The Evolution of Particle Accelerators & Colliders

by WOLFGANG K. H. PANOFSKY

WHEN J. J. THOMSON discovered the electron, he did not call the instrument he was using an accelerator, but an accelerator it certainly was. He accelerated particles between two electrodes to which he had applied a difference in electric potential. He manipulated the resulting beam with electric and magnetic fields to determine the charge-to-mass ratio of cathode rays. Thomson achieved his discovery by studying the properties of the beam itself—not its impact on a target or another beam, as we do today. Accelerators have since become indispensable in the quest to understand Nature at smaller and smaller scales. And although they are much bigger and far more complex, they still operate on much the same physical principles as Thomson's device.

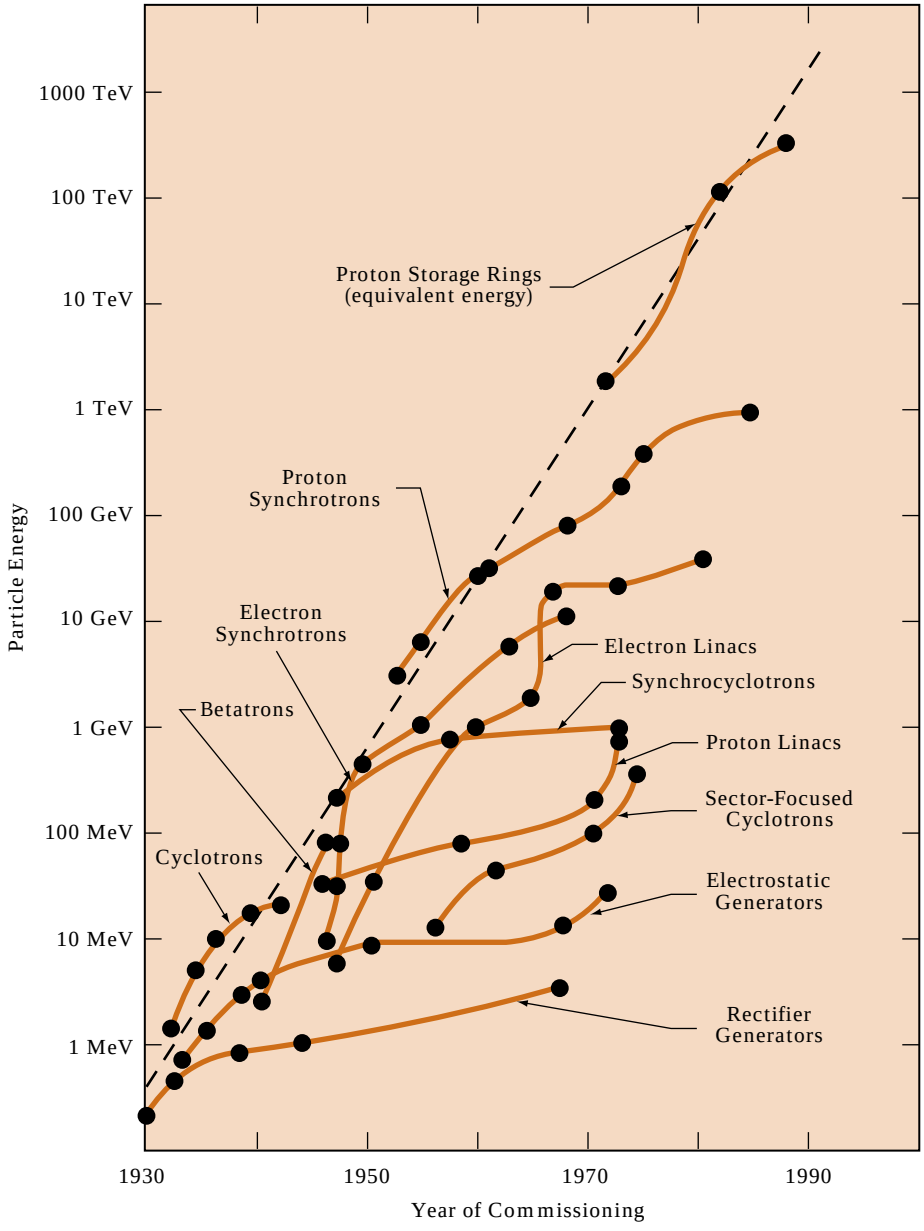
It took another half century, however, before accelerators became entrenched as the key tools in the search for subatomic particles. Before that, experiments were largely based on natural radioactive sources and cosmic rays. Ernest Rutherford and his colleagues established the existence of the atomic nucleus—as well as of protons and neutrons—using radioactive sources. The positron, muon, charged pions and kaons were discovered in cosmic rays.

