such waves can have a longitudinal electric field component; that the group velocity of such waves are below the speed of light, but that the phase velocity is above the speed of light. Thus, a regular waveguide is also an unacceptable device for acceleration of beams.

The problem with the regular waveguide is that, with a phase velocity exceeding c , the accelerating phase of the wave will overtake the particles which are to be accelerated, and ultimately the

Table 3 Mode Patterns in Circular Waveguide.

Wave Type	TM ₀₁	TM ₀₂	TM ₁₁	TE ₀₁	TE ₁₁
Field distributions in cross-sectional plane, at plane of maximum transverse fields			Distributions below along this plane		Distributions below along this plane
Field distributions along guide					
Field components present	E_z, E_r, H_ϕ	E_z, E_r, H_ϕ	$E_z, E_r, E_\phi, H_r, H_\phi$	H_z, H_r, E_ϕ	$H_z, H_r, H_\phi, E_r, E_\phi$

7-98
8355A213

Figure 1: Field patterns for several TE and TM modes in circular waveguide.

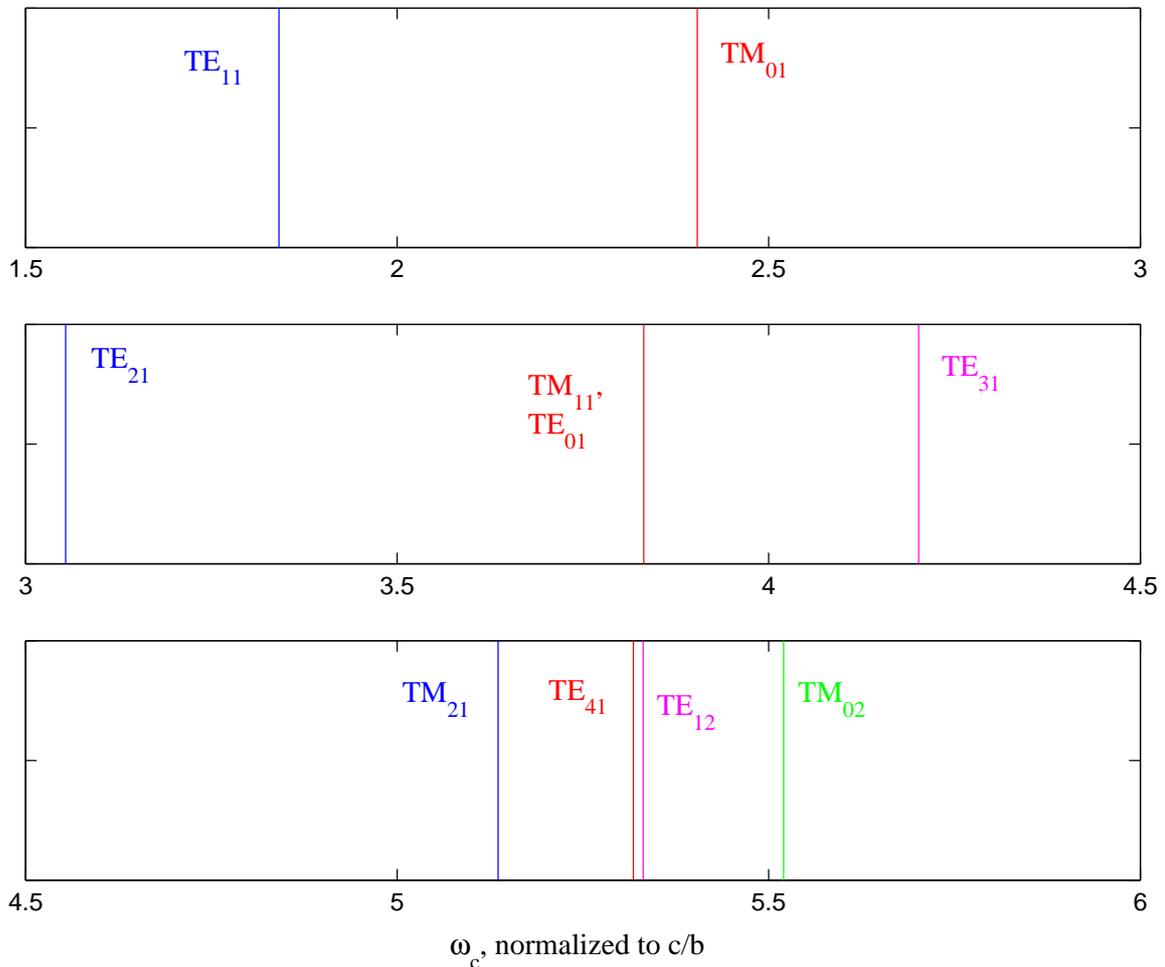


Figure 2: Cutoff frequencies ω_c normalized to c/b for the lowest 10 modes in circular waveguide.

decelerating phase will overlap those particles. One solution, then is to consider using the accelerating phase to accelerate the particles, and then to separate the wave from the particles before the decelerating phase can interact with them. This, in turn means applying a boundary condition to the only “free” axis present in the problem – the z axis.

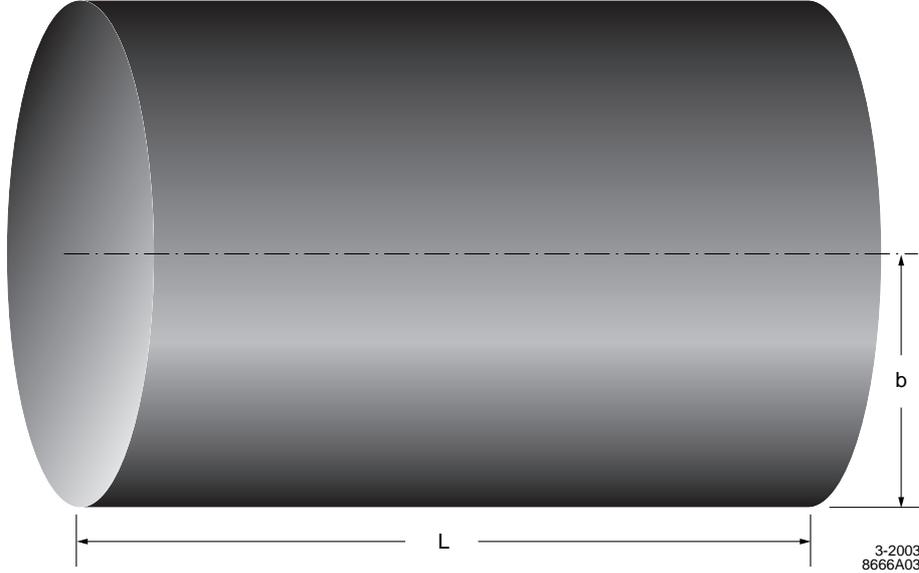


Figure 3: Perfectly-conducting right-circular cylinder with radius b and length L .

Consider a perfectly conducting right-circular cylinder of radius b and length L . It is immediately clear that any solution to Maxwell’s equations must satisfy the boundary conditions for a regular circular waveguide of radius b . In addition, the conducting ends (“endcaps”) of the cylinder imply additional boundary conditions:

- At $z = 0$ and $z = L$, the transverse electric field components, E_r and E_θ , must go to zero, since the electric field must be normal to any conducting boundary.
- At $z = 0$ and $z = L$, the longitudinal magnetic field component must go to zero, since the magnetic field must be tangential to any conducting boundary.
- Since the transverse components of \vec{E} are identically zero on the endcaps, the transverse components of $\vec{\nabla} \times \vec{H}$ must also be zero there. We also know that H_z is identically zero at all of these locations. From Equation 16, therefore, we deduce that $\partial H_\theta / \partial z$ and $\partial H_r / \partial z$ must be zero at the endcaps.

What sort of solution will satisfy all of these requirements? Let us consider first a solution which is based on the TM_{01} mode: from inspection of Equation 39, we see that H_z , E_θ , and H_r are identically zero for this mode, so already we have satisfied the E_θ , H_z , and $\partial H_r / \partial z$ requirements on the endcaps, and of course the TM_{01} mode automatically satisfies the boundary conditions on the barrel.

Suppose that we now superimpose two TM_{01} modes within our cavity: one which is rightward-propagating, with $k \geq 0$, and another which is leftward-propagating, with a k value equal and opposite to the first. The resulting values of E_z , E_r , and H_θ become:

$$E_z = J_0(k_{c,01}r) \cos(kz) \exp(i\omega t), \quad (40)$$

$$\begin{aligned}
E_r &= \frac{-k}{k_{c,01}} J_0'(k_{c,01}r) \sin(kz) \exp(i\omega t), \\
H_\theta &= \frac{-i\omega\epsilon}{k_{c,01}} J_0'(k_{c,01}r) \cos(kz) \exp(i\omega t).
\end{aligned}$$

There are a few things worth noting about our solution, Equation 40. The first is that Equations 35 through 38 no longer apply – those relations were valid only in the case of a single mode, and in this case we have superimposed two modes. The second point is that at $z = 0$, the fields in Equation 40 automatically satisfy the conducting boundary requirements: $E_r = \partial H_\theta / \partial z = 0$. The only remaining conditions are the conducting boundary conditions at $z = L$. These in turn can be satisfied if $kL = j\pi$, where j is an integer. The solution in Equation 40 becomes a *standing wave*, and in this case is known as a TM_{npj} mode: the first index, n , gives the azimuthal periodicity (full symmetry, in $n = 0$, or one oscillation in $n = 1$ case, etc.); the second index, p , gives the number of field nulls in the radial dimension between $r = 0$ and $r = b$; the third index, j , tells the number of field nulls in the longitudinal dimension between $z = 0$ and $z = L$.

Another feature of the TM_{npj} mode is that, unlike the TM_{np} mode, only certain discrete frequencies can be sustained in the cavity. While the TM_{np} mode will support any frequency above $\omega_c \equiv z_{np}/b\sqrt{\mu\epsilon}$, the TM_{npj} mode will only support TM_{np} modes for which $kL = j\pi$. We can make use of Equations 28 through 30 to find an expression for the allowed frequency of the TM_{npj} mode in a cavity of radius b and length L :

$$\omega_{npj} = \frac{1}{\sqrt{\mu\epsilon}} \sqrt{\left(\frac{z_{np}}{b}\right)^2 + \left(\frac{j\pi}{L}\right)^2}. \quad (41)$$

It would appear that, by applying conducting boundary conditions in the longitudinal degree of freedom, we at last have created a time-dependent electric field which is useful for acceleration of particles. By inspection of Equation 39, we can write a general solution to the time-dependent fields in a TM_{npj} mode:

$$\begin{aligned}
E_z &= J_n(k_{c,np}r) \cos(n\theta) \cos(k_j z) \exp(i\omega t), \\
H_z &= 0, \\
E_r &= \frac{-k_j}{k_{c,np}} J_n'(k_{c,np}r) \cos(n\theta) \sin(k_j z) \exp(i\omega t), \\
E_\theta &= \frac{-k_j n}{k_{c,np}^2 r} J_n(k_{c,np}r) \sin(n\theta) \sin(k_j z) \exp(i\omega t), \\
H_r &= \frac{-i\omega\epsilon n}{k_{c,np}^2 r} J_n(k_{c,np}r) \sin(n\theta) \cos(k_j z) \exp(i\omega t), \\
H_\theta &= \frac{-i\omega\epsilon}{k_{c,np}} J_n'(k_{c,np}r) \cos(n\theta) \cos(k_j z) \exp(i\omega t),
\end{aligned} \quad (42)$$

where $k_j \equiv j\pi/L$. As advertised, the E_r and E_θ components are zero at the endcaps; H_z is zero everywhere, including at the endcaps; and the partial derivatives of H_r and H_θ in the z direction are also zero.

Note that the series expansion of $J_n(\mathcal{R})$:

$$J_n(\mathcal{R}) = \sum_{m=0}^{\infty} \frac{(-1)^m (\mathcal{R}/2)^{2m+n}}{m!(m+n)!} \quad (43)$$

indicates that only the $n = 0$ modes will permit a longitudinal electric field on the axis of the cavity. This makes the $n = 0$ modes the preferred ones for actual acceleration.

2.1 Transit Time Effect

Let us consider once again the (unacceptable) TM_{0p} mode. We can express the electric field on axis as simply:

$$E_{z,0p} = E_0 \exp[i(\omega t - kz)], \quad (44)$$

where ω and k are understood to be non-negative. What is the energy gain received by the beam when it interacts with this wave over a distance L ? If the beam is moving at the speed of light, then we can write $t = t_0 + z/c$, and express the energy gain $e dV$ in a distance dz ,

$$e dV = e E_{z,0p} dz = e E_0 \exp[i\omega t_0 + i\omega z/c - ikz] dz, \quad (45)$$

and integrate from $z = 0$ to $z = L$:

$$eV = e E_0 \Re \left\{ \exp(i\omega t_0) \int_0^L dz \exp[i(\omega/c - k)z] \right\}, \quad (46)$$

where we have explicitly required that only the real portion of our heretofore complex quantities can effect any meaningful changes on particles in the real world. Equation 46 can be evaluated:

$$eV = e E_0 \Re \left\{ \exp(i\omega t_0) \frac{\exp[iL(\omega/c - k)] - 1}{i(\omega/c - k)} \right\}. \quad (47)$$

The quantity $\psi \equiv L(\omega/c - k)$ is known as the *transit angle*, and represents the amount the RF phase varies during the passage of the particles through it. We can rewrite Equation 47:

$$eV = e E_0 \Re \left\{ \exp(i\omega t_0) \frac{\exp(i\psi) - 1}{i\psi/L} \right\}. \quad (48)$$

Equation 48 can be written in an even-more useful form by: first, converting the $1/L$ in the denominator to a factor of L outside the brackets; second, factoring $\exp i\psi/2$ out of the numerator; and third, recognizing that the resulting expression has a startling resemblance to the expression for $\sin \psi/2$:

$$\begin{aligned} eV &= e E_0 L T \Re \{ \exp[i(\omega t_0 + \psi/2)] \}, \text{ where} \\ T &\equiv \frac{\sin(\psi/2)}{\psi/2}. \end{aligned} \quad (49)$$

Equation 49 has a number of fascinating properties. First, it shows that the energy gain of passing particles is maximized when t_0 is selected such that $\omega t_0 + \psi/2$ is zero – in other words, the time-varying component of the electric field should achieve a maximum when the particle has travelled a distance $L/2$ from its starting point. Second, even if this optimum value is chosen, the energy gain of the particle will be reduced by a factor of T from what would be achieved if a DC field of E_0 was used for acceleration. Thus, the “efficiency” of acceleration (relative to a DC field) is maximized by minimizing the transit angle.

We can use Equation 49 to calculate the energy gain from a TM_{0pj} mode in a cylindrical cavity. Recall that such a mode is composed of a left-travelling wave and a right-travelling wave superimposed within the cavity. Thus,

$$eV = e \frac{E_0 L}{2} \{ T_L \cos(\omega t_0 + \psi_L/2) + T_R \cos(\omega t_0 + \psi_R/2) \}, \quad (50)$$

where ψ_L and ψ_R represent the transit angle for leftward- and rightward-travelling waves, T_L and T_R their respective transit angle factors, and we have replaced the real portion of the complex

exponential with a cosine function. We can express the transit angles ψ_L , ψ_R as functions of the wave number k_j :

$$\psi_{L,R} = L \left(\frac{\omega}{c} \pm k_j \right), \quad (51)$$

where the + sign corresponds to leftward-travelling and the – sign to rightward travelling. We can combine our expression for $k_j = j\pi/L$ with the expression for ω_{npj} in Equation 41 to expand the transit angles:

$$\psi_{L,R} = \frac{1}{c\sqrt{\mu\epsilon}} \sqrt{\left(\frac{z_{0p}L}{b}\right)^2 + (j\pi)^2} \pm j\pi. \quad (52)$$

From Equation 52, we can see that, for leftward-travelling waves, the transit angle will be $2j\pi$ plus $z_{0p}L/b$ added in quadrature. For most combinations of parameters, this will give a negative value for $\sin(\psi_L/2)$ – that is to say, a decelerating wave. Apparently, if we want both the leftward- and rightward-travelling waves to contribute to the acceleration, the best mode choice is $j = 0$, corresponding to a standing wave with no longitudinal dependence (since the latter goes as $\cos(k_j z)$, and $k_{j=0} = 0$). Note that, for this solution, $\psi_L = \psi_R$, and the two cosine terms in Equation 50 can be maximized simultaneously by an appropriate choice of t_0 .

2.2 R/Q

In the preceding sections, we established that a TM_{npj} mode is acceptable for acceleration of particles; that our requirement of a nonzero accelerating field on the axis of the cavity corresponds to a requirement that $n = 0$; that in order to maximize the effective accelerating gradient for a given electric field applied we require $j = 0$ – thus, we seek to use a TM_{0p0} mode for acceleration.

In order to maintain an accelerating field in the cavity, it will be necessary to store electromagnetic energy in the cavity. Since that energy ultimately comes from the electrical grid and has to be paid for, a useful quantity to calculate is the net acceleration achieved for a given quantity of stored energy. The energy density \mathcal{U} is given by [3, 4]:

$$\mathcal{U} = \frac{1}{2} \left(\vec{E} \cdot \vec{D}^* + \vec{B} \cdot \vec{H}^* \right). \quad (53)$$

For a TM_{0p0} mode, only E_z and H_θ exist, and we can write:

$$\begin{aligned} \vec{E} \cdot \vec{D}^* &= \epsilon [E_0 J_0(k_{c,0p}r)]^2 \cos^2 \omega t, \\ \vec{B} \cdot \vec{H}^* &= \mu \left[\frac{\omega \epsilon E_0}{k_{c,0p}} J'_0(k_{c,0p}r) \right]^2 \sin^2 \omega t. \end{aligned} \quad (54)$$

If we select $t = 0$, Equation 54 requires that the magnetic energy density be zero throughout the cavity and we can compute the total stored energy using only the electric field components. Since total energy is conserved, and the E and H components in Equation 54 are out of phase with one another, one might suspect that the stored energy is oscillating back and forth between the electric and the magnetic components, and that therefore one could find the stored energy by selecting either the electric or magnetic component at an appropriate time. This suspicion is in fact accurate, and often textbooks will note that the time-averaged magnetic energy equals the time-averaged electric energy, and that the time-averaged electric energy is half the peak electric energy; they therefore continue to state that therefore the total stored energy is half of twice the peak stored electric energy, which is a somewhat convoluted means of arriving at the same conclusion.

In any event: we now wish to evaluate the total stored energy:

$$U = \frac{\epsilon}{2} E_0^2 \int_0^b dr \int_0^L dz \int_0^{2\pi} r d\theta J_0^2(k_{c,0p} r). \quad (55)$$

The longitudinal and angular components can be evaluated by inspection, and one is left with a radial integration:

$$U = \pi \epsilon E_0^2 L \int_0^b r dr J_0^2(k_{c,0p} r). \quad (56)$$

We can simplify Equation 56 by introducing dimensionless variable $\mathcal{R} \equiv r/b$:

$$U = \pi \epsilon E_0^2 b^2 L \int_0^1 \mathcal{R} d\mathcal{R} J_0^2(z_{0p} \mathcal{R}). \quad (57)$$

Finally, we can use the fact that $\int_0^1 \mathcal{R} d\mathcal{R} J_0^2(z_{0p} \mathcal{R}) = J_1^2(z_{0p})/2$ [5]:

$$U = \frac{\pi}{2} \epsilon E_0^2 b^2 L J_1^2(z_{0p}). \quad (58)$$

Since $U \propto E_0^2$, we can sensibly form the ratio of the square of the effective accelerating voltage, $V^2 = (E_0 L T)^2$, to the stored energy required to achieve that acceleration:

$$\begin{aligned} \frac{V^2}{U} &= \frac{2E_0^2 L^2 T^2}{\pi \epsilon E_0^2 b^2 L J_1^2(z_{0p})} \\ &= \frac{2T^2}{\pi \epsilon J_1^2(z_{0p})} \frac{L}{b^2}. \end{aligned} \quad (59)$$

Equation 59 can be usefully transformed by replacing one of the factors of b with $z_{0p}/(\omega_{0p0} \sqrt{\mu \epsilon})$, and moving the resulting factor of ω_{0p0} to the LHS:

$$\frac{V^2}{\omega_{0p0} U} = \frac{2T^2}{\pi z_{0p} J_1^2(z_{0p})} \frac{L}{b} \sqrt{\frac{\mu}{\epsilon}}. \quad (60)$$

The quantity $\sqrt{\mu/\epsilon}$ is the impedance of the medium, with units of ohms in MKSA, and is usually abbreviated Z , thus:

$$\frac{V^2}{\omega_{0p0} U} = \frac{2}{\pi z_{0p} J_1^2(z_{0p})} \frac{T^2 L}{b} Z. \quad (61)$$

We can make one additional simplification by noting that the term $2/[\pi z_{0p} J_1^2(z_{0p})]$ is equal to 0.982 for $p = 1$ and asymptotically approaches 1 for larger p values [6]. Thus, to excellent approximation, we can state that:

$$\frac{V^2}{\omega_{0p0} U} = \frac{T^2 L}{b} Z. \quad (62)$$

The quantity $V^2/(\omega_{0p0} U)$ is generally known as R/Q (read ‘‘R over Q’’ or ‘‘R upon Q’’), for reasons which will be discussed below. Although it is referred to as the ratio of two quantities, R/Q is actually a purely geometric quantity of a given accelerating cavity, and it relates the resonant frequency, the achievable acceleration, and the stored energy which is required for operation. As we can see from Equation 62, energy-efficient acceleration is achieved by maximizing the cavity length, transit-time factor, and frequency, while minimizing the cavity radius.

Unfortunately, these requirements are somewhat in conflict, since maximizing L will also increase the transit angle, thus reducing T . We can optimize Equation 62 by noting that $T^2 = \sin^2(\psi/2)/(\psi^2/4)$ contains 2 factors of $1/L^2$ (in the $1/(\psi^2/4)$ component), thus:

$$\frac{R}{Q} = Z \frac{L}{b} \frac{\sin^2(\psi/2)}{\psi/2} \frac{2c}{L\omega_{0p0}}. \quad (63)$$

Recalling that $b = z_{0p}/\omega_{0p0}\sqrt{\mu\epsilon}$, we find:

$$\frac{R}{Q} = Z \frac{2}{z_{0p}} \frac{\sin^2(\psi/2)}{\psi/2} c\sqrt{\mu\epsilon}. \quad (64)$$

Equation 64 shows that the accelerating efficiency of a cavity is maximized when the $p = 1$ mode is used, since z_{0p} monotonically increases with p . It also shows that there is an optimum transit angle, since $\psi = 0$ results in a zero R/Q , and $\psi = 2\pi$ also yields zero R/Q . The value of $\sin^2(\psi/2)/\psi/2$ is plotted in figure 4. It shows that R/Q is maximized at a transit angle of roughly $3\pi/4$ (actually, 133.56°); at this angle, $\sin^2(\psi/2)/\psi/2 = 0.725$. One can also see from Figure 4 that R/Q is a relatively weak function of the transit angle: angles from 90° to 180° will yield values that are within 12% of the maximum value. Finally, we note that for a right circular cylindrical cavity, R/Q can be expressed as a function of transit angle alone, without any other dimensions or free parameters entering. Considering the definition of R/Q , one therefore surmises that accelerating voltage per unit of energy is optimized by maximizing the accelerating frequency and optimizing the transit angle.

