SLAC-PUB-3221
September 1983
(A)

ELEMENTARY PRINCIPLES OF LINEAR ACCELERATORS*

Gregory A. Loew
Stanford Linear Accelerator Center
Stanford University, Stanford, California 94305

and

Richard Talman
Laboratory of Nuclear Studies
Cornell University, Ithaca, New York 14853

TABLE OF CONTENTS

1. Introduction 2
   1.1 The Accelerating Structure and the Beam 2
   1.2 Plan of the Lectures 4
2. Some Historical Milestones 7
3. Proton Linear Accelerators 11
   3.1 General Features 11
   3.2 Transit Time Factor 11
   3.3 Shunt Impedance 14
   3.4 Cavity Q, r/Q and Filling Time 15
   3.5 Phase Stability and Adiabatic Damping 16
   3.6 Transverse Defocusing 22
   3.7 Applications and Some Practical Aspects 25
   3.8 Recent Developments 30
4. Particle Acceleration by Guided Waves 37
   4.1 Motivation 37
   4.2 Wave Confined by Parallel Planes 38
   4.3 Circular Waveguide 45
   4.4 Cylindrical Resonator 46
   4.5 Wave Propagation in Coupled Resonator Chains 48
   4.6 Periodically Loaded Structures 52
   4.7 Space Harmonics 56
   4.8 Traveling Waves or Standing Waves? 61
   4.9 Resonant Coupling 64
5. Electron Linear Accelerators 67
   5.1 The Main Elements 67
   5.2 Structure Design, Choice of Parameters and Energy Gain 70
   5.3 Beam Current, Emittance, Bunch Length and Beam Loading . . . 78
   5.4 Recent Developments and Future Challenges 87

* Work supported by the Department of Energy, contract DE-AC03-76SF00515.

(Presented at the Second Summer School on High Energy Particle Accelerator. Stanford Linear Accelerator Center, Stanford, California, August 2-13, 1982.)

DR- 2148·2

I-13565

1. Introduction

The field of linear accelerators has grown to be so voluminous that compressing it into a few elementary lectures, as is our intention, requires one to be selective in some arbitrary way. In this first section, we will list some of the broad topics which the subject of linear accelerators encompasses and then indicate which of these will be covered in some detail.

1.1 The Accelerating Structure and the Beam

Linear accelerators consist of two basic elements: the accelerating structure and the particle beam.

The accelerating structure depends on the type of linac. The main types of linacs are:

1. DC linacs, like Van de Graafs, in which the structure consists of some kind of column of electrodes. These electrodes sustain a DC electric field which accelerates a continuous stream of particles. DC linacs are limited to a few tens of MeV.

2. Induction linacs in which the accelerating electric fields are obtained, according to Faraday's law, from changing magnetic fluxes. These changing magnetic fluxes are generated by large pulsed currents driven through linear arrays of magnetic toroids. The beam path, along which the electric field develops, can be considered as the single turn secondary of a transformer. Induction linacs are generally used in medium- to high-current pulsed applications.

3. RF linacs, the type on which we will concentrate here, can be categorized in a number of ways: low frequency (UHF), microwave frequency (L, S, C, or X-band), laser frequency; CW or pulsed, traveling-wave or standing-wave; room temperature or superconducting. In all these cases, the structure is a conducting array of gaps, cavities or gratings along which rf waves with an electric field parallel to the beam can be supported and built up through some resonant process. RF linacs are used for a wide spectrum of applications from injectors into circular accelerators, to entire high-energy accelerators such as SLAC, medical accelerators, and many others.

While the accelerating structure can be considered the heart of each individual machine, it cannot work without its associated systems such as the power source, the vacuum, cooling, support and alignment, and instrumentation and control systems. It is important to note that a technological breakthrough in any one of these associated systems can have profound effects on the main design of the accelerator.

The second basic element of the linac is, of course, the beam and the particles which compose it. The vast majority of linacs today are electron (or positron) machines; they number about 1400 accelerators of which a large percentage are commercial radiation therapy machines. The other linacs accelerate protons ($H^{\pm}$) and in a few cases ions. There are about 50 proton or ion linacs in existence in the world.

Very little will be said in these lectures about the sources of these particles, guns, ion sources, duoplasmatrons, polarized beams, strippers, positron radiators, etc. This is an encyclopedic subject by itself with many specialties and sub-specialties which cannot profitably be summarized here. The reader should not conclude that because the subject has been left out, it is not of crucial importance to the design and operation of a given linac. Not only does the source have an effect on how well the accelerator can perform its function, but in some cases it determines how a new concept can or cannot be approached. An example of this is the electron source for the linear collider for which a conventional electron gun cannot create a beam with a sufficiently small emittance and must be followed by a damping ring to "cool" it down.

The fundamental problems in beam dynamics are:

1. Longitudinal bunching and stability,
2. Focusing and transverse stability, and
3. Steering and transport to a target or interaction area.

As long as the number of particles or current density is not too high, the analysis of many of these problems can be carried out by following single particle trajectories and then formulating the behavior of the beam in terms of its envelope. The important parameters are transverse emittance, bunch length and energy spectrum.

Far more difficult are the beam dynamics associated with collective effects such as emittance growth, beam break-up, beam-wall and wakefield effects in general. These problems are treated in other chapters of this book.

1.2 PLAN OF THE LECTURES

These lectures come in five sections. The first is this introduction. The second is a short chronology of what are viewed as important milestones in the field.

The third covers proton linacs. It introduces elementary concepts such as transit time, shunt impedance, Q. But defers a more systematic discussion of periodic accelerating structures to the next section. Critical issues such as phase stability and transverse forces are discussed. Finally, examples of existing facilities are given and modern developments are mentioned.

The fourth section contains an elementary discussion of waveguide accelerating structures. It can be regarded as an introduction to some of the more advanced treatments of the subject, such as are contained in the book on *Linear Accelerators*, edited by P. Lapostolle and A. Septier in 1970, which, except for some recent developments, is undoubtedly the best single source of detailed information on the subject. The encyclopedic nature of this volume, to which we will henceforth refer to as [L.A.], partially excuses the sparsity of other references given in these lectures.

The final section is devoted to electron accelerators. Taking SLAC as an example, various topics are discussed such as structure design, choice of parameters, frequency optimization, beam current, emittance, bunch length and beam loading. Recent developments and future challenges are mentioned briefly.

Since proton and electron machines will be discussed separately, it may be useful, before starting, to take a broad look at the full spectrum of linear accelerator applications shown in Fig. 1. In this figure, the kinetic energy E_K, which is the variable of greatest interest for most physics applications, is shown in the horizontal axis and the normalized velocity (in the form $\beta = v/c$), which determines much of the linac design is plotted along the vertical axis. For particles of a given mass m_0, these variables are related by the formulae:

$$E_K = \frac{m_0c^2}{\sqrt{1-\beta^2}} - m_0c^2 \tag{1.1}$$

$$\cong \frac{1}{2}m_0v^2 \ (\text{non} - \text{relativistically}) \tag{1.2}$$

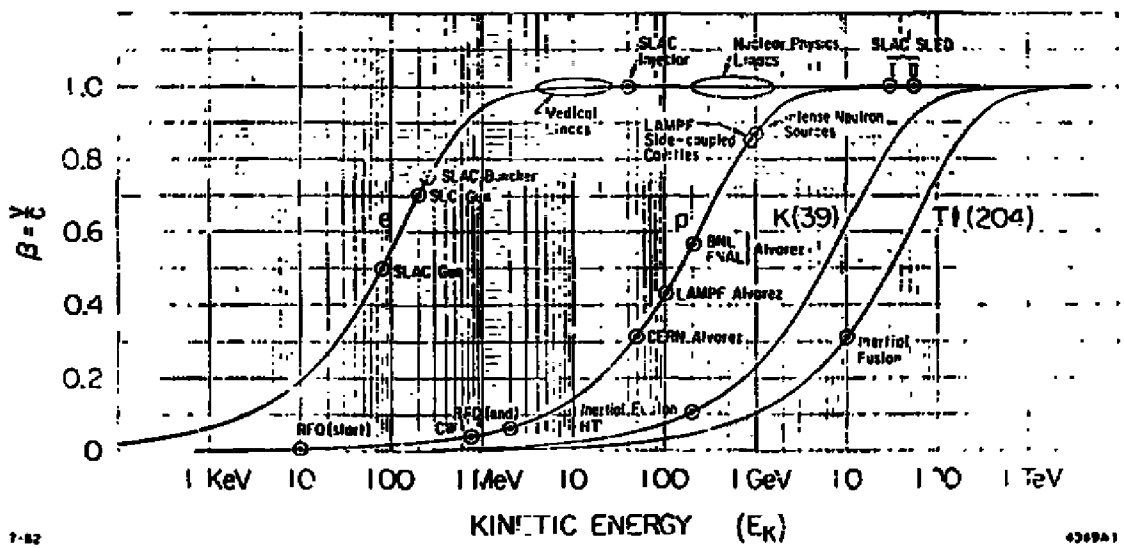


Fig. 1. The full spectrum of linear accelerator applications, displayed at appropriate energies and velocities.

The curves which result for the various particles are plotted on the figure and facilities of successively higher energies are indicated. It can be seen that electrons are almost always relativistic while protons (not to mention heavier ions) almost never are. It is for this reason that proton and electron accelerators tend to be described in different terms.

DISCLAIMER

2. Some Historical Milestones

The history of linear accelerators is briefly described by J. Blewett in the book L.A. A technical discussion of some early projects is contained in the *"Linear Accelerator Issue"* of the Review of Scientific Instruments (February 1955).

Here we have chosen to enumerate only some of the important milestones. The choice has led them to be spread more or less uniformly in time emphasizing the remarkable continuity of progress over a period of more than half a century which promises to continue into the future. This uniform spread results perhaps in the slighting of some of the early work when parallel progress was being made at many laboratories.

The history of linacs can be viewed as a sequence of attempts to fool charged particles so that they see cumulatively acting voltages across linear arrays of gaps. A chronology follows. References can be traced through L.A.

1924 A theoretical paper by G. Ising, Stockholm, describes a method for accelerating positive ions (canal rays) by applying the electrical wavefront from a spark discharge to an array of drift tubes via transmission lines of successively greater lengths.

1928 An experimental paper (including the theory of the betatron) by R. Wideröe, Switzerland, describes the successful acceleration of potassium ions to 50 kV. Figure 2 indicates schematically the setup in which the ions pass successively through three drift tubes: the first and last are grounded, the center one is attached to a 1 MHz oscillator with a voltage of 25 kV. The distance d between gaps is adjusted so that

$$\frac{d}{v} = \frac{1}{2f} \quad \text{or} \quad d = \frac{\beta\lambda_0}{2}$$

where f is the frequency and λ_0 the free space wavelength at that frequency. The potassium ions travel from one gap to the next in one-half an RF period. Since higher frequency oscillators did not exist at the time, lighter particles traveling faster could not be accelerated.

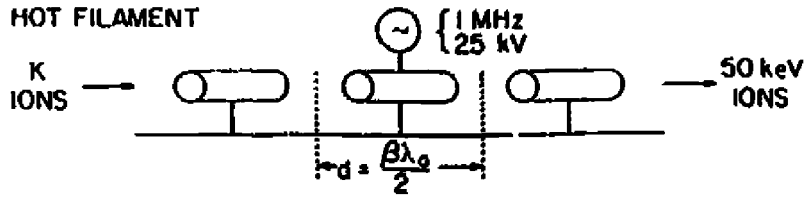


Fig. 2. The Wideröe linac.

1931-35 K. Kingdon (G.E.), L. Snoddy (Univ. of Virginia) et al., accelerate electrons from 28 keV to 2.5 MeV by applying progressive wavefronts to a drift tube array.

1931-34 E. Lawrence, D. Sloan et al., (U.C. at Berkeley) build a Wideröe-type linac (i.e., $d = \beta\lambda_0/2$) with 30 drift tubes, oscillating to 42 kV, driven by 7 MHz oscillators. Mercury ions are accelerated to 1.26 MeV. Oscillators of high enough frequency for protons are still not available. Similar work continues at Cornell, in Japan and in England.

1937-45 W. Hansen and the Varian brothers invent the klystron (at first a low power device) at Stanford. Subsequently, the high power magnetron (2 MW pulsed) is developed in Great Britain for radar purposes as part of the war effort.

1945-47 L. Alvarez, W. Panofsky et al. (U. C. Berkeley) build a 32 MeV proton drift tube linac (Fig. 3) three-feet in diameter, forty feet long, powered by 200 MHz war surplus radar equipment. As indicated, the Alvarez structure differs from the Wideröe structure in that all tubes are contained in one large cylindrical tank and are powered at the same phase; the distance between drift tubes is arranged so that the particles are shielded from the fields when they are in the decelerating phase. Adequate beam acceptance required that the accelerating field not have much variation with radius, thus precluding operation at higher frequency. As will be explained later, longitudinal phase stability turned out to be satisfactory but transverse focusing was problematical.

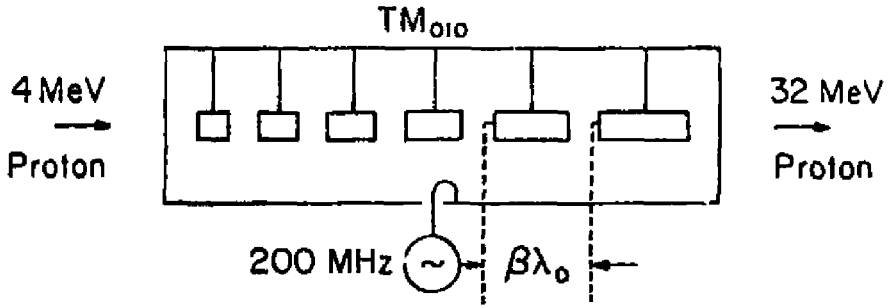


Fig. 3. An Alvarez linac.

1947-48 At Stanford, W. Hansen, E. Ginzton, W. Kennedy et al., build the Mark I disk-loaded linac yielding 4.5 MeV electrons in a nine-foot structure powered to 1 MW at 2856 MHz. It is the first of a series: Mark II (40 MeV); Mark III (1.2 GeV); and SLAC (30 GeV). Parallel efforts take place in Great Britain, France and the USSR, and at M.I.T. and Yale in the U.S.A.

1952 J. Blewett at BNL shows that alternating-gradient focusing works with quadrupole coils inserted in the drift tubes, solving the transverse focusing problem for protons. The Alvarez linac (with some modifications) from then on serves as the model for most subsequent proton and ion linacs (up to 200 MeV).

1965-66 P. Wilson, A. Schwettman et al., at HEPL, Stanford, report the first successful operation of a superconducting linac producing 500 keV electrons with three lead-coated cells.

1967-69 V. Sarantsev et al., at Dubna, USSR, build a linear induction accelerator, as do D. Keefe et al., at LBL, shortly thereafter, both groups with the intent of accelerating electron rings.

1971 R. Koontz, G. Loew, and R. Miller at SLAC for the first time accelerate a single electron bunch through the 3-km linac and show experimentally that beam loading is energy independent.

1972 D. Nagle, E. Knapp et al., at LASL, Los Alamos, successfully operate their 800-MHz side-coupled cavity linac, LAMPF, and produce 800 MeV protons.

1973 P. Wilson, D Farkas and H. Hogg at SLAC invent the rf energy compression scheme called SLED which in the next five subsequent years gets installed on the 3-km linac, boosting its energy up to 30 GeV.

1980 R. Stokes et al., at LASL, successfully test RF quadrupoles (up to 2 MeV) following a 1970 suggestion of I. Kapchinskii and V. Teplyakov, ITEP, Moscow.

1982 H. Grunder, F. Selph, et al., at LBL, use the HILAC and Bevatron to accelerate U^{+238} with charge state ~ 50.

3. Proton Linear Accelerators

3.1 General Features

As was mentioned previously, the early linear accelerators were restricted to protons and heavier particles. Vacuum tubes with inferior high frequency characteristics and various discharge devices were the only RF power sources available. In a single cycle of a 1 MHz oscillator, a highly relativistic particle travels 300 m a length hard to conceive for a piece of apparatus in the early days of accelerator physics. A frequency more typical of modern proton accelerators is 200 MHz. The corresponding free space wavelength λ_0 is

$$\lambda_0 = \frac{c}{f} = 1.5 \text{ m} \quad . \tag{3.1}$$

If the input velocity of a proton is such that

$$\frac{v}{c} = 0.05$$

corresponding to a kinetic energy

$$E_K = 1.17 \text{ MeV} \quad ,$$

then the distance it travels in one cycle is

$$d = \frac{v}{c}\lambda_0 = 7.5 \text{ cm} \quad .$$

If we use drift tubes to shield the proton from decelerating forces, their length must be at least half this distance. Precise parameters for the FNAL 200 MeV proton linac injector are given in Table 1. It can be seen that, even at the output end, the particles are not highly relativistic. This means that the much higher frequency periodic structures typical of electron accelerators are not yet appropriate (for reasons of inefficiency and radial effects to which we will return).

3.2 Transit Time Factor

Consider a proton passing through a pill-box cavity as shown in Fig. 4, in which the maximum potential difference is V_0. If the gap length is g and the instantaneous electric field E_z in the gap is independent of the longitudinal coordinate z, the maximum electric field is given by

Table 1. FNAL 200-MeV Linear Accelerator Specifications

Output energy	200.30 MeV
Output momentum spread, total for 90% of beam	2.7×10^{-3}
Peak Beam current	100 mA
Emittance at 200 MeV (each transverse mode)	1.5 – 3.0 mrad-cm
Beam pulse length	100 μsec
Pulse repetition rate	15 pps
Cavity resonant frequency	201.25 MHz
RF pulse length, variable to	400 μsec
RF duty factor, maximum	.006
Synchronous phase angle, from rf peak	−32°

Cavity number	1	2	3	4	5	6	7	8	9
Proton energy in (MeV)	0.75	10.42	37.54	66.2	92.6	116.5	139.0	160.5	181.0
Proton energy out (MeV)	10.42	37.54	66.18	92.60	116.5	139.0	160.5	181.0	200.3
Cavity length (m)	7.44	19.02	16.53	15.68	15.58	15.54	15.63	15.88	15.73
Cavity diameter (cm)	94	90	88	88	84	84	84	84	84
Drift-tube diameter (cm)	18	18	18	18	18	18	18	18	
Bore-hole diameter (cm)	2.0 2.5	3.0	3.0	3.0	4.0	4.0	4.0	4.0	4.0
Cell length L									
(first cell) (cm)	6.04	22.2	41.1	53.3	61.8	68.2	73.3	77.6	81.3
(last cell) (cm)	21.8	40.9	53.8	61.5	67.9	73.1	77.4	81.1	84.3
Gap length G									
(first cell) (cm)	1.30	4.4	12.2	19.5	22.6	27.1	30.9	34.3	37.3
(last cell) (cm)	6.70	12.7	19.3	25.1	28.9	30.8	34.2	37.1	39.7
G/L									
(first cell)	0.21	0.20	0.30	0.37	0.37	0.40	0.42	0.44	0.45
(last cell)	0.31	0.31	0.36	0.41	0.40	0.42	0.44	0.46	0.47
Axial transit-time factor									
(first cell)	0.64	0.86	0.82	0.75	0.73	0.68	0.64	0.61	0.59
(last cell)	0.81	0.81	0.75	0.69	0.69	0.65	0.61	0.58	0.55
Effective shunt impedance									
(first cell) (MΩ/m)	27.0	53.5	44.5	35.0	30.6	24.8	21.5	18.9	16.7
(last cell) (MΩ/m)	47.97	44.8	35.2	28.5	25.0	21.7	19.0	16. 8	14.9
Drift space following cavity (m)	0.22	0.5	0.75	1.0	1.0	1.0	1.0	1.0	-
Number of full drift tube	55	59	34	28	23	21	20	19	18
Average axial field (MV/m)	1.60 2.31	2.0	2.6	2.6	2.56	2.56	2.56	2.56	2.56
Average gap field									
(first cell) (MV/m)	7.82	10.0	8.7	7.03	6.6	6.4	6.1	5.8	5.6
(last cell) (MV/m)	7.45	8.45	7.2	6.3	6.4	6.1	5.8	5.6	5.4
Peak surface field									
(first cell) (MV/m)	8.9	12.6	13.1	12.9	14.0	14.1	14.2	14.3	14.5
(last cell) (MV/m)	10.2	9.7	12.9	13.2	14.1	14.2	14.3	14.5	14.6
Cavity excitation power (MW)	0.61	1.38	2.245	2.45	2.49	2.33	2.65*	2.7*	2.75*
Total power per cavity for 100 mA (MW)	1.58	4.09	5.11	5.12	4.88	4.58	4.81	4.75	4.85

Total accumulated length	144.5 m
Total number of unit cells	288.0
Total number of full drift tubes	277.0
Total cavity excitation power	19.6 MW
Total linac rf power for 100 mA	39.6 MW

*Estimated

$$E_{z0} = \frac{V_0}{g} \tag{3.2}$$

and the field varies as

$$E_z = \frac{V_0}{g} \cos \omega t \tag{3.3}$$

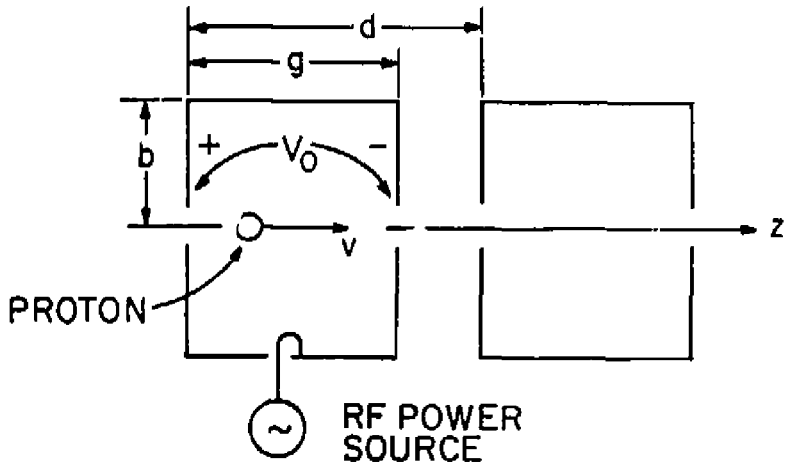


Fig. 4. Pill-box resonator accelerating cells.