One might argue that the second subatomic particle discovered at an accelerator was the neutral pion, but even here the story is more complex. That it existed had already been surmised from the existence of charged pions, and the occurrence of gamma rays in cosmic rays gave preliminary evidence for such a particle. But it was an accelerator-based experiment that truly nailed down the existence of this elusive object.

There followed almost two decades of accelerator-based discoveries of other subatomic particles originally thought to be elementary, notably the antiproton and the vector mesons. Most of these particles have since turned out to be composites of quarks. After 1970 colliders—machines using two accelerator beams in collision—entered the picture. Since then most, but certainly not all, new revelations in particle physics have come from these colliders.

IN CONSIDERING the evolution of accelerator and collider technology, we usually think first of the available energy such tools provide. Fundamentally, this is the way it should be. When the study of the atomic nucleus stood at the forefront of "particle physics" research, sufficient energy was needed to allow two nuclei—which are positively charged and therefore repel one another—to be brought close enough to interact. Today, when the components of these nuclei are the main objects of study, the reasons for high energy are more subtle. Under the laws of quantum mechanics, particles can be described both by their physical trajectory as well as through an associated wave whose behavior gives the probability that a particle can be localized at a given point in space and time. If the wavelength of a probing particle is short, matter can be examined at extremely small distances; if long, then the scale of things that can be investigated will be coarser. Quantum mechanics relates this wavelength to the energy (or, more precisely, the momentum) of the colliding particles: the greater the energy, the shorter the wavelength.

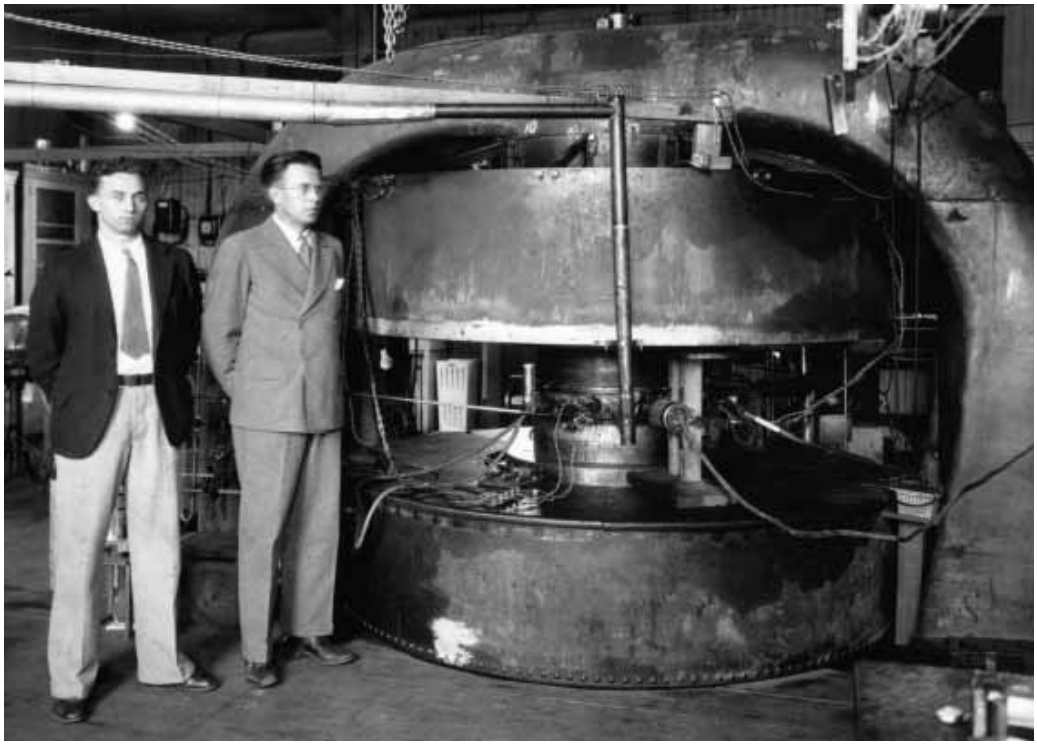
A "Livingston plot" showing accelerator energy versus time, updated to include machines that came on line during the 1980s. The filled circles indicate new or upgraded accelerators of each type.