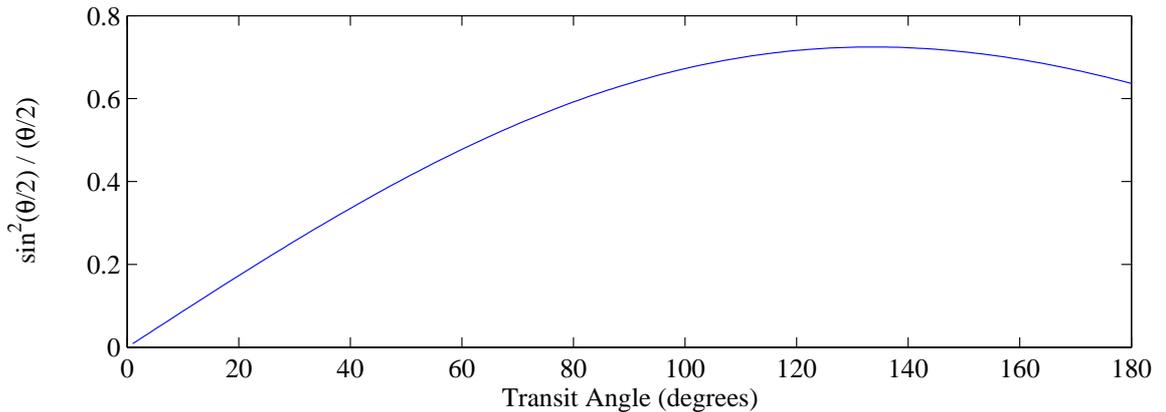


Figure 4: Value of the transit angle factor in R/Q as a function of the transit angle.

2.3 Q of a Cavity and Shunt Impedance

In the previous section we discussed the optimization of a cavity in terms of maximizing the accelerating voltage per unit energy stored in the cavity. The stored energy can indeed be the parameter that limits performance, but more commonly the limitation is input power – the power source for the cavity can supply only so many megawatts, and when the number of megawatts going from the power source into the cavity equals the number of megawatts lost in the cavity, that sets the limit of accelerating field.

We can understand the point at which this occurs by estimating the amount of power dissipated by a cavity. Thus far we have assumed that the walls of the cavity have zero resistance (or infinite

conductivity), which required the electric and magnetic fields to vanish within the conductor. In the case of finite conductivity, it can be shown [2] that the tangential magnetic field must be continuous at the boundary between the vacuum and the conductor. This implies that the magnetic field within the conductor is nonzero, and by the curl equations, the electric field must also be nonzero. The electric field in the conductor can be shown to be [8]

$$\vec{E}_c \approx \sqrt{\frac{\mu_c \omega}{2\sigma}} (-1 - i)(\hat{n} \times \vec{H}_s) \exp(-\xi/\delta) \exp(-i\xi/\delta), \quad (65)$$

where the subscript “c” refers to fields or permeability of the conductor, the subscript “s” refers to the field at the surface of the conductor, \hat{n} is the normal vector pointing from the surface into the vacuum, ξ is a coordinate that points into the surface, and δ is the skin depth of the material at frequency ω :

$$\delta \equiv \sqrt{\frac{2}{\mu_c \omega \sigma}}. \quad (66)$$

The total power dissipated per unit area of the cavity can be computed by computing the ohmic loss per unit volume and integrating from the surface to infinite depth. The electric field is given in Equation 65, and the current is the product of the electric field and the conductivity σ . When the integration is completed, a rather convenient result appears:

$$\frac{dP}{dA} = \frac{\mu_c \omega \delta}{2} |\vec{H}_s|^2 \quad (67)$$

We can eliminate the skin depth in Equation 67:

$$\frac{dP}{dA} = \sqrt{\frac{\mu_c \omega}{2\sigma}} |\vec{H}_s|^2. \quad (68)$$

Thus, the total power dissipated in ohmic losses per unit area of a conducting RF cavity can be estimated by integrating the magnetic field over all of the surfaces and applying Equation 68.

2.3.1 Superconducting Cavity Losses

A useful transformation to Equation 68 is:

$$\begin{aligned} \frac{dP}{dA} &= R_s |\vec{H}_s|^2, \text{ where} \\ R_s &\equiv \sqrt{\frac{\mu \omega}{2\sigma}} = \frac{1}{\sigma \delta}. \end{aligned} \quad (69)$$

The quantity R_s is the “surface resistance” of the material, which increases as the square root of frequency for a conventional conductor.

It is often fashionable these days to design and construct accelerators in which the cavities are made of a superconducting material, usually elemental niobium. For such cavities, we can use Equation 69, but with a different expression for the surface resistance [39]:

$$R_s[\Omega] \approx 9 \times 10^{-5} \frac{\nu^2[\text{GHz}]}{\mathcal{T}[\text{K}]} \exp\left(-\alpha_{sc} \frac{T_c}{\mathcal{T}}\right) + R_{\text{res}}, \quad (70)$$

where \mathcal{T} is the operating temperature, T_c is the critical temperature of 9.2 kelvin, $\alpha_{sc} = 1.92$, and R_{res} is the residual resistance from impurities, which is typically at the level of $10^{-8}\Omega$. Figure 5 shows the surface resistance as a function of frequency for copper ($\sigma = 5.98 \times 10^7$ mho/m), as compared to niobium at 2 kelvin and 4.5 kelvin. Superconductivity is vastly superior in any application where wall losses are the sole limitation.

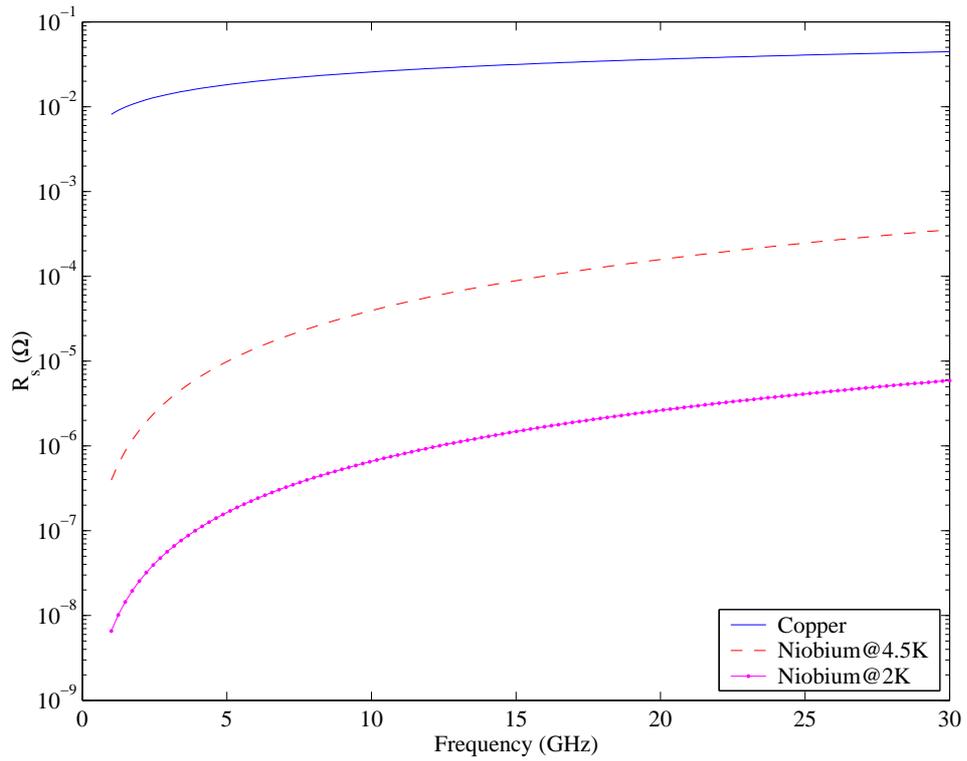


Figure 5: Surface resistance R_s as a function of frequency for copper and for niobium at two different temperatures.

2.3.2 Q of a Cavity

We can now compute the power dissipation in a cavity operating in the TM_{010} mode, with length L and radius b . All that needs to be done is the integration of H_θ^2 over all of the surface. If we choose a time such that H_θ is maximized, then Equation 42 shows

$$|H_\theta| = E_0 \frac{\epsilon\omega_{010}}{k_{c,01}} J'_0(k_{c,01}r). \quad (71)$$

In the TM_{010} mode, $\omega_{010} = k_{c,01}/\sqrt{\mu\epsilon}$. Thus, we can eliminate the factor of $\epsilon\omega/k$ in Equation 71, and replace them with a factor of $1/Z$. We can calculate the total power loss in two integrals, one for the endcap and one for the barrel. The barrel integral can be performed by inspection:

$$\begin{aligned} P_{\text{barrel}} &= R_s \frac{E_0^2}{Z^2} \int_0^L dz \int_0^{2\pi} b d\theta [J'_0(k_{c,01}b)]^2 \\ &= \frac{2\pi E_0^2 R_s L b}{Z^2} [J'_0(z_{01})]^2 \\ &= \frac{2\pi E_0^2 R_s L b}{Z^2} J_1^2(z_{01}), \end{aligned} \quad (72)$$

where we have made use of the fact that $J'_0(\mathcal{R}) = -J_1(\mathcal{R})$ [5]. The endcap integral,

$$P_{\text{endcap}} = R_s \frac{E_0^2}{Z^2} \int_0^b r dr \int_0^{2\pi} d\theta [J'_0(k_{c,01}r)]^2, \quad (73)$$

requires a bit more. We can make use of three Bessel function identities [9]

$$\begin{aligned} \int_0^1 \mathcal{R} d\mathcal{R} [J_n(\mathcal{Q}\mathcal{R})]^2 &= \frac{1}{2} \left\{ [J'_n(\mathcal{Q})]^2 + \left(1 - \frac{n^2}{\mathcal{Q}^2}\right) [J_n(\mathcal{Q})]^2 \right\} \\ J'_n(\mathcal{Q}) &= \frac{1}{2} [J_{n-1}(\mathcal{Q}) - J_{n+1}(\mathcal{Q})], \\ J_{n+1}(\mathcal{Q}) + J_{n-1}(\mathcal{Q}) &= \frac{2n}{\mathcal{Q}} J_n(\mathcal{Q}), \end{aligned} \quad (74)$$

and the fact that $J'_0 = -J_1$ to find the surprising result:

$$\int_0^b r dr [J'_0(z_{01}r/b)]^2 = \frac{b^2}{2} J_1^2(z_{01}). \quad (75)$$

Thus we can solve the endcap integral,

$$P_{\text{endcap}} = \frac{\pi E_0^2 R_s b^2}{Z^2} J_1^2(z_{01}). \quad (76)$$

We can put this all together, remembering that there are two endcaps and that the time-averaged power loss is 1/2 of the losses calculated at the peak of the H -field, to find:

$$P_{\text{cav}} = \frac{\pi E_0^2 R_s}{Z^2} J_1^2(z_{01}) b(L + b). \quad (77)$$

Comparison of Equation 77 to Equation 58 shows that both the stored energy in the cavity and the loss per unit of time are proportional to E_0^2 , the peak on-axis electric field. Thus, we can now predict the destiny of any such cavity hooked up to a power-limited energy source: the stored energy in the cavity will rise as the cavity is filled by the source, and the power lost in the walls

will also increase (since it is simply proportional to the stored energy). At some point, the stored energy will be so large that the power lost in the walls will equal the power entering the cavity from the source, and that will determine the maximum accelerating voltage available to the cavity.

A neatly-dimensionless quantity is the ratio of the stored energy to the power dissipated in 1 RF cycle (actually, 2π RF cycles), which is a quantity known as the “wall Q ”:

$$\frac{\omega U}{P} = \frac{z_{01} Z L}{2R_s(L+b)} \equiv Q_w. \quad (78)$$

Since $P \equiv dU/dt$, we can use Equation 78 to determine the time-evolution of the stored energy in the absence of an external power source:

$$U = U_{t=0} \exp(-\omega t/Q_w). \quad (79)$$

2.3.3 Shunt Impedance

In Equation 64, we saw that R/Q , the ratio of accelerating voltage to stored energy in a cavity, was a function of the cavity length through the dependence on transit angle ψ . Equation 78 shows that the wall Q , the ratio of stored energy to power loss, is a function of both the cavity length L and its radius b . We can combine these “figures of merit” to find the more useful ratio of the accelerating voltage to the power loss:

$$\begin{aligned} R_{\text{cav}} &= Q_w \frac{R}{Q} \\ &= \frac{V^2}{P} \\ &= \frac{Z^2}{R_s} \frac{2c^2 \mu \epsilon}{z_{01}} \frac{\sin^2 \psi/2}{1 + c\sqrt{\mu \epsilon} \frac{2}{z_{01}} \frac{\psi}{2}}. \end{aligned} \quad (80)$$

The quantity R_{cav} is called the “shunt impedance,” and knowledge of the shunt impedance allows the acceleration properties of a cavity to be treated, mathematically, like a resistor: given an input power P and a shunt impedance R_{cav} , the accelerating voltage will satisfy $P = V^2/R_{\text{cav}}$.

Figure 6 shows the shunt impedance form factor (the term in Equation 80 with the ψ 's in it) as a function of ψ . The shunt impedance is optimized for a transit angle which is slightly larger than the one which optimizes R/Q – R_{cav} is maximized for a transit angle of 158° , compared to 133° for the R/Q . For a transit angle of 90° , the shunt impedance is reduced by about one-third from its value at 158° , while a transit angle of 120° represents only an 11% loss in shunt impedance.

3 Multi-Celled Accelerating Structures

in 2.3.3, we saw that the power required to achieve a certain accelerating voltage varies as the square of that voltage, and inversely with the shunt impedance of the cavity which delivers that voltage. Relatively little insight is required to see that, given a cavity with a shunt impedance R_{cav} , one can achieve a voltage of $2V$ with one cavity and a power source of $4V^2/R_{\text{cav}}$, or else with two cavities that each have a power source of V^2/R_{cav} . The latter choice requires twice as many cavities and power sources, but only half as much power, as the former case.

Just to get a feeling for the numbers, consider a cavity operating at the optimal transit angle and a frequency of 1 GHz, and a desired acceleration of 100 GeV. At this frequency, the surface resistance of copper is about 8 m Ω , yielding a shunt impedance of about 6.6 M Ω . The power

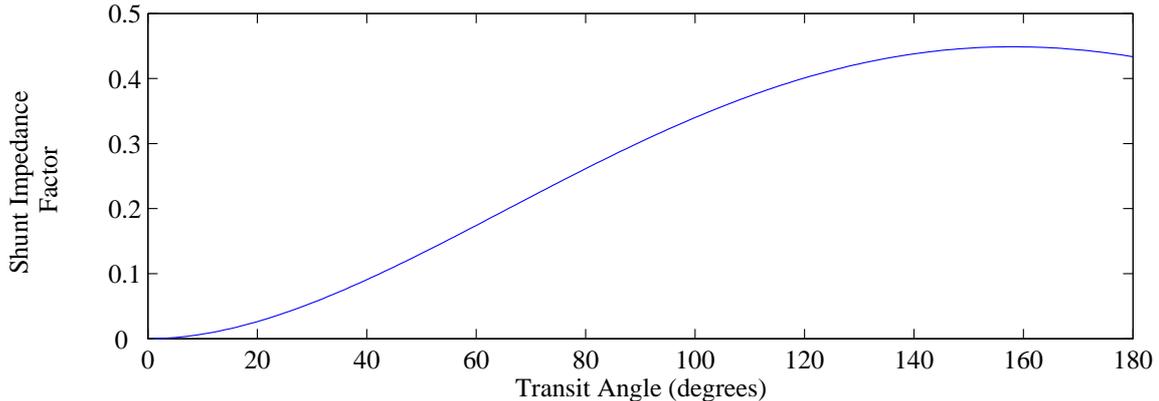


Figure 6: Shunt impedance form factor as a function of transit angle.

required in such a cavity to yield an accelerating voltage of 100 GeV is around 1.5×10^{15} watts. By comparison, the typical power generation capacity available to the state of California is about 4×10^{10} watts. Even if the required power capacity was available, and a solution for cooling the cavity (removing the energy lost in the walls) could be found, the stored energy in such a cavity would be roughly 7 gigajoules, yielding a stored energy density of 1.3 terajoules per cubic meter. Such an energy density is certainly not acceptable to modern ES & H practices!

One option for achieving large accelerations in a multi-cavity system is to have a large number of cavities which are each independently powered. This would require a large number of power sources (one per cavity), and a large number of waveguides which would transport power from the power sources to the cavities. Fortunately, this proves to be unnecessary. The necessary act of cutting a hole in each endcap of the cavity to allow the beam to pass through (!) will also allow us to solve the problem of providing power to each cavity.

3.1 Longitudinal Periodicity in Accelerating Structures

Consider a system shown in Figure 7: a cylindrical waveguide with inner radius b contains a series of cylindrical plates which are evenly spaced a distance d apart in the guide; each plate has a thickness h , and each plate has a hole of radius a in the center. We assume that $a \ll b$, that $h \ll d$, and that the wall conductivity is infinite.

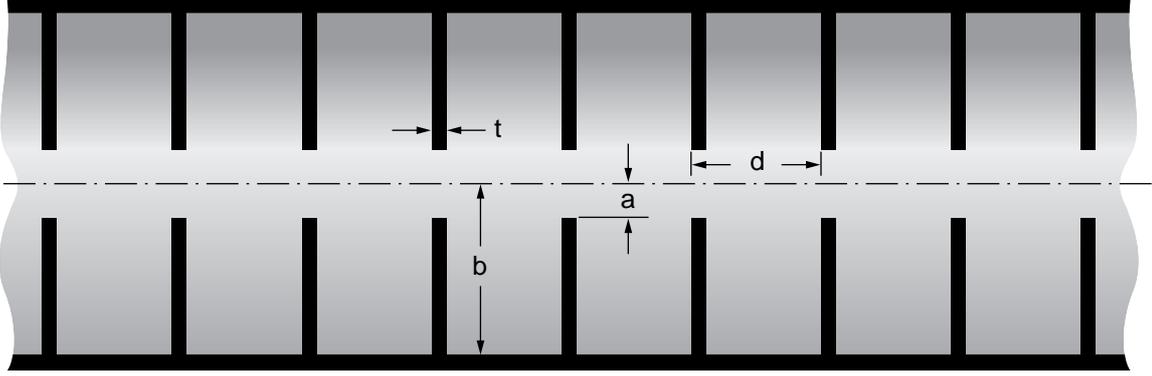
In the limit of $a \rightarrow 0$, we expect the solution for the field in between each pair of plates to reduce to the single-cavity form which was determined in Section 2. We also know from Floquet's theorem that the system with holes will take a solution which satisfies [10]:

$$\vec{E}(r, \theta, z + d, t) = \vec{E}(r, \theta, z, t) \exp[d(-\alpha + ik_z)], \quad (81)$$

where α and k_z are real numbers and α is non-negative. A similar relation holds for \vec{H} .

Let us call the solution for the fields in the cavity in the absence of holes \vec{E}_1, \vec{H}_1 , and the fields in the presence of the holes \vec{E}_2, \vec{H}_2 , and consider the surface integral over the cavity's inner surface $\oint d\vec{A} \cdot (\vec{E}_1 \times \vec{H}_2^* - \vec{E}_2 \times \vec{H}_1^*)$ [11], where \vec{A} points *out* of the cavity. We can apply the divergence theorem to find:

$$\oint d\vec{A} \cdot (\vec{E}_1 \times \vec{H}_2^* - \vec{E}_2 \times \vec{H}_1^*) = \int d\text{Vol} \vec{\nabla} \cdot (\vec{E}_1 \times \vec{H}_2^* - \vec{E}_2 \times \vec{H}_1^*). \quad (82)$$



3-2003
8666A02

Figure 7: Longitudinal section of a cylindrical waveguide “loaded” with periodically-spaced discs.

We can then apply a vector calculus identity to find:

$$\oint d\vec{A} \cdot (\vec{E}_1 \times \vec{H}_2^* - \vec{E}_2 \times \vec{H}_1^*) = \int d\text{Vol} [\vec{H}_2^* \cdot (\vec{\nabla} \times \vec{E}_1) - \vec{E}_1 \cdot (\vec{\nabla} \times \vec{H}_2^*) + \vec{E}_2 \cdot (\vec{\nabla} \times \vec{H}_1^*) - \vec{H}_1^* \cdot (\vec{\nabla} \times \vec{E}_1)]. \quad (83)$$

The curl operators in the above equation will precipitate the usual combination of i , ω , μ , and ϵ but will convert the vector operations to dot products of $\vec{H}_1^* \cdot \vec{H}_2$, and related permutations. If we require that the perturbation be small, then we can say that $\vec{H}_1 \approx \vec{H}_2$ and $\vec{E}_1 \approx \vec{E}_2$. We can thus write:

$$\oint d\vec{A} \cdot (\vec{E}_1 \times \vec{H}_2^* - \vec{E}_2 \times \vec{H}_1^*) = i(\omega_2 - \omega_1) \int d\text{Vol} (\mu \vec{H}_1^* \cdot \vec{H}_1 + \epsilon \vec{E}_1^* \cdot \vec{E}_1). \quad (84)$$

The quantity on the RHS is simply twice the stored energy in the unperturbed cavity. Thus we can derive the change in resonant frequency of the cavities with the holes:

$$\omega_2 - \omega_1 = \frac{-i \oint d\vec{A} \cdot (\vec{E}_1 \times \vec{H}_2^* - \vec{E}_2 \times \vec{H}_1^*)}{2U_1}. \quad (85)$$

By inspection, we can eliminate the first term on the RHS of equation 85. This is because the integral is over the surface of the unperturbed cavity, and on that surface \vec{E}_1 is normal to the surface at all points; therefore, $d\vec{A} \cdot (\vec{E}_1 \times \vec{H}_2^*)$ must vanish for all points on the surface. The quantity \vec{H}_1^* is already known from Equation 42, and $\vec{E}_2 \approx \vec{E}_1$ outside the region of the hole between cavities. Thus, we can estimate the change in the cavity frequency by estimating the radial electric field component in the vicinity of the hole.

3.1.1 Radial Electric Field at the Cavity Hole

The present problem is to determine the electric field in the hole between two resonant accelerating cavities. In the absence of the hole, the field would of course be zero within the conducting material between cavities, and the electric field in the left and right cavities would be related by Floquet’s theorem: $E_R = E_L \exp[d(-\alpha + ik_z)]$; in both cavities, the field would be purely longitudinal (assuming that a TM_{010} mode was selected). In the presence of the hole, we can require that the total solution be a superposition of the hole-free system and a system containing only a conducting sheet with a hole. In the latter system we specify that the longitudinal electric field should go to E_L at $z = -\infty$ and to E_R at $z = +\infty$.