If the proton has an average velocity v and passes through the center of the cavity at $t = 0$, its coordinate is given by

$$z_p = vt \tag{3.4}$$

and its kinetic energy gain while passing through the cavity is

$$\Delta E_K = \int_{-g/2}^{g/2} \frac{eV_0}{g} \cos \frac{\omega z_p}{v} dz_p$$

$$= eV_0 \frac{\sin \omega g/2v}{\omega g/2v} \tag{3.5}$$

$$= eV_0 T_{tr} \tag{3.6}$$

where the familiar $\sin x/x$ term is called the transit time factor T_{tr}.

If the proton passes through a series of cavities, each phased so that the field is maximum when the particle goes through the center, then cumulative acceleration can take place. We see, however, that if each resonator gap g is equal to $\beta\lambda_0/2$ where λ_0 is the free space wavelength and $\beta = v/c$, then T_{tr} is only equal to 0.64 $[(\sin \pi/2)/(\pi/2)]$, appreciably less than unity. To improve upon this situation, for a given V_0 or power per unit length, it is clearly advantageous to reduce the gap length g: this leads naturally to the configurations shown in Fig. 5.

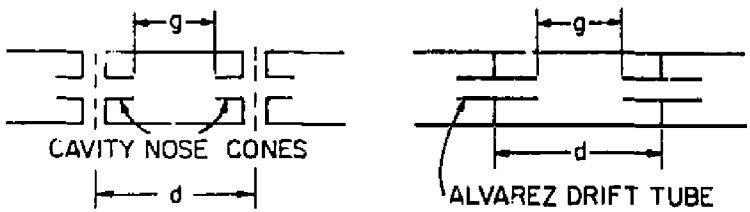


Fig. 5. Structures which increase the transit time factor T_{tr} by decreasing g.

The cavity nose cone and the length of the gaps between Alvarez drift tubes can both be adjusted to bring T_{tr} closer to unity. Ultimately, reducing g leads to sparking due to excessive field gradients. There is little to be gained in T_{tr} by reducing g to less than, say $\beta\lambda_0/4$. Table I shows how the gaps between successive drift tubes in the nine FNAL Alvarez tanks have been adjusted to handle these problems.

3.3 Shunt Impedance

The next question that arises is: How much power do we need to obtain a given amount of energy gain? Power is consumed in basically four areas: the cavity walls, the beam, the transmission line between the source and the cavity, and in reflections which decrease the power delivered to the cavity. For now, let us neglect all but the first. Unless one uses a superconducting cavity, appreciable power is dissipated in the walls of a cavity to maintain high fields in it. The power lost in one cell, P_{lost}, is proportional to V_0^2. The shunt impedance R of the cavity is defined as:

$$R = \frac{V_0^2}{P_{lost}} \tag{3.7}$$

If the cavity is of length ℓ, then for a multi-cavity accelerator it is also customary to

define a shunt impedance per unit length r where

$$r = \frac{R}{d} = \frac{(V_0/d)^2}{P_{lost}/d} \tag{3.8}$$

For a negligible beam intensity and a lossless matched line between the power source and the cavity, one has by definition:

$$P_{source} = P_{lost} \tag{3.9}$$

and the achievable value of V_0 becomes

$$V_0 = \sqrt{R\, P_{source}} \tag{3.10}$$

where the peak energy gain, from (3.6) is

$$\Delta E_K = e\sqrt{R\, T_{tr}^2\, P_{source}} \tag{3.11}$$

Referring again to Table 1 for a numerical example, it can be seen that at the injection end of the FNAL linac, the first tank containing 55 drift tubes has an average shunt impedance

$$R = 38 \times 10^6 \times 7.4 = 281.2 \text{ Megohms}$$

and

$$T_{tr} = 0.75 \ .$$

Then with $P_{source} = 0.6$ MW

$$\Delta E_K = e\sqrt{281.2 \times (0.75)^2 \times 0.6} = 0.74 \text{ MeV} \ .$$

3.4 Cavity Q, r/Q and Filling Time

To calculate P_{lost} it is necessary to know the quality factor Q of the cavity which is defined by

$$Q = \frac{\omega w_{st}}{P_{lost}} \tag{3.12}$$

where w_{st} is the maximum stored energy and it is assumed that the only net flow of power into the cavity exactly replenishes P_{lost}. Combining (3.7) and (3.12) one obtains the important ratio

$$\frac{R}{Q} = \frac{V_0^2}{\omega w_{st}} \tag{3.13}$$

or per unit length,

$$\frac{r}{Q} = \frac{(V_0/d)^2}{\omega(w_{st}/d)} \tag{3.14}$$

Notice that, at a given frequency, this ratio is purely a property of the cavity geometry and does not depend on the wall material or condition, or on the quality of welds, joints, etc. It can be measured using dielectric beads on threads or rods. On the other hand, the separate quantities R, r and Q depend on all these factors.

We will return later to a discussion of power flow along a chain of resonators, but for now let us notice that the natural damping of a free excited single cavity, by (3.12), is

$$\frac{dw_{st}}{dt} = -\frac{\omega}{Q} w_{st} \tag{3.15}$$

As a result, the time for the electric field to decay to $1/e$ of its initial value is

$$t_F = \frac{2Q}{\omega} \tag{3.16}$$

3.5 Phase Stability and Adiabatic Damping

When a bunch of protons is injected into a linac, it inevitably has a spread of velocities. Superficially this suggests that in a sufficiently long accelerator, all protons will eventually drop out of step with the accelerating fields. Fortunately, phase stability prevents this from occurring.

The important idea is that a particle whose velocity is somewhat too low gradually drops back in the bunch, thereby passing the structure at a phase where the acceleration is greater, permitting it eventually to catch up and even overtake the center of the bunch. It is presupposed that the drift tubes, gaps and fields are so arranged that one can speak of a reference or synchronous particle which is accelerated through the

structure without any longitudinal oscillation and defines the bunch center. In practice, this is achieved by empirically adjusting the phases of the drive to the successive individual sections of the linac.

It turns out to be simpler to use z, the distance along the accelerator as the independent variable rather than time. In this section we will use capital letters to represent the dependent variables for a specific particle. This will permit us to use lower-case letters to represent deviations away from the variables of the reference particle. For example, let the particle pass point z at time T. At this time the phase of the RF accelerating field is $\Phi = \omega T$. The synchronous particle passes the same point at time T_s and phase $\Phi_s = \omega T_s$. The time and phase differences are

$$\begin{aligned} t &= T - T_s \\ \phi &= \Phi - \Phi_s = \omega(T - T_s) = \omega t \quad . \end{aligned} \tag{3.17}$$

Referring back to Eq. (3.3), the rate of increase of energy of the synchronous particle is given by

$$\frac{dE_{Ks}}{dz} = q\ \hat{E}_{z0} \cos \Phi_s$$

where $\hat{E}_{z0}$ represents the average electric field in the vicinity of z, including the transit time correction. For maximum acceleration, Φ_s should be adjusted to zero but this would not yield phase stability as we shall see. For the general particle the rate of energy gain is

$$\frac{dE_K}{dz} = q\ \hat{E}_{z0} \cos \Phi \quad . \tag{3.18}$$

The energy offset $e = E_K - E_{Ks}$ satisfies

$$\dot{e} = \frac{de}{dz} = q\ \hat{E}_{z0} [\cos(\Phi_s + \phi) - \cos \Phi_s] \tag{3.19a}$$

$$\simeq -q\ \hat{E}_{z0}\, \phi \sin \Phi_s \tag{3.19b}$$

where the dot notation is intended to remind the reader that z enters the equations the way time does in regular mechanics. Note that we are using q for charge to free up the symbol e for energy offset.

Our intention here (and elsewhere in these lectures) is not to give the most general formulation but rather to explain the ideas using formulae that are as simple as possible. Hence we will emphasize the linearized equation (3.19b) rather than the general, nonlinear equation (3.19a). For the same reason we will use non-relativistic mechanics even though the full relativistic treatment is straightforward (see the article by H. G. Hereward in L.A.). As it happens, for the critical early stages of a proton accelerator where these ideas are most important, the non-relativistic approximation is very good.

Equation (3.19b) shows that there is an energy correction proportional to the phase discrepancy ϕ but this is only half the story. There is also a phase correction due to the energy offset e since a more energetic particle has a greater velocity which allows it to gain in phase. T and T_s are given in terms of the velocities V and V_s by

$$T = \int_0^z \frac{dz'}{V} \; ; \; T_s = \int_0^z \frac{dz'}{V_s} \quad . \tag{3.20}$$

From these and (3.17) we get

$$\dot{\phi} = \frac{d\phi}{dz} = \omega\left(\frac{1}{V} - \frac{1}{V_s}\right) \quad . \tag{3.21}$$

It was the simplicity of this relation and the natural occurrence of z in (3.18) which motivated the choice of z as an independent variable.

Velocities and energies are related by

$$E_K = \frac{1}{2} m V^2 \; ; \; E_{Ks} = \frac{1}{2} m V_s^2 \tag{3.22}$$

and as a result

$$e = \frac{1}{2} m (V^2 - V_s^2) \tag{3.23a}$$

$$\simeq m V_s (V - V_s) \quad . \tag{3.23b}$$

In practice, the energy spreads tend to be very small and the linearized form (3.23b) is a satisfactory approximation (more so, for example, than the linearized form (3.19b) which won't quite do in some cases). Combining (3.21) and (3.23b) we obtain

$$\dot{\phi} = -\frac{\omega}{m V_s^3} e \quad . \tag{3.24}$$

We obtain "Newton's law" by differentiating (3.24) and substituting from (3.21)

$$\ddot{\phi} = \text{"}F\text{"}$$

$$= -\frac{\omega}{mV_s^3} q E_{z0} [\cos(\Phi_s + \phi) - \cos\Phi_s] \qquad (3.25a)$$

$$\simeq \frac{\omega}{mV_s^3} q E_{z0} \phi \sin\Phi_s \quad . \qquad (3.25b)$$

Even the non-linear form (3.25a) has the simple feature that the "force", "F", depends only on the displacement ϕ: this motivates one to introduce an "effective potential energy function"

$$\begin{aligned} \text{"}V\text{"} &= -\int \text{"}F\text{"}\, d\phi \\ &= \frac{\omega}{mV_s^3} q E_{z0} [\sin(\Phi_s + \phi) - \phi\cos\Phi_s] \quad . \end{aligned} \qquad (3.26)$$

The effective force is plotted in Fig. 6a and the effective potential is plotted in Fig. 6b. Clearly Eq. (3.25b) represents an oscillatory system provided that $\sin\Phi_s$ is negative. The "frequency" of small oscillations is given by

$$\frac{2\pi}{\lambda_z} = \sqrt{\frac{\omega q E_0 |\sin\Phi_s|}{mV_s^3}} \quad . \qquad (3.27)$$

Here, since the independent variable is z, it is the wave number $2\pi/\lambda_z$ which appears rather than ω. λ_z is the distance along the accelerator for one cycle of longitudinal oscillation to occur. Note that λ_z becomes large as V_s increases. Relativistically this effect is even more pronounced. The correct relativistic expression differs from (3.27) by the replacement $V_s \to \gamma_s V_s$ where γ_s is the usual relativistic factor $(1-\beta^2)^{-\frac{1}{2}}$.

As usual with oscillators, it is enlightening to write the equations of motion in Hamiltonian form and to discuss the motion in phase space for which the axes are ϕ and e as in Fig. 6c. The Hamiltonian is given by

$$H = -q E_0 [\sin(\Phi_s + \phi) - \phi\cos\Phi_s] - \frac{\omega}{2mV_s^3} e^2 \qquad (3.28a)$$

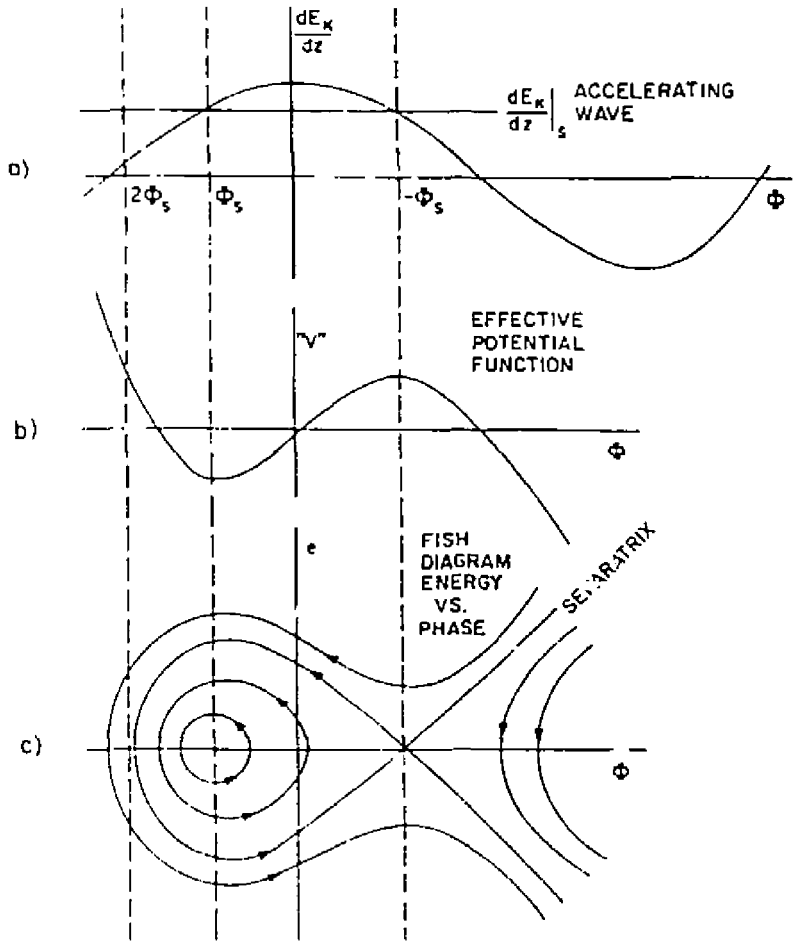


Fig. 6. Phase stability graphs.

and the equations of motion (3.19) and (3.24) are

$$\dot{e} = -\frac{\partial H}{\partial \phi} \tag{3.29}$$

$$\dot{\phi} = \frac{\partial H}{\partial e} \tag{3.30}$$

For small oscillations, and dropping a constant term, H reduces to

$$H = -\frac{1}{2} q \dot{E}_{z0} |\sin \Phi_s| \phi^2 - \frac{\omega}{2mV_s^3} e^2 \quad . \tag{3.2-b}$$

When a Hamiltonian does not depend on the independent variable, it is a constant of the motion. In that case a phase space trajectory is an ellipse with the equation

$$H(\phi, e) = H_1 \quad \text{(a constant)} \quad . \tag{3.31}$$

Just as with a pendulum for which the maximum velocity can be calculated if the maximum height is known, or vice versa, we can obtain

$$H_1 = -\frac{1}{2} q \dot{E}_{z0} |\sin \Phi_s| \phi_{max}^2 = -\frac{\omega}{2mV_s^3} e_{max}^2 \quad . \tag{3.32}$$

In other words, the aspect ratio of the phase space ellipse is determined by the Hamiltonian

$$\frac{e_{max}}{\phi_{max}} = \sqrt{\frac{q \dot{E}_{z0} |\sin \Phi_s| m V_s^3}{\omega}} \quad . \tag{3.33}$$

But H is not independent of z. The very fact that we are discussing an accelerator implies that V_s varies with z and the ideas presented here cannot be applied without further discussion. If the rate of variation is slow enough, the motion in phase space will still follow an almost elliptical path close to one of those shown in Fig. 6c. Slow (or adiabatic) in this context means that the fractional variation in V_s should be small as the particle travels a distance λ_z. It is useful to evaluate λ_z for the data describing the FNAL linac and given in Table 1. You should find that the adiabatic condition is quite well satisfied.

But in such an adiabatically varying situation, the value of $H(\phi, e)$ does not remain even approximately constant. On the other hand, it is not hard to show (e.g., Landau and Lifshitz, *Mechanics*) that the area of the phase space ellipse is an "adiabatic invariant." In other words,

$$e_{max} \phi_{max} = \text{constant} \quad . \tag{3.34}$$

Taken together equations (3.33) and (3.34), by fixing both the product and ratio of ϵ_{max} and ϕ_{max}, fix them individually. It follows that their variation is given by

$$\epsilon_{max} \propto \left[E_{z0} \, |\sin \Phi_s| \, V_s^3 \right]^{1/4} \tag{3.35}$$

$$\phi_{max} \propto \left[E_{z0} \, |\sin \Phi_s| \, V_s^3 \right]^{-1/4} . \tag{3.36}$$

The latter relation shows that the bunch naturally shrinks down and becomes short compared to the phase acceptance of the linac sections. Even the energy spectrum is, in a manner of speaking, damped since one is normally interested in the fractional energy spread ϵ_{max}/E_{ks}.

We have not said much about the motion for large amplitudes where (3.28a) must be used. Nor will we, as it is assumed that the reader has already encountered phase space plots like Fig. 6c elsewhere. On injection, particles inside the separatrix are captured. Those outside are largely lost although even some of them can be captured because of adiabatic damping.

3.0 Transverse Defocusing

There is one undesirable consequence of the phase stability considerations of the last section. We saw, following Eq. (3.25b), that longitudinal stability requires that $\sin \Phi_s$ be negative. Referring back to Eqs. (3.2) and (3.3), the time rate of change of the electric field is

$$\frac{dE_z}{dt} = - E_{z0} \, \omega \cos \omega t$$

which, for the synchronous particle becomes

$$\frac{dE_z}{dt} = - E_{z0} \, \omega \sin \Phi_s \quad . \tag{3.37}$$

This expression is positive if there is longitudinal stability. Looking at Fig. 7 it can be seen that fringe fields in the gap region lead to transverse forces. As the proton enters the gap it feels a focusing force and, as it leaves, a defocusing force. There is a net defocusing effect since from (3.37) the magnitude of the field is increasing with time.

A quick way of seeing the inevitability of this effect is to consider the motion in a moving frame of reference in which the synchronous particle is at rest. In this frame there is only a time-independent electric field since the phase velocity of the field is equal to the particle velocity. In this frame, Earnshaw's theorem which applies to DC fields and potentials excludes the possibility of a potential minimum, and motion in at least one of the three dimensions must be unstable. Since the longitudinal motion has been adjusted to be stable, the transverse motion will not be.

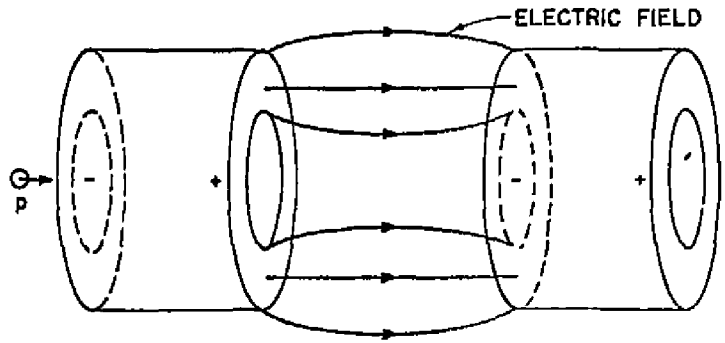


Fig. 7. Transverse focusing fields in accelerating gaps.

It is possible to show this effect mathematically. So far we have not explicitly given the z-dependence of the accelerating field other than to say that the protons are inside drift tubes where there is no field except at times when the field is in the right direction. As mentioned above, we can represent this by a traveling wave given by

$$E_z(z,t) = \hat{E}_{z0}\cos\omega\left(t - \int_0^z \frac{dz'}{V_s}\right) \quad . \tag{3.38}$$

The argument has been arranged so that a particle of speed V_s stays at a fixed phase of the traveling wave. In Sec. 4 we will see how, by Fourier transformation, such a traveling wave can be obtained for a general accelerating structure, but for now let us accept (3.38) as the dependence on z and t separately, for example near the center of one of the accelerating gaps.

We will now use this expression of the field to find the form of the radial Lorentz force equation. Taking r as a transverse radial displacement and θ as an azimuthal angle, we can show that, for small r, the form of (3.38) and Maxwell's equations fix the

other field components. We also assume that H_z vanishes and, assuming rotational symmetry, that $E_\theta = H_r = 0$. In the absence of free charge, Gauss's law makes $\nabla \cdot \vec{E}$ vanish, i.e.,

$$\frac{1}{r}\frac{\partial}{\partial r}(rE_r) = -\frac{\partial E_z}{\partial z} \quad . \tag{3.39}$$

For small r, E_r will be proportional to r, and we get

$$E_r = \frac{1}{2}\frac{\partial}{\partial r}(rE_r) = -\frac{r}{2}\hat{E}_{z0}\frac{\omega}{V_s}\sin\omega\left(t - \int_0^z \frac{dz'}{V_s}\right) \quad . \tag{3.40}$$

By Ampère's law

$$\frac{1}{r}\frac{\partial}{\partial r}(rH_\theta) = \epsilon_0\frac{\partial E_z}{\partial t} \tag{3.41}$$

and from this we obtain

$$H_\theta = -\frac{r\omega}{2}\hat{E}_{z0}\,\epsilon_0\sin\omega\left(t - \int_0^z \frac{dz'}{V_s}\right) \quad . \tag{3.42}$$

We can now write the radial Lorentz force equation for the synchronous particle

$$\begin{aligned}\frac{d}{dt}(m\dot{r}) &= q(E_r - V_s\mu_0 H_\theta) \\ &= \frac{q\omega\hat{E}_{z0}}{2V_s}(1-\beta_s^2)\,r\sin\Phi_s\end{aligned} \tag{3.43}$$

where it has been recognized that the phase factors in (3.38), (3.40) and (3.42) are the same as in (3.25b), i.e., Φ_s. But since the sign in (3.43) is opposite to that in (3.25b), one or the other must lead to unstable motion which is what we set out to demonstrate. It should be noticed though, by the presence of the factor $1-\beta_s^2$, that the defocusing vanishes in the relativistic limit.