This relationship can be expressed quantitatively. To examine matter at the scale of an atom (about 10^{-8} centimeter), the energies required are in the range of a thousand electron volts. (An electron volt is the energy unit customarily used by particle physicists; it is the energy a particle acquires when it is accelerated across a potential difference of one volt.) At the scale of the nucleus, energies in the million electron volt—or MeV—range are needed. To examine the fine structure of the basic constituents of matter requires energies generally exceeding a billion electron volts, or 1 GeV.

But there is another reason for using high energy. Most of the objects of interest to the elementary particle physicist today do not exist as free particles in Nature; they have to be created artificially in the laboratory. The famous $E = mc^2$ relationship governs the collision energy E required to produce a particle of mass m. Many of the most interesting particles are so heavy that collision energies of many GeV are needed to create them. In fact, the key to understanding the origins of many parameters, including the masses of the known particles, required to make today's theories consistent is believed to reside in the attainment of collision energies in the trillion electron volt, or TeV, range.

Our progress in attaining ever higher collision energy has indeed been impressive. The graph on the left, originally produced by M. Stanley Livingston in 1954, shows how the laboratory energy of the particle beams produced by accelerators has increased. This plot has been updated by adding modern developments. One of the first things to notice is that the energy of man-made accelerators has been growing exponentially in time. Starting from the 1930s, the energy has increased—roughly speaking—by about a factor of 10 every six to eight years. A second conclusion is that this spectacular achievement has resulted



M. Stanley Livingston and Ernest O. Lawrence, with their 27-inch cyclotron at Berkeley Radiation Laboratory. (Courtesy Lawrence Berkeley National Laboratory)

from a succession of technologies rather than from construction of bigger and better machines of a given type. When any one technology ran out of steam, a successor technology usually took over.

In another respect, however, the Livingston plot is misleading. It suggests that energy is the primary, if not the only, parameter that defines the discovery potential of an accelerator or collider. Energy is indeed required if physicists wish to cross a new threshold of discovery, provided that this threshold is defined by the energy needed to induce a new phenomenon. But there are several other parameters that are important for an accelerator to achieve—for example, the intensity of the beam, or the number of particles accelerated per second.