This problem has been solved in excruciating detail by Jackson [12], and here we quote only the result

$$E_r = \frac{E_L - E_R}{\pi} \frac{r}{\sqrt{a^2 - r^2}}. \quad (86)$$

3.1.2 Computing the Frequency Shift

We can now solve for the frequency shift in Equation 85. For $E_{2,r}$ we can write $E_0(1 - \exp[d(-\alpha + ik_z)])J_0(k_cr)r/\sqrt{a^2 - r^2}$; for $H_{1,\theta}^*$ we can write $i\omega_1\epsilon/k_cE_0^*J'_0(k_cr)$; put it all together, and we find:

$$\begin{aligned} 2U_1(\omega_2 - \omega_1) &= i \oint d\vec{A} \cdot \vec{E}_2 \times \vec{H}_1^* \\ &= -\frac{\epsilon\omega_1}{2\pi k_c} E_0 E_0^* \{1 - \exp[d(-\alpha + ik_z)]\} \int_0^a r^2 dr \int_0^{2\pi} d\theta \frac{J_0(k_cr)J'_0(k_cr)}{\sqrt{a^2 - r^2}}, \end{aligned} \quad (87)$$

where we have included the fact that, due to sinusoidal oscillations, $\vec{E}_2 \times \vec{H}_1^* = (E_2 H_1^*)/2$ if we define $\vec{E}_2 \equiv \hat{E}_2 E_2 \exp(i\omega t)$, etc. Without too much controversy we can require that E_0 be real, that only the real component of the RHS of Equation 87 be used, and we can at the same time perform the angular integral:

$$2U_1(\omega_2 - \omega_1) = -\frac{\epsilon\omega_1}{k_c} E_0^2 [1 - \exp(-\alpha d) \cos(k_z d)] \int_0^a dr \frac{r^2 J_0(k_cr)J'_0(k_cr)}{\sqrt{a^2 - r^2}}. \quad (88)$$

Performing the Horrendous Integral: Equation 88 contains an integral,

$$\mathcal{I} = \int_0^a dr \frac{r^2 J_0(k_cr)J'_0(k_cr)}{\sqrt{a^2 - r^2}}, \quad (89)$$

which looks pretty difficult. It can be completed by switching to a normalized variable, $\mathcal{R} \equiv r/a$, which recasts \mathcal{I} as follows:

$$\mathcal{I} = -a^2 \int_0^1 d\mathcal{R} \frac{\mathcal{R}^2 J_0(k_ca\mathcal{R})J_1(k_ca\mathcal{R})}{\sqrt{1 - \mathcal{R}^2}}, \quad (90)$$

where we have also replaced J'_0 with $-J_1$. The new form does not appear to be much of an improvement, until one realizes that this permits a trigonometric substitution: $\int d\mathcal{R} f(\mathcal{R})/\sqrt{1 - \mathcal{R}^2} \equiv \int d\mathcal{X} f(\sin \mathcal{X})$, where $\sin \mathcal{X} = \mathcal{R}$. Replacing a bunch of Bessel functions with Bessel functions of trigonometric functions seems like a losing proposition, but one can then use the series expansion of the Bessel functions [13] to find:

$$\mathcal{I} = -a^2 \int_0^{\pi/2} d\mathcal{X} \sin^2 \mathcal{X} \left[1 - \frac{1}{4}k_c^2 a^2 \sin^2 \mathcal{X} + \frac{1}{64}k_c^4 a^4 \sin^4 \mathcal{X} - \dots \right] \left[\frac{1}{2}k_c a \sin \mathcal{X} - \frac{1}{16}k_c^3 a^3 \sin^3 \mathcal{X} + \dots \right]. \quad (91)$$

Since $k_c a$ is proportional to a/b , we can neglect any term with a power of $k_c a$ above that of the lowest order present. Neglecting all but the lowest order in $k_c a$ transforms our integral to:

$$\mathcal{I} \approx -a^2 \int_0^{\pi/2} d\mathcal{X} \frac{1}{2}k_c a \sin^3 \mathcal{X}, \quad (92)$$

and this form is convenient for “integration by table,” yielding [14]:

$$\mathcal{I} \approx -\frac{1}{3}k_c a^3. \quad (93)$$

Putting it All Together: We can now return to Equation 88:

$$U_1(\omega_2 - \omega_1) = -\frac{\epsilon\omega_1}{k_c} E_0^2 [1 - \exp(-\alpha d) \cos(k_z d)] \left(-\frac{1}{3} k_c a^3\right). \quad (94)$$

Our expression for U_1 , Equation 58, can be used to find:

$$\omega_2 - \omega_1 = \frac{2}{3\pi} \frac{z_{01}}{J_1^2(z_{01})} \frac{a^3}{b^3 d} \frac{1}{\sqrt{\mu\epsilon}} [1 - \exp(-\alpha d) \cos k_z d], \quad (95)$$

where we have included the fact that there are two holes in each cavity, and therefore the frequency shift is twice what was originally calculated for one hole.

We can make a more useful expression by replacing the exponential decay, $\exp(-\alpha d)$, with the amount of attenuation expected at frequency ω_1 when passing through a hole of radius a and length h . Assuming that the hole is small (so that ω_1 is well below cutoff), the exponential term becomes $\exp(-z_{01}h/a)$. Finally, since $\omega_1 = z_{01}/(b\sqrt{\mu\epsilon})$, we can write:

$$\omega = \frac{z_{01}}{b\sqrt{\mu\epsilon}} \left(1 + \frac{2}{3\pi J_1^2(z_{01})} \frac{a^3}{b^2 d} [1 - \exp(-z_{01}h/a) \cos k_z d]\right). \quad (96)$$

3.2 Properties of the Multi-Celled Structure

We have previously examined the wave propagation properties of: free space (which will propagate any frequency but only if the electric, magnetic, and momentum vectors are all mutually perpendicular); a simple waveguide (which will propagate any frequency above its cutoff, and supports a longitudinal electric field, but which has a phase velocity greater than light speed); and a cylindrical cavity (which will permit only discrete frequencies above its cutoff but which allows a longitudinal electric field which can be used to accelerate a particle).

The multi-celled cavity described by Equation 96 will propagate frequencies above the cutoff of the outer waveguide, but will not propagate frequencies above an upper limit, given when $\cos k_z d = -1$. Like the simple waveguide, the permitted frequencies form a continuous spectrum, rather than the discrete spectrum of the single cylindrical cavity with solid endcaps. Since the mode of each cell in the multi-cell structure is a perturbation of the TM_{010} mode of the cylindrical cavity, we expect that the structure will support longitudinal electric fields.

Will the fields be suitable for acceleration of particles? We can use Equation 96 to find the phase and group velocities of the resulting waveform:

$$\begin{aligned} v_{\text{ph}} = \frac{\omega}{k_z} &= \frac{z_{01}}{k_z b \sqrt{\mu\epsilon}} \left(1 + \frac{2}{3\pi J_1^2(z_{01})} \frac{a^3}{b^2 d} [1 - \exp(-z_{01}h/a) \cos k_z d]\right) \\ v_{\text{gr}} = \frac{d\omega}{dk_z} &= \frac{z_{01}}{b\sqrt{\mu\epsilon}} \frac{2}{3\pi J_1^2(z_{01})} \frac{a^3}{b^2} \exp(-z_{01}h/a) \sin k_z d. \end{aligned} \quad (97)$$

With four free parameters – a , b , d , and h – it appears that we can pick a frequency, a group velocity, and a phase velocity, and still have one degree of freedom left for designing our structure. To make this more concrete, let us consider a waveguide with an inner radius b of 10 cm, a spacing between discs d of 10 cm, and a hole radius a of 5 cm; assume that h is approximately zero. Figure 8 shows ω as a function of k for (a) a circular waveguide with no discs (b) a single cavity with endcaps but not holes (c) an infinitely long multi-cell structure with discs and holes. Also shown is the speed of light line. (Note: the present parameters do not satisfy the requirement that a be small compared to b and d and thus a perturbation; the larger value is more illustrative because

the small- a behavior is preserved but in a way that plots better.) As Figure 8 shows, the line representing the speed of light crosses the $\omega - k_z$ curve of the multi-celled structure, implying that there exists a frequency for which the phase velocity in the structure is c , which is essential for particle acceleration. The intersection occurs at $k_z \approx 27.6 \text{ m}^{-1}$, indicating a wavelength of 22.7 cm and a frequency of 1.33 GHz. The factor $k_z d$ is 0.88π , or 158° ; Equation 97 tells us that the group velocity is 6.3% of the speed of light.

3.2.1 Finite-Length versus Infinite-Length Structures

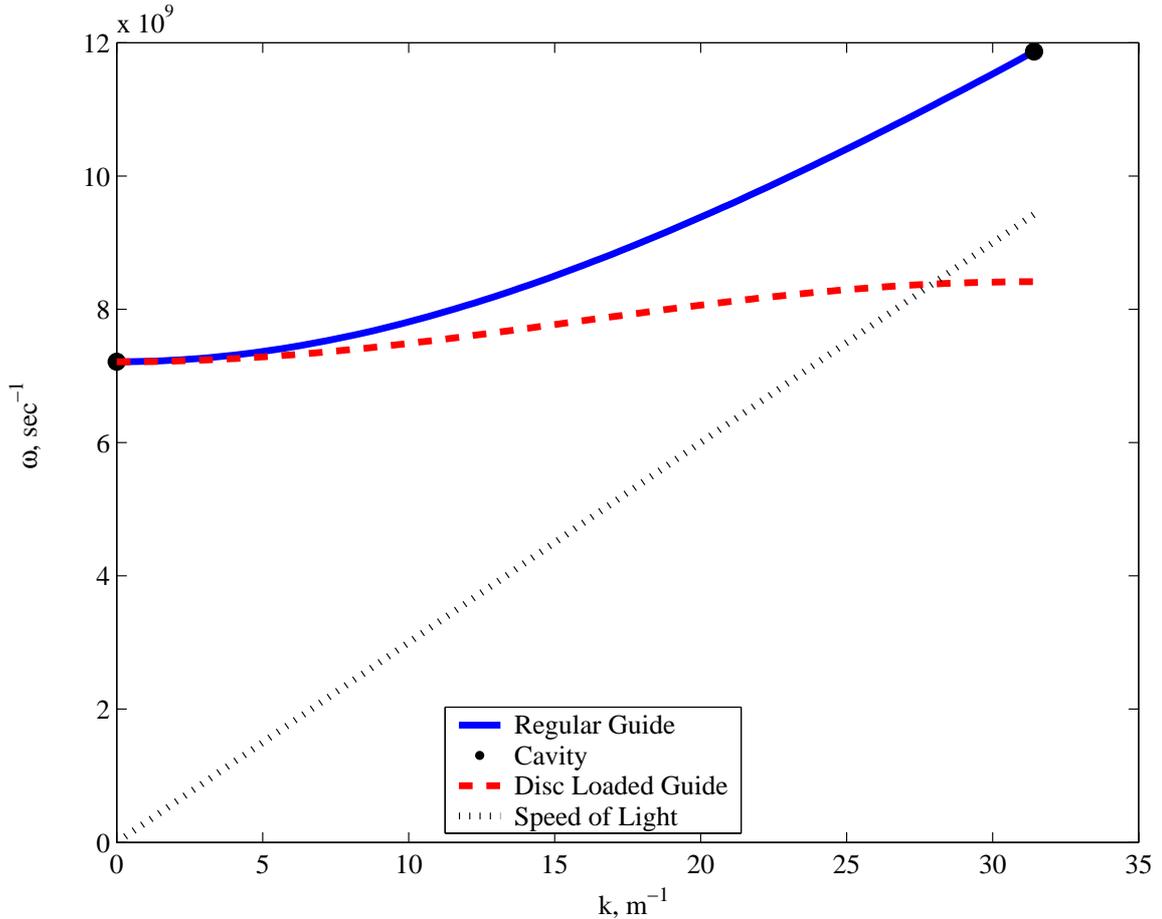


Figure 8: Diagram of ω vs k_z (“dispersion diagram”) for regular waveguide, a single cavity with solid endcaps, and a disc-loaded waveguide. The line representing the speed of light is also shown.

Figure 8 implies that an accelerating structure will propagate TM_{010} -like waves of any frequency between the lower and upper cutoff frequencies. This is true in the limit of an infinitely-long accelerating structure, but not in the case of a structure with a finite number of cells.

For a finite number of cells N_{cell} , the structure acts like a set of N_{cell} coupled oscillators, each of which has a resonant frequency equal to the lower cutoff frequency. Such a system has a total of N_{cell} normal modes of oscillation; these modes are uniformly distributed in $k_z d$, from $k_z d = 0$ to $k_z d = \pi$, as shown in Figure 9. Because of the sinusoidal structure of the $\omega - k_z$ curve, the resonant frequencies are closely-spaced at the 0-mode and π -mode ends of the curve and more widely spaced

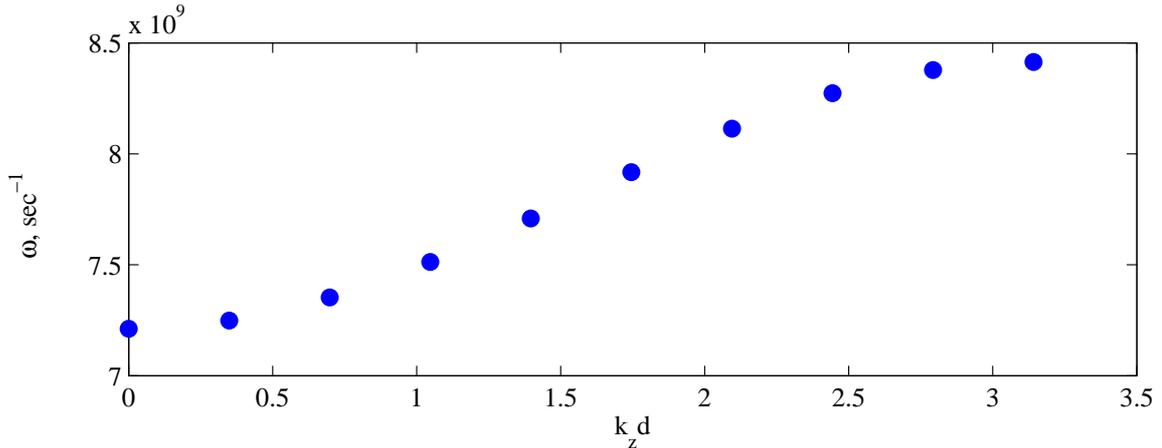


Figure 9: Normal modes of a disc-loaded waveguide with parameters identical to those used in Figure 8, but only 10 cells (rather than an infinite number of cells).

in the center. Each mode has a width equal to the Q value for the structure; thus, the structure will only do a reasonable job of propagating a wave if its frequency is within $1/Q$ of one of the structure's normal modes. Since Q is typically in the thousands or tens of thousands for a copper structure (and can be in the billions for a superconducting structure), and very few structures have significantly more than 100 cells, in a practical system it is not accurate to treat the $\omega - k_z$ curve as truly continuous. This will prove to be important for standing-wave structures (especially superconducting ones), as we will see later.

3.2.2 The Meaning of $k_z d$

What do we really mean when we say that the “phase velocity” of the structure is set by Equation 97? Consider a cavity with some set of parameters a , b , d , and h , which is excited by a power source with frequency ω . Equation 96 allows us to compute the k_z value which corresponds to the driving frequency and the cavity dimensions, and Equation 97 allows us to compute the phase velocity, ω/k_z .

Now consider a particle which passes through the cavity with velocity $v_e = \omega/k_z$ (for now, disregard the possibility that the required velocity might exceed c). If the phase of the electric field in the cavity is zero when the particle enters, then the phase will be $\omega t = \omega d/v_e = k_z d$ when the particle exits. The quantity $k_z d$, then, is apparently equivalent to the transit angle of the particle passing through a single cavity (although in a multi-cell accelerating structure the usual nomenclature is “phase advance per cell”).

Now consider the relation between the field in two adjacent cells, given by Floquet's theorem in Equation 81. If we assume a simple sinusoidal variation of E as a function of time, we find:

$$E(r, \theta, z + d, t) = E(r, \theta, z + d) \exp(i\omega t) = E(r, \theta, z) \exp(i\omega t) \exp[d(-\alpha + ik_z)]. \quad (98)$$

If we neglect the attenuation represented by α , we can rewrite the preceding relation as follows:

$$E(r, \theta, z + d) \exp[i(\omega t - k_z d)] = E(r, \theta, z) \exp(i\omega t), \quad (99)$$

which means that the field at $z + d$ at a time $t + k_z d/\omega$ will be the same as the field at z and time t . The phase velocity of the structure, then, is the velocity a particle needs if it is to encounter the

same RF phase in each cavity of the structure (such a particle is typically called a “synchronous particle,” or alternately the field with a phase velocity matched to the particle’s velocity is called the “synchronous mode”).

We can now intuit that, if the phase velocity refers to the apparent cell-to-cell propagation of waves in the structure, the group velocity must be the actual speed of energy flow through the structure. Furthermore, we see that the coefficient $k_z d$ is equivalent to the transit angle of a single-cell cavity, and one begins to suspect that we can apply all of the single-cell concepts (R/Q , shunt impedance, etc.) to a multi-cell structure with relatively little effort.

3.2.3 $k_z d$ Values Over π

So far, nothing in the formalism of multi-cell structures prohibits the value of $k_z d$ from exceeding π , corresponding to a phase advance greater than 180° per cell. We can certainly imagine extending the dispersion diagram, Figure 8, to such large values, as in Figure 10, in which the ω versus k_z curve is extended to $k_z d = 3\pi$. As required, the dispersion diagram is sinusoidal. As a consequence, the accelerating structure actually contains an infinite number of so-called “space harmonics” at any given frequency; half of these space harmonics are so-called “forward-wave” modes, where a positive slope of the $\omega - k_z$ curve requires that the phase velocity and the group velocity of the mode have the same sign, while the other half are “backward-wave” modes. Furthermore, although all the modes at a given frequency have the same group velocity, only one forward-wave and one backward-wave can have a given phase velocity (typically c in electron applications).

The electric field in the accelerating structure is usually represented as a summation over space harmonics:

$$E_z(z) = \sum_{n=0}^{\infty} E_n \exp[i(k_{zn}z - \omega t)], \quad (100)$$

(apologies are offered for re-using n as an index variable here, but we won’t be dealing with it for long). The wave number is defined as $k_{zn} = k_{z0} + 2\pi n/d$, where k_{z0} is the lowest harmonic in the series (and is usually the one of interest). When RF energy at a given frequency ω is stored in the accelerating structure, it populates all of the space harmonics to varying degrees (since they all oscillate at the same frequency); only the excitation which is synchronous with the beam will provide acceleration. Since the energy stored in other space harmonics does not provide acceleration, the shunt impedance which is relevant to the beam is reduced by a factor of (stored energy in synchronous mode)/(total stored energy).

What does all this mean? What are these “other space harmonics,” and what determines the relative excitation of the various space harmonics? To answer this, let us consider two cells of an accelerating structure which operates in the π mode with a phase velocity $v_{\text{ph}} = c$ for the $n = 0$ space harmonic. This means that $k_{z0} = \pi/d$ and that $\omega/k_{z0} = \omega d/\pi = c$. Since the structure operates in π mode, the electric field in consecutive cells will be equal and opposite at a given time t : $E_1 = E \sin \omega t$, $E_2 = -E \sin \omega t$. Let us consider a particle which enters cell 1 at time $t = 0$ and has velocity $v_e \leq c$. The energy gain in the two cells can be computed by integrating the time-varying electric field experienced by the moving particle (and assuming that the velocity is a constant during this process):

$$\begin{aligned} \Delta U_1 &= eE \frac{v_e}{\omega} \left[1 - \cos \left(\frac{d\omega}{v_e} \right) \right], \\ \Delta U_2 &= eE \frac{v_e}{\omega} \left[\cos \left(\frac{2d\omega}{v_e} \right) - \cos \left(\frac{d\omega}{v_e} \right) \right]. \end{aligned} \quad (101)$$

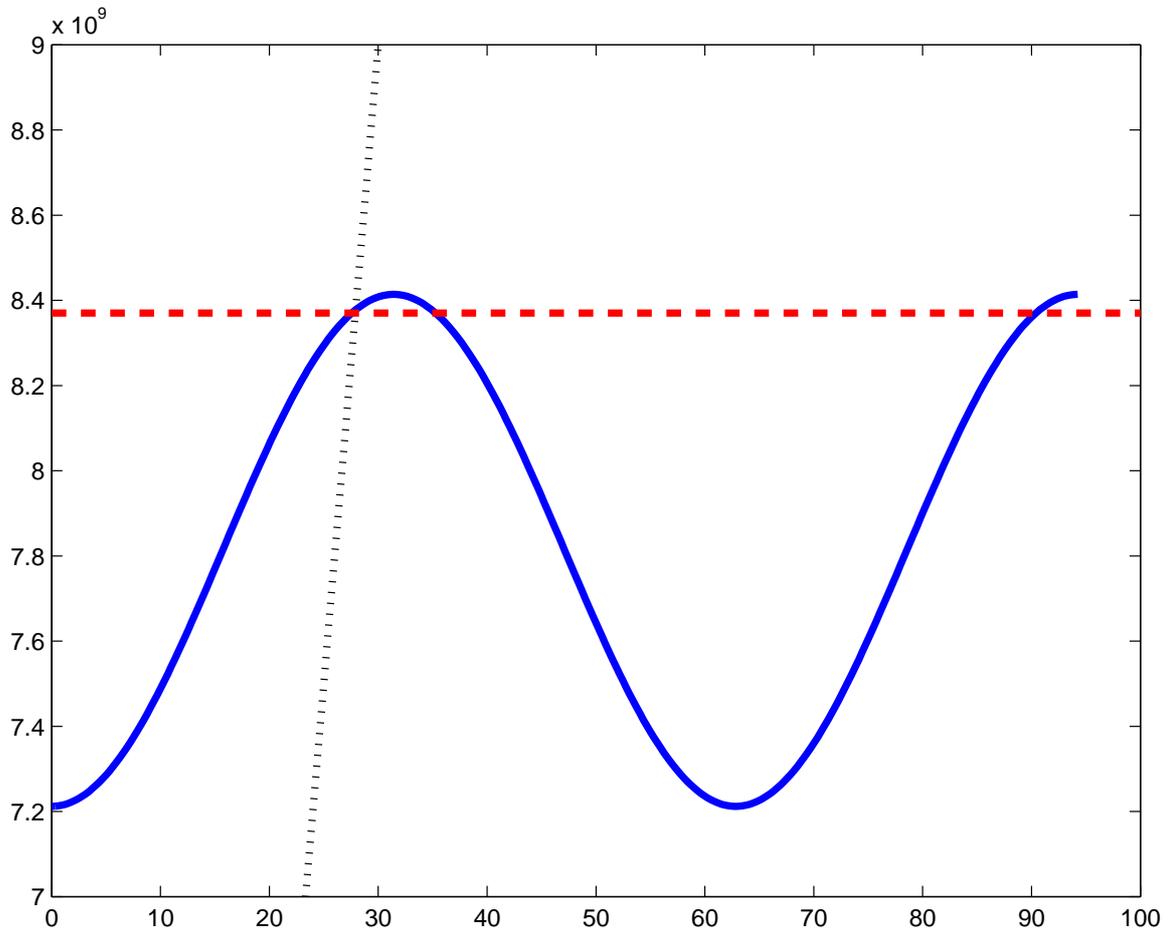


Figure 10: Dispersion diagram for a disc-loaded waveguide, extended out to $k_z d = 3\pi$. The red line indicates that there are 3 space harmonics in this region which will propagate in the structure, but only one has a phase velocity equal to the speed of light.