Space charge forces also tend to enlarge the transverse beam size but the effect just described limits operation even at low beam current. Early corrective schemes using electrostatic grids were not very satisfactory. It was the discovery of alternating-gradient focusing by Blewett, Courant, Livingston and Snyder and the introduction of quadrupole magnets inside the drift tubes which solved the problem.

3.7 Applications and Some Practical Aspects

For high energy physics, the main application of proton and ion linacs is as injectors into circular machines, but by themselves they also have numerous applications in nuclear research and medicine. Included are meson factories (LAMPF), heavy ion research (SUPER HILAC, UNILAC, RILAC) and intense neutron sources.

A typical layout of a proton linac is shown in Fig. 8. The source is generally a metal envelope containing a gas (hydrogen or other) which is ionized into a plasma. The protons (or ions) are formed into a jet through a cylindrical nozzle.

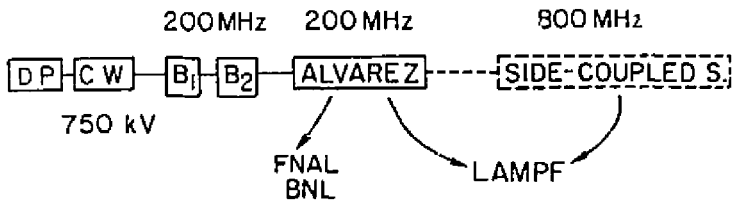


Fig. 8. Typical proton or ion linac layout.

This system, called a duoplasmatron (DP), floats at a large negative voltage supplied by a Cockcroft-Walton supply and emits the protons (or ions) into a high-gradient accelerator column (CW). This results in particles with an energy of as much as 1 MeV (typically 750 keV) but near ground potential. Photographs of the BNL versions of these components are shown in Figs. 9 and 10.

This pulse of protons forms a DC current of perhaps 100 or 200 mA for 100-200 microseconds. The protons then become bunched in the drift space following the low power, velocity-modulating, 200 MHz buncher labelled B_1. The second buncher, B_2, used for longitudinal phase space matching, injects the protons into the Alvarez linac where the capture efficiency at a synchronous phase $\Phi_s = -30°$ is as much as 80%. Figures 11, 12 and 13 show successively more magnified views of the BNL linac. After 150 or 200 m the protons have been accelerated to about 200 MeV with a pulse repetition rate of 10 or 15 pps. The parameters of other injectors, such as that at FNAL, are similar. At LAMPF this stage is less than half as long and achieves 100 MeV with a greater repetition rate (120 pps) and a 500 μsec pulse length.

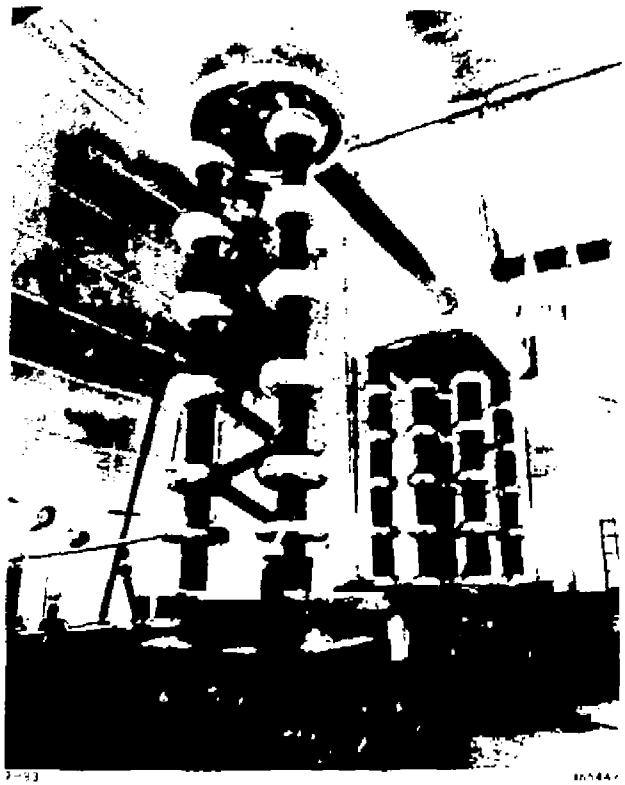


Fig. 9. Cockcroft-Walton pre-injector (750 kV)
(BNL-HAEFELE)

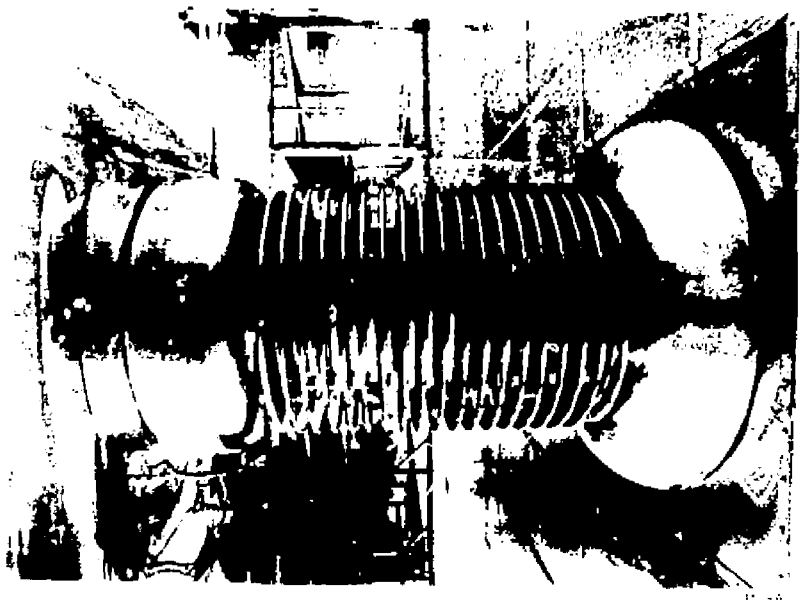


Fig. 10. High-gradient accelerating column (750 kV).
(BNL)

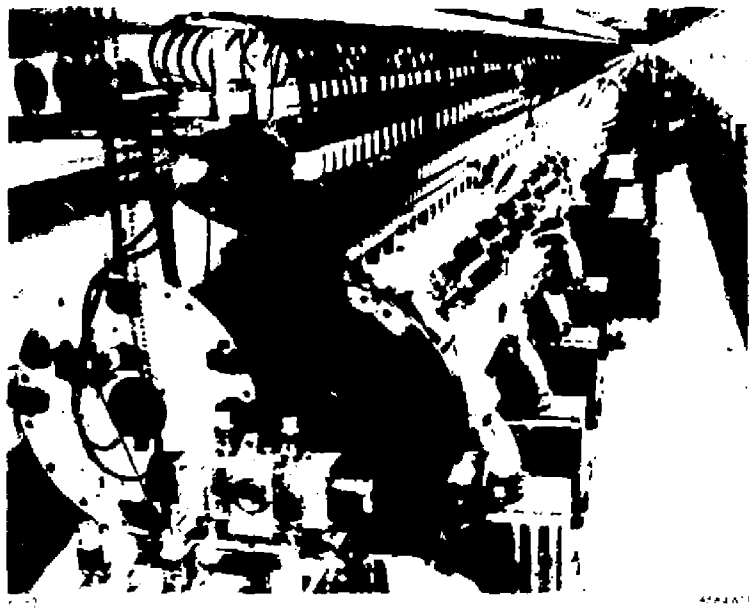


Fig. 11. BNL Alvarez linac from low-energy end.

Fig. 12. BNL Alvarez linac showing drift tubes, multi-stem arrangement, tuning bars, and slugs for field stabilization.

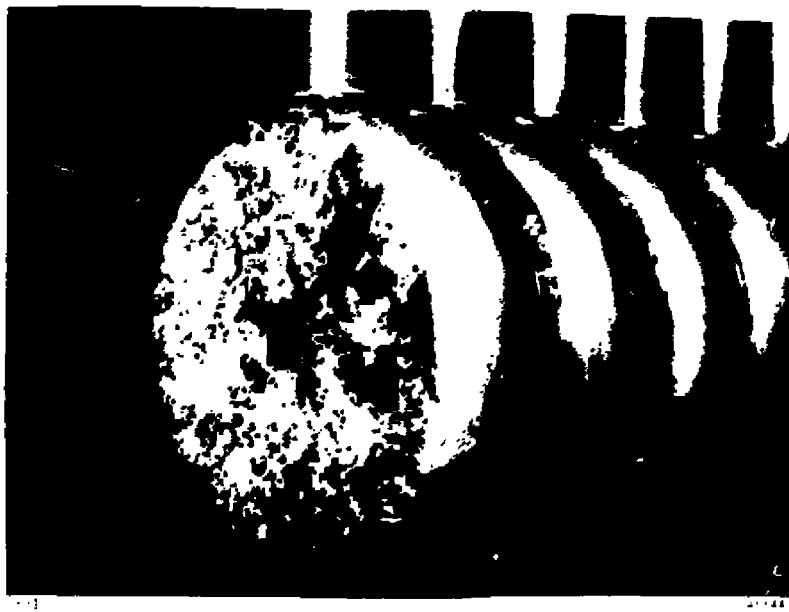


Fig. 13. First Alvarez drift tubes (notice sparking).

In all these Alvarez linacs, typical average axial fields are in the range of 1 to 2.5 MV/m, with gap fields at 6 to 10 MV/m. There are many tanks, each containing many drift tubes (refer again to, e.g., Table 1). Each tank is supplied with a peak power of several MW, typically $\sim$ 5 MW.

To obtain high [illegible] y, the remainder of the LAMPF accelerator which goes up to 800 MeV u[illegible] ucture at higher frequency, 800 MHz. This structure uses side-coupled cav[illegible] will defer explanation of this important development until a more systematic for[illegible]m for periodic structures has been developed in Sec. 4.

3.8 Recent Developments

One recent invention, first proposed in 1970 by I. M. Kapchinskii and V. A. Teplyakov at ITEP in the Soviet Union, is to replace the front end of the linac structure by a device in which the RF fields that are used for bunching and acceleration can also serve for transverse focusing. In the USA, the structure which can perform

these combined functions is called the RFQ (Q for quadrupole). As mentioned in Sec. 2, RFQ's have now operated successfully at several laboratories. Table 2 gives parameters for the designs of three such devices.

Table 2. Design Parameters for Three Different RFQ's

	Los Alamos	CERN	BNL2 Polarized
Ion	H^-	H^+	H^-
Frequency (MHz)	425	202.56	201.25
Input energy (MeV)	0.10	0.05	0.02
Output energy (MeV)	2.00	0.52	0.75
Nominal current (mA)	100	100	0.1
Transmission efficiency (%)	91.4	93.6	100
Output emittance[a]	0.021	0.05	0.011
Length (cm)	288.6	138.2	123.4
Peak rf power (kW)[b]	1020	304	66

[a]RMS normalized area/π in cm mrad units.

[b]Includes power required for the rf manifold and for the beam.

To understand how the RFQ works, refer to the early model built at ITEP (Fig. 14) and the simplified sketch of the vanes which make up the four opposing arrays of poles (Fig. 15).

The fields and currents are said to resonate in a modified TE_{210} mode. At any particular distance along the resonant structure, the poles are shaped and powered to give an electric quadrupole field. As the fields vary sinusoidally in time, an alternating gradient with net transverse focusing results. However, since the poles are also given a sinusoidal-like radial variation, this configuration leads to longitudinal acceleration and adiabatic bunching as well.

One reason for which this device is of such great interest is that in a small space, it can replace the very large Cockroft-Walton section of an ordinary injector, as can be inferred by noting the typical input and output energies in Table 2. Although the peak RF powers are relatively high, the RFQ's hold the promise of far better capture efficiency and reduced emittance growth at reduced cost.

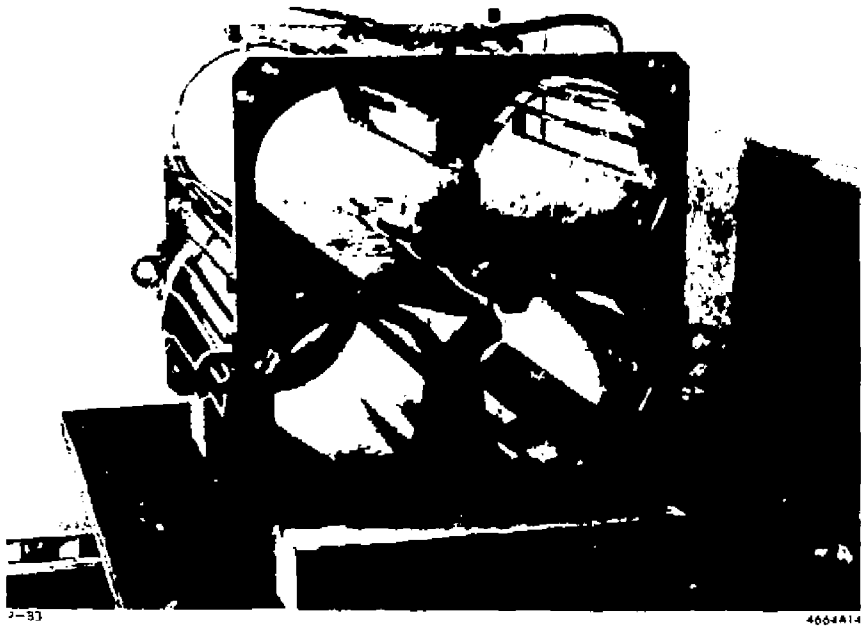


Fig. 14. RFQ structure built at ITEP, Moscow.

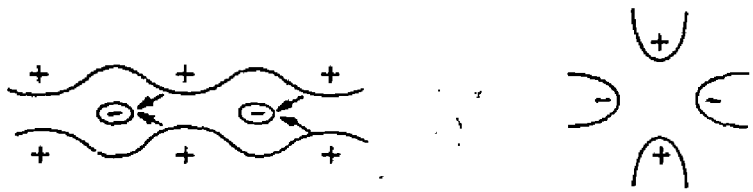
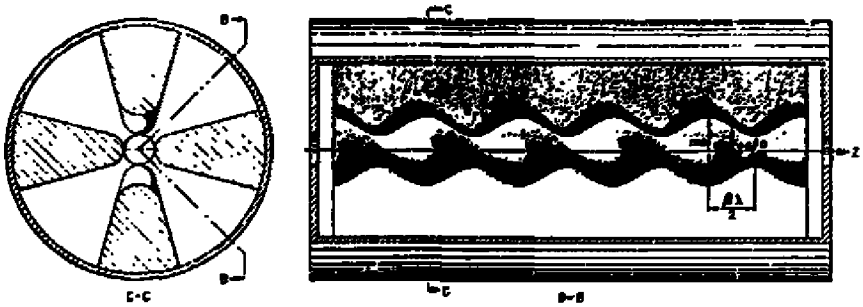


Fig. 15. Transverse and longitudinal views of the poles of an RFQ.

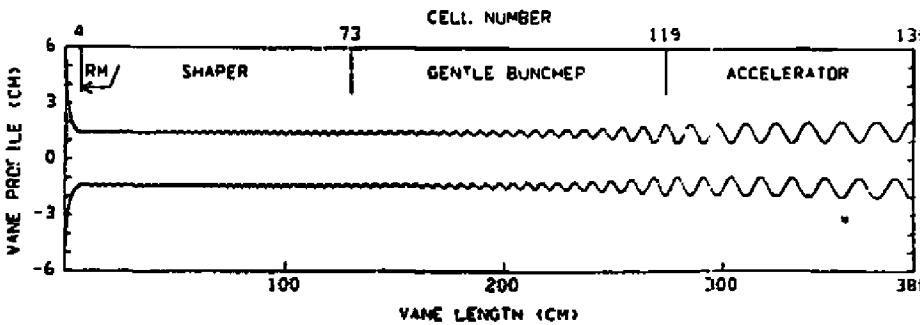


Schematic drawing of RFQ resonator.

FMIT Prototype RFQ

Ion: Deuteron						
Frequency: 80 MHz	W(MeV)	0.075	0.075	0.12	0.50	2.00
Nominal Current: 100 mA	E_0(MV/m)	3.01	17.6	17.6	17.6	17.6
Current Limit: 205 mA	B	1.00	1.00	1.12	2.13	2.13
PARMTEQ RESULTS	a(cm)	8.31	1.42	1.35	0.89	1.00
Input Current: 106 mA	r_0(cm)	8.31	1.42	1.42	1.42	1.63
Output Current: 100 mA	ϕ_s(deg)	-	-90	-72	-30	-30
Input (90%): 0.070	E_0(MV/m)	0	0	0.40	1.97	1.31
Oueput (90%): 0.160	V(kV)	185	185	185	185	185
Input (RMS): 0.017	A	0.20	6.82	6.82	6.82	5.20
Output (RMS): 0.016	L(m)	0	6.7	130	274	388

Beltrance



FMIT prototype RFQ pole-tip shape.

Fig. 16. Characteristics of RFQ for FMIT linac (LANL).

Figure 16 gives detailed design information for the RFQ for the LANL FMIT project to accelerate deuterons. The cavity containing the 3.88 m long vane shows the pole profile in the four successive sections of the device: the radial matching section (RM), the bunch shaper (S), the gentle buncher (GB) and the accelerator (A).

Another interesting modern development is the linear induction accelerator mentioned in Sec. 2. The principle of this accelerator is illustrated by the module shown in Fig. 17.

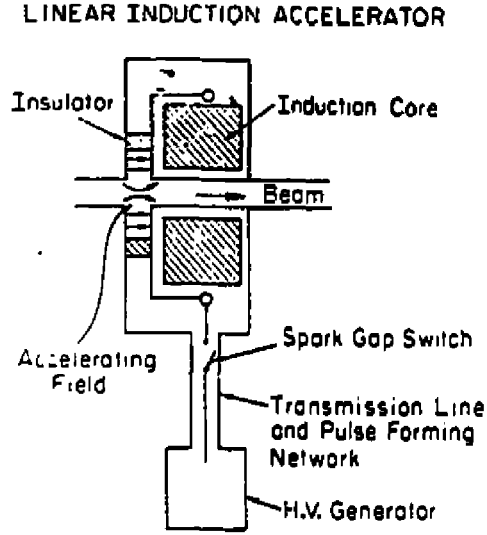


Fig. 17. Linear induction accelerator module.

As stated earlier, the idea of the induction linac is based on Faraday's law. The accelerating action can be described as analogous to a transformer. The primary of the transformer is excited by a high voltage source connected to a pulse forming network, switch and transmission line. The increasing current flowing through the primary when the switch is fired creates a rapidly changing magnetic flux in the toroidal core (thin-laminated iron or ferrite) which in turn induces a voltage across the accelerating gap. The beam then can be seen as the secondary winding. The accelerator itself consists of an array of such modules which can be triggered in succession at a rate commensurate with the velocity of the particles, their particular q/m and the mean

accelerating field. Accelerating capability, expressed as the product of gap voltage times pulse length, is limited by the saturation of the ferromagnetic material, which in turn is determined by the amount and type of material present. Typical average gradients of 2.5 MV/m can be reached. The technology of the induction linac seems to be well suited for very high beam currents (100 to 10,000 A) with short pulses (10 to 2000 nsec) at repetition rates up to 100 Hz.

The first version of the induction linac was the Astron injector designed and completed by N. Christofilos at LLNL in 1963. More advanced versions were built in 1971 at LBL and Dubna for their respective electron-ring accelerator (ERA) projects. The idea of the ERA was to form and accelerate rings of electrons and to embed protons or ions in the potential well of the electrons circulating in the rings. By accelerating the rings to relativistic velocities without "spilling" the protons, it was hoped that extremely high-gradient proton machines could be built. Unfortunately, it was discovered after some time that instabilities would prevent this scheme from being realizable at high energy. As a result, the ERA projects lost their momentum although at Dubna this type of accelerator is still used at low energy for ions. In the process, however, the feasibility of practical induction linacs for very high currents was established. Table 3 gives the parameters of some induction linacs in existence. All of them happen to be for electrons.

In recent years, the induction linac has been proposed for a new application: heavy ion fusion. In this application, the linac which is called the driver is to produce pulses of ions such as, for example, Krypton or Thallium (see Fig. 1) on the order of three megajoules. This pulsed beam energy, to be achieved by accelerating the ions up to a maximum of 20 GeV and currents of 15,000 A, is to be focused simultaneously from several directions onto a deuterium-tritium pellet to produce what is called inertial confinement fusion. The pulse length of the induction linac modules can be reduced along the machine as the energy of the ions increases and the time of passage through them decreases. Typically, the pulse length might be 2000 nsec at the front end and 20 nsec at the output. Progressive beam bunching is obtained by shaping the current pulses from the pulse forming networks so as to produce a positive voltage ramp within each module. Thus, later particles with higher energies catch up with earlier ones at lower energies. Even at an output energy of 20 GeV, the particles are still

non-relativistic and final power amplification up to several tera-watts can be obtained by longitudinal bunch compression in the transport lines leading to impact on the pellet.