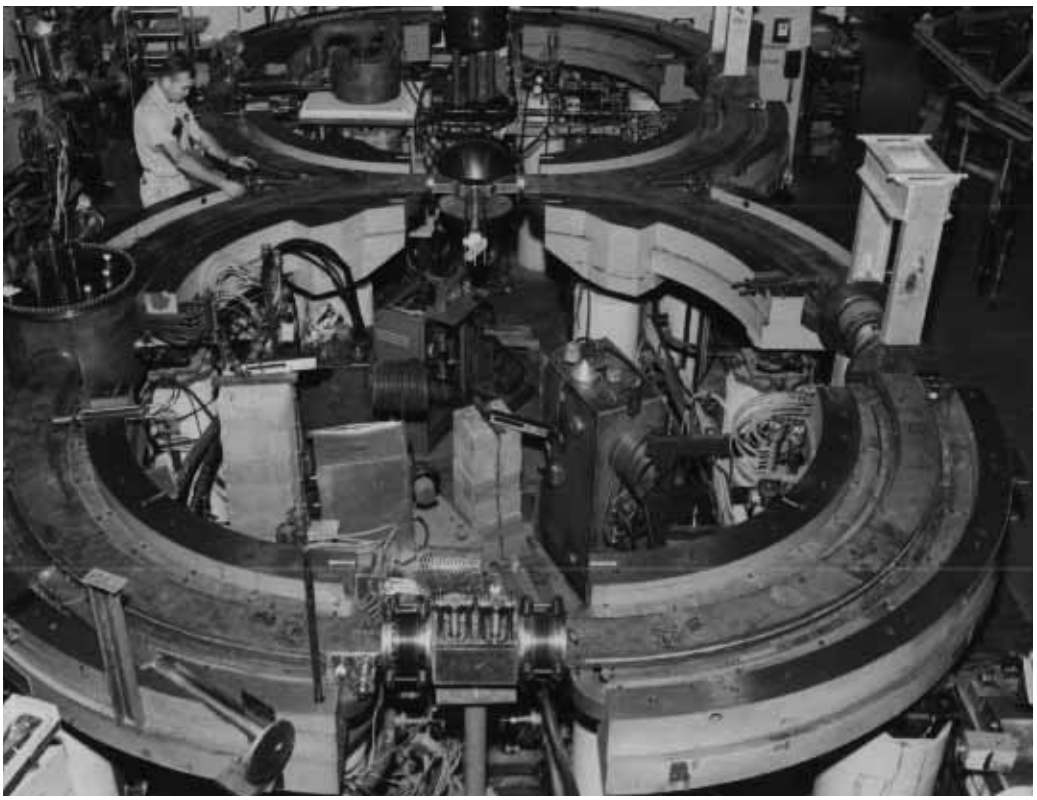
When the beam strikes a target, its particles collide with those in the target. The likelihood of producing a reaction is described by a number called the cross section, which is the effective area a target particle presents to an incident particle for that reaction to occur. The overall interaction rate is then the product of the beam intensity, the density of target particles, the cross section of the reaction under investigation, and the length of target material the incident particle penetrates. This rate, and therefore the beam intensity, is extremely important if physicists are to collect data that have sufficient statistical accuracy to draw meaningful conclusions.

Another important parameter is what we call the duty cycle—the percentage of time the beam is actually on. Unlike Thomson's device, most modern accelerators do not provide a steady flow of particles, generally because that would require too much electric power; instead, the beam is pulsed on and off. When physicists try to identify what reaction has taken place, one piece of evidence is whether the different particles emerge from a collision at the same time. Thus electronic circuits register the instant when a particle traverses a detector. But if the accelerator's duty cycle is small, then all the particles will burst forth during a short time interval. Therefore a relatively large number of accidental coincidences in time will occur, caused by particles emerging from different individual reactions, instead of from real coincidences due to particles emerging from a single event. If time coincidence is an important signature, a short duty cycle is a disadvantage.

Then there is the problem of backgrounds. In addition to the reaction under study, detectors will register two kinds of undesirable events. Some backgrounds arise from particles generated by processes other than the beam's interaction with the target or another beam—such as with residual gas, from "halo" particles traveling along the main beam, or even from cosmic rays. Other backgrounds stem from reactions that are already well understood and contain no new information. Accelerators differ in terms of the presence or absence of both kinds of backgrounds; their discovery potential differs accordingly. The amount and kinds of background are directly related to the ease of data analysis, the type of detector to be built, or whether the desired results can be extracted at all.

In general, as the energy increases, the number of possible reactions also increases. So does the burden on the discriminating power of detectors and on the data-analysis potential of computers that can isolate the "wheat" from the "chaff." With the growth in energy indicated by the Livingston plot, there had to be a parallel growth in the analyzing potential of the equipment required to identify events of interest—as well as a growth in the number of people involved in its construction and operation.

And finally there is the matter of economy. Even if a planned accelerator is technically capable of providing the needed energy, intensity, duty cycle, and low background, it still must be affordable and operable. The resources required—money, land, electric power—must be sufficiently moderate that the expected results will have



The first colliding-beam machine, a double-ring electron-electron collider, built by a small group of Princeton and Stanford physicists. (Courtesy Stanford University)

commensurate value. Of course "value" has to be broadly interpreted in terms not only of foreseeable or conjectured economic benefits but also of cultural values related to the increase in basic understanding. In view of all these considerations, the choice of the next logical step in accelerator construction is always a complex and frequently a controversial issue. Energy is but one of many parameters to be considered, and the value of the project has to be sufficiently great before a decision to go ahead can be acceptable to the community at large.