For a particle with $v_e = c$, we can replace ω/v_e with π/d , and we find that $\Delta U_1 = \Delta U_2 = eEd/2\pi$, which we recognize as the expected energy gain for the particle when the transit time factor of $2/\pi$ is included. For a particle with $v_e = c/2$, we find that $\Delta U_1 = \Delta U_2 = 0$, and the particle gains no energy in either cell of the structure. This is expected when we realize that a phase velocity of c and a particle velocity of $c/2$ means that the particle is in each cell for 1 full oscillation of the RF, and it therefore gains energy in the first half of the cell and loses energy in the second half of the cell.

What happens if the particle has a velocity of $c/3$? In this case, the particle experiences an energy gain of $2eEd/3\pi$ in each cell. So a particle which is not synchronous with the cell phase velocity can gain some energy in each cell. Qualitatively, this result means that for the first $2/3$ of each cell the energy gain of the particle cancels out (since it takes 1 full oscillation to travel through the first $2/3$ of the cell), while for the last $1/3$ of the cell the particle gains energy (since this is only half a full oscillation). Similarly, particle velocities of $c/5, c/7, \dots, c/(2m+1)$ will lead to a net energy gain in every cell.

Now consider a particle with a velocity of $v_e = 2c/5$. This particle will have an energy gain of $2eEd/5\pi$ in the first cell and $-2eEd/5\pi$ in the second. This particle will accelerate in cell 1 and decelerate in cell 2, to achieve no net energy gain over the length of the structure. By extending the logic of this calculation, one finds that if there are an infinite number of cells in the structure, the energy gain and loss in the various cells will cancel out unless the particle velocity is $v_e = c/(2m+1)$, in which case the energy gain per cell will be $eEd/(2m+1) \cdot 2/\pi$.

Since only discrete particle velocities result in net acceleration, we can re-conceptualize the problem by stating that the electric field pattern in the structure contains components which are synchronous with particles at $v_e = c/(2m+1)$ and oscillate at frequency ω . This implies that $v_{\text{ph}} = \omega/k = v_e = c/(2m+1)$, or that $k = k_{z0}(2m+1)$, where $m \geq 0$. For the structure in question, then, the k values deduced above correspond to the k_{zn} values for the various space harmonics.

We can make the situation even more explicit by considering that the electric field is an infinite series of unit-steps with period $2d$. The Fourier expansion of such a series is [36]:

$$E(z, t) = \sin \omega t \frac{4E}{\pi} \sum_{n=0}^{\infty} \frac{1}{2n+1} \sin \left[\frac{(2n+1)\pi z}{d} \right]. \quad (102)$$

If we take the $n = 0$ term and compute the energy gain from this term on a synchronous ($v_e = c$) particle, we find that it is $eEd \cdot 2/\pi$ per cell, which is exactly what was computed using the square-wave representation of the field. Similarly, if we consider the $n = 1$ component, the amplitude of this component is $1/3$ as large as the $n = 0$ component and the velocity required to be synchronous with the $n = 1$ component is $v_e = c/3$. Thus, our previous calculation – that a particle with $v_e = c/3$ will achieve an energy gain $1/3$ as large as that of a $v_e = c$ particle in this system – can be performed by inspection of the Fourier expansion of the accelerating field. From this we can make the following conclusions:

- The “space harmonics” of an accelerator structure correspond to the Fourier series representation of the field (ie, the decomposition of the repetitive square-wave of the electric field into sinusoidal components with the correct periodicity)
- This decomposition implicitly includes the transit angle factor and the fact that the structure can accelerate particles which are not synchronous with the phase velocity of the square wave (which always matches the phase velocity of the lowest space harmonic)
- The energy which is used to maintain the $n > 0$ space harmonics is useless for accelerating

beams which are synchronous with the $n = 0$ space harmonic; the shunt impedance and R/Q are reduced by the presence of the higher harmonics

- the presence of higher harmonics is inevitable, since no single harmonic satisfies the boundary conditions but the summation of the space harmonics does.

In our example above, the ratio of the energy in the $n = 0$ space harmonic to the total energy is $1/\sum[1/(2n + 1)^2]$, or approximately 81%. This means that the effective shunt impedance per unit length of this structure will be about 81% of what is calculated using the formalism described in 2.3.3.

3.2.4 Modes Other than TM_{01}

The analysis applied to the TM_{01} mode, as extended to the multi-cell accelerating structure, can also be applied to the other modes that a single-cell cavity can support. Like the TM_{01} mode, each cavity mode can be extended to a continuous spectrum of structure modes within a “pass band” of allowed frequencies; at each frequency there are an infinite number of modes with identical group velocities but varying phase velocities.

This has some rather interesting implications. Imagine that a structure which was built for linear acceleration is powered from a source which is at a frequency far above the TM_{01} pass band. If the source frequency falls into the pass band of one of the other modes (TM_{11} or TE_{01} , for example), then it will excite those modes. This allows accelerating structures to be used for purposes other than simple linear acceleration. Some examples include RF deflectors, which are accelerating structures operated at a frequency corresponding to a mode which includes a deflecting field at $r = 0$ – essentially, the device can be used as a steering element with a high-frequency, periodic deflecting field. A related implication is that these other modes can be excited by the beam, since a short beam contains Fourier components up to extremely high frequencies. We shall examine this possibility later.

3.2.5 Calculation of Shunt Impedance and Q For One Cell

We have previously encountered Equations 78 and 80, which permit the calculation of the shunt impedance and quality factor of an accelerating cavity. In the context of a real accelerating structure, it is necessary to include a few caveats to achieve an accurate estimate of these quantities.

Disc Thickness Correction: In computing the Q for a cell, the correct cell length to use in Equation 78 is the disc-face to disc-face distance, not the center-to-center distance. Similarly, the transit angle used in Equation 80 should be the disc-face to disc-face transit angle, not the transit angle per cell. These corrections will reduce both the shunt impedance per cell and the Q for structures with relatively thick discs. This factor is in addition to the correction for the $n > 0$ space harmonics discussed earlier.

Iris Aperture Correction: the presence of a hole in the center of the disc between cells reduces the shunt impedance per cell even further. This factor is not straightforward to compute analytically, but can be estimated from a fit to simulations [37]:

$$R_{\text{cell}} \approx \frac{R_{\text{cav}}}{1 + 30.5(a/\lambda)^2}, \quad (103)$$

where R_{cav} is the single-cavity shunt impedance calculated without the correction for the hole, a is the radius of the hole, and λ is the RF wavelength.

4 Travelling-Wave Accelerator Structures

In the previous Section, we determined that a cylindrical waveguide which is “loaded” with conducting discs set periodically along its length is suitable for accelerating particles, in that it is possible to design such a structure with a phase velocity of the longitudinal electric field which is equal to the velocity of the particles to be accelerated. So far so good – but we still do not have much insight into what constitutes a “good” accelerator structure (or even an acceptable one).

Such structures – usually known as *disc-loaded waveguides*, or “DLWG’s” – can be designed in two fundamental flavors – travelling-wave type or standing-wave type. In this section we will explore the parameters of the more common travelling-wave structure type, seeking insight into what constitutes a “good” DLWG.

Figure 11 shows a schematic of a travelling-wave DLWG: RF power at frequency ω is introduced at an input coupler at the upstream end, propagates through the structure in the form of accelerating fields to the downstream end, and exits through an output coupler. Immediately upstream of the input coupler and downstream of the output coupler are cutoff irises; these are thick discs with very long holes which prevent (via evanescence) any significant RF power from escaping from the structure.

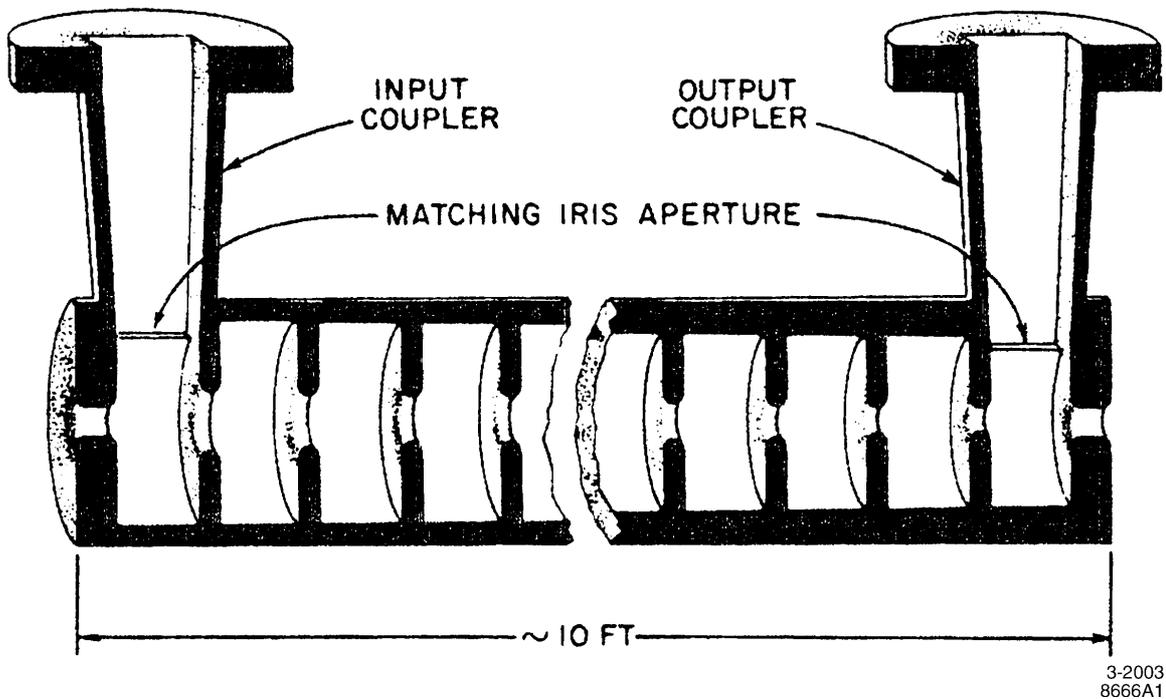


Figure 11: Schematic of a travelling-wave disc-loaded waveguide, in which RF power propagates from the input coupler to the output coupler through a series of accelerating cavities.

Let us imagine that in the steady-state the structure’s stored energy per unit length is $U'(z) \equiv dU/dz$, the power lost into the walls per unit length is given by $p_w(z) \equiv dP_w/dz$, that the power from the source is given by P_0 , and that the power flow at a point along the structure is given by $P(z)$. By conservation of energy, we can require that at every point z along the structure, the time rate of change of stored energy per unit length must be equal to the power dissipated into the walls per unit length, plus the power flow from upstream to z , minus the power flow from z to

downstream:

$$\frac{dU'(z)}{dt} + p_w(z) + \frac{dP(z)}{dz} = 0. \quad (104)$$

From Equation 78 we know that we can relate the wall losses P_w to the stored energy through the wall Q and the frequency. Similarly, the power flow $P(z)$ must be simply equal to the energy per unit length multiplied by the group velocity. Finally, in steady state the stored energy per unit length U' is constant in time. Thus, Equation 104 reduces to:

$$\frac{dP(z)}{dz} = -p_w(z) = -\frac{\omega U'(z)}{Q_w} = -\frac{\omega P(z)}{v_{\text{gr}} Q_w}. \quad (105)$$

Equation 105 has the advantage of being easy to solve for $P(z)$, since it can easily be rearranged into the form $dP/P = -K ds$, yielding

$$P(z) = P_0 \exp\left(-\frac{\omega z}{v_{\text{gr}} Q_w}\right). \quad (106)$$

Equation 106 is usually simplified by introducing an attenuation coefficient, α_0 (not related to the α in Equation 96). By definition, $\alpha_0 \equiv \omega/(2Q_w v_{\text{gr}})$ (with units of ‘‘nepers per meter,’’), thus $P(z) = P_0 \exp(-2\alpha_0 z)$.

As shown in Equation 80 relates the shunt impedance of a cavity to its wall losses. Here we can define a shunt impedance per unit length (sometimes called ‘‘normalized shunt impedance’’), $r_l(z) \equiv dR_{\text{struc}}/dz$, and an accelerating gradient $G_0 \equiv dV/dz$ (where it is implicit that we have selected the relative phase between the beam and the RF to maximize V and thus G_0). If we require that at each point along the structure the relationship between accelerating voltage, shunt impedance, and wall losses should hold, then:

$$\begin{aligned} r_l(z) dz &= \frac{(G_0(z) dz)^2}{p_w(z) dz}, \text{ or} \\ p_w(z) &= \frac{G_0^2(z)}{r_l(z)}. \end{aligned} \quad (107)$$

We can replace $p_w(z)$ with the relationship between $P(z)$ and $p_w(z)$, and include the relationship between P_0 and $P(z)$, to find:

$$G_0^2(z) = 2\alpha_0 r_l(z) P_0 \exp(-2\alpha_0 z) = \frac{\omega}{v_{\text{gr}}} \frac{r_l}{Q_w} P_0 \exp(-2\alpha_0 z). \quad (108)$$

4.1 Constant Impedance Structure

At this point, let us assume that the accelerating structure is made of a set of cells that are identical to one another. In this case, the group velocity, attenuation factor, wall Q , and normalized shunt impedance are all constant throughout the structure. Such a structure is called a *constant impedance structure*. Equation 108 shows that in such a structure, the electric field is higher at the input end than at the output.

We can calculate the total voltage in a constant-gradient structure by integrating the square root of Equation 108 over the length of the structure:

$$V = \sqrt{\frac{2}{\alpha_0} r_l P_0 [1 - \exp(-\alpha_0 L)]}. \quad (109)$$

A commonly used accelerator parameter is the attenuation factor, $\tau \equiv \alpha_0 L$. We can use this to eliminate α_0 :

$$V = \sqrt{r_l L P_0} \sqrt{2\tau} \frac{1 - \exp(-\tau)}{\tau}. \quad (110)$$

Finally, we define the filling time, which is the time required for RF power to travel from the input coupler to the output coupler:

$$t_f = \frac{L}{v_{\text{gr}}} = \frac{2Q}{\omega} \tau. \quad (111)$$

It is interesting to note that, for a structure with a fixed length L , the maximum accelerating voltage is attained for $\tau = 1.26$ [15]. The reason for this is that, for a fixed structure length, cell configuration, and frequency, the value of τ is inversely proportional to the group velocity. For a very low τ , the group velocity is too high; this translates to too little stored energy, since the stored energy per meter in the absence of losses is P/v_{gr} . As the group velocity is increased the stored energy and the wall losses both increase (the latter being proportional to the former, as we have seen), until at some point the wall losses become so large that they limit the maximum stored energy density and hence the accelerating voltage.

4.2 Constant Gradient Structure

The accelerating structure described above has the benefit of simplicity: once a single cell of the structure has been designed to satisfaction, one simply makes a whole lot of them and brazes them together, add input and output coupler, add pumping and cooling, and a completed accelerator structure is available for use. Unfortunately, there are several drawbacks to the constant impedance structure, most notably that the accelerating gradient in the upstream end of the structure is larger than the gradient in the downstream end of the structure; this can be seen clearly in Equation 108. This means that if a structure's performance is limited by its accelerating gradient, it is the front of the structure that will be limiting performance – the body of the structure will be forced to operate at a gradient far lower than the limiting gradient. The ratio of the peak accelerating gradient, $G_0(z = 0)$, to the average gradient, V/L , is $\tau/(1 - e^{-\tau})$ [16]. For a structure with the optimal $\tau = 1.26$, the peak gradient is 1.7 times the average; this is clearly a serious issue.

A more suitable approach to linear acceleration would force the accelerating gradient to be roughly constant along the length of the structure. To see how this can be accomplished, recall that $G_0^2(z)/r_l(z) = p_w(z)$. If $G_0(z)$ is constant with z , then $r_l(z)p_w(z)$ must also be constant with z . Equation 80 shows that, for an ideal cavity, the shunt impedance is a function of the frequency (through R_s) and the phase advance per cell (through the transit angle ψ). To good approximation, this fact remains true for a cell of an accelerating structure; since all the cells must have the same frequency, and phase advance per cell is not an attractive parameter to play with, the value of r_l is to good approximation a constant along the structure. We are thus left to require that the power lost into the walls $p_w(z)$ also be a constant along the structure. Equation 105 thus shows that $dP(z)/dz$ must be a constant to ensure that $p_w(z)$ is constant. Since

$$\frac{dP(z)}{dz} = -\frac{\omega P(z)}{v_{\text{gr}} Q_w}, \quad (112)$$

and Q_w , like r_l , is very close to a constant for cells with equal frequency and phase advance, we conclude that a constant gradient structure requires a gradual tapering of v_{gr} from front to back. Qualitatively, we can see that the constant-gradient structure requires a stored energy density which is constant along the structure. Since the power flow is being attenuated by wall losses, this can

only be achieved by slowing down the power flow at the back of the structure relative to the flow in the front of the structure.

Since v_{gr} is now a function of z along the structure, so must the attenuation factor α_0 be. Nonetheless, we can without controversy define an attenuation factor τ which is analogous to the parameter for constant-impedance structures: $P(L) = P_0 e^{-2\tau}$. Since dP/dz is constant, it must be equal to the output power minus the input power divided by the structure length, or

$$\frac{dP}{dz} = \frac{P(L) - P_0}{L} = -\frac{P_0}{L}(1 - e^{-2\tau}). \quad (113)$$

We can also use Equation 105 to relate dP/dz to the frequency, group velocity and wall Q of the structure:

$$\frac{dP}{dz} = -\frac{P_0}{L}(1 - e^{-2\tau}) = -\frac{\omega P(z)}{v_{\text{gr}}(z)Q_w}. \quad (114)$$

Since dP/dz is constant, $P(z)$ must decrease linearly with z , and thus v_{gr} must also decrease linearly with z . On the other hand, the phase velocity must remain constant. From Equation 97, we see that the group velocity depends upon a^3/b^3 ; a term with the same dependence appears in the phase velocity term, but is added to a term which is proportional to $1/b$. Thus, to accomplish both goals we must reduce both b and a^3/b^3 along the length of the structure.

The expressions above can be combined to yield the required group velocity profile along a constant gradient structure [17]:

$$v_{\text{gr}}(z) = \frac{\omega}{Q_w} \frac{L - (1 - e^{-2\tau})z}{(1 - e^{-2\tau})}. \quad (115)$$

The filling time can be computed by integrating $dz/v_g(z)$ over the structure length, which miraculously reproduces the filling time relationship for constant-impedance structures. Finally, Equations 107 and 114 can be combined to yield the accelerating gradient and the total accelerating voltage:

$$\begin{aligned} G_0 &= \sqrt{p_w(z)r_l} = \sqrt{-r_l dP/dz} = \sqrt{r_l P_0(1 - e^{-2\tau})/L}, \\ V &= \sqrt{r_l P_0 L(1 - e^{-2\tau})}. \end{aligned} \quad (116)$$

Unlike a constant-impedance structure, the constant-gradient structure has no optimal τ value.

5 Standing-Wave Accelerator Structures

In the case of a travelling-wave accelerator, RF power is injected into the structure at an upstream input coupler, propagates down the length of the structure, and exits via an output coupler that is remarkably similar to the input coupler. Alternatively, one could inject RF power into both upstream and downstream couplers, or one could replace the downstream coupler with a short that would reflect the RF power back towards the upstream. In either of these cases, the resulting accelerator – with power propagating in both upstream and downstream directions – develops some standing-wave properties which alter the structure’s accelerating behaviors.

5.1 A Simple Example

Let us begin by considering a travelling-wave accelerating structure, in this case one designed to operate in the $\pi/2$ mode, which is supplied with equal amounts of RF power at both the input and the output coupler. Let the structure be perversely defined to have $N_{\text{cell}} + 1$ cells, with a cell index

j running from 0 to N_{cell} . The accelerating field due to the power from the input coupler, which we define to be propagating in the *forward* direction (i.e., in the same direction as the beam) is given by:

$$E_j^+ = E \sin\left(\omega t - \frac{\pi}{2}j\right). \quad (117)$$

We can see by inspection that at time $t = 0$ the field in cell 0 is zero and the field in cell 1 is $-E$, and both are rising as time advances such that when the field in cell 0 reaches E the field in cell 1 reaches zero; thus a $\pi/2$ travelling-wave which propagates from cell 0 to cell N_{cell} . Similarly, the accelerating field due to the power from the output coupler is given by:

$$E_j^- = E \sin\left[\omega t - \frac{\pi}{2}(N_{\text{cell}} - j)\right]. \quad (118)$$

We can see that, as advertised, this field has a $\pi/2$ character and propagates from cell N_{cell} to cell 0.

Let us assume that the particle is moving in the nominal forward direction (ie, in the direction of the E^+ wave), and that it arrives at the center of cell j at a time defined as t_j . Let us further assume that the arrival time is set so that the forward-propagating field is maximized at that time, ie, $\omega t_j = \pi(j + 1)/2$. The net accelerating field is thus:

$$E = E \left[\sin\left(\frac{\pi}{2}\right) + \sin\left(\frac{\pi}{2}(2j + 1) - \frac{\pi}{2}N_{\text{cell}}\right) \right]. \quad (119)$$

From Equation 119, we can see that the forward-going wave always accelerates the beam, while the backward-going wave will accelerate in some cells, decelerate in others, and provide no net acceleration in yet others. The net result of the backward wave when all cells are considered is zero acceleration. Thus, the shunt impedance per cell of the structure, when supplied with RF power in this manner, is half as much as an equivalent travelling-wave structure, since the power which maintains the backward-travelling wave does not contribute to acceleration.