Table 3. Parameters for Typical Induction Accelerators

Accelerator	Astron Injector Livermore 1963	ERA Injector Berkeley 1971	NEP 2 Injector Dubna 1971	ATA Livermore 1983
Kinetic energy, MeV	3.7	4.0	30	50
Beam current on target, A	350	900	250	10,000
Pulse duration, ns	300	2-45	500	50
Pulse energy, kJ	0.4	0.1	3.8	25
Rep rate, pps	0-60	0-5	50	5
Number of switch modules	300	17	750	200

4. Particle Acceleration by Guided Waves

We will now present a somewhat more general discussion of particle acceleration by guided waves. As we have seen in the previous section, a series of appropriately phased accelerating gaps, separated by drift spaces, can be analyzed simply to give a satisfactory description of a proton linac. The present discussion will include such structures as a special case but will focus on the guided wave itself rather than on the details of the structure. In the interest of simplicity we will concentrate mainly on particles with speeds near the velocity of light (i.e., electrons) although the same formalism applies also to slower particles. Rather than building on our previous discussion of proton accelerators, we will begin again and consider the general problem of particle acceleration by an electromagnetic wave.

4.1 MOTIVATION

As mentioned before, we are concentrating on particles with velocity near c and this makes it natural to attempt to use an electromagnetic wave in free space since it travels at the same speed and might have a cumulative effect on the particles. Unfortunately, such a wave has the property that its E and H-field are orthogonal in space and have components only in the transverse plane perpendicular to the direction of propagation. Since, in addition, the wave is bound to be of limited transverse extent, only one of two outcomes is possible. Either the particles start along a direction exactly collinear with the wave, in which case they see no force at all (i.e., the transverse electric and magnetic forces cancel exactly); or they start at a slight angle, in which case the transverse component of the E-field produces some acceleration along the direction of motion of the particles but the effect is not cumulative because the particles soon find themselves outside of the field of the wave.

There are two ways around this difficulty: either one keeps bending the beam back into the wave or one reflects the wave back into the beam. The first is the principle of the Inverse Free-Electron Laser which is discussed elsewhere in these Proceedings. The second one is the principle of all linacs using rf waveguides. An appropriate waveguide can redirect the electromagnetic wave to keep it superimposed on the particle trajectory. At the same time, it can also be such that it has a longitudinal electric field in the direction needed to impart energy to the particles. To have a cumulative effect,

however, the particles and the wave must be synchronous. As we will now see, it is the need for the simultaneity of these conditions which presents the main challenge to the accelerator structure designer

We will now go through a short eview of guided waves to clarify these points.

4.2 Wave Confined by Parallel Planes

Figure 18(a) shows two plane waves propagating in free space at angle θ on either side of the z-axis, and with their E-fields lying in the xz plane (the plane of the paper). The lines represent a snapshot of wavefronts of maximum $|\vec{E}|$ at a fixed time. They are labelled $+$, $-$, $+$, etc., to indicate whether $\vec{E}$ points left or right. The free space wavelength λ_0 is shown. It is given by

$$\lambda_0 = c\frac{2\pi}{\omega} \quad . \tag{4.1}$$

At a few of the intersections of the wavefronts the vector sums of the fields due to the two waves are shown. On the dashed lines the E-field is transverse to the z-axis. Along these lines it is permissible to place a mirror (i.e., a conducting plane) without affecting the fields.

Now the two waves can be thought of as bouncing back and forth between the two conducting planes. Changing our perception somewhat, we can see in Fig. 18(b) that there is a single wave having a longitudinal component of $\vec{E}$ and propagating in the direction of the z-axis. The guide wavelength λ_g is shown. It is the distance along the z-axis after which the pattern repeats and it is given by

$$\lambda_g = \frac{\lambda_0}{\cos\theta} \tag{4.2}$$

as can be deduced from the geometry of the figure

We are now confronted by one of the obstacles which tends to foil our efforts. If we wait for a time $2\pi/\omega$ which is one period of oscillation, the pattern will have advanced a distance λ_g. That is, the phase velocity is

$$v_p = \lambda_g\frac{\omega}{2\pi} = \lambda_g\frac{c}{\lambda_0} = \frac{c}{\cos\theta} \quad . \tag{4.3}$$

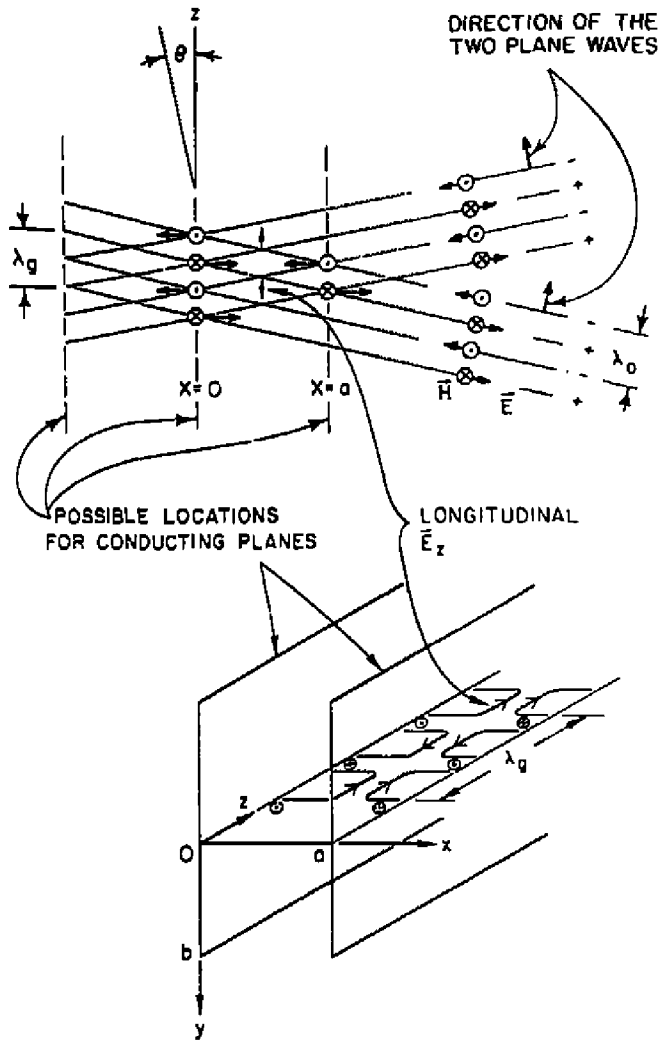


Fig. 18. (a) The origin of longitudinal electric fields in guided waves, and (b) their configuration between two parallel conducting planes.

Since $\cos\theta$ is less than 1, the phase velocity is greater than c. Since the particle to be accelerated is necessariiy traveling at a velocity lower than c, our wave cannot yet be used for cumulative acceleration.

Before addressing this problem, we should discuss some other preliminary questions regarding waveguides. The conducting wall could be moved out to any of the other planes on which E_z vanishes and this introduces the possibility of different waveguide modes. Also, it is natural to "close" our waveguide, for example by making it rectangular through the addition of conducting walls parallel to the plane of the paper: this makes even more modes possible. In this case it is still possible, although harder, to visualize our wave as made up of plane waves reflecting off the walls - four waves must be used. Rather than attempting this, we shall withdraw to a more abstract treatment based on Maxwell's equations. The wave we are considering has a transverse E_x and a longitudinal E_z component but only a transverse H_y magnetic component, which makes it a so-called TM wave. All field components can be derived from E_z which satisfies the wave equation

$$\frac{\partial^2 E_z}{\partial x^2} + \gamma^2 E_z = -\frac{\omega^2}{c^2} E_z \tag{4.4}$$

where we have already built in some assumptions which we will now spell out. We have assumed a time dependence of the form $e^{j\omega t}$ and a z-dependence $e^{-\gamma z}$ where γ is complex:

$$\gamma = \alpha + j\beta \quad . \tag{4.5}$$

Also, for simplicity, we are limiting ourselves to the case of Fig. 18 with no y-dependence. By inspection, we guess a solution satisfying (4.4) and the boundary conditions at $x = 0$ and $x = a$, with amplitude C,

$$E_z = C \sin\left(\pi \frac{x}{a}\right) e^{-\gamma z} \quad . \tag{4.6}$$

For now, a is the guide width. Substituting (4.6) into (4.4) we obtain, for this simplest possible case, the relation between guide wavelength and frequency

$$-\left(\frac{\pi}{a}\right)^2 + \gamma^2 = -\frac{\omega^2}{c^2} \quad . \tag{4.7}$$

For other modes the first term is changed, but for simplicity we will not generalize the notation to cover all cases. The reader should be aware of the fact that some of the following formulae, though correct in form, may need appropriate modification for other modes.

If there is no attenuation, α vanishes and (4.7) yields

$$\beta = \sqrt{\left(\frac{\omega}{c}\right)^2 - \left(\frac{\pi}{a}\right)^2} \tag{4.8}$$

where β is the wave number corresponding to the guide wavelength λ_g:

$$\beta = \frac{2\pi}{\lambda_g} \quad . \tag{4.9}$$

This β (which should really be called β_g or k_g) must not be confused with the usual $\beta = v/c$ used for relativistic particles. Expression (4.8) is called a dispersion relation. It can be rewritten in terms of the "cut-off frequency," a term to be justified shortly, given by

$$\omega_c = \frac{\pi}{a} c \tag{4.10}$$

whereby it becomes

$$\beta^2 = \frac{1}{c^2}(\omega^2 - \omega_c^2) \quad . \tag{4.11}$$

The graph of ω versus β in this field is called a Brillouin diagram. It is easily recognized to be a hyperbola (see Fig. 19). When ω drops below ω_c, β becomes imaginary and there is no more propagation. That is why ω_c is called the cut-off frequency.

Following the usual analysis of a dispersion relation, the phase velocity is given by

$$v_p = \frac{\omega}{\beta} = \frac{c}{\sqrt{1 - (\omega_c/\omega)^2}} \tag{4.12}$$

which, as expected, is greater than c, as can also be noted from the slopes shown in Fig. 19.

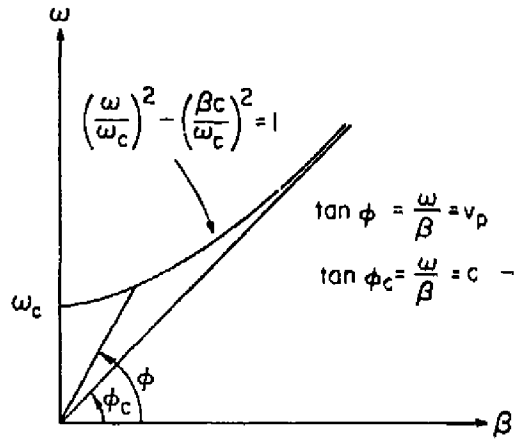


Fig 19. Brillouin diagram for guided waves in an unloaded guide.

It is useful to think about the situation at the extremes of this graph. At high ω, (4.11) becomes the free space linear relation between frequency and wave number. In Fig. 18 this corresponds to wavefronts that are very close together ($\lambda_g \to \lambda_0$, $\beta \to \omega/c$) and the wave travels freely down the guide, like a flashlight beam in a hollow tube. We have already seen that, although the speed v_p is close to desired, the wave is not suitable for acceleration because its E-field is transverse.

At the low frequency extreme, the lines in Fig. 18(a) are as wide apart as they can be while still allowing the picture to be completed; that is, the spacing from + to − should be a,

$$\frac{\lambda_0}{2} = a$$

which agrees with (4.10). The wave number β is very small, corresponding to large guide wavelength and large phase velocity. In this case the wave bounces back and forth at right angles to the guide; it is not useful for acceleration because of the high value of v_p.

Since accelerator guides are generally circular, perhaps more time than is justified has already been devoted to rectangular guides and parallel planes in particular. But there are a few more points which can illustrate our ideas in this simple case, free from

the intrusion of complicated geometry and mathematics. These have to do with power flow and dissipation in the walls.

The non-vanishing transverse field components between infinite parallel planes corresponding to E_z in (4.6) are given, using Maxwell's equations, by

$$E_x = -\frac{\gamma a^2}{\pi^2}\frac{\partial E_z}{\partial x} = -\frac{\gamma a}{\pi} C \cos\left(\frac{\pi x}{a}\right) e^{-\gamma z}$$
$$H_y = -j\frac{\omega\epsilon_0 a^2}{\pi^2}\frac{\partial E_z}{\partial x} = -j\frac{\omega\epsilon_0 a}{\pi} C \cos\left(\frac{\pi x}{a}\right) e^{-\gamma z} \quad . \tag{4.13}$$

The power flow P_{tr} is obtained by integrating the average Poynting vector over a transverse plane

$$P_{tr} = \frac{1}{2} Re \int E_x H_y^* dx dy \quad . \tag{4.14}$$

Since we are assuming that there is no y-dependence, i.e., the parallel planes are infinite in y, we don't have to satisfy the boundary conditions on any bottom or top plane. For a height b, with $\gamma = j\beta$ taken as pure imaginary, we get

$$P_{tr} = C^2 \frac{ab}{4\pi^2} \epsilon_0 \omega a^2 \beta \quad . \tag{4.15}$$

The standard method for calculating power loss in the walls (an issue of dominant importance in linear accelerators) is to take the fields obtained above for the lossless case to calculate the wall currents. In our case, the loss is in the side walls where, by Ampère's law, the longitudinal current I_z per height b is given by

$$I_z = H_y b \quad .$$

Although the model is not conceptually quite correct, one obtains the correct answer if one assumes that this current flows uniformly in a skin depth δ given by

$$\delta = \left(\frac{2}{\mu_0 \sigma \omega}\right)^{1/2} \tag{4.16}$$

where $\mu_0 = 4\pi \times 10^{-7}$ henry/m and σ is the conductivity in mhos/m. The average power lost in a short length dz of height b of the two side walls is of the form $2 \times \frac{1}{2}|I|^2 R$, i.e., given by

$$\frac{dP_{lost}}{dz} = \frac{|H_y b|^2}{\sigma \delta b} = \frac{\omega^2 \epsilon_0^2 a^2 C^2 b}{\sigma \delta \pi^2} \tag{4.17}$$

where $(\sigma\delta)^{-1}$ is the skin resistance R_s. To account for this loss, the propagation constant must have a small real part given by

$$\alpha = \frac{dP_{lost}/dz}{2P_{tr}} = \frac{1}{\beta a}\,\frac{4\pi}{\lambda_0}\,\frac{R_s}{\sqrt{\mu_0/\epsilon_0}} \quad . \tag{4.18}$$

Essentially the same information can be quantified by defining a Q factor for the guide

$$Q = \frac{\omega\, w_{st}}{dP_{lost}/dz} \tag{4.19}$$

where w_{st} is the stored energy per unit length. This quantity can be calculated from the maximum electric field, averaged over a wavelength:

$$\begin{aligned} w_{st} &= \frac{\epsilon_0}{4}\left[|E_x|^2 + |E_z|^2\right] ab \\ &= \frac{\epsilon_0}{4}\left[\left(\frac{\beta a}{\pi}\right)^2 + 1\right] C^2 ab \end{aligned} \tag{4.20}$$

which yields

$$Q = \frac{\pi\lambda_0\left[\left(\frac{\beta a}{\pi}\right)^2 + 1\right]\sqrt{\mu_0/\epsilon_0}}{8\,R_s\,a} \quad . \tag{4.21}$$

Finally, there is one more relation between these quantities. If the energy travels with the group velocity v_g, then

$$P_{tr} = v_g\, w_{st} \tag{4.22}$$

where

$$v_g = \frac{d\omega}{d\beta} = \frac{P_{tr}}{w_{st}} \quad . \tag{4.23}$$

Note that in Fig. 19, the group velocity is the slope of the $\omega - \beta$ diagram. It is equal to zero at cutoff ($\omega = \omega_c$) and equal to c when $\omega \to \infty$. The reader can verify that both definitions of v_g are equivalent and that Q and α are related by the equation $\alpha = \omega/2v_gQ$.

For a practical accelerator structure, all of the above quantities would have to be calculated. In what follows we will not attempt this as it can become quite complicated. Conceptually, however, these calculations are not different from those which have just been done. The exact formulae, if they can be obtained in closed form, resemble the ones we have just derived as far as dimensional factors are concerned. The numerical factors, of course, depend on the exact geometry.

4.3 Circular Waveguide

Most linear accelerator structures have configurations with circular cross sections. To simplify matters, let us take the case where there is just one boundary, $r = b$. If we limit ourselves to TM modes with no θ-dependence, the wave equation of (4.4) can be replaced by

$$\frac{\partial^2 E_z}{\partial r^2} + \frac{1}{r}\frac{\partial E_z}{\partial r} + \gamma^2 E_z = -\frac{\omega^2}{c^2} E_z \quad . \tag{4.24}$$

The other non-vanishing fields are given in terms of E_z by

$$H_\theta = -\frac{j\omega\epsilon_0}{k_c^2}\frac{\partial E_z}{\partial r} \tag{4.25}$$

$$E_r = -\frac{\gamma}{k_c^2}\frac{\partial E_z}{\partial r} \tag{4.26}$$

where

$$k_c^2 = \gamma^2 + \left(\frac{\omega}{c}\right)^2 \quad . \tag{4.27}$$

The simplest solution of (4.24) is the so-called TM_{01} mode where the first subscript (0) denotes zero-θ variation and the second (1) denotes one radial variation. For this mode the fields become

$$\begin{aligned} E_z &= C\ J_0(k_c r)e^{-\gamma z} \\ H_\theta &= \frac{j\omega\epsilon_0}{k_c}\ C\ J_1(k_c r)e^{-\gamma z} \\ E_r &= \frac{\gamma}{k_c}\ C\ J_1(k_c r)e^{-\gamma z} \end{aligned} \tag{4.28}$$

If one compares expressions (4.28) with those of (4.6) and (4.13), one sees that the main changes are that Bessel functions have replaced sine and cosine functions. The waveguide radius becomes involved because of the requirement that E_z vanish at $r = b$, which means that $k_c b$ must be the first zero of $J_0(k_c r)$:

$$k_c b = 2.405 \quad . \tag{4.29}$$

The guide dispersion relation takes the form of (4.11) if the cutoff frequency is defined by:

$$\omega_c = k_c c = 2.405 \frac{c}{b} \quad . \tag{4.30}$$

Note that, as claimed earlier, this expression differs from (4.10) only by a dimensionless numerical factor, in this case 2.405 instead of π. All the other waveguide properties such as P_{tr}, P_{lost}, α, Q can be calculated as for the case described in Sec. 4.3. We will not make the effort to do this because the TM_{01} mode still suffers from the shortcoming that the phase velocity, as given by (4.12), is greater than c, making it useless for cumulative acceleration. However, having obtained the field expressions for the circular waveguide, we can profitably return to our cylindrical resonator originally discussed in Secs. 3.2, 3.3 and 3.4.

4.4 CYLINDRICAL RESONATOR

In the circular waveguide, the dispersion relation (4.11) determines the longitudinal propagation constant as a function of ω:

$$\beta = \frac{\omega}{c}\left(1 - \frac{\omega_c^2}{\omega^2}\right)^{1/2} \quad . \tag{4.31}$$

If however the guide is capped off at both ends by conducting planes as shown in Fig. 20, a constraint is placed on β, and ω can only assume discrete values corresponding to resonances. As in Fig. 18, the boundary conditions at the plane surfaces can be satisfied by superimposing two waves of appropriate wavelength, one going left, one right. The result is a standing wave. A simple possibility is to take the mode where $E_r = 0$ at the plane surfaces and

$$d = \frac{\lambda_g}{2}$$

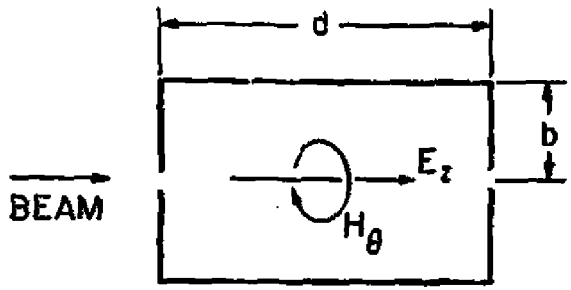


Fig. 20. Cylindrical resonator.

i.e.,

$$\beta = \frac{\pi}{d} \ .$$

Using (4.11) and (4.30) this yields for the resonant frequency

$$\omega_r = c\left[\left(\frac{2.405}{b}\right)^2 + \left(\frac{\pi}{d}\right)^2\right]^{1/2} \ . \tag{4.32}$$

This is the so-called TM_{011} mode which has one variation in the z-coordinate. It turns out, however, that in a cylindrical waveguide, there is an even lower mode (the one used in Alvarez linacs) in which E_r vanishes not just at the ends, as required by the conducting planes, but everywhere. For this mode λ_g is infinite, which makes β zero. The resonant frequency is

$$\omega_r = 2.405\frac{c}{b} \tag{4.33}$$

which is independent of d. This mode is called the TM_{010} mode. The expressions for the field given by (4.28) are reduced to

$$\begin{aligned} E_z &= C\ J_0(k_c r) \\ H_\theta &= -j\,\frac{C}{\eta}\,J_1(k_c r) \end{aligned} \tag{4.34}$$

where $\eta = (\mu_0/\epsilon_0)^{1/2} = 377$ ohms. From these, using the definitions of the shunt impedance R (3.7) and Q (3.12) or (4.19), it is possible to show that

$$R = \frac{\eta^2 d^2}{R_s \pi b(d+b)\,J_1^2(k_c b)} \tag{4.35}$$

and

$$Q = \frac{\eta}{R_s}\,\frac{2.045}{2(1+b/d)} \tag{4.35}$$

where $R_s = 1/\sigma\delta$ is as defined previously. It would be instructive for the reader to verify these expressions and to get a feeling for the magnitudes involved by working out some numerical examples for R, Q, R/Q and T_f at 200 MHz ($d = 8$m) and 2856 MHz ($d = 3.5$cm).