All these comments may appear fairly obvious, but they are frequently forgotten. Inventions that advance just one of the parameters—in particular, energy—are often proposed sincerely. But unless the other parameters can be improved at the same time, to generate an overall efficient complex, increasing the energy alone usually cannot lead to fundamentally new insights.

THE ENERGY that really matters in doing elementary particle physics is the collision energy—that is, the energy available to induce a reaction, including the creation of new particles. This collision energy is *less* than the laboratory energy of the particles in a beam if that beam strikes a stationary target. When one particle hits another at rest, part of the available energy must go toward the kinetic energy of the system remaining after the collision. If a proton of low energy E strikes another proton at rest, for example, the collision energy is $E/2$ and the remaining $E/2$ is the kinetic energy with which the protons move ahead. At very high energies the situation is complicated by relativity. If a particle of total energy E hits another particle of mass M, then the collision energy is given by $E_{coll} \sim (2Mc^2E)^{1/2}$, which is much less than $E/2$ for E much larger than Mc^2.

If two particles of equal mass traveling in opposite directions collide head on, however, the total kinetic energy of the combined system after collision is zero, and therefore the entire energy of the two particles becomes available as collision energy. This is the basic energy advantage offered by colliding-beam machines, or colliders.

The idea of colliding-beam machines is very old. The earliest reference to their possibility stems from a Russian publication of the 1920s; it would not be surprising if the same idea occurred independently to many people. The first collider actually used for particle-physics experiments, built at Stanford in the late 1950s, produced electron-electron collisions (see photograph on the left). Other early machines, generating electron-positron collisions, were built in Italy, Siberia and France. Since then there has been a plethora of electron-positron, proton-proton and proton-antiproton colliders.

There is another problem, however. If the particles participating in a collision are themselves composite—that is, composed of constituents—then the available energy must be shared among these constituents. The threshold for new phenomena is generally defined by the collision energy in the constituent frame: the energy that becomes available in the interaction between two individual constituents. Here there are major differences that depend on whether the accelerated particles are protons, deuterons, electrons or something else. Protons are composed of three quarks and surrounded by various gluons. Electrons and muons, as well as quarks and gluons, are considered pointlike, at least down to distances of 10^{-16} centimeter. Recognizing these differences, we can translate the Livingston plot into another chart (top right, next page) showing energy in the constituent frame versus year of operation for colliding-beam machines.

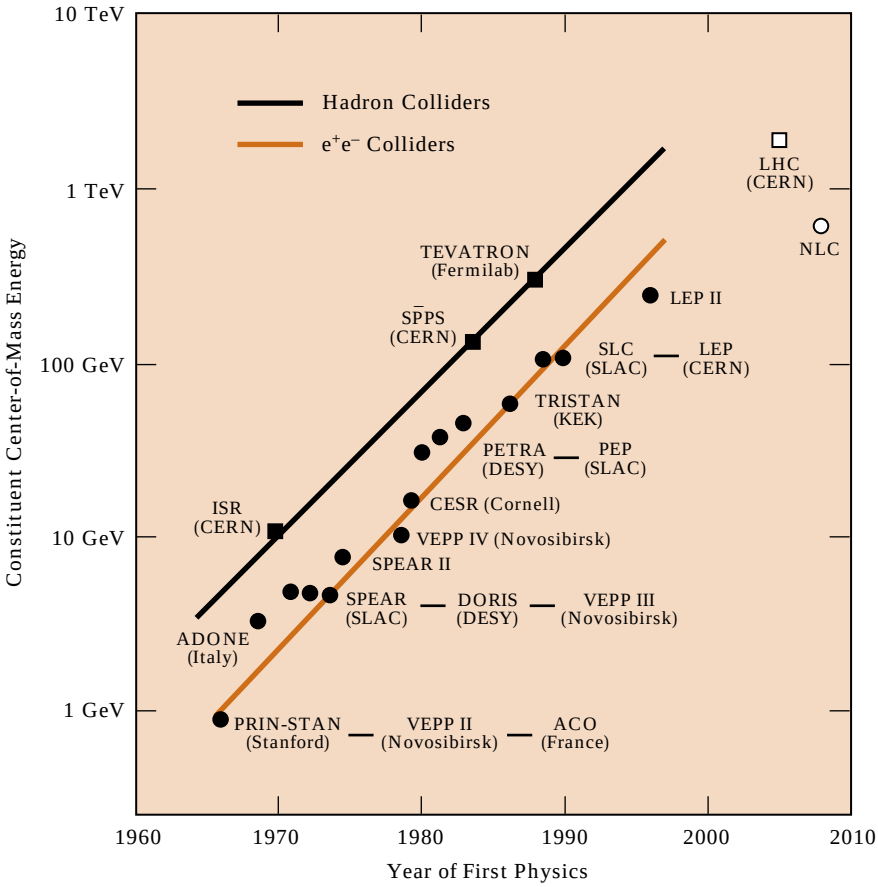
But the idea of generating higher collision energy via colliding beams