So why would anyone ever use a standing-wave structure? Let us now replace the structure above with one that operates in the π mode. Now the total accelerating field is given by:

$$E = E \left[\sin\left(\frac{\pi}{2}\right) + \sin\left(\frac{\pi}{2} + 2\pi j - \pi N_{\text{cell}}\right) \right]. \quad (120)$$

From Equation 120, we see that the backwards wave can accelerate a forward-travelling particle, provided that the number of cells is odd (recall that the cells are numbered from 0 to N_{cell}). If the number of cells is even, the forward- and backward-travelling waves both still provide acceleration, but in this case the phase of the input power of one of the feeds must be flipped by 180° relative to the other. As a consequence, the shunt impedance of a π -mode standing wave structure is equal to what would be expected in a travelling-wave structure.

What are the properties of a standing-wave structure that would make it attractive in some cases? For one thing, the standing-wave structure is bi-directional: it can accelerate particles which are velocity-matched to it in either direction, while a travelling-wave structure can only accelerate particles in one direction. In addition, we shall see that a standing-wave structure can be slowly “charged up” with RF energy from a long, low-power pulse. This can pose a distinct advantage, especially when superconducting structures are contemplated.

5.1.1 Use of π Mode

Equation 97 shows that, in a long multi-celled accelerator structure, the group velocity of the π -mode is zero. How can a multi-celled structure operating in the π -mode be filled with energy from a single input coupler?

Let us recall the coupled-oscillator approach to the multi-celled accelerating structure: for $N + 1$ cells numbered 0 to N , there are $N_{\text{cell}} + 1$ normal modes, and their frequencies are given by Equation 96 if $k_z d$ is replaced by $\pi j / N_{\text{cell}}$ (where j also runs from 0 to N_{cell}) [42]. Each mode has a distinct wave function (pattern of excitation amplitude and phase versus cell number) in addition to its distinct frequency. When a near-monochromatic source of RF excitation is tuned to the frequency of the π -mode, it excites the entire mode. Thus, the energy which enters the structure at the input coupler and excites the π -mode will “instantaneously” be supplied to all of the cells, and thus the entire structure is filled despite the zero group velocity of the π -mode.

The explanation above is lacking a little something, and that something is called, “causality” – the idea that the incoming RF energy is transmitted with infinite velocity to all the cells of a standing-wave structure is a rather severe violation of Special Relativity. In fact, the explanation above is actually accurate as far as it goes: by definition, the excitation amplitude of the π -mode is equal in all of the cells, because in a low-loss structure the wavefunction of the π -mode calls for equal excitations in all cells and alternating phases. What is missing from the explanation above is that the structure with $N + 1$ cells contains N modes other than the π -mode. These modes are excited by the incoming RF power as well, but since their frequencies are different they behave like resonators which are driven far off-resonance (which indeed they are). When the wavefunctions for these other modes, their oscillation when driven off-resonance by the incoming RF power, and their decay rates are considered, one finds that the superposition of all the mode oscillations results in an effective group velocity which is lower than the speed of light.

Nonetheless, when the acceleration of particles is concerned, it can be shown that the net acceleration of a particle can be estimated fairly accurately by computing the excitation of the π -mode alone and neglecting the remaining modes. This can be explained in either the time domain or the frequency domain. In the frequency domain, only the π -mode is synchronous with the beam, and we have already seen that non-synchronous modes will tend to cancel out over the length of a multi-celled accelerator structure. In the time domain, it is true that the superposition of all the modes may result in zero actual voltage in the downstream cells of the structure (especially if the beam is injected very soon after the RF power arrives at the input coupler); however, in this case the acceleration provided by the other, non-synchronous modes, is such that the total acceleration achieved in the energized cells is nearly equal to what would be obtained from a pure π -mode excitation which fills all of the cells. In other words, for all practical purposes we can ignore causality and solve our problems as if the entire length of the structure is uniformly excited by the π -mode instantaneously, and that is what we will do for (almost) the rest of this discussion.

Actually, there is one additional potential source of trouble in standing-wave structures which contributes to their design properties. As mentioned, the modes are uniformly distributed in k_z , which means that the frequencies of the structure modes near the π -mode are very closely spaced. For very long standing-wave structures, this can result in the π -mode and nearby travelling-wave modes overlapping one another. If this happens, energy will be transferred from the standing-wave to the travelling-wave modes. Since this is an undesirable state of affairs, standing-wave structures are typically constructed with a relatively small number of cells (to maximize the frequency spacing at the π -mode) and a high Q (low-frequency and/or superconducting structures).

5.1.2 Input Coupler, Loaded Q , and Filling Time

Another major difference between a standing-wave structure and a similar travelling-wave structure is in the design of the input coupler. A travelling-wave structure has an input coupler at its upstream end and an output coupler at its downstream end; both couplers are designed to be “well-matched,” which in this case means that RF power can flow into and out of the structure with almost zero

reflection. For a standing-wave structure, this would be counterproductive, since it would allow the stored energy to flow back out of the structure efficiently; what we want in a standing-wave structure, however, is to efficiently *trap* the energy in the structure. This implies that the input coupler must have a strong tendency to reflect incident RF power.

To understand more fully the implications of this, we follow the approach of Nantista [38]: a monochromatic power source supplies a wave with amplitude E_{in} , which is incident upon a resonant cavity of some kind; the boundary between the cavity and the input waveguide is an iris with reflection coefficient Γ . In general Γ is real and negative, which causes a reflected wave which is out of phase with the incident wave, $E_{\text{ref}} = \Gamma E_{\text{in}} = -|\Gamma|E_{\text{in}}$. Meanwhile, there is also an electromagnetic wave, E_e , which is the wave emitted from the stored energy in the cavity (ie, the stored energy is “leaking out” through the iris). The net amplitude flowing backwards, from the iris towards the power source, is $E_{\text{out}} = E_e + \Gamma E_{\text{in}}$.

Now let us require that energy be conserved, which on an instantaneous basis means that RF power must also be conserved. The amount of power incident on the iris from the power source must be equal to the sum of the power in the reverse wave, the power lost in the walls of the resonant cavity, and the instantaneous increase in the stored energy in the cavity:

$$P_{\text{in}} = P_{\text{out}} + P_c + \frac{dU_c}{dt}. \quad (121)$$

Let us define the proportionality between the power in the waveguide and the resulting square of the field amplitude as \mathcal{K}_{wg} , such that $P_{\text{in}} = \mathcal{K}_{\text{wg}}E_{\text{in}}^2$, and similarly $P_{\text{out}} = \mathcal{K}_{\text{wg}}(E_e + \Gamma E_{\text{in}})^2$. In the absence of incoming power, obviously $P_{\text{out}} = \mathcal{K}_{\text{wg}}E_e^2$. From the definition of wall Q , Equation 78, we know that $P_c = \omega U_c/Q_w$. It can be shown that, in the absence of incoming power, the outgoing power is also proportional to the stored energy and thus to the power lost into the cavity walls. We can define the ratio of power emitted from the cavity to power lost in the walls as the cavity coupling coefficient, β_c :

$$\beta_c \equiv \frac{P_{\text{out}}}{P_c}, \quad P_{\text{in}} = 0. \quad (122)$$

Putting all this together allows us to write:

$$U_c = \frac{Q_w}{\omega} P_c = \frac{Q_w}{\omega} \frac{\mathcal{K}_{\text{wg}} E_e^2}{\beta_c}. \quad (123)$$

Let us now recast Equation 121, with some substitutions:

$$\begin{aligned} \mathcal{K}_{\text{wg}} E_{\text{in}}^2 &= \mathcal{K}_{\text{wg}} (E_e + \Gamma E_{\text{in}})^2 + \frac{\mathcal{K}_{\text{wg}}}{\beta_c} E_e^2 + \frac{2\mathcal{K}_{\text{wg}} Q_w}{\omega \beta_c} E_e \frac{dE_e}{dt}, \text{ or} \\ E_{\text{in}}^2 &= (E_e + \Gamma E_{\text{in}})^2 + \frac{1}{\beta_c} E_e^2 + \frac{2Q_w}{\omega \beta_c} E_e \frac{dE_e}{dt}, \end{aligned} \quad (124)$$

where we have taken the time derivative of Equation 123. If we assume, as discussed above, that $\Gamma \approx -1$, then we can write a simple differential equation which describes the emitted wave amplitude:

$$\begin{aligned} 2E_{\text{in}} &= E_e \left(1 + \frac{1}{\beta_c}\right) + \frac{2Q_w}{\omega \beta_c} \frac{dE_e}{dt}, \text{ or} \\ \frac{\omega \beta_c}{Q_w} E_{\text{in}} &= \frac{dE_e}{dt} + E_e \frac{\omega}{2Q_L}, \end{aligned} \quad (125)$$

where $Q_L \equiv Q_w/(1 + \beta_c)$ is called the *loaded Q* (ie, the Q – the proportionality between the RF period and the e -folding time – when the cavity is “loaded” by the coupler as well as the wall losses).

If we define the characteristic time of the system, $t_c \equiv 2Q_L/\omega$, then we can rewrite Equation 125 in an even-more compact and useful form:

$$t_c \frac{dE_e}{dt} + E_e = \frac{2\beta_c}{1 + \beta_c} E_{in}. \quad (126)$$

If we assume that $E_{in} = 0$ for $t < 0$ and is a constant for $t > 0$, we can solve Equation 126:

$$E_e = \frac{2\beta_c}{1 + \beta_c} E_{in} \left(1 - e^{-t/t_c}\right). \quad (127)$$

Similarly, we can now solve for $E_{out} = E_e - E_{in}$:

$$E_{out} = \frac{2\beta_c}{1 + \beta_c} E_{in} \left(1 - e^{-t/t_c}\right) - E_{in}. \quad (128)$$

Note that, for $\beta_c = 1$, the backwards-wave amplitude E_{out} goes asymptotically to zero, as the reflected and emitted waves become equal and opposite. The characteristic “e-folding” time for the cavity with input coupler is t_c , which is the filling time of a standing-wave cavity.

How do the stored energy and the accelerating voltage of a standing-wave structure vary with time? Equation 123 defines the stored energy as a function of the emitted wave amplitude, which we solved for in Equation 126. With appropriate simplifications,

$$U_c(t) = t_c P_{in} \frac{2\beta_c}{1 + \beta_c} \left(1 - e^{-t/t_c}\right)^2. \quad (129)$$

Similarly, by the definition of R/Q , we find:

$$V(t) = \left(1 - e^{-t/t_c}\right) \sqrt{\frac{R}{Q} \omega t_c P_{in} \frac{2\beta_c}{1 + \beta_c}}. \quad (130)$$

Note that a large value of β_c will result in a small value of Q_L and a correspondingly short fill-time, but also will result in a low equilibrium voltage in the structure.

6 Superconducting Accelerator Structures

With the formalism of various kinds of structures available to us, it is worthwhile to consider the circumstances under which superconductivity can be used to improve the performance of an accelerating structure.

Let us first consider a constant-impedance travelling-wave RF structure, with a length of 1 meter and a frequency of 1 GHz (nice round numbers). As mentioned previously, the optimal τ value for such a structure is 1.26, yielding a voltage of $0.90\sqrt{r_l L P_0}$. From Equation 80, we see that the shunt impedance per cell is inversely proportional to the surface resistance of the structure. For copper this resistance is 8.1 m Ω , while for niobium at 4 degrees Kelvin it is 271 n Ω , a factor of 30,000 smaller. This implies that, for a given voltage, the RF power required for a superconducting travelling-wave structure is smaller by a factor of 170 compared to a copper structure. Since the copper surface resistance varies as the square root of frequency, while the niobium value varies as the square of frequency, this scale factor goes as the $-3/2$ power of frequency; for frequencies above roughly 30 GHz, therefore, copper requires less RF power than 4 K niobium, while for frequencies below 30 GHz niobium wins out.

Unfortunately, the design of the structure becomes somewhat complicated due to the very low group velocity and very large filling time which is required. A factor of 30,000 improvement in

shunt impedance, for a given geometry, implies a factor of 30,000 increase in Q as well. Since the attenuation constant τ is fixed at its optimum value, and $\tau = \omega L / (2Q_w v_{\text{gr}})$, the group velocity must be reduced by a factor of 30,000 compared to the copper structure. This implies a very small hole radius a , which poses a variety of problems. In addition, the structure fill time $t_f = 2\tau Q_w / \omega$, also increases by a factor of 30,000. This poses an interesting dilemma – it takes less RF power to maintain the accelerating voltage, but more wall-plug energy (due to the long filling time) to attain that voltage in the first place.

An alternate approach to the superconducting constant-gradient structure is to use a very long structure, but again the factor of 30,000 which must be made up is formidable.

What about a copper structure and a superconducting structure which have the same length and the same group velocity? Let us once again assume that the copper structure has an optimal τ value of 1.26. The definition of τ indicates that the superconducting structure has a τ value which is 30,000 times smaller than the copper structure, very close to zero, and Equation 110 implies that:

$$V = \sqrt{2r_l L P_0 \tau}. \quad (131)$$

Since $\tau \propto 1/Q_w$, the expression above contains a factor of R/Q , and other factors which are geometrical only and hence the same for copper and for niobium. This implies that, for the same RF power and cell construction, the niobium structure achieves a higher voltage by a factor of $1.26/[1 - \exp(-1.26)]$, which is a factor of 1.76. While a 76% improvement is not to be ignored, it hardly justifies the use of niobium at liquid helium temperatures.

The operating mode in which superconductivity really comes into play is standing-wave accelerating structures. The standing-wave structure has similar properties to the constant-impedance, in that the power required to maintain a given accelerating voltage is low but the total energy required to attain that voltage is large due to the long structure filling time. This disadvantage can be overcome by filling the structure relatively less often and keeping the structure filled for longer periods (i.e., pulsing the linac at a lower repetition rate and having a longer RF pulse). In addition, since the π mode group velocity is not a function of the cavity aperture, the small a dimension required for an efficient travelling-wave superconducting structure is not needed for standing-wave.

7 Frequency Scaling of Cavity Parameters

This summary of the scaling of single-cavity parameters follows Wangler's approach [20].

Parameter	Normal Conducting	Superconducting
R_s	$\omega^{1/2}$	ω^2
R/Q	ω^0	ω^0
Q_w	$\omega^{-1/2}$	ω^{-2}
R_{cav}	$\omega^{-1/2}$	ω^{-2}
r_l	$\omega^{1/2}$	ω^{-1}

8 Limitations of Accelerator Structures

Given our present knowledge, the recipe for an accelerator which maximizes accelerating voltage for a certain input power is to maximize r_l and, in the case of travelling-wave accelerators, minimize v_{gr} . The former is accomplished by increasing the frequency for normal-conducting designs and

decreasing the frequency for superconducting designs; the latter is accomplished by minimizing the size of the hole between cells in the travelling-wave case. All of these optimizations increase the filling time required to achieve the desired voltage; as discussed above, this can be countered by reducing the number of fills per second and increasing the length of each fill.

In the real world, there are other limitations that prevent us from arbitrarily increasing the shunt impedance and decreasing the group velocity.

8.1 Steady-State Beam Loading

So far we have not even discussed the effect of the beam on the accelerator. The stored electromagnetic energy in the accelerator produces a longitudinal electric field; the electrons in the beam will be accelerated by that field and remove energy from it, thus reducing the accelerating field. We can rewrite Equation 105 to include this term:

$$\frac{dP}{dz} = -p_w(z) - I_{\text{beam}}E(z), \quad (132)$$

where I_{beam} is the average beam current during the beam pulse. We can now make use of the fact that the wall losses are related to the RF power and also to the gradient and the shunt impedance to require that

$$\frac{G_0^2(z)}{r_l} = \frac{\omega P(z)}{v_{\text{gr}}(z)Q_w}. \quad (133)$$

We now take a z derivative of both sides of Equation 133. In the case of a constant-impedance structure, the only component of the RHS which is a function of z is $P(z)$, thus

$$2G_0(z)\frac{dG_0}{dz} = \frac{r_l\omega}{v_{\text{gr}}Q_w}\frac{dP}{dz}. \quad (134)$$

This leads to the following set of substitutions:

$$\begin{aligned} 2G_0(z)\frac{dG_0}{dz} &= \frac{r_l\omega}{v_{\text{gr}}Q_w}(-p_w(z) - I_{\text{beam}}G_0(z)) \\ &= -\frac{r_l\omega}{v_{\text{gr}}Q_w}(I_{\text{beam}}G_0(z) + \frac{G_0^2(z)}{r_l}) \\ \frac{dG_0}{dz} &= -r_l\alpha_0 I_{\text{beam}} - \alpha_0 G_0(z). \end{aligned} \quad (135)$$

The resulting accelerating field along the cavity and total accelerating voltage are given by [21]:

$$\begin{aligned} G_0(z) &= \sqrt{2\alpha_0 r_l P_0} \exp(-\alpha_0 z) - I_{\text{beam}} r_l [1 - \exp(-\alpha_0 z)], \\ V &= \sqrt{r_l L P_0} \sqrt{2\tau} \frac{1 - \exp(-\tau)}{\tau} - r_l L I_{\text{beam}} \left[1 + \frac{1}{\tau} (1 - \exp(-\tau)) \right]. \end{aligned} \quad (136)$$

The first term on the RHS of the structure voltage equation can be recognized as the no-load accelerating voltage; the second term is the beam loading term. The coefficient in square brackets goes to 2 for $\tau \rightarrow 0$, and goes to 1 for $\tau \rightarrow \infty$.

A similar but more involved mathematical procedure can be performed on the constant-gradient structure, which yields an expression for the structure accelerating voltage [22]:

$$V = \sqrt{r_l L P_0} \sqrt{1 - \exp(-2\tau)} - \frac{1}{2} r_l L I_{\text{beam}} \left(1 - \frac{2\tau e^{-2\tau}}{1 - e^{-2\tau}} \right). \quad (137)$$

In this case, for small values of τ the coefficient in the parentheses on the RHS goes to 2τ , while for large values it goes to 1.

Note that, if $P_0 \rightarrow 0$, the first term on the RHS of equation 137 goes to zero, but the second term does not, and thus the beam is decelerated in an unpowered structure. Furthermore, the deceleration efficiency of the structure is determined by the length, shunt impedance, and attenuation factor. It would appear that a structure that efficiently generates an accelerating voltage from an external power source also efficiently generates a decelerating voltage from the beam itself.

8.1.1 Power Transfer Efficiency

Since electric power costs money, it's worthwhile to maximize the RF power transfer efficiency of an accelerating structure, which is defined to be the ratio of the power taken out of the structure by the beam to the power put into the structure by the power source. The former is $I_{\text{beam}}V$, while the latter is P_0 , thus we can compute the efficiency of a constant-gradient structure η :

$$\eta \equiv \frac{VI_{\text{beam}}}{P_0} = I_{\text{beam}} \sqrt{\frac{r_l L}{P_0}} \sqrt{1 - e^{-2\tau}} - \frac{1}{2} I_{\text{beam}}^2 r_l \frac{L}{P_0} \left(1 - \frac{2\tau e^{-2\tau}}{1 - e^{-2\tau}} \right). \quad (138)$$

The efficiency is conveniently quadratic in I_{beam} and concave-down, implying that a maximum efficiency occurs at

$$I_{\text{opt}} = \sqrt{\frac{P_0}{r_l L}} \frac{(1 - e^{-2\tau})^{3/2}}{1 - (1 + 2\tau)e^{-2\tau}}. \quad (139)$$

Though it is not exactly obvious from Equation 139, the optimal current varies from ∞ at $\tau = 0$ to $\sqrt{P_0/(Lr_l)}$ at $\tau = \infty$. Thus, the current-carrying capacity of the accelerator can be increased by increasing the input power (which also increases the gradient), decreasing the shunt impedance (which decreases the gradient), or decreasing the attenuation (which also decreases the gradient). At zero current the efficiency is obviously zero; at twice the optimal current the efficiency is again zero, since the beam loading is then so high that the beam achieves no net acceleration. For even higher currents (or for any current at zero input power) the efficiency is negative – the beam loses energy passing through the structure. As we shall see later, this can be a useful thing under some circumstances.

The maximum efficiency, which is the efficiency obtained at the maximum current, is:

$$\eta_{\text{max}} = \frac{1}{2} \frac{(1 - e^{-2\tau})^2}{1 - (1 + 2\tau)e^{-2\tau}}. \quad (140)$$

For small values of τ the maximum efficiency is 1, while for large values it asymptotically approaches 1/2. For a *pulsed* linac, there is an additional factor which contributes to the efficiency, which is the ratio of the beam time to the structure filling time. This factor is important because the structure must contain adequate RF energy to establish its accelerating field at any time the beam is present, including the time when the last electron passes through the structure. Since the group velocity in the structure is typically much slower than the beam velocity, the last electron passes through the structure in a tiny fraction of the fill time, but the structure still contains one fill time worth of stored energy, which will now never be used for acceleration. Thus the overall efficiency, including fill time considerations, is:

$$\eta_0 = \eta_{\text{max}}(\tau) \left[2 \frac{I_{\text{beam}}}{I_{\text{opt}}} - \frac{I_{\text{beam}}^2}{I_{\text{opt}}^2} \right] \frac{t_{\text{beam}}}{t_{\text{beam}} + t_f}. \quad (141)$$

8.1.2 Disadvantage of High Beam Loading

The accelerating structure is optimally loaded for a beam current I_{opt} ; at this current, all of the RF power goes either into the beam or into wall losses, and none is sent to the output coupler of the structure. This leads to an obvious question: why would anyone choose to run their accelerator at any current but I_{opt} ?

In order to answer the question, consider the expression for the loaded voltage of a constant-gradient structure, Equation 137. If the beam current varies, the accelerating voltage will also vary; the relevant coefficient, dV/dI_{beam} , is given by:

$$\frac{dV}{dI_{\text{beam}}} = -\frac{1}{2}r_l L \left(1 - \frac{2\tau e^{-2\tau}}{1 - e^{-2\tau}} \right). \quad (142)$$

If, however, we wish to see the *relative* change in the accelerating voltage given a certain *relative* change in the beam current, this is given by:

$$\frac{dV/V}{dI_{\text{beam}}/I_{\text{beam}}} = -\frac{1}{2}r_l L \left(1 - \frac{2\tau e^{-2\tau}}{1 - e^{-2\tau}} \right) \frac{I}{V}. \quad (143)$$

From this expression, we see that if we want a stable beam energy at the end of the linac, the relative stability of the beam current must be greater for larger absolute currents.

We can make the relationship even more explicit by noting that, for an accelerator which is operated with its optimal current, the unloaded voltage is exactly twice the loaded voltage. Thus, the loaded voltage and the loading voltage are equal and opposite. We can thus rewrite the expression for the voltage sensitivity:

$$\frac{dV/V}{dI_{\text{beam}}/I_{\text{beam}}} = -\frac{I}{I_{\text{opt}}}. \quad (144)$$

At the optimal current, a 1% fluctuation in the beam current leads to a 1% fluctuation in the accelerating voltage. At 1/5 of the optimal current, the same relative fluctuation (1%) leads to only 0.2% fluctuation in the accelerating voltage. The disadvantage of loaded linacs, then, is that they are “touchy,” or sensitive to small variations in the beam current, as compared to less-loaded linacs.