Note that for this calculation we have neglected the effect of the holes at the ends of the resonator, needed to let the particles pass. As we discussed earlier, a linear accelerator can be built by stringing many such cavities together and adjusting their phases to the beam to give cumulative energy gain. However, rather than assuming that these cavities are independently driven, we will now try to analyze the model where they are connected together into a periodic chain of coupled resonators.

4.5 Wave Propagation in Coupled Resonator Chains

For simplicity we will start with the simplest possible resonator, namely an LC circuit. The pillbox resonator in the TM_{010} mode is indeed little more than an LC circuit where C_s is the capacity between the end plates and L is the inductance of the toroidal outer region where the magnetic field is predominant. Representing the coupling between resonators by a parallel capacitance $2C_p$, the circuit diagram for the chain is shown in Fig. 21.

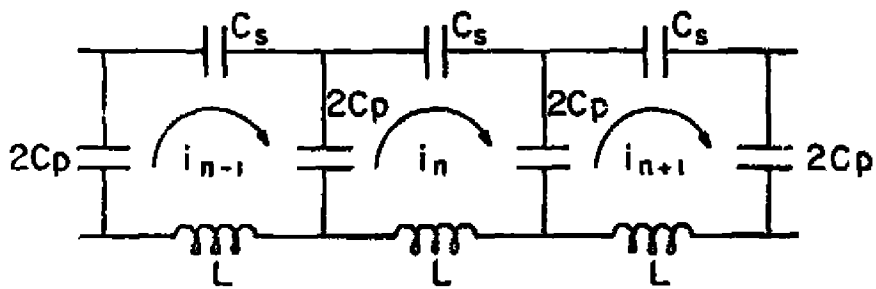


Fig. 21. Chain of lightly coupled resonators.

Note that in the limit of large C_p, the resonators are uncoupled. For the pillbox resonators this would correspond to having no beam hole. To insure that we are close

to that situation, we will assume that

$$C_p \gg C_s \ . \tag{4.37}$$

We do not intend to imply that representing the coupling by a capacitance is an obvious model for the actual coupling, but it is simple and leads to the same qualitative results as more complicated models.

It was L. Brillouin who first systematized the analysis of such periodic chains, but the essential results we require can be derived in a few lines. Applying Kirchoff's current law to the n*th* loop in Fig. 21 yields

$$\left(\frac{2}{j\omega 2C_p} + \frac{1}{j\omega C_s} + j\omega L\right) i_n - \frac{1}{j\omega 2C_p}(i_{n+1} + i_{n-1}) = 0$$

or

$$i_{n+1} - 2\cos(\beta d) i_n + i_{n-1} = 0 \tag{4.38}$$

where

$$\cos(\beta d) = 1 + \frac{C_p}{C_s} - \omega^2 C_p L \ . \tag{4.39}$$

From the point of view of the present discussion we have simply introduced a new symbol (βd) which is a certain combination of ω and the circuit parameters. But, naturally, our choice of symbol anticipates identifying this quantity with the previously defined quantities β and d.

Two solutions of the difference equation (4.38) are

$$i_n = \begin{matrix}\sin \\ \cos\end{matrix}(n\beta d + \phi_0) \qquad (n = 0, 1, 2, \ldots) \tag{4.40}$$

as can be checked with standard trigonometric identities. If β as defined by (4.39) is real, then (4.40) will represent standing wave solutions which the structure can support. Remember that it is implicit in an impedance equation such as (4.39) that the time dependence $e^{j\omega t}$ is assumed and the real part is to be taken. By superimposing the two solutions (4.40) with proper phase, unattenuated traveling waves going in either direction can be formed. The condition for β to be real is that

$$-1 \leq \cos \beta d \leq 1 \tag{4.41}$$

which can be re-expressed as a band of propagating frequencies

$$\omega_0 < \omega < \omega_\pi \ . \tag{4.42}$$

The frequencies ω_0 and ω_π defining the limits of the band are obtained by combining (4.39) and (4.41) to give

$$\omega_0 = \frac{1}{\sqrt{LC_s}} \tag{4.43}$$

$$\omega_\pi \simeq \frac{1}{\sqrt{LC_s}}\left(1+\frac{C_s}{C_p}\right) \ . \tag{4.44}$$

Approximation (4.37) was used in obtaining ω_π. Note that the subscripts on ω_0 and ω_π are the corresponding values of the angle βd at the band edges.

The Brillouin diagram for this structure is shown in Fig. 22 which is a graph of (4.39). The band of propagating frequencies includes the resonant frequency $1/\sqrt{LC_s}$ of the uncoupled resonators. For weak coupling (C_p much larger than C_s) the band is narrow, and for strong coupling (C_p closer to C_s) it is broad. In fact, we have

$$\frac{\Delta\omega}{\omega_0} = \frac{\omega_\pi - \omega_0}{\omega_0} = \frac{C_s}{C_p} \ . \tag{4.45}$$

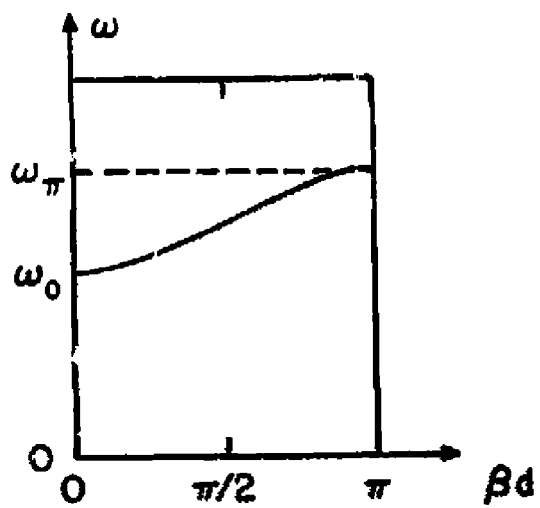


Fig. 22. Brillouin diagram for a chain of coupled LC circuits.

Certain features of this Brillouin diagram are important properties of periodic structures in general. It is even in βd and periodic in βd with period 2π since these are both properties of the function $\cos \beta d$. The dashed extensions in Fig. 23 convey this effect. But this extension is superfluous for the system we are discussing since, in (4.40), shifting βd by 2π does not alter the solution. This is due to the fact that our single cells are fully described by one variable (the loop current) and continuous variation of the argument is not defined. On the other hand, our single cell could be more complicated. For example, suppose it had a mode of oscillation at some higher frequency as well. This would lead to another propagating band as shown in Fig. 23. For linear circuits, these modes do not influence each other. For reasons to which we will return, we have chosen to emphasize the band $-\pi < \beta d < 0$ as the principal band for the higher mode.

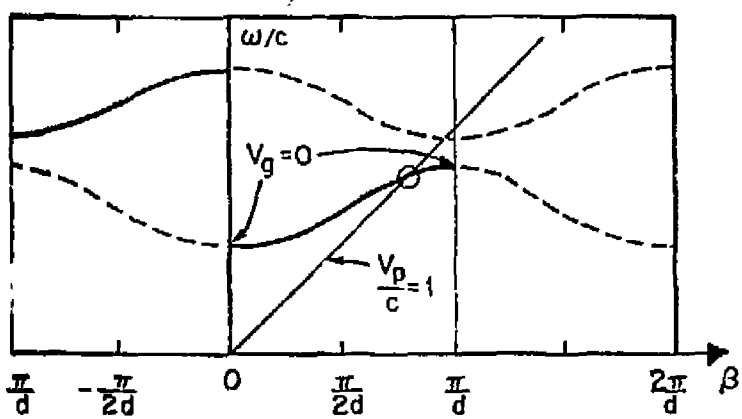


Fig. 23. Brillouin diagram for a chain of cells, each having two modes.

As usual with dispersion relations, the Brillouin diagram can be used to extract phase and group velocities. Taking d as the physical length of one cell and with ω/c as the vertical axis, the slope of a 45° line, as shown, corresponds to a phase velocity equal to c, which is required for an electron accelerator. This condition fixes the operating frequency at the intersection point. Another general feature is that the group velocity (given by the slope $d\omega/d\beta$) is always less than c and vanishes at ω_0 and at ω_π.

4.6 Periodically Loaded Structures

We now have to face up to the fact that a linear accelerator is more complicated than a chain of one-dimensional resonators. The individual cells support not just one, but many modes and it may not be justified to treat the coupling as weak. While the latter fear is somewhat unfounded, it still makes sense to attack the problem from yet another point of view – analyzing the structure as a periodically-loaded waveguide.

This brings us back to the question which was left open in section 4.2: how do we slow down the phase velocity of a waveguide? For the chain of pillbox resonators the trick was easy since in the lowest mode the frequency is independent of the length d and in the weak coupling limit the resonators can be phased independently. There are many other ways of slowing down the phase velocity, some of which are shown in Fig. 24.

Of these structures, the simplest is the iris-loaded or disk-loaded waveguide. This is just another name for the pillbox resonator chain which we have been discussing, except that now the iris hole radius, a, has become larger. Treating the irises as perturbations, we can ask what effect they have on the waveguide modes. If there is a wave propagating along the guide in the $+z$ direction, there will be a reflection at each iris. For a band of frequencies, these reflections will interfere destructively because of the phase shifts resulting from the spatial separation of the irises. As a result, at these frequencies the irises will have only a minor effect on the propagation and the dispersion relation will resemble Fig. 19.

But there are certain frequencies, $\omega_{n\pi}$, for which the reflected waves from successive irises are exactly in phase. At such frequencies, unattenuated propagation is impossible since reflections necessarily result in comparable amplitudes for forward and backward going waves. Standing waves are possible, however. The condition for this "Bragg" reflection is

$$\beta = \frac{\pi}{d}\, n \qquad n = 0, 1, 2, \ldots \tag{4.46}$$

since this makes the phase advance in going ahead to the next iris and then back an integral multiple of 2π. This situation is again characterized by a vanishing group velocity.

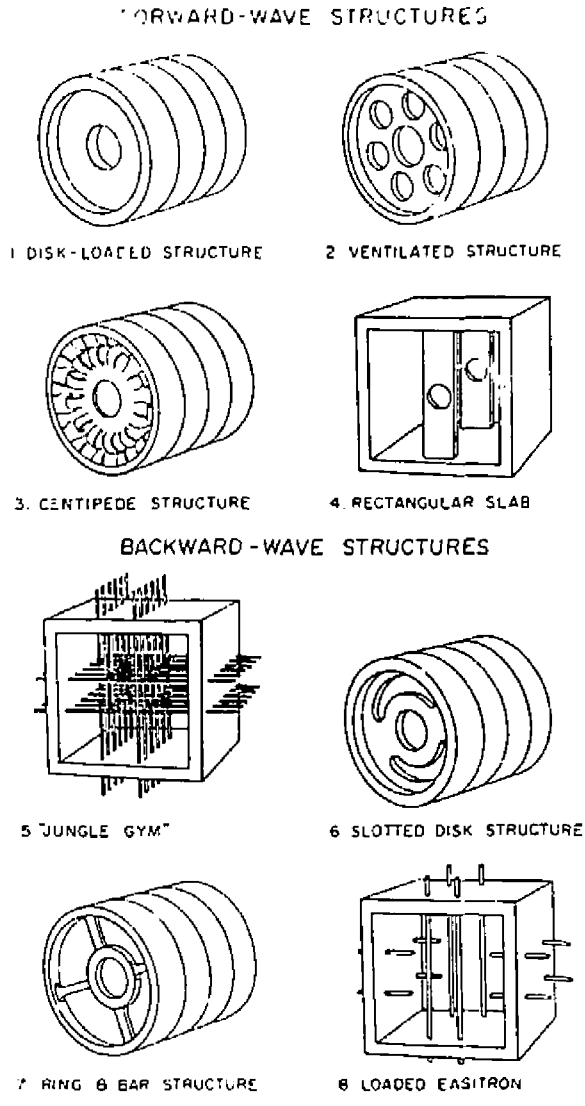


Fig. 24. Some slow-wave structures.

We can also anticipate that for frequencies near $\omega_{n\pi}$ there will be a strong effect of the irises on the dispersion relation between ω and β. To investigate this, consider Fig. 25 in which we have guessed two standing-wave field configurations of opposite symmetry relative to the irises. In the figures, lines and arrows indicate electric fields, and circles indicate magnetic fields. It can be seen that in case (b), since E_r vanishes at the locations of the irises, the presence or absence of the irises does not have much effect. For this case then, the relation between ω and β will be very similar to that for a waveguide without irises. But in case (a) the fields cannot be as drawn since the presence of the irises must force E_r to vanish there, at least at the radial positions occupied by the metal irises. In other words, currents flowing in the irises cause the standing wave of wavelength $2d$ to have a different (actually lower) frequency than if the irises were not there.

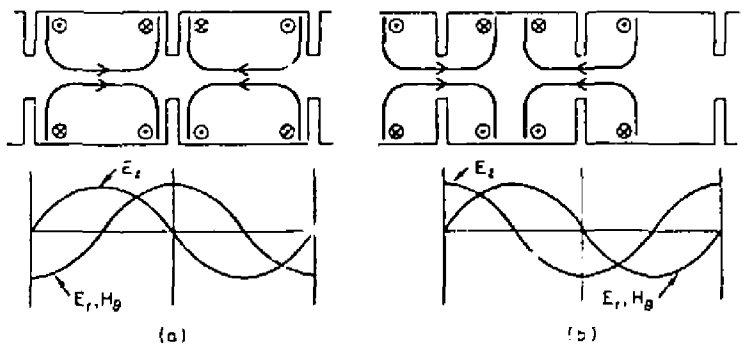


Fig. 25. "Guessed" standing-wave field configurations of opposite symmetry with respect to the irises at the stop band. Lines and arrows indicate electric fields, circles indicate magnetic fields.

We now have in hand all the elements required to complete qualitatively the Brillouin diagram for the disk-loaded waveguide (Fig. 26). We can start with Fig. 19 for the unloaded guide. As β increases away from zero, ω will initially follow the hyperbola but will then flatten out and become horizontal at $\beta = \pi/d$ where the group velocity must vanish. We have just seen that there is another frequency giving this same wave number and that, furthermore, it lies on the hyperbola for the unloaded guide. This allows us to complete the branch between $\beta = \pi/d$ and $\beta = 2\pi/d$.

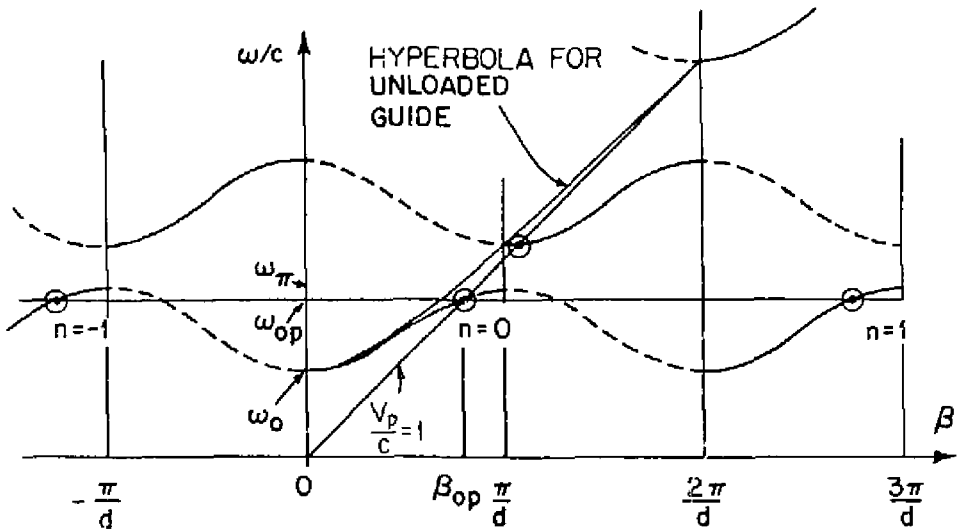


Fig. 26. Brillouin diagram for a disk-loaded waveguide.

Similar considerations lead to successively higher branches. Again, for an electron accelerator, possible operating points lie on the 45° line. Clearly, the loading disks are capable of lowering the branches to intersect this line as shown. For normal operation, the intersection with the lowest branch is chosen. It is indicated by ω_{0p}, β_{0p}.

4.7 Space Harmonics

A more quantitative understanding and completion of Fig. 26 requires that one find a method to calculate the exact dispersion relation and the corresponding electric and magnetic field expressions for a given slow-wave structure. To match the periodic boundary conditions shown in Fig. 27, it is customary to invoke a theorem ascribed to the French mathematician, Floquet. This theorem states that for a given mode of propagation at a given steady-state frequency, the fields at one cross section differ from those one period away only by a complex constant.

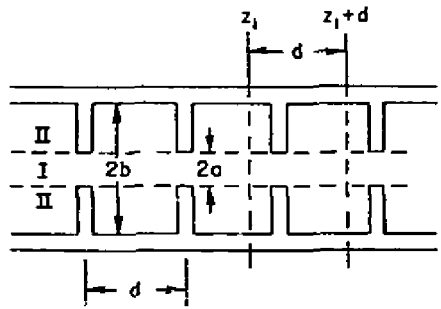


Fig. 27. Definition of periodic planes (z_1) and ($z_1 + d$) for the application of Floquet's Theorem, and regions I and II for the formulation of the boundary value problem.

The proof of the theorem lies in the fact that when a structure of infinite length is displaced along its axis by one period, it cannot be distinguished from its original self. (For a more formal proof, see, e.g., Landau and Lifshitz, *Mechanics*).

Suppose we write

$$E(r, z, t) = F(r, z)\, e^{-j\beta_0 z}\, e^{j\omega t} \tag{4.47}$$

where $F(r, z)$ is a periodic function of z with period d, $\beta_0 d \leq \pi$, θ-dependence is excluded, and propagation is assumed in the positive z-direction. Referring to planes

(z_1) and $(z_1 + d)$ in Fig. 27, expression (4.47) becomes:

$$\begin{aligned} &\text{At} \quad z = z_1 \quad E(r, z_1, t) = F(r, z_1)\, e^{-j\beta_0 z_1}\, e^{j\omega t} \quad , \\ &\text{at} \quad z = z_1 + d \quad E(r, z_1 + d, t) = F(r, z_1 + d)\, e^{-j\beta_0 (z_1 + d)}\, e^{j\omega t} \quad . \end{aligned} \tag{4.48}$$

But by definition

$$F(r, z_1 + d) = F(r, z_1) \quad . \tag{4.49}$$

Thus

$$E(r, z_1 + d, t) = E(r, z_1, t)\, e^{-j\beta_0 d} \quad , \tag{4.50}$$

quite independently of where z_1 is taken, which indicates that the form of (4.47) indeed obeys Floquet's Theorem. Since $F(r, z)$ is periodic, we can expand it in a spatial Fourier series of the form:

$$F(r, z) = \sum_{n=-\infty}^{n=+\infty} C\, a_n(r)\, e^{-j\frac{2\pi n}{d} z} \tag{4.51}$$

where n is an integer which can take all positive and negative values. It can be shown that under rotational symmetry, the solution given in (4.47) becomes

$$E(r, z, t) = \sum_{n=-\infty}^{n=+\infty} C\, a_n\, J_0(k_{rn} r)\, e^{j(\omega t - \beta_n z)} \tag{4.52}$$

where

$$\beta_n = \beta_0 + \frac{2\pi n}{d} \quad , \tag{4.53}$$

$$k_{rn}^2 = k^2 - \beta_n^2 \quad , \tag{4.54}$$

C is the usual amplitude factor, and a_n is the amplitude of the n^{th} so-called space harmonic.

There are several methods of determining the a_n amplitudes of these space harmonics for a given structure. Before we mention them, we must make a number of observations regarding Eqs. (4.51) through (4.54).

- The fact that the expansion contains all positive as well as negative integers arises from the fact that we have assumed an even structure (with reflection symmetry with respect to the origin, $z = 0$).

- The wave $E_z(r, z, t)$ which meets the boundary conditions of the periodically loaded structure is made up of space harmonic components which each have a different propagation constant (4.53). However, once $\beta_0(\omega)$ is known, all β_n's are known.

  Each space harmonic has a different phase velocity given by

$$v_{pn} = \frac{\omega}{\beta_0 + \frac{2\pi n}{d}} \quad . \tag{4.55}$$

  For any $n < 0$, the phase velocity of the space harmonic is negative, i.e., a particle would have to travel in the negative direction to remain in phase with it. The continuous lines in Fig. 26 all belong to a wave whose energy (or "group") travels in the positive direction, and conversely for the dotted lines. At the ω_{0p} intercepts, since the $\omega - \beta$ curves are simply translations of themselves by $2n\pi/d$, all have the same slope corresponding to the same group velocity.

- In the definition of shunt impedance (3.8) and r/Q (3.14), the values of V_0/d that should be used are those of the space harmonic which is synchronized with the particle. In most cases, that is the fundamental, i.e., a_0.

- From (4.52) and (4.54) it is seen that the terms $a_n J_0(k_{rn} r)$ are not constant as a function of radius. If k_{rn}^2 is negative, k_{rn} becomes imaginary and J_0 changes into I_0, which instead of falling off with r, increases. There is one notable exception which turns out to be extremely fortunate for the operation of linear accelerators. Indeed, when $n = 0$ and $v_p = c$, $\beta_0^2 = k^2$ and $k_{r0} = 0$, which makes $J_0 = 1$ and the accelerating field independent of radius. As a result, acceleration and energy acquired are independent of the radial position of the synchronous particle.

- The actual variation of ω versus β_0 depends on the relative diameters ($2a$ and $2b$) of regions I and II in Fig. 27. When $2a \ll 2b$, the passbands are very narrow until, when $2a = 0$, there is no more passband at all. Conversely, when $2a \to 2b$, the passband becomes very wide until finally, we get back to the pure TM_{01} circular waveguide mode.