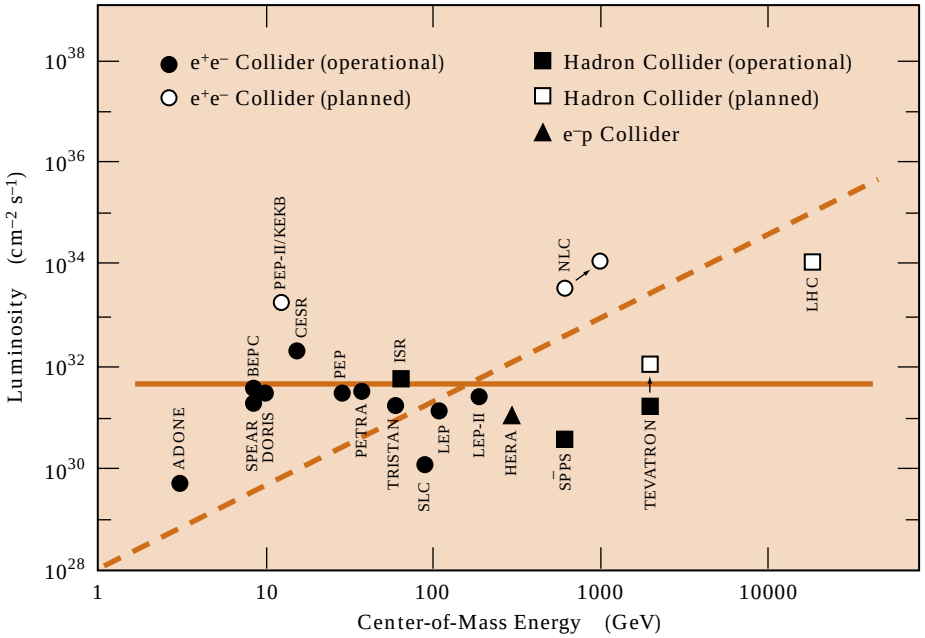
Right: The energy in the constituent frame of electron-positron and hadron colliders constructed (filled circles and squares) or planned. The energy of hadron colliders has here been derated by factors of 6–10 in accordance with the fact that the incident proton energy is shared among its quark and gluon constituents.

is worthless unless (as discussed above) higher interaction rates can be generated, too. To succeed, the density of the two beams must be high enough—approaching that of atoms in ordinary matter—and their interaction cross sections must be sufficient to generate an adequate data rate. In colliding-beam machines the critical figure is the luminosity L, which is the interaction rate per second per unit cross section. The bottom graph on this page illustrates the luminosity of some of these machines. In contrast to the constituent collision energy, which has continued the tradition of exponential growth begun in the Livingston plot, the luminosity has grown much more slowly. There are good reasons for this trend that I will discuss shortly.

Naturally there are differences that must be evaluated when choosing which particles to use in accelerators and colliders. In addition to the energy advantage mentioned for electrons, there are other factors. As protons experience the strong interaction, their use is desirable, at least in respect to hadron-hadron interactions. Moreover, the cross sections involved in hadron interactions are generally much larger than those encountered in electron machines, which therefore require higher luminosity to be equally productive.