8.2 Short-Range Beam Loading

We have already seen that a continuous current I passing through an accelerating structure will induce a decelerating voltage that lowers the available energy gain in the structure. The relations which describe this deceleration in the steady state in a travelling wave structure are only valid 1 structure filling time after the current has been introduced. Other methods are needed to estimate the degree of beam loading which is present during the first structure fill time after beam turn-on.

8.2.1 Transient Beam Loading

Let us assume that our constant-impedance travelling-wave structure has been filled with RF but that the beam is initially absent; then, at time $t = 0$, the beam is introduced with uniform current $I(t > 0) = I_{\text{beam}}$. The decelerating voltage in a constant impedance structure between $t = 0$ and $t = t_f$ is given by [23]:

$$V(t) = \frac{r_l L I_{\text{beam}}}{\tau} \{ (1 + \tau)[1 - \exp(-\alpha_0 v_{\text{gr}} t)] + \alpha_0 v_{\text{gr}} t \exp(-\alpha_0 v_{\text{gr}} t) \}. \quad (145)$$

The above expression becomes equal to the deceleration in Equation 136 as $t \rightarrow t_f$ (at which time $v_{gr}t_f = L$ or $\alpha_0 v_{gr}t_f = \tau$). A similarly disgusting expression can be derived for a constant-gradient structure.

For a heavily-loaded linac, the transient voltage shown above is generally unacceptable because of the resulting energy spread along the bunch train. It is necessary to compensate the transient by making the accelerating voltage in the structure vary with time in such a way as to cancel the loading transient. Popular approaches to this compensation include injecting the beam before the structure is completely full of energy (“delta-t compensation”), or giving the incoming RF energy pulse a non-square shape (“delta-v compensation”). If the beam pulse is short compared to the filling time, the decelerating voltage will be approximately linear in time; in this case, it is possible to use structures with different accelerating frequencies to perform “delta-f” compensation: because the beam will not be synchronized to the off-frequency structures, the different bunches within a train will have different phases with respect to the crest in these structures; this technique can be used to cancel the linear voltage-time relationship.

8.2.2 Single-Bunch Beam Loading

Beam loading also occurs within a single bunch – loading from the head of the bunch decreases the accelerating voltage available to the particles in the tail of the bunch. Such loading is typically described as a *wakefield*, in other words an induced field that is left in the “wake” of the bunch.

Calculation of the longitudinal wakefield is a difficult process in that the bunch is usually short compared to any other dimension in the problem (RF wavelength, cell length, cell diameter, aperture radius), and thus there are very serious speed-of-light and causality issues which enter into the estimate. A convenient formula for estimation of the wakefield is provided by Bane *et. al* [24]: the decelerating electric field a distance z behind a particle of charge q is given by $E_{wf} = qW_L(z)$, where

$$W_L(z) = \frac{Zc}{\pi a^2} \exp\left(-\sqrt{\frac{z}{s_z}}\right), \text{ where} \quad (146)$$

$$s_z \approx 0.41 \frac{a^{1.8} g^{1.6}}{d^{2.4}},$$

and a is the aperture radius of the iris, g is the interior width of the cell, and d is the cell period (i.e., $g = d - h$, where h is the disc thickness); Z is the impedance of the medium, which is the canonical 377Ω for an evacuated accelerator structure.

8.2.3 Self-Loading

In the previous discussion of single-bunch loading, a formula for the decelerating voltage behind a charge is provided. That formula has a nonzero value for $z = 0$, indicating that a single electron can experience a wakefield from its own passage!

To understand this, consider a thought experiment proposed by Wangler [40], in which two particles with charge q and energy U_0 in each pass through a resonant cavity with frequency ν , and there is a longitudinal distance $c/2\nu$ between the two particles. At the start of the experiment there is no stored energy and therefore no voltage in the cavity. The first particle passes through the cavity and induces a voltage $-V_b$ in its wake (decelerating, as indicated by the sign). The voltage can only be established if the cavity stores some amount of energy $R_{cav}V_b^2$, which by conservation of energy must be removed from the first particle; $U_1 = U_0 - R_{cav}V_b^2$. We can also equate the loss of energy in the first particle to a decelerating voltage, $U_1 = U_0 - qV_1$. Note that V_1 is not required to be equal to V_b .

Now the second particle passes through, but at a time difference such that the field in the cavity is now an accelerating field with voltage V_b . Since the second charge is identical to the first, it will leave in its wake a decelerating voltage $-V_b$ which cancels the voltage left by the first particle, so after the second particle is gone the cavity is once again empty of energy and at zero voltage. However, the second particle has also been accelerated by the voltage left by the first, hence $U_2 = U_0 - qV_1 + qV_b$.

Now: since the cavity is empty, and energy is conserved, we know that the total energy in the two particles must be $2U_0$. Therefore, $qV_b - 2qV_1 = 0$, or $V_1 = V_b/2$. That is to say, a point charge experiences half of the energy loss indicated by the $z = 0$ wakefield. This is the *fundamental theorem of beam loading*.

8.2.4 Single-Bunch Loading Compensation

In general the effect of the single-bunch loading can be computed by convolving the wakefield W_L with the RMS bunch shape, but some semi-quantitative understanding of the interaction is always welcome. We can achieve this understanding by modelling the bunch with 2 macroparticles: one with charge $q/2$ at $z = 0$, and another with charge $q/2$ at $z = 2\sigma_z$, thus giving the overall distribution an RMS length of σ_z . The decelerating voltage experienced by each particle is given by:

$$\begin{aligned} V_1 &= \frac{Lq}{4}W_L(0), \\ V_2 &= \frac{Lq}{4}W_L(0) + \frac{Lq}{2}W_L(2\sigma_z), \\ &= \frac{Lq}{4}W_L(0)(1 + 2e^{-\Delta}), \end{aligned} \tag{147}$$

where $\Delta = \sqrt{2\sigma_z/s_x}$ and L is the structure length. The resulting average and RMS energy loss are given by:

$$\begin{aligned} \langle V \rangle &= \frac{Lq}{4}W_L(0)(1 + e^{-\Delta}), \\ \sigma_V &= \frac{Lq}{4}W_L(0)e^{-2\Delta}. \end{aligned} \tag{148}$$

Thus, both the average and the RMS deceleration are worse for short bunches than for long, which makes intuitive sense.

The RMS energy spread term arises from the fact that the second macroparticle is decelerated by the voltage left in the wake of the first macroparticle. We can compensate this effect by adjusting the relative phase of the accelerating field in such a way that the bunch is not riding on the crest of the wave, but slightly ahead of the crest. This will impart a greater voltage to the tail of the bunch than to the head, and through careful adjustment of the phase the effect of single-bunch loading and the effect of the acceleration can be made, in first order at least, to cancel.

What phase is required to compensate the beam loading? If the bunch centroid is accelerated at phase ϕ , then the first macroparticle is at phase $\phi - 2\pi\sigma_z/\lambda$, while the second is at phase $\phi + 2\pi\sigma_z/\lambda$. We can equate the change in energy gain by the two macroparticles to the change in deceleration due to wakefields:

$$V[\cos(\phi - 2\pi\sigma_z/\lambda) - \cos(\phi + 2\pi\sigma_z/\lambda)] = \frac{Lq}{2}W_L(0)e^{-\Delta}. \tag{149}$$

We can apply appropriate trigonometric identities and find:

$$\sin \phi = \frac{LqW_L(0)e^{-\Delta}}{4V} \frac{1}{\sin(2\pi\sigma_z/\lambda)}. \quad (150)$$

A reasonable approximation for $V \gg LqW_L(0)$, $\sigma_z \ll s_0$, and $\sigma_z \ll \lambda$ is:

$$\phi = \frac{LqW_L(0)}{8\pi V} \frac{\lambda}{\sigma_z}. \quad (151)$$

Since running far off-crest reduces the average energy gain even further, it is worthwhile to note that a high voltage, low charge, long bunch, or short RF wavelength all reduce the distance the bunch must be from the RF crest.

8.3 Transverse Wakefields and Beam Break-Up

In the previous section, we saw that the phenomenon of beam loading could be understood in terms of the electron beam exciting modes in the accelerating structure. In the long-term, only the modes with the highest Q values tend to survive (i.e., the fundamental accelerating mode), and therefore the effect can be modelled quantitatively with the r_l and Q values determined for the fundamental TM_{01} mode, while in the shorter term (such as the length of 1 bunch) a very large number of modes are present, interactions occur on a time scale which is short compared to the speed-of-light travel times about the RF structure, and a different model must be used to study beam loading.

We also know from previous discussion that the RF structure has an infinite number of passbands, corresponding to the fundamental TM_{01} mode and also to a wide variety of other modes. In particular, the structure has passbands for modes with a dipole characteristic – modes which have a nonzero deflecting field at $r = 0$. These include the TM_{11} mode, which has either a nonzero B_x or B_y component at the center of the structure, depending on the mode polarization. In general the frequency passbands for the dipole modes do not overlap those of the accelerating mode, and thus they are not excited by the RF power source. They can, however, be excited by the passage of a beam which has a nonzero dipole moment; a beam which passes off-axis through the structure fulfills this requirement. Therefore, a beam which passes off-axis through a structure will excite modes that produce a transverse deflection.

Like the longitudinal effect of beam loading, the transverse deflection effect (known as a *transverse wakefield*) has a long-term and a short-term characteristic. The long-term character is typically dominated by a small number of modes with large Q values, while the short-term includes a vast number of modes with low Q values.

8.3.1 The HEM “Hybrid” Mode

We have previously considered the travelling-wave modes in a cylindrical cavity which have no longitudinal magnetic field (TM modes), and those which have no longitudinal electric field (TE modes). Each mode has a distinct cutoff wave number, k_c , which depends upon the waveguide radius b , the mode number (corresponding to a particular Bessel function or derivative of a particular Bessel function), and the conducting-wall boundary conditions. For example, the TM_{11} mode requires that $k_c = z_{11}/b$, where $z_{11} \approx 3.832$, while the TE_{11} mode requires that $k_c = z'_{11}/b$, where $z'_{11} \approx 1.841$, and z'_{11} is the first zero of J'_1 . Because $z'_{11} \neq z_{11}$, it is generally impossible in regular waveguide to construct a wave in which TM_{11} and TE_{11} modes have the same dispersion relation (i.e, the same group velocity and phase velocity for a given frequency).

If, however, we restrict our consideration to the region $r \leq a$ in a disc-loaded waveguide, we see that the boundary conditions are somewhat different. Specifically, since we exclude the region

which includes $r = b$, we have excluded the boundary condition that $E_{r,\theta} = 0$ at $r = b$, but we must impose a boundary condition that $E_\theta = 0$ at $r = a$. The TM_{11} and TE_{11} expressions for E_θ are, respectively [25]:

$$\begin{aligned} E_{\theta,\text{TM}} &= \frac{ik}{k_c^2 r} E_0 J_1(k_c r) \sin \theta, \\ E_{\theta,\text{TE}} &= -\frac{\omega \epsilon}{k_c} H_0 J_1'(k_c r) \sin \theta. \end{aligned} \quad (152)$$

By inspection of Equation 152, it is evident that, in this special case, we can satisfy the boundary conditions of the system with a TM and TE mode which share a common cutoff wave number if the *amplitudes* of the two waves are forced to a particular linear relationship, specifically:

$$H_0 = E_0 \cdot \frac{ik}{\omega \epsilon k_c a} \frac{J_1(k_c a)}{J_1'(k_c a)}, \quad (153)$$

and the usual relationship between ω , k , and k_c obtains. The resulting mode – a mixture of TE and TM modes – is known as a “hybrid” mode, usually denoted HEM_{11} .

The expressions found so far for the HEM_{11} mode will apparently permit any value of k_c to satisfy the boundary conditions for $r \leq a$. In order to find the restrictions on k_c , it is necessary to construct a solution for Maxwell’s equations for $a < r \leq b$, within which the endcap and barrel boundary conditions are applied, and then to match the fields at $r = a$. When this is completed, the result is that the HEM_{11} field behaves rather like a TM_{11} mode, with a cutoff wave number equal to z_{11}/b . Thus, for any given accelerating structure, the frequency of the lowest-order dipole mode will be approximately 3.832/2.404 times the cutoff frequency for the accelerating TM_{01} mode. The hybrid character of the HEM mode can be seen in the fact that, for some structures, the HEM dispersion curve is not monotonic between zero and π .

The HEM mode is amenable to the formalism of shunt impedance, Q factors, etc., that was developed for the accelerating mode. For the deflecting modes, the definition of R_\perp/Q is altered slightly from the form in Section 2.2, since we are interested now in a deflecting voltage, thus [26]:

$$\frac{R_\perp}{Q} \equiv Z \frac{c^3}{\omega^3} \frac{|\int (\partial E_z / \partial x) \cos(\omega z / c) dz|^2}{\int E^2 dV}. \quad (154)$$

Rather than evaluate the integrals in Equation 154 in their monstrous glory, we simply note that in general $R_\perp/Q \approx 3.2 \times 10^{-8} \Omega \text{ sec } \omega_\perp$. Similarly, one can reasonably estimate that Q will be comparable to Q for an accelerating mode with the same frequency. Note that, unlike R/Q for the accelerating mode, the R/Q for the dipole mode actually increases with frequency.

8.3.2 Multi-Bunch Beam Break-Up

Consider a constant-impedance accelerating structure for which the lowest-order dipole mode has a transverse shunt impedance R_\perp/Q . The deflecting voltage induced a time t after the passage of a point charge q with transverse offset x is given by [27]

$$V_x(t) = \frac{xq\omega_\perp^2}{2c} \frac{R_\perp}{Q} \exp(i\omega_\perp t) \exp(-\omega_\perp t/2Q). \quad (155)$$

Now consider two bunches, separated by a time t_0 , which enter a long, quadrupole-focused accelerator with identical offsets x_0 . The first bunch will to lowest order approximation execute an unperturbed betatron oscillation – that is, the oscillation amplitude x/σ_x will remain a constant.

The second bunch will receive an impulse determined by Equation 155 from the first RF structures in the accelerator, which will become an offset 90° away in betatron phase. After 180° in betatron phase, the first and second bunch will have offsets given by $-x_0/\sigma_{x0}$, and the second bunch will receive an impulse proportional to this offset; however, since the initial impulse has also changed sign, the first and second impulses add coherently. As the beam continues along the linac, the second bunch will receive continued kicks from the first bunch, and the oscillations of the two bunches will always be phase locked such that the sign of the latest impulse is the same as the sign of the sum of all previous impulses. The second bunch will therefore experience runaway amplification of an initial offset. This is called the *beam break-up instability*.

In an actual multi-bunch system, the beam break-up phenomenon is somewhat more complicated than described above, for two correlated reasons. First, each bunch delivers an impulse which is experienced by trailing bunches according to $\exp[i\omega_\perp(t_2 - t_1)]$; second, the bunch spacing is typically harmonically related to the fundamental mode frequency, while the dipole mode frequencies are not. Therefore, the impulse delivered to a given bunch usually contains a large number of contributions which are in arbitrary phase with respect to one another. For very long bunch trains this can lead to a saturation, in which each bunch effectively only experiences the kick from the previous bunch and therefore all of the bunches past the saturation point get approximately the same kick.

It is clear that, for linear colliders, if the trailing bunches receive deflections which are comparable to the bunch sizes, then the “effective” size of the entire bunch train becomes too large to deliver good luminosity. Thus, multi-bunch beam break up must be limited to acceptable levels. Several mechanisms are available for such limitations:

Wait for the dipole mode to decay. The dipole mode, like the fundamental mode, decays with a time constant of $2Q/\omega_\perp$; after sufficient time, the dipole mode amplitude will be completely converted into heat in the structure. Unfortunately, this is about the same length of time required for the fundamental mode to decay into heat, so it’s almost never practical to wait that long between bunches.

Use a low accelerating frequency. Equation 155 shows that the deflection is proportional to $\omega_\perp^2 R_\perp/Q$, which in turn means that it is ultimately proportional to ω_\perp^3 ; and in general $\omega_\perp \propto \omega$. Therefore, reduction of the fundamental mode frequency will drastically reduce the impact of MB-BBU.

Use a low charge. This will also reduce the MB-BBU problem, but only linearly with charge.

Limit the injection jitter. Since the growth is a function of the injection jitter, but the acceptable end-linac amplitude is a function of the beam size, limiting the injection jitter to a tiny fraction of σ will keep MB-BBU to acceptable levels. Unfortunately, in practice it’s almost never possible to keep the injection jitter down to less than about 0.1σ , so the other approaches to limiting the problem must ensure that such a level is acceptable.

Detune the dipole mode frequencies. The synchronous dipole mode frequency is a function of the cell parameters d , a , b , and h . While d is typically fixed in an accelerator, the other parameters can be adjusted to cause the different cells to have different values of ω_\perp .

Consider for example a structure in which the parameters of the cells have been adjusted such that the distribution of dipole mode frequencies is Gaussian with standard deviation σ_ω . This will cause the dipole mode in the time domain to approximately obey:

$$V(t) \propto \exp(i\bar{\omega}_\perp t) \exp(-t^2/2\sigma_t^2), \quad (156)$$

where $\bar{\omega}_\perp$ is the mean dipole mode frequency and $\sigma_t = 1/\sigma_\omega$. If we consider for example a structure with a fundamental mode frequency of 1 GHz, and dipole mode frequencies around 1.6 GHz, then if we detune the dipole modes with an RMS bandwidth of 10% (0.16 GHz), then the amplitude of

the resulting wakefield will fall to $1/e$ of its initial value in $\sqrt{2}\sigma_t = \sqrt{2}/(2\pi \cdot 0.16\text{GHz})$, or about 1.4 nanoseconds. This is much shorter than the decay time of the fundamental mode due to wall losses, so one could imagine waiting a few times σ_t between bunches.

The detuning technique has a few fundamental limitations. First, it imposes strict construction tolerances on the RF structure. This is because the detuning basically relies upon the fact that, in a time t , the different frequencies have different phase advances; thus, a bunch with a positive initial offset will leave a deflecting impulse that is initially positive in all cells, and at a time t later some cells will have positive deflections and others will have negative deflections. Cell misalignments which are comparable to the bunch offset will break the resulting cancellation between deflections – if the driving bunch has a positive offset from the center of one cell and a negative offset from the center of another, the phase slippage between the two cells could cause them both to have positive deflections. Similar tight tolerances apply to the cell frequencies.

Second, a structure that is detuned in this manner cannot also be a constant-gradient structure. This is because the degrees of freedom usually used to achieve constant-gradient performance are the same ones needed for detuning. Structures that have been detuned for wakefield control usually have an unloaded gradient in the back of the structure that is higher than the gradient in the front (sometimes this is called “over-constant gradient”).

Third, a real accelerating structure has a finite number of cells. This means that a Gaussian detuning can be achieved only by having a large frequency spacing for frequencies far from the mean, and a narrow spacing for frequencies close to the mean. Since the frequency spectrum is discrete, there will be a minimum mode spacing $\Delta\omega, \text{min}$. A consequence of this is that the wakefield from one bunch will regenerate or recombine at a time $\Delta t \approx 1/\Delta\omega, \text{min}$ after the bunch passes through the structure. One way of improving the performance of detuned structures is to perform the detuning over several structures – if N_{struc} structures with N_{cell} cells each are used, it allows $N_{\text{struc}}N_{\text{cell}}$ dipole frequencies for detuning. This can only give a performance improvement if the N_{struc} structures are a small fraction of a betatron wavelength.

Damp the dipole modes. If the Q of the dipole modes can be artificially lowered below the wall-loss limit, then the wait time between bunches can be correspondingly reduced. Damping of high-frequency modes is usually accomplished by cutting slots into the structure which are cut off to the accelerating mode but not to the dipole mode (due to its higher frequency). The slots lead to matched waveguides that propagate the dipole modes away from the acceleration section of the structure, or to lossy material which can absorb the electromagnetic energy in the dipole mode excitations.

Accelerating structures with dipole-mode damping have been constructed with damping factors of a few (ie, Q is reduced by a factor of a few) to damping factors of several hundred. Damping an accelerating structure usually requires more complex design work than for undamped structures, and potentially requires introduction of lossy material of unknown pedigree into the ultra-clean accelerator environment. Furthermore, we have already seen that RF “cutoff” is a matter of degree rather than an absolute, on/off phenomenon. Some amount of the fundamental mode will pass through the cutoff slots in a damped structure; thus, a structure with dipole mode damping will also experience a reduction in “ Q ” for the fundamental mode, and will be required to absorb some amount of fundamental mode power in the damping system. Greater degrees of dipole damping obviously lead to larger losses in the fundamental mode, which can lead to unacceptably low shunt impedance or excessive heating of the dipole mode loads.

Strengthen the Focusing Lattice. By strengthening the quads in the accelerator, the RMS beam offsets are reduced and the RMS beam angular divergence is increased. Both of these will decrease the size of wakefield deflections relative to the natural angular size of the beam. Unfortunately, increasing the focusing strength of the quads also leads to more dispersive emittance growth from

the energy-dependent focusing of the quads.

8.3.3 Single Bunch Beam Break-Up

As with beam loading, transverse wakefields can have an impact on beam dynamics within a single bunch, and that impact is difficult to calculate from first principles. A convenient formula is that a point charge q with an offset x from the center of an accelerating structure with length L generates a deflecting voltage a distance z behind the charge given by $V = qxW_{\perp}(z)L$, where

$$W_{\perp}(z) \approx \frac{4Zcs_0}{\pi a^4} \left[1 - \left(1 + \sqrt{\frac{z}{s_{\perp}}} \right) \exp \left(-\sqrt{\frac{z}{s_{\perp}}} \right) \right], \text{ where} \quad (157)$$

$$s_{\perp} \equiv 0.169 \frac{a^{1.79} g^{0.38}}{d^{1.17}},$$

and a , g , d are defined as in Equation 146. Note that, at $z = 0$, $W'_{\perp} \equiv \partial W_{\perp} / \partial z \approx 2Zc / \pi a^4$.

To see quantitatively the effect of the short-range wakefield, we can follow the harmonic oscillator approach of Wangler [29]: once again, we imagine that the beam contains 2 macroparticles, with charge $q/2$ in each, with a separation between macroparticles of $2\sigma_z$. Let the transverse coordinates of the two particles be represented by $y_1(s)$ and $y_2(s)$, where s is the coordinate along the accelerator (thus distinguished in this case from z , the coordinate along the bunch); for simplicity, let us assume that the beam energy is a constant in the linac (ie, the system is configured such that the RF power and phase exactly compensates the beam loading). To complete our model, let k_{β} represent the focusing of the accelerator, and let us assume that it is a constant (as opposed to the actual, discrete focusing provided by quads in between accelerator elements), and let us assume that the bunch is so short that we can represent the transverse wakefield with $W_{\perp}(s) \approx W'_{\perp}(0)s$.