The choice of the phase shift per period for the fundamental is generally made on the basis of where the highest shunt impedance per unit length can be obtained. Snapshots of electric field configurations for the cases where $\beta_0 d = 0$, $\pi/2$, $2\pi/3$ and π are shown in Fig. 28. Somewhat loosely, these 0, $\pi/2$, $2\pi/3$ and π configurations are also called modes, not to be confused with *TM*, *TE* modes or superpositions thereof. Note that these field configurations have been chosen at such an instant in time that they can interchangeably represent a traveling wave or the maximum of a standing wave. As will be seen later, for traveling-wave accelerators of the disk-loaded type, the optimum choice is $\beta_0 d = 2\pi/3$ (120°). For standing-wave accelerators, it is $\beta_0 d = \pi$ or $\pi/2$, as shown in Sec. 4.9).

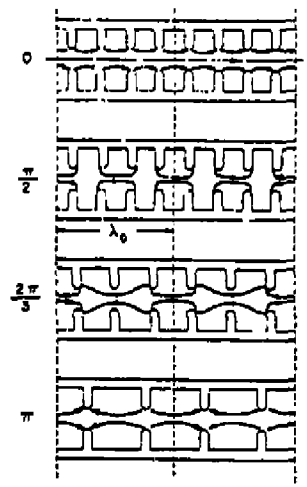


Fig. 28. Snapshots of electric field configurations for disk-loaded structures with various phase shifts per period (also loosely called "modes")

The $\omega - \beta$ diagram can be determined exactly, point-by-point, by obtaining the resonant frequencies of stacks of n cavities, which have $(n + 1)$ resonances, one for each value of $\beta_0 d$. For example, a stack of three cavities can resonate for $\beta_0 d = 0$, $\beta_0 d = \pi/3$, $\beta_0 d = 2\pi/3$ and $\beta_0 d = \pi$. The frequencies can be obtained either experimentally or by means of computer programs such as SUPERFISH or LALA

which find the resonances and calculate the fields for standing waves. A conversion formula can then be used to obtain the fields for the traveling waves. The fields can also be found experimentally by means of bead perturbation techniques in cavity structures. An example of a field profile for the $\beta_0 d = \pi/2$ "mode" is shown in Fig. 29.

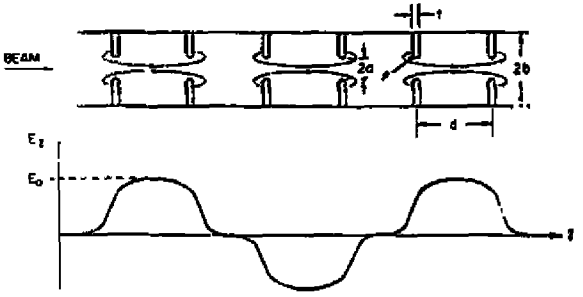


Fig. 29. Example of approximate axial E_z-field amplitude for $\beta_0 d = \pi/2$ " mode." This is a snapshot of a traveling wave at a given time.

Yet another theoretical approach to obtain the fields is to set up general solutions in the two regions (I) and (II) shown in Fig. 27 (sometimes called "tube" and "slot" fields) and then to match these solutions along the common boundary. This approach was first used successfully in the early days (1948) by W. Walkinshaw, in England, long before the advent of the modern field-mapping computer codes.

Once the fields are obtained, a simple Fourier-type analysis can be used to get the space harmonic amplitudes. The other structure parameters such as σ, Q, v_p, v_g, r and r/Q can then also be calculated from the definitions and by the techniques introduced in Sec. 4.2. This entire subject is discussed in detail in a later chapter by E. Keil.

Before we leave the subject of space harmonics and Brillouin diagrams, we should return briefly to the second passband shown in Figs. 23 and 26. As we hinted earlier, this passband has its origin in the second "normal mode" of oscillation of the cells making up the chain. From what we have learned about the lowest passband, the $n = 0$ space harmonic of this higher band propagates in the negative direction ($d\omega/d\beta < 0$). To call attention to this fact in Fig. 26, we have drawn this branch as a dashed line.

Because of this property, this passband is often referred to belonging to a backward wave, i.e., the fundamental has its phase velocity and group velocity going in opposite direction.

The existence of this and higher passbands can be responsible for serious instabilities in the operation of linear accelerators, even though the RF sources driving them do not in principle supply power at these frequencies. What happens is that the short bunches of relativistic electrons can excite the structure at frequencies at which these higher modes are synchronous or quasi-synchronous with the bunches, whereby their amplitude can grow, either through feedback or cumulative interaction. This interaction can lead to deleterious effects such as beam deflection and ultimate beam breakup. Such effects can limit the achievable beam current in multi-bunch beams and cause emittance growth in single-bunch beams. These problems are treated in a later lecture on instabilities by A. Chao.

4.8 Traveling Waves or Standing Waves?

Until now, not much has been said about how the fields are established in the linac and where the rf power is fed into the structure. There are basically three ways in which this can be done.

In the first, the structure is of the traveling-wave type, in which case the power is fed in at one end, propagates through the structure with some attenuation and the balance is absorbed in a load (Fig. 30a). Steady-state is reached when the structure is filled with energy after one pass. For a length ℓ, the filling time is $t_F = \ell / v_g$.

In the second, the structure is of the standing-wave type, in which case there is only one coupler, either at one end or in the middle of the structure (Fig. 30b). The fields build up through multiple reflections, the number of reflections of any consequence being determined by the wall losses. If the structure is long and the wall losses are high, the number of reflections that add any significant field is small, and vice versa. If the structure is matched to the power source, there is a so-called Q_L or loaded Q associated with it which is equal to half the Q_0 associated with the wall losses. The maximum field build-up is attained with an exponential time constant $t_F = 2Q_L/\omega$ or Q_0/ω, as defined in expression (3.16). In three filling times, the fields attain 95% of their ultimate value. One of the questions that must inevitably be asked is: aren't the

reflected waves useless since they have their phase velocity in the wrong direction and hence do not deliver any net energy to the beam? The answer is that this is indeed true for all values of $\beta_0 d \neq \pi$. However, when $\beta_0 d = \pi$, the $n = 0$ space harmonic of the forward wave coalesces with the $n = -1$ space harmonic of the backward wave and both contribute energy to the beam if it is in synchronism. As a result, all standing-wave linacs are built in the so-called π-mode, or in a variation thereof, as we will see in the next section.

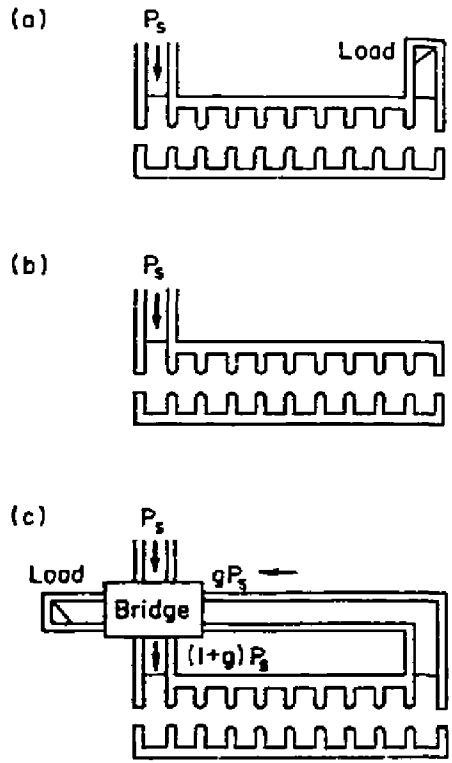


Fig. 30. Linacs with (a) traveling-wave, (b) standing-wave and (c) traveling-wave with feedback structures and source power P_s.

The third method, shown in Fig. 30c, which was proposed in England in 1949 by R.-Shersby-Harvie and L. Mullett, is to feed the power left over at the output end of

the traveling-wave structure back to the input via a low-loss waveguide. At this point this power can be recombined at the correct phase with the source power by means of a bridge. In this scheme, the backward-wave problem is circumvented at the cost of considerable complexity. This scheme has few advantages over the standing-wave scheme and has been used rarely.

The question of how one chooses between a traveling-wave structure and a standing-wave structure depends on the desired application and a number of other constraints such as the available rf power, energy and length, and the required pulse length and duty cycle.

The following general comments and comparisons can be made:

1. Since the standing-wave structure is generally built with a fixed coupling, it cannot be matched under all conditions. One can show that if the power source has a relatively short rf pulse length (measured in a few t_F's), at best 80% of the rf energy can be delivered to the structure during the pulse. The remaining 20% is reflected to the source. This mismatch can under certain conditions be destabilizing to the source, causing frequency pulling, phase oscillations and fluctuations in output power. Furthermore, since the fields take several filling times ($2Q_L/\omega$) to build up to their final value, the beam will vary in energy throughout the rf pulse if it is injected after, say one filling time, and the total pulse is only a few filling times long. If the application calls for a precisely defined particle energy, this is then an unacceptable disadvantage. Therefore, in all such cases, the traveling-wave linac, which reaches steady-state in one filling time, is favored over its standing-wave counterpart.

2. On the other hand, standing-wave structures such as the side-coupled or the disk-and-washer structure can be designed to have shunt impedances as much as 50% higher than traveling-wave structures at the same frequency. This advantage also prevails for phase velocities in the range $0.3 < v/c < 0.8$ where disk-loaded structures are relatively inefficient. If the pulse length is long or infinite, i.e., the machine is CW, whatever energy is wasted during the filling time is negligible. Conversely, the traveling-wave structure, for a long pulse, always wastes in a load that power which is left at the output end. Thus, long-pulse proton linacs like the LAMPF accelerator, and all CW linacs use standing-wave

structures. Two examples of CW machines are the room temperature structures proposed for microtrons, and the superconducting structures which reach shunt impedances 10^4 to 10^5 higher than at room temperature. A third example is the structure used in all $e^{\pm}$ storage rings. In this case, the standing-wave structure meets the extra requirement that it can accelerate particles of opposite charges in opposite directions.

3. There is one case which falls between those described above. That is the case where one wants to obtain a high gradient in a short length with a source of relatively low power (e.g., a magnetron) and a pulse length of a few filling times. Typical applications of this case are the short medical accelerators and special sections for injectors and positron sources. The standing-wave structure here is at an advantage because the one-way wall losses are low and the large number of reflected waves can build up the fields to a high level. Were the reflected energy to be dissipated in an external load after a single pass, most of the power would be wasted. The only way to make the traveling-wave structure competitive in this situation would be to greatly increase the filling time of the short section by lowering its group velocity accordingly. This in turn would increase the energy density and the fields. However, to achieve this very low group velocity, one would have to reduce the iris diameter considerably. This would lead to beam transmission problems and also to difficulty with dimensional tolerances. Consequently, in these cases, designers are willing to live with the disadvantages of the standing-wave structure discussed earlier in order to achieve the required high energy within a very short length.

4.9 Resonant Coupling

There is one problem in the above discussion of standing-wave structures that still needs to be elaborated because it has led to important innovations. We have seen that using the π-mode is indeed the way to make use of both forward and backward waves. The problem, however, is that if we resonate a structure of reasonable length, say 50 cavities, at ω_{op} where $\beta_{op}d = \pi$, it will contain 50 half-wavelengths at this frequency. Adjacent to it, however, there will be a frequency at which it will contain 49 half-wavelengths or for which $\beta d = 49/50\ \pi$. Since the $\omega - \beta$ curve has zero slope at $\beta_{op}\pi = d$, the neighboring resonance at $\beta d = 49/50\ \pi$ will be extremely

close to ω_{op}, to the point where any small cavity mistuning, frequency modulation or beam loading will excite the neighboring mode (or modes). In the jargon, the π-mode is said to be very unstable. If, on the other hand, ω_{op} were in the middle of the pass-band, i.e., at $\beta_{op}d = \pi/2$, the slope of the $\omega - \beta$ curve would be greatest, and maximum frequency separation (or mode stability) would be achieved. This line of reasoning led E. Knapp at LANL to invent the side-coupled cavity structure in the early 1960's. Figure 31 makes the invention understandable. Figure 31a shows the standing-wave structure resonating in the $\pi/2$-mode. Clearly, this configuration is inefficient compared to the π-mode since under steady-state conditions, every second cavity is unexcited and produces no acceleration. Figure 31b shows a modification where the excited cavities have been stretched and the unexcited cavities have been shrunk. Every cavity however, short or long, is still tuned to the same frequency. Thus the $\pi/2$ mode and its inherent stability are preserved. Finally, Fig. 31c illustrates the discontinuous step in the invention: the empty "useless" cavity has been placed on the side, off the beam line. From the rf point of view, the mode remains $\pi/2$ but for the beam, it has become π-like. The side-cavities which do the coupling are empty but resonant, the feature which is responsible for the name of "resonant coupling." Figure 31d shows the physical embodiment of the idea. The accelerating cavities are shaped and provided with nose cones for maximum shunt impedance, and the coupling cavities are staggered to reduce asymmetries introduced by the slots.

Another way of looking at what has been done is to consider the $\omega - \beta$ plots which accompany Fig. 31a, b and c. The double periodicity (in 31b) would normally lead to a break and a stop-band in the $\omega - \beta$ curve at the $\pi/2$ point: two branches are formed as in Fig. 28. However, through the process of resonant coupling, what is called "confluence" between the two branches is achieved, as shown in Fig. 31c, re-establishing the stability of the $\pi/2$ mode. Thus in conclusion, at the cost of some extra complexity, the side-coupled cavity chain and a number of other structures based on the same principle have yielded the best of two worlds over a wide range of particle velocities: good efficiency of the π-mode, and insensitivity to fabrication errors and beam loading of the $\pi/2$-mode. The resonant posts or slugs used in the Alvarez structure shown earlier in Fig. 12 are another example of the same idea at lower frequency.

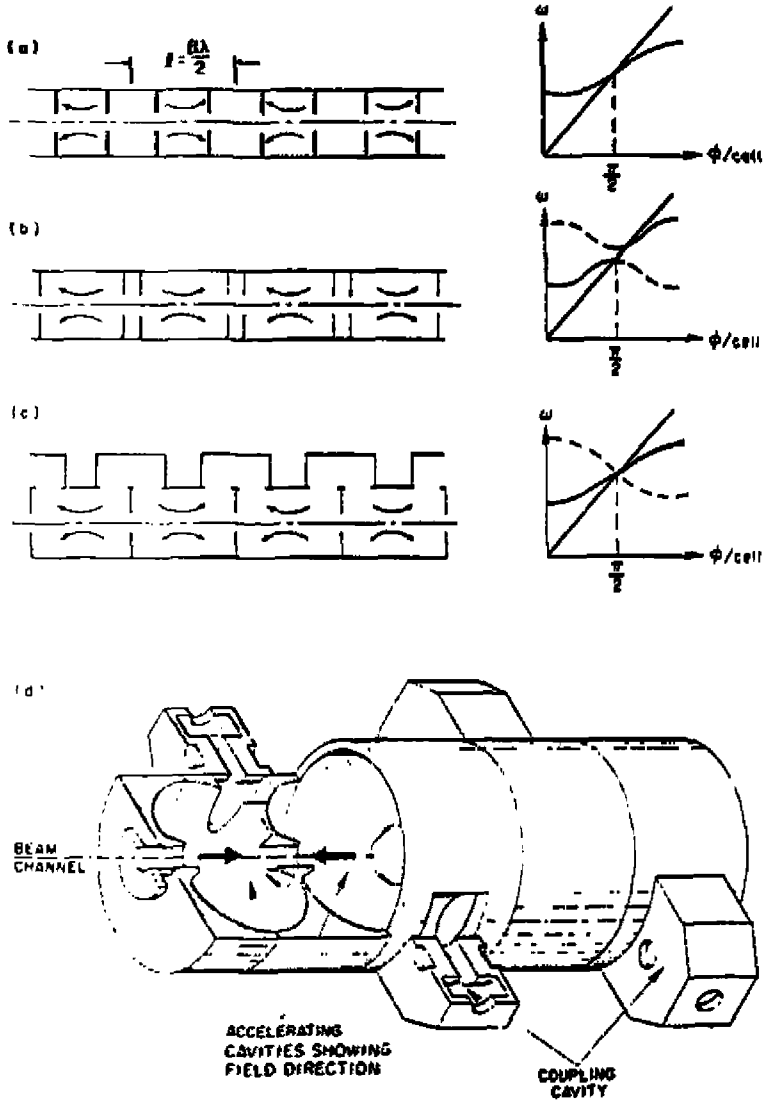


Fig. 31. Evolution of $\pi/2$-mode iris-loaded structure to π-mode side-coupled structure.

5. Electron Linear Accelerators

In this last section we will cover in very broad terms a few of the major design and operating features specific to electron linacs. A number of references will be made to the SLAC linear accelerator as an example. For more details, the reader should consult the book on the *Stanford Two-Mile Accelerator*, R. B. Neal, General Editor, W. A. Benjamin Inc., 1968, hereafter abbreviated S.T.M.A., as well as L.A. referred to earlier.

5.1 The Main Elements

The main elements of an electron linac are shown in Fig. 32. The electron gun is a triode consisting of a thermionic cathode, a wire-mesh control grid and an anode. The potential difference between the cathode and the anode is typically 80 kV which gives the electrons a velocity of $c/2$. The grid is normally held at a voltage somewhat negative ($\sim$ 50 V) with respect to the cathode so that the gun is biased off except when a beam pulse is needed. To trigger the gun, a positive pulse, typically of a few hundred volts, is applied for the desired length of time. Fast transistor pulse circuitry has improved in the last few years to the point where beam pulses as short as one or two nanoseconds are now achievable. Most electron linacs operate with beam pulses in the 1-10 μsec range at repetition rates up to 500 and even 1000 pps. The continuous stream of electrons out of the gun first passes through a prebuncher. This is a single low-Q re-entrant cavity supplied with typically $\sim$ 2 kW of rf power at the fundamental frequency. The function of this prebuncher is to velocity-modulate the electrons with peak gap voltages up to 10-15 kV so that early particles are slowed down and late particles are speeded up. As a result of this process, longitudinal compression or bunching results.

At a distance of 30 cm downstream, one can get on the order of 70% of the electrons into a phase interval of 70 electrical degrees. At this point, a refinement which is not shown in the figure is sometimes used to discard a large fraction of the bunches. This is a transverse electric field rf beam chopper which can act as an additional gate (say for single bunches every 12.5 nsec if the chopper frequency is 40 MHz) by deflecting all bunches except the ones that go through at zero-crossing. Yet another variation is to combine the prebuncher and the chopper into one rf cavity at a subharmonic frequency

of the accelerator. This scheme then produces a train of more widely spaced and longer bunches with a potential of packing more charge into each of them. Combined with a fast grid pulser, it can generate a single bunch. This scheme is being used for the SLAC Linear Collider (SLC) injector.

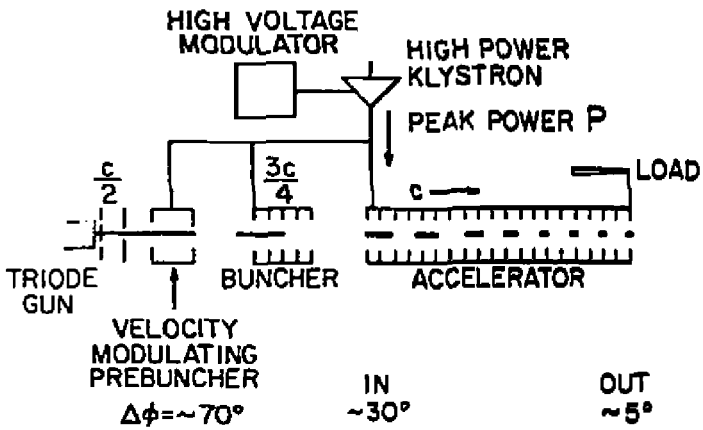


Fig. 32. An electron linac (or the beginning of one!)

The next element in Fig. 32 is the buncher. At SLAC this is a 4-cell-long, $2\pi/3$-mode, traveling-wave section with a constant phase velocity of 0.75 c and a maximum power input of 2 MW . The bunches emerge from it with a kinetic energy of about 250 keV in a phase interval of 30°. Bunching takes phase here because on the average the entering bunches are made to ride ahead of the crest and early electrons at first get slowed down while later ones get accelerated. Being slower than the wave when they enter the section, they then all rise towards the crest and gain energy. In addition to this average motion, they perform approximately one quarter-wave oscillation with respect to the central electron: the buncher acts like a quarter-wave impedance matching transformer in the $e - \Phi$ space of Fig. 6, between the prebuncher on the one side and the accelerator section on the other. The accelerator section is a traveling-wave structure with phase velocity c. In the SLAC case, as will be described later, it is an S-Band, constant-gradient, $2\pi/3$-mode, 3-meter long structure. The 30° long bunches are caused to enter it around the field null ahead of crest, and they asymptotically approach the crest of the wave as they are accelerated. Unlike the protons at lower

velocities, they do this monotonically without oscillations. A simple expression for this asymptotic motion will be given in sub-section 5.3. At the output of the first accelerator section, the electron bunches are 5° long and their energy is typically on the order of 30 MeV for an rf power input P of 16 MW. This power is typically supplied by a high power klystron which in turn is pulsed by a high voltage modulator. RF power is also supplied to the buncher and to the prebuncher.

In addition to these accelerating and bunching fields, the electrons are acted upon by their own space charge forces, radial rf fields from the structures, stray magnetic fields, and in case of very high currents, wakefields which they themselves induce and leave behind. For all these reasons, steering, degaussing and focusing devices, not shown in Fig. 32, are also incorporated into the design of the linear accelerator. As an example, in the SLAC injector, there are two thin magnetic lenses to focus the beam. The first lens is located downstream of the gun, and the prebuncher gap is placed at the beam waist formed by this first lens. The second lens is located after the prebuncher and produces a second waist at the entrance of the buncher. These waists, i.e., points of minimum beam diameter, are created at those points where one wants to minimize the effects of radial fields which generally grow linearly with radius.