Proton accelerators are generally much more efficient than electron machines when used to produce secondary beams of neutrons, pions, kaons, muons, and neutrinos. But electrons produce secondary beams that are sharply concentrated in the forward direction, and these beams are less contaminated by neutrons.





Peak luminosities achieved at existing colliders and values projected for planned or upgraded machines. The dashed line indicates luminosity increasing as the square of the center-of-mass energy Note that the rated machine energy has been used in calculating the abscissa. (Data updated courtesy Greg Loew, SLAC)

Far left: William Hansen (right) and colleagues with a section of his first linear electron accelerator, which operated at Stanford University in 1947. Eventually 3.6 m long, it could accelerate electrons to an energy of 6 MeV. (Courtesy Stanford University)

Left: Ernest Lawrence's first successful cyclotron, built in 1930. It was 13 cm in diameter and accelerated protons to 80 keV. (Courtesy Lawrence Berkeley National Laboratory)

When discussing the relative merits of electron and proton colliders, the background situation is complex because the factors that cause them are quite different. When accelerated, and especially when their path is bent by magnets, electrons radiate X rays in the form of synchrotron radiation. Protons usually have more serious interactions with residual gas atoms, and those that deviate from the nominal collider orbit are more apt to produce unwanted backgrounds from such causes.

A much more difficult—and to some extent controversial—subject is the comparison of the complexities of events initiated by electrons with those induced by hadrons in general, and protons in particular. Today particle physicists are usually, but not always, interested in the results of "hard" collisions between the elementary constituents (by which I mean entities considered to be pointlike at the smallest observable distances). Because protons are composite objects, a single hard collision between their basic constituents will be accompanied by a much larger number of extraneous "soft" collisions than is the case for electrons. Thus the fraction of interesting events produced in an electron machine is generally much larger than it is for proton machines. So the analysis load in isolating the "needle" from the "haystack" tends to be considerably more severe at hadron machines.

Beyond this problem is the matter of interpretability. When we use electrons to bombard hadron targets, be they stationary or contained in an opposing beam, we are exploring a complex structure with an (as-yet) elementary object whose behavior is well understood. Thus the information about the structure of the proton resulting from electron-proton collisions, for example, tends to be easier to interpret than the results from proton-proton collisions. All the above observations are generalities, of course, and there are numerous and important exceptions. For instance, if neutrinos or muons—copiously produced as secondary beams from proton machines—are used to explore the structure of hadrons, the results are complementary to those produced by electron beams.

Everything I have said about electrons is also true of muons. The use of muon beams offers significant advantages and disadvantages relative to electrons. The two lightest charged leptons, the electron and muon, experience essentially the same interactions. But muons, being heavier, radiate far less electromagnetic energy than do electrons of equal energy; therefore backgrounds from radiative effects are much lower. On the other hand, muons have a short lifetime (about 2 microseconds), whereas electrons are stable. Colliding-beam devices using muons must be designed to be compatible with this fact. In addition, the remnants of the muons that decay during acceleration and storage constitute a severe background. Thus, while the idea of muon colliders as tools for particle physics has recently looked promising, there is no example as yet of a successful muon collider.

BUT THERE is an overarching issue of costs that dominates the answer to the question, "How large can accelerators and colliders become, and what energy can they attain?" The relationship of size and cost to energy is determined by a set of relations known as scaling laws. Accelerators and colliders can be broadly classified into linear and circular (or nearly circular) machines. With classical electrostatic accelerators and proton or electron radio-frequency linear accelerators, the scaling laws imply that the costs and other resources required should grow about linearly with energy. Although roughly true, linear scaling laws tend to become invalid as the machines approach various physical limits. The old electrostatic machines became too difficult and expensive to construct when electrical breakdown made it hard to devise accelerating columns able to withstand the necessary high voltages. And radio-frequency linear accelerators indeed obey linear scaling laws as long as there are no limits associated with their required luminosity.