If the first macroparticle has an initial offset of y_0 and an initial angle of zero, then the equation of motion

$$\ddot{y}_1 + k_{\beta}^2 y_1 = 0 \quad (158)$$

can be solved by inspection: $y_1(s) = y_0 \cos(k_{\beta}s)$. Assuming that the same initial conditions are applied to the second particle, its equation of motion is different due to the presence of the transverse wakefield driving term:

$$\ddot{y}_2 + k_{\beta}^2 y_2 = y_1 \frac{W'_{\perp} q \sigma_z}{\gamma} \frac{e}{m_e c^2}. \quad (159)$$

Since $y_1 \propto \cos(k_{\beta}s)$, this equation of motion represents an undamped harmonic oscillator which is driven on resonance, with a solution

$$y_2(s) = y_0 \left[\cos(k_{\beta}s) + s \sin(k_{\beta}s) \frac{W'_{\perp} q \sigma_z}{2\gamma k_{\beta}} \frac{e}{m_e c^2} \right]. \quad (160)$$

As a function of s , then, the second macroparticle initially executes the same oscillation as the first, but it develops an oscillation in the opposite betatron phase which grows linearly with s . The oscillation can be reduced by reducing the wakefield, increasing the beam energy and gradient, or increasing the focusing strength of the lattice (represented here by the k_{β} coefficient).

8.3.4 BNS Damping: Smooth Focusing Approximation

A technique to defeat single-bunch beam break-up was proposed by Balakin, Novokhatsky, and Smirnov, and is thus known as ‘‘BNS Damping’’ [30]. Consider a system similar to the one described above, except with the exception that the two particles oscillate with different betatron wave

numbers $k_{\beta 1, \beta 2}$ in the linac's focusing. In this case the equation of motion of the second particle is given by

$$\ddot{y}_2 + k_{\beta 2}^2 y_2 = \frac{W'_\perp q \sigma_z}{\gamma} \frac{e}{m_e c^2} y_0 \cos(k_{\beta 1} s). \quad (161)$$

The equation of motion is now that of an undamped oscillator which is driven at a frequency other than its resonant frequency. The general solution is

$$y_2(s) = y_0 \cos(k_{\beta 2} s) + y_0 \frac{W'_\perp q \sigma_z}{\gamma} \frac{e}{m_e c^2} \frac{\cos(k_{\beta 1} s) - \cos(k_{\beta 2} s)}{k_{\beta 2}^2 - k_{\beta 1}^2}. \quad (162)$$

The trajectory difference between the first and second particle can be expressed as

$$y_2(s) - y_1(s) = y_0 \left(1 - \frac{W'_\perp q \sigma_z}{\gamma} \frac{e}{m_e c^2} \frac{1}{k_{\beta 2}^2 - k_{\beta 1}^2} \right) [\cos(k_{\beta 2} s) - \cos(k_{\beta 1} s)]. \quad (163)$$

In order to ensure that the emittance growth from wakefields is minimized, we need to arrange that $y_2 - y_1 = 0$ at the end of the linac. By inspection of Equation 163, this can be done in two ways: either $\cos(k_{\beta 2} s) - \cos(k_{\beta 1} s) = 0$ or else the expression in parentheses on the RHS of Equation 163 must be equal to zero.

“Beating” Regime of BNS Damping. The solution

$$\cos(k_{\beta 2} s) - \cos(k_{\beta 1} s) = 0 \quad (164)$$

is the “beating” regime of BNS damping. In this regime, the different oscillation frequencies of the two particles cause the wakefield to initially drive the second particle's oscillation, and then to cancel it again (as the phase relationship between the two particles changes sign). Note that the beating solution requires a variation in betatron wave numbers which is independent of the amplitude of the driving wakefield term:

$$k_{\beta 2} - k_{\beta 1} = \frac{2\pi n}{L_{\text{linac}}}, \quad (165)$$

where L_{linac} is the linac length and n is an integer. For the beating solution, the trailing particle can have either a shorter or a larger betatron wavelength.

“Autophasing” Regime of BNS Damping. If we require

$$1 - \frac{W'_\perp q \sigma_z}{\gamma} \frac{e}{m_e c^2} \frac{1}{k_{\beta 2}^2 - k_{\beta 1}^2} = 0, \quad (166)$$

then the wakefield driving will be locally cancelled (as opposed to the beating regime, in which the wakefield is only cancelled at the end of the linac or at a few discrete locations along the linac). In this case the change in betatron wavelength is a function of the wakefield strength, and the trailing particle must have a larger value of k_β (and thus a shorter betatron wavelength) in order to cancel the wakefield effect. This regime is known as the “autophasing” regime of BNS damping, and the wave number relationship is:

$$k_{\beta 2}^2 - k_{\beta 1}^2 = \frac{W'_\perp q \sigma_z}{\gamma} \frac{e}{m_e c^2}. \quad (167)$$

Achieving the Betatron Wavelength Variation. Any means of achieving a smooth variation in the focusing strength of the linac over the length of the bunch may be used to introduce BNS damping. One much-discussed method is RF Quadrupoles (RFQ's), which produce a transverse

focusing which varies in time. By phasing the RFQ such that the bunch's longitudinal centroid passes through on the RFQ's zero-crossing, the head and tail may be differentially focused.

Since RFQ's require additional space in the linac, another solution which preserves the linac packing fraction is desired. Typically, BNS damping is achieved by running the bunch off the RF crest, which introduces an energy spread which is correlated between the head and tail of the bunch.

If the bunch is run off-crest such as to produce a fractional RMS energy spread σ_E/E_{beam} , then particle 1 will have an energy given by $E_{\text{beam}} + \sigma_E$, and particle 2 will have an energy given by $E_{\text{beam}} - \sigma_E$. This means that, in terms of the design focusing strength, the two particles will experience focusing given by:

$$\begin{aligned} k_{\beta 1}^2 &= k_{\beta 2}(1 - \sigma_E/E_{\text{beam}}), \\ k_{\beta 2}^2 &= k_{\beta 2}(1 + \sigma_E/E_{\text{beam}}). \end{aligned} \quad (168)$$

Assuming that the required energy spread is small, we can rewrite the beating BNS condition as:

$$\sigma_E/E_{\text{beam}} = \frac{\pi n}{L_{\text{acc}} k_{\beta}}. \quad (169)$$

We can simplify this expression by replacing k_{β} with $2\pi/\lambda_{\beta}$, where λ_{β} is the betatron wavelength; we can further simplify by replacing λ_{β} with $L_{\text{cell}}/\nu_{\beta}$, where ν_{β} is the fractional tune of the linac betatron cell (i.e., $\nu_{\beta} = 0.25$ for 90° per cell betatron phase advance) and L_{cell} is the length of the cell; and we can replace L_{linac} with $L_{\text{cell}}N_{\beta}$, where N_{β} is the number of focusing cells in the linac:

$$\sigma_E/E_{\text{beam}} = \frac{n}{2N_{\beta}\nu_{\beta}}. \quad (170)$$

We can similarly estimate the energy spread required for the autophasing BNS condition:

$$\sigma_E/E_{\text{beam}} = \frac{1}{8\pi^2} \frac{W'_{\perp} q \sigma_z}{E_{\text{beam}}} \frac{L_{\text{cell}}^2}{\nu_{\beta}^2}. \quad (171)$$

As advertised, the autophasing condition requires that the energy of the tail be lower than the energy of the head, and the required energy spread is reduced when the quad spacing is reduced, the betatron phase advance per cell is increased, or the bunch length is reduced. Note that the combination of longitudinal and transverse wakefields requires some willingness to compromise, since the former are mitigated by long bunches and positioning the beam ahead of the RF crest, while the latter are mitigated by short bunches and positioning the beam behind the RF crest.

The derivation of the autophasing condition above is based on the assumption of smooth focusing, which is equivalent to an infinitely-long quadrupole with a constant magnetic field gradient. This is different from the reality in two important ways: first, a constant focusing gradient in one plane would constitute a constant *defocusing* gradient in the other plane, which would impractically drive the beam in the defocusing plane along a hyperbolic-cosine trajectory; second, a constant focusing gradient would require that the quadrupole be equal in length to the accelerator, which would not leave much room for the accelerating structures. Both of these issues can be addressed by replacing the assumed focusing system with discrete-focusing quadrupole elements in a FODO configuration. In such a situation, the autophasing condition is that the chromatic focusing of a FODO cell cancels the two-particle wakefield effect in both the horizontal and the vertical planes. The autophasing condition is slightly different under this circumstance [31]:

$$\sigma_E/E_{\text{beam}} = \frac{1}{16} \frac{W'_{\perp} q \sigma_z}{E_{\text{beam}}} \frac{L_{\text{cell}}^2}{\sin^2(\pi\nu_{\beta})}. \quad (172)$$

Note that, for small values of the phase advance per cell, the FODO expression for the autophasing energy spread is close to half that required for the smooth-focusing approximation.

8.3.5 Loading and Efficiency for Off-Crest Acceleration

In the preceding sections, we have seen that it is often advantageous or essential to accelerate beams off-crest in order to manage single-bunch effects. How does this influence the multi-bunch beam loading and the overall efficiency of acceleration?

In Section 8.1, we considered the case of accelerating a continuous train of on-crest bunches, and used conservation of energy to arrive at expressions for the loaded gradient and structure voltage. We found that both the gradient and the voltage could be expressed as the sum of two terms: the unloaded gradient/voltage (the term which is proportional to $\sqrt{P_0}$), and the beam loading term (which is proportional to I_{beam}). If we change the relative phase of the beam and the supplied RF power, ϕ , then, from the beam's point of view, the first term is reduced by a factor of $\cos \phi$, while the second term remains the same (since the beam is always in phase with itself). For a constant gradient structure, then:

$$\begin{aligned} V_{\text{beam}} &= V_{\text{noload}} \cos \phi + V_{\text{loading}} \\ &= \cos \phi \sqrt{r_l L P_0} \sqrt{1 - \exp(-2\tau)} - \frac{1}{2} r_l L I_{\text{beam}} \left(1 - \frac{2\tau e^{-2\tau}}{1 - e^{-2\tau}} \right). \end{aligned} \quad (173)$$

So what happens to the resulting efficiency? We can follow the formalism of Section 8.1 find the optimum current I_ϕ and maximum efficiency η_ϕ for off-crest acceleration in terms of the equivalent on-crest quantities I_{opt} and $\eta_{\text{max}}(\tau)$:

$$\begin{aligned} I_\phi &= I_{\text{opt}} \cos \phi, \\ \eta_\phi(\tau) &= \eta_{\text{max}}(\tau) \cos^2 \phi. \end{aligned} \quad (174)$$

Finally, since the efficiency remains a concave-down parabola with zeroes at $I_{\text{beam}} = 0$ and $I_{\text{beam}} = 2I_\phi$, we can write:

$$\eta = \eta_{\text{max}}(\tau) \cos^2 \phi \left[2 \frac{I_{\text{beam}}}{I_{\text{opt}} \cos \phi} - \frac{I_{\text{beam}}^2}{I_{\text{opt}}^2 \cos^2 \phi} \right] \frac{t_{\text{beam}}}{t_{\text{beam}} + t_f}, \quad (175)$$

where, again, I_{opt} and $\eta_{\text{max}}(\tau)$ are the optimum beam current and maximum efficiency for on-crest acceleration.

Note that we can now write the beam voltage, V_{beam} , and the variation of the voltage over the beam's longitudinal extent, $V'_{\text{beam}} \equiv dV_{\text{beam}}/dz$, as follows:

$$\begin{aligned} V_{\text{beam}} &= \cos \phi V_{\text{noload}} + V_{\text{loading,mb}} + V_{\text{loading,sb}} \\ V'_{\text{beam}} &= k_z \sin \phi V_{\text{noload}} + V'_{\text{loading,sb}}, \end{aligned} \quad (176)$$

where $V_{\text{loading,sb}}$ is the single-bunch loading discussed in Section 8.2.2. Because the multi-bunch loading is always in phase with the beam (ie, $\phi_{\text{loading,mb}} = \pi$), and the derivative of the voltage V' is proportional to the sine of the phase angle, it follows that there is no V' contribution from the multi-bunch loading. The interesting result of this is the following: when computing the phase angle required for BNS damping or single-bunch loading compensation, it is the *unloaded* voltage or gradient which should be used. This is because the actual goal of running off-crest is to introduce a certain value of V' from the supplied RF power, and we see from the relation above that V' is proportional to the unloaded voltage.

8.4 Limitations to the Accelerating Gradient

Because real estate costs money (and, in most places an accelerator physicist would like to live and work, it can be a great deal of money), there is an inevitable evolutionary pressure to ever-higher accelerating gradients (and thus shorter accelerators). Consequently, understanding the limitations in gradient for an accelerator are critical to selecting a design.

We have already seen that, for a given accelerator, increasing the gradient requires an increase in the input power; thus, the gradient can be limited by economic considerations (cost of electrical energy, or “wall-plug power,” from the grid, or cost of RF power sources and infrastructure). Indeed, for a given accelerator structure design there is an optimum gradient at which the costs which are linear in system length are equal to the costs which are linear in system power. Even disregarding the economics of the situation, there are physical phenomena which limit the gradient which is practically achievable in an accelerating structure.

8.4.1 Field Emission and Dark Current Capture

Field emission is the emission of electrons from the surface of a metal exposed to a strong electric field. The electrons near the surface of the accelerating structure are, according to classical physics, trapped: the binding energies of the electrons are on the order of eV per angstrom (or 10^4 MeV/meter), while the typical electric fields in the accelerator are a few tens of MeV per meter. Thus, classical physics indicates that no field emission should occur.

The fact that field emission does occur can be attributed to two phenomena [34]:

- The surface of the accelerator structure contains a number of microscopic imperfections (bumps, scratches, etc.). Because the conducting surface is an equipotential, the electric field at a small imperfection will be enhanced by a factor (usually denoted, unfortunately, as β) which can be as large as a few hundred.
- The phenomenon of quantum tunneling permits electrons in the walls of the structure to be emitted by surface fields which are much smaller than those which are classically required.

The combination of these features leads to an average field-emitted current given by [34]:

$$\bar{I}_{\text{FE}} \approx \frac{5.7 \times 10^{-12} \times 10^{4.52/\sqrt{\Phi}} A_e (\beta E_s)^{2.5}}{\Phi^{1.75}} \times \exp\left(-\frac{6.53 \times 10^9 \Phi^{1.5}}{\beta E_s}\right), \quad (177)$$

where \bar{I}_{FE} is the field-emitted current in amperes, A_e is the emitter area in meters, E_s is the unenhanced surface field in eV/meter, Φ is the metal’s work function in eV (about 4.5 eV for copper and about 4.3 eV for niobium).

Electrons liberated by the field emission process typically have fairly low energies and velocities, but they are able to interact with the accelerating field of the structure. This field will accelerate the liberated electrons, just as it does the beam electrons. If the accelerating gradient is low, the electrons will not build up much velocity and will be overtaken by the accelerating wave (which has a phase velocity of c), and thus decelerated. If, however, the gradient is sufficiently large, the field-emitted electrons will be accelerated to near-light speed in a single half-cycle of the RF or even less. In this case, the field-emitted electrons can become synchronous with the RF, in which case they are called dark current. The accelerating gradient required to capture dark current is given by:

$$E_{\text{capture}} = 1.6 \text{ MeV} \frac{\omega}{2\pi c}. \quad (178)$$

Even captured dark current can rarely propagate far down a linear accelerator. This is because the dark current from a given structure will get an acceleration comparable to that experienced by the design beam, but the initial energy of the dark current is almost zero, while the initial energy of the beam in a given structure is several (or many) GeV; therefore, in the main linac of a linear collider, the beam energy is very different from the dark current energy at the linac quadrupole magnets, and the dark current will be over-focused and “blown out” of the accelerator. Nonetheless, no accelerator can operate with excessive dark current: it can drive wakefields, blind beam diagnostics, and cause radiation (and radiation damage). In a superconducting accelerator, dark current is even worse: it absorbs the energy stored in an RF cavity and then, after being “blown out,” dumps it into the cryo system, thus enlarging the heat load on the cryo system. Linac structures for linear colliders must typically be operated below the dark current capture gradient, or must be extremely clean and smooth to suppress formation of dark current, or both.

8.4.2 RF Breakdown

RF breakdown is a phenomenon in which a plasma discharge (or spark) forms in an accelerating structure. The discharge causes an impedance mismatch in the structure; as a result, the RF power propagating from the input coupler is reflected at the spark location.

It goes without saying that an RF structure cannot provide decent acceleration under breakdown conditions, since the incoming RF power is reflected back out the input coupler. In addition, such breakdowns are correlated with small pits that appear in the structure; therefore, excessive breakdowns can lead to permanent structure damage.

Interestingly, much of the physics of RF breakdown is still not very well understood. It is generally agreed that RF breakdown is initiated by field emission: if the field-emitted current becomes large enough, the heat dissipated by the current flow in the structure will cause vaporization of material, leading to plasma formation and breakdown [32]. Furthermore, breakdowns are typically accompanied by a dramatic increase in field-emission current. It is also observed that the rate and severity of breakdowns increases rapidly as the accelerating gradient of a structure is increased, which is consistent with the rapid rise in field-emission current as a function of electric field.

Because RF structures usually arrive from manufacturing in a state which encourages RF breakdowns – highly gas-loaded surfaces, lots of nucleation points for field emission, etc. – the structures are usually prepared for operation through a semi-controlled set of RF breakdowns known as “RF processing.” During RF processing the total pulse energy (peak power and pulse length) is set to a low value and slowly increased. At a given RF power level, a number of breakdowns occur, and eventually the breakdown rate is lowered; the pulse energy is then increased, and the cycle is repeated. It is believed that, by using RF pulses with low energy, field-emission nucleation sites are vaporized when they are just barely above the breakdown threshold, and thus a minimum amount of energy is dissipated in the breakdown event; if the breakdown occurred when a larger RF pulse was used, the resulting increase in energy transferred to the breakdown would do more damage to the surface of the structure.

Although in principle the procedure above should permit any structure to be processed to just about any gradient if sufficient processing time is allocated, in practice it is observed that any given structure has a maximum field above which processing is ineffective. This may be due to the formation of secondary nucleation points when a primary nucleation point is vaporized; a breakdown at high power will tend to absorb more energy than one at low power, which can cause molten or vaporized material to be expelled from the breakdown site and re-solidify elsewhere; if the newly-formed site is as large or larger than the original one, then the achievable gradient is lower after the breakdown than before, and processing cannot make further progress once this becomes

the statistical norm. A rough empirical formula for the maximum achievable surface electric field in a structure is [32]:

$$E_{s, \max} \approx 195 \text{ MV/m} [\nu(\text{GHz})]^{1/2}. \quad (179)$$

The accelerating gradient is usually a factor of 2-4 lower than the peak surface field.

It is not understood why the breakdown limit should have a frequency dependence, since the basic mechanism of field emission is independent of frequency. It has also been observed that RF breakdown can be mitigated through use of lower group velocity structures, or a reduced pulse length (the breakdown limit seems to scale roughly as $t^{-1/4}$ [33]).

Because of the very low surface resistance and temperature of superconducting structures, true RF breakdown is almost exclusively a phenomenon of normal-conducting structures.

8.4.3 Pulsed Heating

When RF power is introduced into an accelerating structure, the resulting current flows in the structure walls will cause the temperature of the surface conducting layer to rise. Typically the surface layer's temperature will be restored to its original value by thermal conductivity between RF pulses. Thus, the structure will experience continual thermal cycling during pulsed operation. Such thermal cycles can cause structure damage through fatigue, which causes cracking and roughening of the surface. One might expect, then, that a structure which is continually operated with a large pulsed temperature rise will begin to develop RF breakdown problems, and indeed there is an observed relationship between these two effects.

The temperature at the surface of a structure after a time t is usually dominated by the heat deposition per unit area and the diffusion distance into the material (rather than the skin depth). This temperature rise is given by [35]:

$$\Delta T(t) = \frac{1}{\mathcal{D}c_\epsilon\sqrt{\pi\alpha_d}} \int_0^t \frac{dt'}{\sqrt{t-t'}} \frac{dP(t')}{dA}, \quad (180)$$

where \mathcal{D} is the density, c_ϵ is the specific heat at constant strain, k_T is the thermal conductivity, and $\alpha_d \equiv k_T/(\rho c_\epsilon)$ is the thermal diffusivity; for copper, these quantities are $8.95 \times 10^3 \text{ kg/m}^3$, 385 J/kg/K , 390 W/m/K , and $1.13 \times 10^{-4} \text{ m}^2/\text{second}$, respectively. The power dissipated per unit area is given by Equation 69. If we consider a single-cell cavity operating in the TM_{010} mode, the peak temperature rise in the cavity will be given by:

$$\Delta T(t)_{\max} = \sqrt{t} \frac{0.0587\epsilon^2}{T^2 \mathcal{D}c_\epsilon\sqrt{\pi\alpha_d}} b^2 \omega^2 R_s(\omega) G^2, \quad (181)$$

where G and T are the on-crest *unloaded* accelerating gradient and the transit factor, respectively.

How severe a limitation is Equation 181 in an accelerating structure? Let us consider again our example structure with $\omega = 2\pi \times 1.33 \text{ GHz}$, $b = 10 \text{ cm}$, and $\psi = 158^\circ$ (yielding a transit-angle factor of 0.7119); a 1 msec pulse with an unloaded gradient of 10 MV/m will yield a temperature rise of approximately 2.9 degrees Kelvin. Because b and ω are generally inversely proportional to one another, all of the remaining frequency dependence is in the surface resistance term. Thus, we expect that the surface heating will increase with the square root of frequency as well as the square root of pulse length.

8.4.4 Gradient Limit in Superconducting Structures

The gradient of a superconducting cavity is typically set by the acceptable heat load in the cryo system: due to its low but finite surface resistance, a superconducting cavity dissipates a small

amount of heat when energy is stored in it, and the heat load scales as the square of the gradient. Furthermore, as the gradient in such a cavity is increased, the Q value of the cavity begins to drop for reasons not well known (although increased dark current is a possible candidate). This “high-gradient Q droop” causes the heat load to increase faster than the square of the gradient above some threshold. Note, however, that the heat load is a “soft” limit to the gradient, in principle amenable to correction by adding more cooling capacity.

A harder gradient limit is set by the distressing tendency of superconductors to enter the normal conducting state (“quench”) when a high surface field is applied to them. For elemental niobium the critical magnetic field is 0.2 Tesla. For a TM_{010} cavity the relationship between the accelerating and the surface magnetic field is given by Equation 71. For a pillbox cavity, therefore, $H_{\max} = E_0/ZJ'_{0,\max}$, and since $J'_{0,\max} = 0.582$, the maximum accelerating field achievable in a superconducting cavity is given by:

$$E_{\max} = 1.72ZH_c. \quad (182)$$

For $H_c = 0.2 \text{ T}/\mu = 1.6 \times 10^5$ amperes per meter, one obtains an estimate of 104 MV/meter as a limiting field, independent of frequency. When the transit-time effect is taken into effect (assuming a standing-wave structure operating in π -mode), the limiting gradient is around 66 MV/m, and more sophisticated calculations which take into account the actual shape of the cavity and the iris apertures arrive at an estimate of about 50 MV/m as a gradient limit. In principle, the peak magnetic field could conceivably exceed the DC critical field by 20% or so [41], but in practice no superconducting linear accelerator structure has even achieved the level of 50 MV/meter implied by the DC limit.