Along the buncher and the accelerator section there is almost always a solenoidal DC magnet field of a few kilogauss. The required focusing action of the axial magnetic field is obtained because there are radial DC fields where the beam enters the solenoid. This produces an azimuthal force causing the beam to begin to spiral. The azimuthal velocity and the axial field produce the desired inward force.

The linear accelerator which has been described so far either ends at this point, in which case the beam is extracted for whatever use it is destined; or it just constitutes the injector or beginning of a longer, multisection machine such as the Orsay, Saclay, Kharkov, KEK, MIT, NIKHEF, SLAC and many other linacs in the world. The design that has been described is fairly standard. The only major variance is in the buncher which in a number of short accelerators is designed with a tapered phase velocity which gently forms and "escorts" the bunches up to the velocity of light. The buncher and the accelerator in this case are constructed in one section powered at the upstream end. In the low energy (4-6 MeV) medical linacs, the rf source is a 2MW magnetron and the structure, as discussed in Sec. 4.8, is generally of the standing-wave type,

under one meter of length. Solenoidal focusing can sometimes be avoided by using the alternating-gradient focusing effect of the π-mode radial fields in each cavity.

Referring back to Fig. 1, it should be noted that in all these machines, when the electrons have acquired a kinetic energy of 3 MeV, they have already reached a velocity of 0.99 c, i.e., they are totally relativistic. Hence, the beam dynamics "drama" described earlier for protons in Secs. 3.5 and 3.6, which takes place over hundreds of meters in a machine such as LAMPF, is played in an electron linac in the first meter!

The longer, higher energy electron linacs simply use many separately powered sections. The trade-off between rf power and length is generally resolved on the basis of economics, technological factors, the availability of land and the cost of AC power to operate the machine over a number of years. In addition to the rf structures, klystrons and modulators, these longer machines use periodic focusing systems in the form of FODO arrays, quadrupole doublets or triplets. Of course they also have their associated vacuum, mechanical support, cooling, instrumentation and control, and other support systems. For positron operation, they generally have a tungsten target which is inserted at a point where the electrons have achieved a few hundred MeV, or at SLAC 5-10 GeV. The positrons are generated in a wide forward cone which must be focused by a high field solenoid so that the beam can be contained within the transverse admittance of the subsequent accelerator sections.

All these subjects are described in much greater detail in L.A. and S.T.M.A.

5.2 Structure Design, Choice of Parameters and Energy Gain

Having described the main elements of a "generic" electron linac, we will now return once more to the structure and to some of the practical choices involved in its design. The first choice is to select the general type of structure. In what follows we will concentrate on the simple traveling-wave disk-loaded structure. Only a few comments will be made on standing-wave structures for the sake of comparison. The most critical parameters to be chosen are the frequency, the phase-shift per cell, and the attenuation per section. On the attenuation, there is in addition a sub-choice between two possibilities. Either the attenuation is kept constant as a function of length in the section and the structure is uniform, in which case the power decays exponentially: the structure is called "constant impedance." Or the attenuation is increased as a func-

[illegible] by varying the dimensions of the structure so that the power decays [illegible] and the field is kept constant: the structure is called "constant gradient." To understand the effect of these parameters, it is best to calculate the electron energy gain in terms of the shunt impedance per unit length, redefined once more as

$$r = \frac{E^2}{-dP/dz} \tag{5.1}$$

where E is now the electric field amplitude of the fundamental, synchronous space harmonic.

The other quantities of interest defined earlier are:

$$\frac{r}{Q} = \frac{E^2}{\omega\, w_{st}} \tag{5.2}$$

$$Q = \frac{\omega\, w_{st}}{-dP/dz} \tag{5.3}$$

$$\alpha = \frac{-dP/dz}{2P} \tag{5.4}$$

$$P = v_g\, w_{st} \tag{5.5}$$

from which one can get the relation

$$\alpha = \frac{\omega}{2 v_g Q} \tag{5.6}$$

For a section of length ℓ, the kinetic energy gain per section for a synchronous electron riding on the wave crest is then

$$\Delta E_K = e \int_0^{\ell} E(z)\, dz \quad . \tag{5.7}$$

For the constant-impedance case, α is not a function of z and from (5.1) we get

$$E^2(z) = 2\alpha r\, P(z) \tag{5.8}$$

whereby the integration of (5.7) leads to

$$\Delta E_K = e\left(\frac{1 - e^{-\tau}}{\tau}\right) (2\tau)^{1/2}\, (P_0 r \ell)^{1/2} \tag{5.9}$$

where P_0 is the input power to the section and

$$\tau = \alpha \ell \ . \tag{5.10}$$

For the constant-gradient case, the form of $\alpha(z)$ that leads to a constant $E(z)$ must be determined. To do this in closed form, we will assume that we can neglect the variations of τ and Q that occur if the structure dimensions are modified to adjust $\alpha(z)$, and that, in fact, the only quantity that changes rapidly is the group velocity in (5.6). Since from (5.1)

$$E^2(z) = -\tau \frac{dP(z)}{dz} \ , \tag{5.11}$$

if one wants to make $E(z)$ constant, $dP(z)/dz$ must be kept constant. Then redefining (5.10) as

$$\tau = \int_0^{\ell} \alpha(z) dz \ , \tag{5.12}$$

the linear decaying profile of $P(z)$ must be given by

$$P(z) = P_0 \left[1 - \frac{z}{\ell}(1 - e^{-2\tau}) \right] \tag{5.13}$$

whereby

$$\alpha(z) = \frac{1 - e^{-2\tau}}{2\ell \left[1 - \frac{z}{\ell}(1 - e^{-2\tau}) \right]} \tag{5.14}$$

and

$$v_g(z) = \frac{\omega \ell}{Q} \frac{\left[1 - \frac{z}{\ell}(1 - e^{-2\tau}) \right]}{(1 - e^{-2\tau})} \ . \tag{5.15}$$

We can see that to get a constant-gradient profile, we must let the group velocity decrease linearly with the same slope as the power. This is obvious if we refer to (5.5) since then w_{et} is also constant. We can now obtain the kinetic energy gain for the constant-gradient section. Expression (5.7) is simply $\Delta E_K = eEt$ and since, from (5.13)

$$\frac{dP}{dz} = -\frac{P_0}{\ell}(1 - e^{-2\tau}) \ , \tag{5.16}$$

using (5.11) we obtain

$$\Delta E_K = e(1 - e^{-2\tau})^{1/2}(P_0\, r\, \ell)^{1/2} \quad . \tag{5.17}$$

Using expressions (5.9) and (5.17), we can now draw some very general conclusions.

Both expressions depend on the factor $(P_0\, r\, \ell)^{1/2}$. The trade-off between P_0 and ℓ mentioned earlier is immediately apparent. For maximum energy gain, one wants to maximize r and the factors of τ. From Fig. 33, we see the behavior of r as a function of the number n of disks per wavelength. The solid curves are calculated for an array

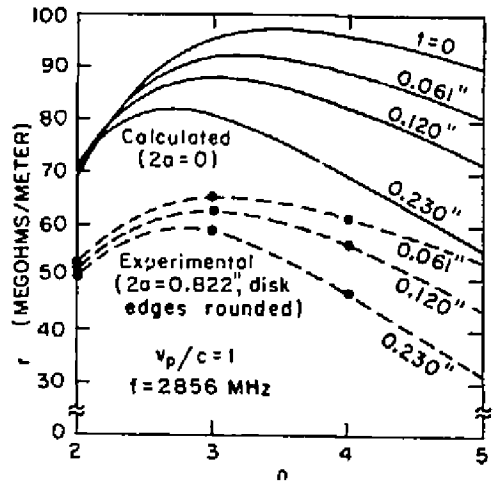


Fig. 33. Shunt impedance per unit length versus number n of disks per wavelength for various thicknesses t.

of pillboxes of diameter $2b$ with zero-iris diameter ($2a = 0$). For this reason, they are somewhat higher than the experimental values shown by the dashed curves. Both sets are taken for three values of disk-thickness t, $v_p/c = 1$ and $f = 2856$ MHz. It is seen that a broad maximum is obtained for $n = 3$ disks/λ_0, i.e., for the $2\pi/3$-mode discussed earlier. For this reason, this mode was selected for the SLAC linac. The thicker value of t (0.230") was selected for greater mechanical strength and because it permits a greater radius of curvature on the disk edge, which decreases the risk of electrical breakdown.

The factors of τ in Eqs. (5.9) and (5.17) are compared in Fig. 34. It is seen that for small τ, they are about equal and that for $\tau \sim 1$, they are close to 0.9. The value of $\tau = 0.57$ was selected for the SLAC linac as a broad compromise between maximizing energy and minimizing the effect of beam loading (see next sub-section) and the filling time ($t_F = 2Q\tau/\omega$). The constant-gradient design was selected because of a number of advantages such as lower ratio of peak-to-average field and uniform power dissipation per unit length. It also turned out *ex-post-facto* that the constant-gradient structure, due to its non-modular construction, is less susceptible to cumulative beam breakup because the effective interaction length with the synchronous transverse deflecting mode is foreshortened.

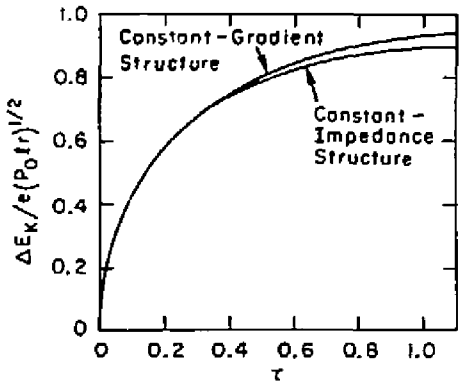


Fig. 34. Normalized energy gain versus τ.

The choice of rf frequency is also made on the basis of a broad set of criteria. Table 4, which is taken from S.T.M.A., shows the frequency dependence of some of the key parameters. The single most important quantity is the shunt impedance r which varies as $f^{1/2}$. Since Q falls as $f^{-1/2}$, $r\omega/Q$ which is a measure of E^2/w_{st} varies as f^2. This is obviously correct since the volume per unit length of structure varies as f^{-2}, all transverse dimensions varying as f^{-1}. Thus, any linac design for which one wants to maximize the beam energy and in which the stored energy is discarded after each pulse, will be favored by high frequency. On the other hand, the frequency cannot be increased indefinitely because the decreased iris diameter ($2a$) eventually becomes a limitation, both from the point of view of beam interception and transverse wakefields which scale roughly as $(2a)^{-3}$. Also, at lower frequency, the larger energy stored makes

the structure's sensitivity to beam loading. Finally, there are factors which, although not as fundamental, can be just as important such as the availability of high power rf sources as a function of frequency, and the absolute machining tolerances. For all these reasons the SLAC frequency was set at 2855 MHz or 10.5 cm wavelength. Most other linacs in the world are built at the same frequency or nearby, in the vicinity of 3000 MHz. The final dimensions of the SLAC sections are shown in Fig. 35, together with the linear profile of the normalized group velocity (v_g/c) and the resulting shunt impedance. As we see, $r(z)$ has an average value of 57 MΩ/m and it varies slowly with z, which makes our original assumptions acceptable. Figure 36 is a photograph of the final 3.05 meter-long SLAC constant-gradient section with its input and output

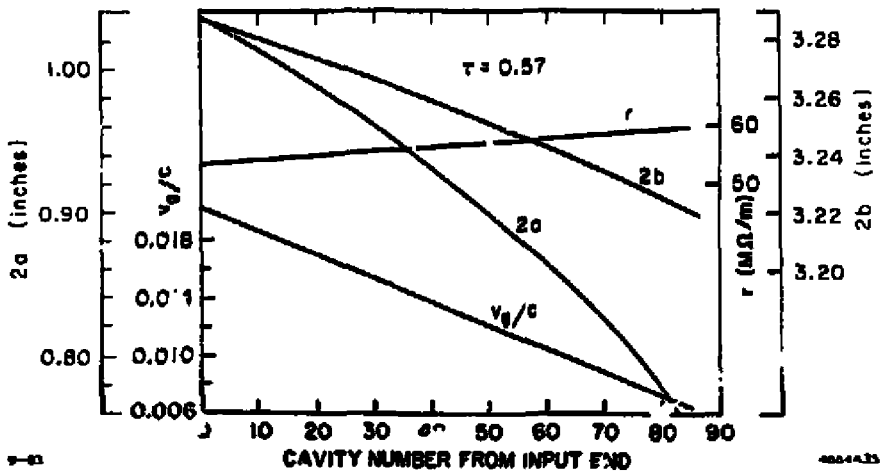


Fig. 35. Variations of 2b, 2a, v_g/c and the shunt impedance r (corrected for the fundamental space harmonic) as a function of cavity number along SLAC 3.05 meter constant-gradient section for $t = 0.230$ in.

Table 4.

Frequency Dependence of Principal Linear Accelerator Parameters

Parameter	Frequency dependence	Frequency preference		Notes
		High	Low	
Shunt impedance per unit length (r)	$f^{1/2}$	X		a
RF loss factor (Q)	$f^{-1/2}$		X	a
Filling time (t_f)	$f^{-3/2}$	X		a, b
Total RF peak power	$f^{-1/2}$	X		a, b, c
RF feed interval (l)	$f^{-3/2}$		X	a, b
No. of RF feeds	$f^{3/2}$		X	a, b, d
RF peak power per feed	f^{-2}	X		a, b, c
RF energy stored in accelerator	f^{-2}	X		a, b, c
Beam loading ($-dV/di$)	$f^{1/2}$		X	a, b, d
Peak beam current at maximum conversion efficiency	$f^{-1/2}$		X	a, b, c, f
Diameter of beam aperture	f^{-1}		X	a
Maximum RF power available from single source	f^{-2}		X	e
Maximum permissible electric field strength	$f^{1/2}$	X		g
Relative frequency and dimensional tolerances	$f^{1/2}$	X		a, b
Absolute wavelength and dimensional tolerances	$f^{-1/2}$		X	a, b
Power dissipation capability of accelerator structure	f^{-1}		X	a, b, d

Notes:

a. For direct scaling of modular dimensions of accelerator structure.
b. For same RF attenuation in accelerator section between feeds.
c. For fixed electron energy and total length.
d. For fixed total length.
e. When limited by cathode emission.
f. When limited by beam loading.
g. Approximate; empirical.

Fig. 36. SLAC 3.05 meter-long constant-gradient accelerator section.

couplers. Using Eq. (5.17), the reader should verify that the energy obtained per section, taking into account $\sim$ 10% feed losses, is approximately

$$\Delta E_K = 10\, e\, \sqrt{P_{MW}}\ \text{MeV} \quad . \tag{5.18}$$

If the power of a 36 MW klystron is split four ways, as it is at SLAC, then

$$\Delta E_K = 30\,\text{MeV/section}$$

or 120 MeV/klystron.

In comparison, a matched standing-wave structure yields a steady-state energy of

$$\Delta E_K = e\,(P_0 \ell r)^{1/2} \tag{5.19}$$

with values of r that can be 50% higher. Thus, in principle, energies that are 25% higher could be obtained for the same values of P_0 and ℓ. In actual fact, because of all

the difficulties listed under Secs. 4.8 and 4.9, such standing-wave structures are only built for the special applications discussed earlier.

5.3 Beam Current, Emittance, Bunch Length and Beam Loading

In the preceding sub-section, not much has been said about the beam itself, its characteristics, its behavior and its effect on the structure. The subject of beam dynamics in an electron linac is again so vast that we cannot hope to do it justice in the space allotted here. Thus only a few aspects of the problem will be touched upon.

A bunch of electrons can generally be characterized by the total number of particles contained in it, the transverse emittance, the bunch length and the energy spectrum. In addition, the beam, which is generally made of a train of bunches, has a pulse length and a repetition rate (unless it is CW), and the notion of beam power can be associated with it. We will now discuss these various properties, illustrating them with some practical examples.

Total number of particles per bunch. Typically, in an electron linac, the total number of electrons per bunch is in the range 10^7 - 10^{11}. For example, the SLAC linac, when it accelerates a beam of 50 mA peak current within a 1.6 μsec pulse, produces a train of $\sim$ 5000 bunches and a total of 5×10^{11} electrons per pulse. Thus, the number of electrons per bunch in this case is 10^8 and the beam energy per pulse at 25 GeV is 2000 joules, or 720 kilowatts at 360 pulses/sec, i.e., close to one megawatt! The goal of the SLC project at SLAC is to obtain single bunches of electrons and positrons of 5×10^{10} particles, i.e., of 500 times greater intensity. Such single bunches with $\sim$ 8 nanocoulombs of charge at 50 GeV will carry an energy of 400 joules, or 72 kilowatts at 180 pulses/sec.

By contrast, a medical linac at 5 MeV, 200 mA peak current and a duty cycle of 10^{-3} can deliver to its x-ray target a power of one kilowatt.

Transverse emittance. Assuming a bunch of Gaussian or other not too unreasonable shape, one can define transverse emittances $\sigma_x\sigma_x'$ and $\sigma_y\sigma_y'$, where σ_x and σ_y are the rms transverse dimensions of the bunch and σ_x' and σ_y' are the corresponding rms angular dimensions; for a single particle $x' = p_x/p_z = p_x/\gamma\beta m_0 c$ and $y' = p_y/p_z = p_z/\gamma\beta m_0 c$. The p_x and p_y's refer to transverse momenta and γ is the usual $(1-\beta^2)^{-1/2}$. Often, the above quantities are defined as $\beta\gamma\sigma_x\sigma_x'$ and $\beta\gamma\sigma_y\sigma_y'$, or simply

$\gamma\sigma_x\sigma'_x$ and $\gamma\sigma_y\sigma'_y$ when $\beta \to 1$. These quantities have the advantage of being invariant with energy. They are measured in many different units of which some typical ones are $m_0c \cdot$ cm or (MeV/c) cm. Since the second unit is twice as big as the first, the emittance measured in (MeV/c) cm is half that in $m_0c \cdot$ cm. In the x, p_x and y, p_y phase-space planes, Gaussian beam envelopes are represented by ellipses and one often quotes the above quantities multiplied by π to denote emittance areas (see Fig. 37). According to Liouville's Theorem, these areas are invariant provided that only non-dissipative forces act on the particles.

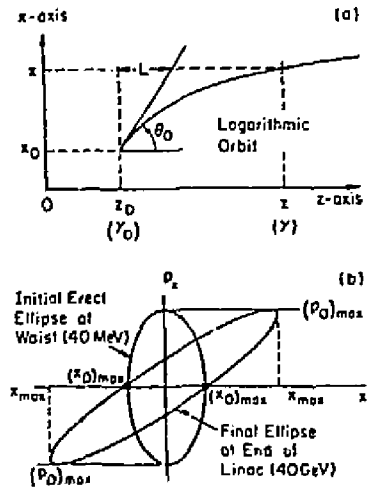


Fig. 37. (a) Logarithmic orbit due to initial deflection θ_0 at x_0 showing contracted length L; (b) Rotation of transverse phase-space ellipse from injector to accelerator output.

As an example, a recent transverse emittance measurement for a single bunch of 5×10^{10} electrons out of the proposed 40 MeV SLC injector gave $\sim 7.5 \times 10^{-3}$ π (MeV/c) cm. This means that if this beam is focused to a waist of 0.75 cm radius, its angular divergence is 0.25 mrad. If there were no transverse emittance growth along the accelerator, at say 40 GeV, the angular divergence would shrink to 0.25 μrad for the same 0.75 cm waist (see proof below). Even though the above emittance from the injector is quite small, the SLC requires that the emittance at the final focus be no greater than 1.5×10^{-3} π (MeV/c) cm or five times smaller. This explains why

damping rings to "cool" the beams downstream of the injector had to be added in the SLC design.

The transverse beam dynamics in the injector of a linear accelerator are complicated by the presence of space charge forces and radial defocusing forces due to waves and particles which are not yet moving at the velocity of light. We have seen earlier from Eq. (3.43) that as $\beta_z \to 1$, the magnetic and electric forces from a synchronous wave cancel each other. Thus, leaving out all other spurious forces, the transverse equation of motion of an electron in the x- (or y)-plane past the injector is given by

$$\frac{dp_x}{dt} = 0 \tag{5.20}$$

but since $dz = c\,dt$, we can rewrite this equation as

$$\frac{d}{dz}\left(\gamma \frac{dx}{dz}\right) = 0 \ . \tag{5.21}$$

Integrating, we obtain

$$\gamma \frac{dx}{dz} = \text{const} \tag{5.22}$$

which confirms the correctness of the assumption we made above.

Since under these conditions the quantity $\gamma\, dx/dz$ is conserved along the length of the linac, we can make it equal to $\gamma_0\, \theta_0$ at z_0, say at the output of the injector. Then a second integration leads to

$$\int_{x_0}^{x} dx = \gamma_0 \theta_0 \int_{z_0}^{z} \frac{dz}{\gamma} \tag{5.23}$$

and if we assume that the energy grows linearly as $\gamma = \gamma' z$ where γ' is constant, then

$$x - x_0 = \frac{\gamma_0}{\gamma'}\, \theta_0\, \ell n \frac{z}{z_0} \tag{5.24}$$

or

$$x - x_0 = \theta_0\, z_0\, \ell n\, \frac{z}{z_0} \ . \tag{5.25}$$

Looking at Fig. 37a, we see that an electron starting at z_0 with energy γ_0 (where actually z_0 is defined as γ_0/γ'), $x = x_0$ and angle θ_0, will end up at z with transverse

displacement $x_0 + \theta_0 z_0 \ell n\, z/z_0$. It is as if the length $z - z_0$ has been contracted to $L = z_0 \ell n\, z/z_0$. Thus if $z_0 = 3$ m, $z = 3000$ m, $\gamma' = 26.66$/m, the 2997 m length looks like it has been contracted to 20.7 m. The orbits instead of being linear are logarithmic.