9 Optimization, or, Linear Accelerator Numerology

With the information collected above, we can begin to consider what the properties of a linear collider’s main linac should be. The task is a daunting one in that no less than 19 parameters are required to specify the linac:

- Global parameters: total energy gain V_{tot} , repetition rate f , and total length L_{linac} of the linac (3)
- RF Power source parameters: Frequency ω and number of power sources N_{klys} (2)
- RF structure parameters: Structure radius b , iris radius a , cell length d , structure length L , iris thickness h (5)
- Beam parameters: Average current $\langle I \rangle$, bunch train duration t_{beam} , charge per bunch q , bunch length σ_z (4)
- Phase parameters: average phase $\bar{\phi}$, RMS deviation from average phase σ_ϕ (2)
- Given the parameters above, the power required for the linac P_{tot} can be deduced (1)
- Quadrupole lattice: quad spacing L_q , phase advance per betatron cell ν_β (2).

The total energy gain V_{tot} is set by experimental requirements on the center-of-mass energy, and can be taken to be a constant at about 240 GeV (since the desired center-of-mass energy is 500 GeV, and the beam enters the linac with a nonzero energy). Similarly, the beam average power must be around 10 MW to provide the desired luminosity; 10 MW/240 GeV yields an average current $\langle I \rangle$ of 42 microamperes. The bunch length σ_z must be approximately matched to the

vertical depth-of-focus at the IP, to avoid unwanted emittance dilution from the “hourglass effect” – this yields a typical RMS bunch length of $200 \mu\text{m}$. The bunch charge q is limited to typical values of 10^{10} particles, or 1.6 nC, to limit detector backgrounds from the beam-beam effect. Finally, for economic reasons we wish to minimize the total linac length L_{linac} and the linac average power P_{tot} , and for now we will ignore the number of power sources N_{klys} . This adds up to 4 “given” values, 2 minimization parameters, and one parameter which can be ignored – out of 19!

In order to add some insight, let us consider the most successful linear accelerator in history: the SLAC linac.

9.1 The SLAC Linac

The SLAC linac was constructed in the early 1960’s and has received a handful of modest upgrades since then. The present configuration includes approximately 300 quadrupole magnets and 1000 constant-gradient accelerating structures. The structure parameters are given in Table 1. Note that the parameters in Table 1 include an additional 6 of the linac parameters (a , b , d , h , ω , L); the remaining parameters in the table are completely determined by the choices made for the aforementioned 6.

Table 1: Parameters of the SLAC constant-gradient travelling-wave structure. From G.A. Loew, R.B. Neal, “Accelerating Structures,” in *Linear Accelerators* (edited by P. Lapostolle and A. Septier (1970)).

Parameter	symbol	Units	Value
Frequency	$\omega/2\pi$	Hz	2856 MHz
Length	L	m	3.048
Cell radius	b	cm	4.17–4.09
Iris radius	a	cm	1.31–0.96
Cell length	d	cm	3.50
Phase shift per cell	ψ	-	$2\pi/3$
Disc thickness	h	cm	0.584
Quality Factor	Q	-	13,000
Shunt impedance per meter	r_l	$\text{M}\Omega/\text{m}$	52–60
Filling Time	t_f	nsec	830
Group Velocity	v_{gr}	$\%c$	2.0–0.65
Attenuation	τ	“nepers”	0.57

The SLAC structure is routinely operated at unloaded on-crest gradients of about 21 MeV/meter, or 63 MV per structure. Assuming a typical shunt impedance of $56 \text{ M}\Omega$, the input power needed to sustain this gradient is about 35 MV per structure, and we will use this as our “baseline design” parameters.

Although the lattice parameters for the linac FODO lattice vary, the typical quadrupole spacing is 12 meters (1 quad per 4 RF structures), with a phase advance per FODO cell that tapers to 45° at the end of the linac. For our purposes today, we will limit ourselves to a quadrupole spacing of 12 meters and a phase advance per FODO cell of 90° .

9.2 SLAC Structure for TeV-Scale Linear Collider

Let us use the example of the SLAC structure in parameterizing a linac for a linear collider.

First, the average current $\langle I \rangle$ is 42 microamperes. The single-bunch charge is 1.6 nC, so the time-averaged bunch rate must be 1 bunch per 38 microseconds. It is clear that this beam cannot be efficiently accelerated by this linac in a CW (continuous-wave) manner, because the inter-bunch time would exceed the structure filling time. Apparently it will be necessary to pulse the linac and to group the bunches closer in time – into bunch trains.

9.2.1 Single Bunch Requirements

Let us assume that the SLAC structure runs at an unloaded on-crest voltage of 63 MV. The longitudinal wake factor, $W_L(0) = Z_0 c / \pi a^2$, can be computed for a typical iris aperture of 1.135 cm, yielding a wake factor of $2.8 \times 10^{14} \text{ V/C/m}$. Equation 151 indicates that the single-bunch beam loading can be compensated, to lowest order, with a mean phase angle of about 26° . This sounds like a lot, but bear in mind that $\cos(26^\circ)$ is about 0.90, so the actual reduction in accelerating field seen by the beam is quite acceptable.

In a similar vein, the slope of the transverse wake potential, given by $2Z_0 c / \pi a^4$, or about $4.3 \times 10^{18} \text{ V/C/m}^3$. Assuming, as we do, that the FODO cell length and phase advance are a constant, the RMS correlated energy spread required for “autophasing” is given by Equation 172 to be 50 MeV. At the 240 GeV end of the linac, this corresponds to 0.02% of the beam energy, while at the 10 GeV end it is 0.5%.

Given the parameters so far discussed, the simplest way to introduce the “autophasing” energy spread is to run the beam behind the RF crest (so that the head will gain more energy than the tail) until such time as a correlated energy spread of 50 MeV is achieved, and henceforward run the beam 26° ahead of the crest (to maintain, more or less, the resulting 50 MeV energy spread). Algebraically, the length of linac required to introduce the autophasing energy spread is given by:

$$L_{\text{BNS-in}} = \frac{\sigma_E}{V} \frac{L_{\text{struc}}}{k_z \sigma_z \sin(\phi_{\text{BNS-in}})}. \quad (183)$$

A further refinement is that the single-bunch loading provides a correlated energy spread with the correct sign for autophasing; thus, for example, running the bunch on-crest will result in an energy spread equivalent to running a zero-loading bunch 26° off-crest. If we consider the expression above, for 26° off-crest and zero loading the required 50 MeV energy spread is introduced in about 450 meters; this means that, given the single-bunch loading we expect, we can run with the beam at the RF crest for 450 meters and achieve the autophasing energy spread, then switch to accelerating 26° ahead of the crest to approximately maintain the autophasing condition.

9.2.2 Bunch Train Requirements

For on-crest operation at 63 MV unloaded voltage, 35 MW input power, equation 139 shows that an average current of 0.81 amperes will fully load the structure. Although this is the most efficient, full loading would make linac operation unstable (recall that at full loading, 1% variation in beam current results in 1% variation in energy gain) and would reduce the loaded voltage by 50% (resulting in an uncomfortably-long linac). The SLAC linac has been operated at 20% loading in recent years, so we can take this as a “baseline” and consider a beam current of about 0.16 amperes.

With a single-bunch charge of 1.6 nC, an average current of 0.16 amperes corresponds to a bunch spacing of 10 nanoseconds. A bunch spacing of 29 RF periods corresponds to 10.1 nsec, so this is probably a good spacing to choose.

The dipole-mode frequency of the SLAC structure is approximately 4.1 GHz. If the dipole modes are gaussian-detuned with a 10% bandwidth and a 3 sigma cutoff, the resulting wakefield

amplitude as a function of time is shown in Figure 12. Although in principle the wakefield should fall to utterly negligible levels in 10.1 nanoseconds, given the parameters above, the small number of modes (43) causes a relatively quick regeneration of the wake, such that the reduction factor from $t = 0$ to $t = 10.1$ nanoseconds is only about a factor of 20. Although a factor of 20 reduction is helpful, it may not be sufficient to guarantee that the multi-bunch emittance growth is acceptable. An additional factor can be achieved, if necessary, by using RF structures with different HOM frequencies. Such an approach effectively increases the total number of modes applied to the detuning process.

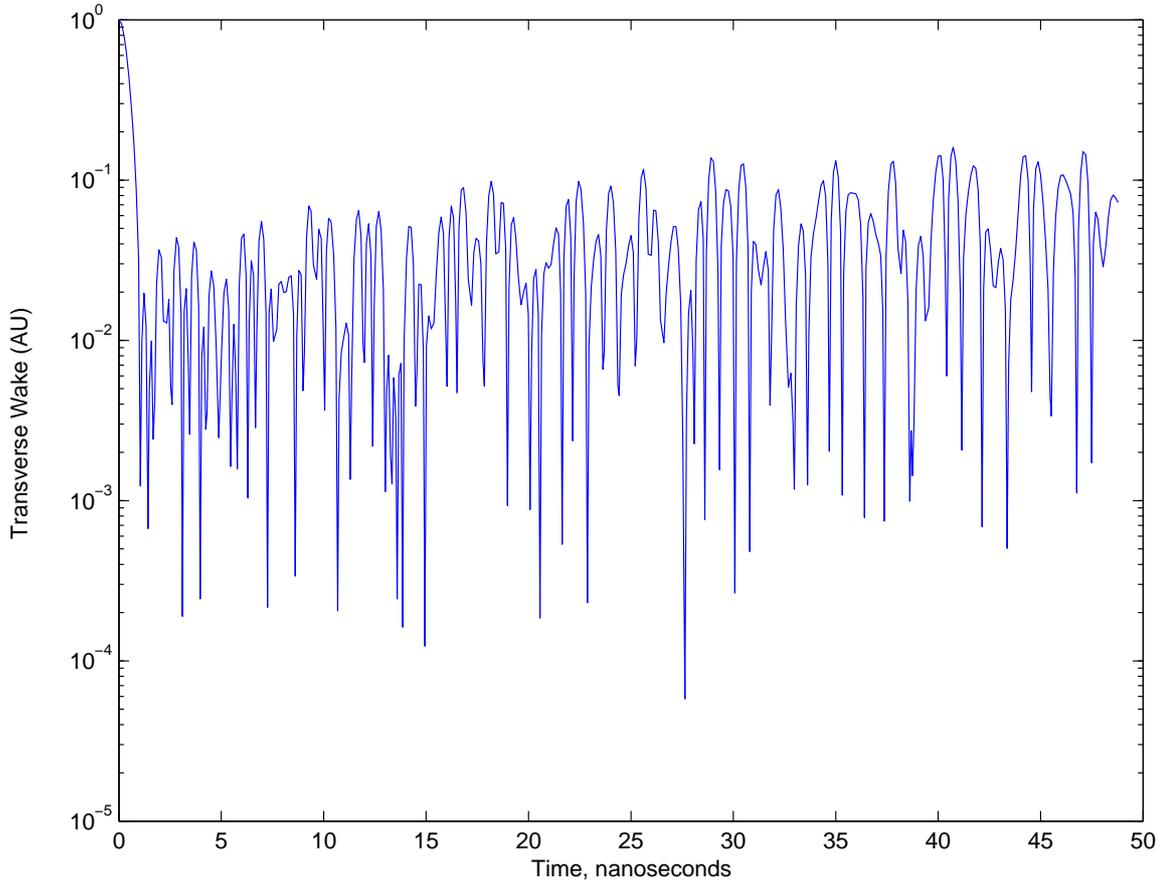


Figure 12: Envelope of the long-range transverse wakefield, relative to the $t = 0$ wake, for an 86-cell S-band structure with HOM detuning parameters as described in the text.

9.2.3 Pulsed Heating, Train Length, and Repetition Rate

Thus far we have determined that the bunch trains in the SLAC structure should consist of bunches separated by approximately 29 RF periods to attain a beam loading factor of 20%. We have not, however, determined how long each train should be. In principle the RF efficiency is maximized when the bunch train length is maximized (since the ratio of the filling time to the beam time is reduced), therefore we want to understand the limitations on the available bunch train length.

In section 8.4.3, we saw that for a 1.33 GHz structure at 10 MeV/meter gradient, a 1 millisecond RF pulse will generate a temperature rise of about 2.9 kelvin. Extrapolating from that structure

to 2.856 GHz and 21 MeV/meter, we find that the same 1 msec pulse yields a temperature rise of about 18 kelvins. This constitutes the ultimate limit to the length of the bunch train, and reduces the energy lost to filling time to negligible levels.

A more drastic limitation on the bunch train length is the available pulse length from a high-powered RF source. The “highest recorded pulse energy in a multi-megawatt klystron” was produced at 3.0 GHz, with 160 MW peak power and 3.0 μ sec pulse length. Assuming we can trade off pulse length for peak power, and that we will accept a number of klystrons equal to the number of RF structures (but not greater), then this klystron could be reconfigured to produce the required 35 MW for 13.7 μ sec. This would allow 830 nanoseconds for filling the structure, followed by 12.9 μ sec for the bunch train. With a bunch spacing of 10.1 nsec, the bunch train would contain about 1280 bunches. Furthermore, at the speed of light, 12.9 μ sec corresponds to 3.8 km train length. This means that a large damping ring circumference is needed, but the size is much less than HERA-e (at 6.3 km circumference), to say nothing of LEP (almost 27 km circumference).

The average bunch rate required for the facility is 1 bunch per 38 μ sec, or 2.6×10^4 bunches per second; with 1280 bunches per train, this means that a repetition rate of about 20 trains per second is required. Of course, the train length and repetition rate can also be traded against one another. At 120 trains per second, the train is reduced to about 210 bunches and the damping ring circumference is reduced to about 2/3 of a kilometer.

9.2.4 Putting it All Together

The “SLAC-LC” uses 3 meter long, 2856 MHz RF structures operating at an unloaded gradient of 21 MeV/meter set in a FODO lattice with 1 quad per 4 RF structures. The power required to achieve the desired gradient is 35 MW/structure. With a beam current of 0.16 amperes (corresponding to 20% loading) and a beam-to-RF phase of 26° (required for single-bunch loading compensation), the loaded gradient in the linac is 20% lower than the unloaded, or about 17 MeV/meter. Half of the reduction is from the off-crest running for single-bunch loading, and half is from steady-state loading. The 0.16 ampere beam current corresponds to 1 bunch of 1.6 nC every 29 RF cycles (10.1 nsec).

For a structure with an attenuation coefficient $\tau = 0.57$, the maximum RF-to-beam efficiency is 73%. In this case, with the off-crest operation and the limited beam loading, the efficiency from steady-state operation is just under 24%. Assuming a maximal bunch train length set by the power source issued discussed above, the beamtime-filltime factor is 0.94; if a shorter train and higher linac pulse repetition rate is selected, the factor is reduced (for example, to 0.72 at 120 linac pulses per second). Thus, the overall efficiency of acceleration is between 17% and 23%.

The main drawback of the S-band linac is its combination of low gradient and low shunt impedance. At 17 MeV/meter, the length of each linac is over 14 km. In order to double the unloaded gradient from 21 MeV/meter to 42 MeV/meter, the input power must quadruple from 35 MW to 140 MW. At this point, the RF source described above becomes limited to about 3 μ sec pulse length, constraining us to use the less-efficient 120 Hz operational model. Furthermore, the beam average current must be doubled to maintain the same loading factor. Since single bunch charge cannot be increased due to beam-beam forces at the IP, we must instead double the number of bunches per train and halve their spacing. This makes the cumulative effects of the long-range wakefields significantly dicier. Furthermore, experience at SLAC suggests that gradients of around 30 MeV/meter are the limit for reliable operation of this structure.

In the interest of a cheaper linear collider, we can go away from the SLAC structure in two directions: we can increase the achievable gradient and shunt impedance by going to a higher frequency structure (yielding in the main a shorter linac), or we can switch to a lower-frequency,

superconducting structure operating in a standing-wave mode (yielding in the main a more energy-efficient linac).

10 Recommended Reading

The “bible” of linacs is *Linear Accelerators*, edited by Pierre Lapostolle and Albert Septier and published in 1970. This volume contains articles published by the luminaries of the field at that time on a vast array of topics in electron, proton, and ion linacs. The book is hard to find and harder to carry due to its size. Warning: several statements and assertions in this book have proved to be untrue.

Another useful book on the topic is *RF Linear Accelerators*, written by Thomas Wangler and published in 1998. This book has the benefit of a single author, so its notation and style are consistent throughout the text, and it is a far more modern book than Lapostolle and Septier. Its main drawback, in my opinion, is that it left more of the formalism to the imagination than a mathematically-weak individual (such as myself) would prefer.

Absolutely indispensable for general accelerator physics is *Particle Accelerator Physics*, by Helmut Wiedemann (now director of USPAS), which includes the main results on waveguides, cavities, and accelerating structures.

General mathematics and physics texts: I recommend *Elementary Applied Differential Equations* by Richard Haberman (1987), *Foundations of Electromagnetic Theory*, by John Reitz, Frederick Milford, and Robert Christy (1980), and of course *Classical Electrodynamics*, By John David Jackson (1975). The latter is especially recommended for its information on the different systems of units and dimensions used in electrodynamics problems, and has an entire section on RF systems written in an ingenious combination of MKSA and Gaussian units. I’ve heard that there is a third edition, published in 1990, which moves even more in the heretical direction of MKSA units for theoretical electrodynamics problems, but I have no experience with it.

11 Acknowledgements

Several accelerator luminaries have given me the benefit of their wisdom, experience, and patience in preparation of this set of notes: Greg Loew, Zenghai Li, Roger Miller, Tor Raubenheimer, Perry Wilson. I am even more indebted to Dave Whittum, who wrote a set of class notes on this topic in 1997 which I have used relentlessly as a reference and which show an unusually large fraction of the amount of mathematics needed to get from equation to solution of any given problem. The well-drawn figures in this document were the handiwork of Terry Anderson of SLAC, whose assistance is gratefully acknowledged.

References

- [1] J. D. Jackson, *Classical Electrodynamics*, 2nd. Ed, p. 818 (1975).
- [2] J. R. Reitz, F. J. Milford, R. W. Christy, *Foundations of Electromagnetic Theory*, pages 344-347 (1980).
- [3] Jackson, *Classical Electrodynamics*, p. 236.
- [4] J. R. Reitz, F. J. Milford, R. W. Christy, *Foundations of Electromagnetic Theory*, p. 365.

- [5] D. Jackson, *Fourier Series and Orthogonal Polynomials*, pages 74 and 81 (1941).
- [6] D. Whittum, private communication.
- [7] D. Halliday and R. Resnick, *Physics*, 3rd Edition, page 886 (1978).
- [8] Jackson, *Classical Electrodynamics*, p. 337.
- [9] D. Jackson, *Fourier Series and Orthogonal Polynomials*, pages 83 and 86 (1941).
- [10] T. Wangler, *RF Linear Accelerators*, page 59 (1998).
- [11] E. L. Chu, *The Theory of Electron Linear Accelerators*, page 56 (1951).
- [12] Jackson, *Classical Electrodynamics*, p. 126.
- [13] R. Haberman, *Elementary Applied Partial Differential Equations*, 2nd. Ed., page 234 (1987).
- [14] *Chemical Rubber Company Handbook of Chemistry and Physics*, 65th Ed., page A-59 (1984).
- [15] G.A. Loew, "Accelerating Structures," in *Linear Accelerators*, P. M. Lapostolle and A.L. Septier, editors, page 53 (1970).
- [16] G.A. Loew in *Linear Accelerators*, page 61 (1970).
- [17] H. Wiedemann, *Particle Accelerator Physics*, page 41 (1993).
- [18] G.A. Loew in *Linear Accelerators*, pages 66-75 (1970).
- [19] G.A. Loew in *Linear Accelerators*, page 71 (1970).
- [20] T. Wangler, *RF Linear Accelerators*, page 59 (1998).
- [21] J.E. Leiss, in *Linear Accelerators*, page 151 (1970).
- [22] H. Wiedemann, *Particle Accelerator Physics*, page 42 (1993).
- [23] J.E. Leiss, in *Linear Accelerators*, page 156 (1970).
- [24] K.L.F. Bane, A. Mosnier, A. Novokhatski, K. Yokoya, "Calculation of the Short-Range Longitudinal Wakefields in the NLC Linac," in *ICAP 98 – Proceedings* (1998).
- [25] B.W. Montague, "Radio-Frequency Particle Separation," in *Progress in Nuclear Technologies and Instrumentation*, F. J. M. Farley, ed., volume 3, page 21 (1968).
- [26] T. Wangler, *RF Linear Accelerators*, page 348 (1998).
- [27] R. H. Helm, G. A. Loew, in *Linear Accelerators*, page 201 (1970).
- [28] K.L.F. Bane, private communication.
- [29] T. Wangler, *RF Linear Accelerators*, pages 342-344 (1998).
- [30] V. Balakin, A. Novokhatsky, V. Smirnov, "VLEPP: Transverse Beam Dynamics," in *Proceedings, High Energy Accelerators*, pages 119-120 (1983).

- [31] V. Tsakanov, “Autophasing and Lattice Scaling in High Energy Linear Accelerators,” *Physical Review Special Topics – Accelerators and Beams* **1** 041001 (1998).
- [32] J.W. Wang and G.A. Loew, “RF Breakdown Studies in Copper Electron Linac Structures,” SLAC-PUB-4866, in *Proceedings of the 1989 Particle Accelerator Conference* (1989).
- [33] R. B. Palmer, “Prospects for High-Energy e^+e^- Linear Colliders,” in *Annual Review of Nuclear and Particle Science*, v.40, page 568 (1990).
- [34] J.W. Wang and G.A. Loew, “Field Emission and RF Breakdown in High Gradient Room Temperature Linac Structures,” SLAC-PUB-7684 (1997).
- [35] D.P. Pritzkau and R.H. Siemann, “Experimental Study of RF Pulsed Heating on Oxygen-Free Electronic Copper,” *Physical Review Special Topics – Accelerators and Beams* **5** 112002 (2002).
- [36] *Chemical Rubber Company Handbook of Chemistry and Physics*, 65th Edition, page A-78 (1984).
- [37] P. Wilson, “High-Energy Electron Linacs: Applications to Storage Ring RF Systems and Linear Colliders,” SLAC-PUB-2884 (1982).
- [38] C. Nantista, “Radio-Frequency Pulse Compression for Linear Accelerators,” SLAC-Report-95-455 (1995).
- [39] T. Wangler, *RF Linear Accelerators*, page 23 (1998).
- [40] T. Wangler, *RF Linear Accelerators*, page 297 (1998).
- [41] T. Wangler, *RF Linear Accelerators*, page 369 (1998).
- [42] T.I. Smith, “Standing Wave Modes in a Superconducting Linear Accelerator,” HEPL-437 (1966).