It is instructive to follow the evolution of the phase-space ellipse from its waist at the output of the injector (40 MeV) to its final configuration at the output of the accelerator (40 GeV). Since p_x cannot grow beyond $(p_0)_{max}$ and the two $(x_0)_{max}$ points represent rays which are parallel to the z-axis and therefore cannot be deflected by an increase in energy, the ellipse has no choice but to get elongated between the two $(p_0)_{max}$ horizontal lines while remaining fixed at the two $(x_0)_{max}$ points. It can be shown that as the ellipse gets elongated, the $(x)_{max}$ coordinate is given by

$$x_{max}^2 = (x_0)_{max}^2 + \left(\theta_0 z_0 \ell n\, \frac{z}{z_0}\right)^2 \tag{5.26}$$

where $\theta_0 = (p_0)_{max}/p_{z0}$. In the example we chose, $x_0 = 0.75$ cm, $\theta_0 = 0.25$ mrad, $z_0 \ell n\, z/z_0 = 20.7$ m, we find that $x_{max} = 0.91$ cm. We see that the radius of the beam, in the absence of any transverse emittance growth, would barely need any focusing to stay within the aperture of the accelerator irises. In actual fact, however, emittance growth does take place because of the effects discussed earlier, and a FODO array or other quadrupole focusing system is indeed required. The description of these systems is beyond the scope of these lectures.

Bunch length. In sub-section 5.1, we have described qualitatively how the bunches are formed in the injector of a linear accelerator. Following the electrons from the gun through a prebuncher and a buncher is a complicated problem which is best done by a computer. However, when the particles finally enter the $v_p = c$ constant-gradient accelerator section, there is a fairly simple expression for the asymptotic phase ϕ_∞ of an electron of velocity $v_0 = \beta_0 c$ entering with phase ϕ_0 relative to the wave:

$$\cos\phi_\infty = \cos\phi_0 - \frac{2\pi}{\alpha}\left(\frac{1-\beta_0}{1+\beta_0}\right)^{1/2} \tag{5.27}$$

Unlike for protons, the phase origin is taken at the field null, 90° ahead of crest (Fig. 38) and $\alpha = e\,E\lambda_0/m_0c^2$.* When the factor $2\pi/\alpha\,[(1-\beta_0)/(1+\beta_0)]^{1/2}$ is made equal to 1 through the proper choice of α ($\alpha = 3.63$ if $\beta_0 = 1/2$, or $\alpha = 2.37$ if $\beta_0 = 3/4$), then an electron, entering the accelerator at $\phi_0 = 0$, asymptotically approaches the crest, $\phi_\infty = -\pi/2$. For small phase extents $\pm\pi_0$ around $\phi_0 = 0$, it can be shown that

$$\phi_\infty \approx -\frac{\pi}{2} - \frac{\phi_0^2}{2} \quad ,$$

i.e., all other electrons end up behind the crest. For example, if $-1/4 \leq \phi_0 \leq 1/4$ radian, then $-[\pi/2 + 1/32] \leq \phi_\infty \leq -\pi/2$ radian, i.e., a 30° bunch ends up in less than a 2° bunch. Typically, at SLAC, bunches over a wide range of charge have been measured to contain most of their charge within 5°. The asymptotic charge distribution is not quite Gaussian, even if that of the entering bunch is (which is generally not the case either). For the SLC, it is planned to have most of the 5×10^{10} electrons within 20°.

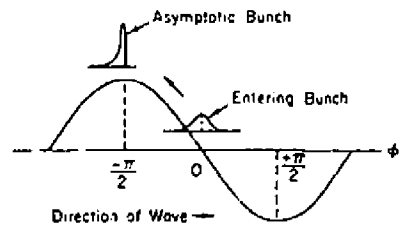


Fig. 38. Asymptotic bunching process in $v_p = c$ constant-gradient accelerator section with value of α optimized for entrance conditions.

Energy spectrum due to bunch length and beam loading. The last beam property of great interest to linac builders and users which we will consider is the energy spectrum, namely, how monochromatic is the beam? Typically, one wants to have all the electrons within an energy spectrum width of 1% or less. There are two effects inherent to how an electron linac works which bear on the spectrum. The first one is the bunch length which by its very nature implies that all electrons within it do not receive the same energy, either because they are not on the same part of the wave or because of their

*Not to be confused with the attenuation/unit length α defined earlier!

own wakefields. The second one is beam loading from bunch to bunch due to energy removal from the structure by each bunch: this effect can reach steady-state after a certain number of bunches. In addition to these two inherent effects, there are a number of other technological effects which can be detrimental to the spectrum width such as poor injection conditions, phase and frequency modulation in the master oscillator, poor amplitude and phase stability of the rf sources, poor accelerator temperature control and many others. We will not examine these.

In a multi-section linac with individually phase controllable klystrons, it is possible to phase each tube so that a bunch of finite length $\Delta\phi$ is centered on the crest. Then it is easy to show that

$$\frac{\delta\,\Delta E_K}{\Delta E_K} = 1 - \cos\frac{\Delta\phi}{2} \approx \frac{(\Delta\phi)^2}{8} \tag{5.28}$$

or $\sim 0.12\%$ for $\Delta\phi = 1/10$ radian. The reader can convince himself that this is the best that can be done for a low current bunch in which wakefields do not play a role. Indeed, this effect results from the shape of the cosine wave and nothing can be done about it, not even rocking the bunch from one side of the crest to the other, which is sometimes suggested! This would only work if the top of the wave were a triangle instead of a curve. On the other hand, if the bunch is highly charged as in the SLC, the energy decrease due to the wakefield effect of the head on the tail can be compensated for by locating the bunch ahead of crest at the proper phase, which to first order can cancel the decrease if the slopes can be matched.

Coming back to the low current case, it is often asked what the effect of misphasing individual klystrons is on the beam. The answer is that to first order, misphasing individual klystrons only decreases the maximum reachable energy but does not affect the spectrum. Indeed, it can easily be shown that the $(\Delta\phi)^2/8$ rule still applies as long as on average, the bunch travels centered on crest.

The beam loading effect, which has been referred to several times earlier, comes about because the rf energy stored in the linac structure is reduced as successive bunches get accelerated and extract energy from it. The effect goes through a transient phase until a new lower equilibrium energy level is attained. In a traveling-wave structure, it can be seen intuitively that each bunch that traverses it sets up instantaneous rf field "packets" in each cavity at the velocity of light (except in the buncher).

These packets then travel down the structure in phase at the group velocity. If the train of bunches continues, there is a build-up of fields versus z, even though the packets get attenuated with distance. Steady-state is reached when the packet generated in the first cavity reaches the last. As it happens, this is exactly the definition of the filling time t_F. Thus if we inject a train of bunches after one filling time, i.e., when the section if filled, it takes an additional filling time before steady-state is reached.

We will give a short derivation of the steady-state case and let the transient case be understood intuitively. If a peak current of intensity i is assumed, our power loss equation (5.4) must be written as

$$\frac{dP}{dz} = \left(\frac{dP}{dz}\right)_{wall} - (Ei)_{beam} \quad . \tag{5.29}$$

assuming that the electrons are synchronous and ride on the crest. Then, differentiating the expression

$$E^2 = 2\alpha r P \tag{5.30}$$

with respect to z, we get

$$E\,\frac{dE}{dz} = r P\,\frac{d\alpha}{dz} + \alpha r\,\frac{dP}{dz} \quad . \tag{5.31}$$

Substituting (5.29) for dP/dz in (5.31), $-E^2/r$ for $(dP/dz)_{wall}$ and $E^2/2\alpha r$ for P, we obtain the general differential equation

$$\frac{dE}{dz} = -\alpha E\left(1 - \frac{1}{2\alpha^2}\,\frac{d\alpha}{dz}\right) - \alpha r i \quad . \tag{5.32}$$

It follows that for the constant-impedance case

$$\frac{dE}{dz} = -\alpha E - \alpha r i \tag{5.33}$$

and for the constant-gradient case, replacing $d\alpha/dz$ by its value given by (5.14),

$$\frac{dE}{dz} = -\alpha r i \quad . \tag{5.34}$$

The solutions are:

Constant-Impedance

$$E = E_0 e^{-\alpha z} - ri(1 - e^{-\alpha z}) \tag{5.35}$$

where $E_0 = (2\alpha r P_0)^{1/2}$,

Constant-Gradient

$$E = E_0 + \frac{ir}{2}\,\ell n\left[1 - \frac{z}{\ell}\,(1 - e^{-2\tau})\right] \tag{5.36}$$

where

$$E_0 = (2\alpha_0 r P_0)^{1/2} = \left[\frac{r P_0}{\ell}\,(1 - e^{-2\tau})\right]^{1/2} . \tag{5.37}$$

The energy equations are then obtained by simple integration, as in (5.7):

Constant-Impedance

$$\Delta E_K = e(P_0 \ell r)^{1/2}\,(2\tau)^{1/2}\,\frac{1 - e^{-\tau}}{\tau} - eir\ell\left(1 - \frac{1 - e^{-\tau}}{\tau}\right) \tag{5.38}$$

Constant-Gradient

$$\Delta E_K = e(P_0 \ell r)^{1/2}\,(1 - e^{-2\tau})^{1/2} - \frac{e\,i\,r\,\ell}{2}\left(1 - \frac{2\tau e^{-2\tau}}{1 - e^{-2\tau}}\right) . \tag{5.39}$$

We can see that when $i = 0$, expressions (5.38) and (5.39) collapse back to (5.9) and (5.17) respectively. The reader can verify that for the SLAC conditions used for (5.18) and $i = 50$ mA, $\tau = 0.57$, the energy reduction per section due to beam loading under steady-state conditions is 2 MeV out of 30 MeV or 6.7%. The steady-state field profiles along a section given by (5.35) and (5.36) are sketched in Fig. 39a. The shaded areas show the reduction due to beam loading. The energies are the integrals under the respective curves. Figure 39b gives the evolution of energy as a function of time. The interval between 0 and t_F shows the energy increase as the section is filling, the second interval between t_F and $2t_F$ shows the transient decrease due to injection of the beam at $t = t_F$, and the third interval ($t > t_F$) shows the steady-state energy.

The beam loading effect, as can be seen, is inherent to the operation of the linac. Once steady-state is reached, it has no effect on the spectrum width, unless the current

itself fluctuates. If the beam pulse length itself is only on the order of $2t_F$, which is the case at SLAC, half the beam pulse is in the transient state. The remedy which is often used to reduce this overall spectrum broadening effect is to inject the beam somewhat earlier than at time t_F so that the first bunches see their energy reduced as well. With a large enough number of klystrons, it is possible to stagger the onset of the rf in successive sections so that the two effects cancel each other down to less than 1% (Fig. 39c). Thus, the steady-state reduction cannot be avoided but at least its effect on the spectrum can be minimized.

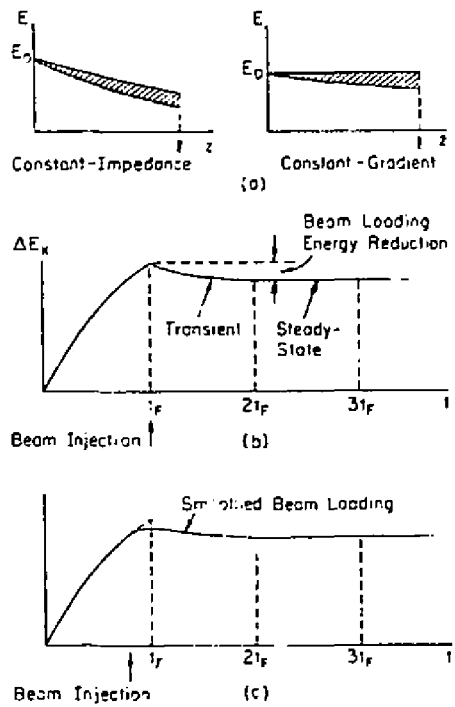


Fig. 39. Effect of beam loading: (a) on E-field profiles in sections, (b) on energy ΔE_K delivered versus time, (c) on energy ΔE_K if beam is injected before $t = t_F$.

In standing-wave accelerators, beam loading is of course also present. However, since the beam-induced waves are caused to bounce back and forth just like the externally supplied waves, the reduced field profile takes several filling times to reach steady-state.

5.4 Recent Developments and Future Challenges

During the past few years, progress in the design of linear electron accelerators has been made on a broad front. Industrial manufacturers have specialized in the production of a large number of medical and radiographic linacs. Their contribution has been strong in the areas high rf-to-beam energy conversion efficiency, high shunt impedance standing-wave structures, minimum external focusing, good packaging, and in some specialized machines excellent emittance characteristics through the design of very high voltage electron guns (up to $\sim$ 400 kV). In some cases, linacs have been built to produce beam pulses in the 10-20 nsec range with currents up to 20 A. In the universities, the accent at laboratories such as MIT, ALS (Saclay) and NIKHEF (Amsterdam) has been on linacs with duty cycles up to 10%. The challenge there has been to obtain reasonable currents ($\sim$ 20 mA) for long pulses (10-40 μsec) with excellent emittances and without the onset of beam breakup.

At laboratories in the U.S. such as HEPL and SLAC (Stanford), University of Illinois and Cornell, in Germany at Karlsruhe and Wuppertal, in Switzerland at CERN, and in Japan at KEK, considerable effort has gone into investigating superconducting structures for linacs, microtrons and storage rings. The hope to build CW linacs with accelerating gradients up to the theoretical limit of 30 MV/m in niobium has not been realized so far. On the other hand, much progress has been made towards understanding and overcoming some of the major obstacles in the field such as the onset of multipactoring and thermo-magnetic breakdown due to impurities in the niobium. Thus, with a few tens of watts of power, several laboratories have been able to obtain gradients of 2-4 MV/m fairly routinely under CW conditions. The entire subject of rf superconductivity in its application to accelerators is dealt by M. Tigner, elsewhere in this book.

Other institutions have built or used a variety of dedicated low duty-cycle linacs for injection of electrons (and positrons) into storage rings. This category includes several laboratories such as Cornell, Frascati, Orsay, DESY, Novosibirsk, the Photon

Factory at KEK. and in the future the LEP injector at CERN and two injector linacs being planned in the P.R.C. (Beijing and Hefei). A large 4 GeV linac, with 200 mA peak current and high repetition rate (1000 pps) is presently being designed as an injector to a stretcher ring for the National Electron Accelerator Laboratory (NEAL) at Newport News (Virginia).

Another area where innovation has taken place is in the field of rf energy compression. The attempt here has been to gain electron energy without increasing average or even klystron peak power. The basic scheme, invented at SLAC by P. Wilson, D. Farkas and H. Hogg, has been called SLED, an acronym for SLAC Energy Development. (For a general reference, see for example, Z. D. Farkas, H. A. Hogg, G. A. Loew and P. E. Wilson, Proceedings of the IXth International Conference on High Energy Accelerators, SLAC (1974), page 576). The principle is illustrated in Figs. 40 and 41.

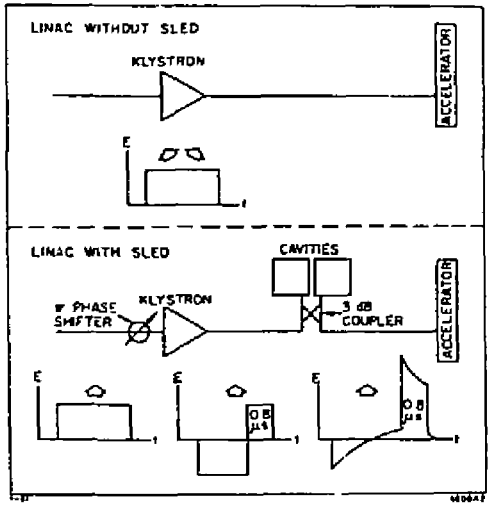


Fig. 40. The SLED principle.

Referring to the top of Fig. 40, we see that under normal non-SLED operation, the rf drive pulse which is amplified by the klystron is directly transmitted to the linac. The SLED system, which is shown at the bottom of Fig. 40, has two major components: a 180° fast phase shifter on the drive side of the klystron and two high-Q ($Q_0 = 100,000$) cavities on the output side of the tube with a 3 db coupler connected as

shown. During the first part of the pulse, the phase of the rf drive signal is reversed and the rf cavities fill up with energy with that phase. Because of the 3 db coupler, no energy is reflected to the klystron. The fields emitted by the cavities add algebraically with the fields reflected by the cavity coupling irises, and the power flows toward the accelerator. Exactly one linac filling time (~ 0.8 μsec) before the end of the pulse, the phase of the drive signal is flipped back by 180°: as a result both cavities discharge their energy into the accelerator. The discharging pulse amplitude is increased by the fact that the klystron pulse adds to it. The output pulse length is tailored exactly to fill the accelerator sections at one instant of time.

SLED OPTIONS

	I	II
PULSE LENGTH	2.5μs	5μs
EFFECTIVE POWER GAIN	2	3.15
ENERGY GAIN	1.4	1.78
REPETITION RATE	360 pps	180 pps

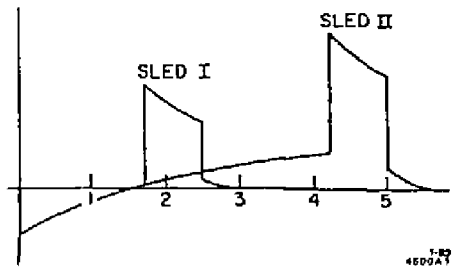


Fig. 41. Comparison of two rf pulse length options for SLED showing how a longer charging time increases the output pulse amplitude.

It can be seen intuitively that the longer the time to charge the cavities, the more energy there is available at the time of the discharge. On the other hand, there is a limit to this process because of the copper losses in the cavities. Figure 41 shows two rf pulse length options in what is obviously a continuum. Much beyond 5 μsec, one reaches a point of diminishing returns. The two cases of 2.5 and 5 μsec with their respective energy gains and pulse shapes are given because the first one equals the pulse length that was initially available on the SLAC accelerator and the second, in

conjunction with a 50 MW klystron, would satisfy the 50 GeV energy requirements of the SLC. The two repetition rates in combination with the corresponding pulse lengths require the same AC power.

Note that while the SLED system is indeed a form of energy compression, it is not perfect because the cavities start out by rejecting the energy, then absorb some of it themselves, and finally do not emit nice rectangular pulses. Thus, for example, the power gain in the 5 μsec case, instead of being 5 μsec/0.8 μsec, i.e., 6.25, is only 3.15. The idea in itself, however, has been applied with considerable success and other laboratories such as CERN and the High Energy Institute at Beijing a.. adopting it for their linac designs.

In the preceding pages, we have referred a number of times to the SLC or Linear Collider project at SLAC. Of all the known applications of electron and positron linacs, this is probably the most challenging because it puts to a test and stretches to a maximum the capabilities of these accelerators. For the SLC to become operational, we must improve our understanding and control of beam generation, injection, emittance, beam centering and focusing to keep the emittance from growing, energy spectrum, phase and amplitude stability of the klystrons, positron generation and many others. These problems are discussed elsewhere in this book by R. Stiening.

The SLC will be the first test-bed for these ideas. Beyond this first 50 GeV-on-50 GeV $e^{\pm}$ linear collider, it is now being asked whether machines of this type could be built in the 1-2 TeV range. This question will of course be much easier to answer after some experience has been obtained with the SLC. However, one problem can already be focused on now. This is the problem of energy consumption. We will end these lectures with a simple example which might stimulate the reader to think further.

It is not clear at this point where the ultimate limits of accelerating gradient will be nor what factors will determine the length of the accelerator. Suppose we want to build a 1 TeV-on-1 TeV $e^{\pm}$ collider. Say we have two choices of gradient: 20 MV/m (conservative, attainable now) or 100 MV/m (less conservative, yet to be tested). The respective lengths would be 50 km and 10 km/linac. Returning once more to Eq. (5.2) and rewriting it as

$$w_{st} = \frac{E^2}{\omega r/Q} \ , \tag{5.40}$$

we see that even if we are able to bring the energy onto the axis of the accelerator at 100% efficiency, it will take a large number of joules to establish the fields. Table 5 summarizes the results, assuming that $r/Q = 5000\Omega/m$ at 2856 MHz. The power requirements are worked out for an example where the two linacs are pulsed 200 times per second to reach the overall desired luminosity. Each pulse is used to accelerate n bunches ($4 \leq n \leq 10$) which will at most remove 20% of the stored energy. Unless something very ingenious is done, the remaining 80% energy will be wasted. Note further that the real power consumption will be at least twice the rf power consumption because of all the inefficiencies involved in converting AC to rf. In a few years, at \$0.10/kW-h, 1/2 GW of power will cost \$50,000/hr. It is quickly seen that there will be a strong incentive to improve upon the numbers in this table. Two avenues suggest themselves. One is to increase the frequency since $\omega r/Q$ scales as ω^2. The other is to invent a method of recovering the energy for some useful purpose. The challenge is yours!

Table 5.
rf Energy and Power Needs for a 1 TeV-on-1 TeV $e^{\pm}$ Linear Collider Assuming No Losses in Establishing Fields

Gradient (MV/m)	20	100
Length/linac (km)	50	10
rf Frequency (MHz)	2856	2856
w_{st} (Joules/m)	4.47	111.73
Total energy stored/linac/pulse (MJ)	0.223	1.117
rf Power/2 linacs/200 pps (MW)	89.36	446.8

Acknowledgements

The authors wish to thank G. R. Lambertson, R. H. Miller, R. B. Neal and Wang Juwen for useful discussions.