The scaling laws for circular machines are more complex. Ernest

Ernest Courant, M. Stanley Livingston, and Hartland Snyder (left to right), who conceived the idea of strong focussing. (Courtesy Brookhaven National Laboratory)

Lawrence's cyclotrons obeyed an approximately cubic relationship between size or cost and the momentum of the accelerated particle. The magnet's radius grew linearly with the momentum, and the magnet gap also had to increase accordingly to provide enough clearance for the higher radio-frequency voltages required to keep particles and the voltage crest synchronized. All this changed in 1945 with the invention of phase stability by Edwin McMillan and Vladimir Vexler. Their independent work showed that only moderate radio-frequency voltages are required in circular machines because all the particles can be "locked" in synchronism with the accelerating fields.

Then came the 1952 invention of strong focusing, again independently by Nicholas Christophilos and by Ernest Courant, Livingston, and Hartland Snyder (see photograph on the right). Conventional wisdom says that a magnetic lens to focus particles both horizontally and vertically cannot be constructed— in contrast to optical lenses, which can. But the principle of strong focusing showed that, while a magnetic lens indeed focuses in one plane and defocuses in the orthogonal plane, if two such lenses are separated along the beam path, then their net effect is to focus in both planes simultaneously. This breakthrough made it possible to squeeze beams in circular (and also linear!) accelerators to much tighter dimensions, thus reducing magnetic field volumes and therefore costs.

Because the basic linear scaling laws apply to linear machines for both electrons and protons, prominent physicists predicted that all future accelerators would eventually be linear. But the question remained, "Where is the crossover in costs between circular and linear machines?" New inventions, particularly strong focusing, raised the predicted crossover to much higher energy. Moreover, strong focusing also made the scaling law for high energy proton synchrotrons almost linear. The transverse dimensions of the beam aperture do not need to grow very much with energy; thus the cost of large circular proton colliders grows roughly linearly with energy.

While the scaling laws for proton machines are not affected significantly by radiation losses (although such losses are by no means negligible for the largest proton colliders), they become the dominant factor for circular electron machines. The radiation energy loss per turn of a circulating electron varies as the fourth power of the energy divided by the machine radius. It is also inversely proportional to the mass of the circulating particle, which tells you why electrons radiate much more profusely than protons. In an electron storage ring, certain costs are roughly proportional to its radius while others are proportional to the radiation loss, which must be compensated by building large and expensive radio-frequency amplifiers. As the energy grows, it therefore becomes necessary to increase the radius. The total cost of the radio-frequency systems and the ring itself will be roughly minimized if the radius increases as the square of the energy.

Such a consideration therefore indicates that linear electron machines should eventually become less expensive than circular ones. But what is meant by the word "eventually?" The answer depends on the details. As designers of circular electron machines have been highly resourceful in reducing the costs of components, the crossover energy between circular and linear colliders has been increasing with time. But it appears likely that CERN's Large Electron-Positron collider (LEP), with its 28 kilometer circumference, will be the largest circular electron-positron collider ever built.

The only reasonable alternative is to have two linear accelerators, one with an electron beam and the other with a positron beam, aimed at one another—thereby bringing these beams into collision. This is the essential principle of a linear collider; much research and development has been dedicated to making such a machine a reality. SLAC pioneered this technology by cheating somewhat on the linear collider principle. Its linear collider SLC accelerates both electron and positron beams in the same two-mile accelerator; it brings these

beams into collision by swinging them through two arcs of magnets and then using other magnets to focus the beams just before collision. In the SLC (and any future linear collider), there is a continuing struggle to attain sufficient luminosity. This problem is more severe for a linear collider than a circular storage ring, in which a single bunch of particles is reused over and over again thousands of times per second. In a linear collider the particles are thrown away in a suitable beam dump after each encounter. Thus it is necessary to generate and focus bunches of exceedingly high density.

An extremely tight focus of the beam is required at the point of collision. There are two fundamental limits to the feasible tightness. The first has to do with the brightness of the sources that generate electrons and positrons, and the second is related to the disruption caused by one bunch on the other as they pass through each other. According to a fundamental physics principle that is of great importance for the design of optical systems, the brightness (by which I mean the intensity that illuminates a given area and is propagated into a given angular aperture) cannot be increased whatever you do with a light beam—or, for that matter, a particle beam. Thus even the fanciest optical or magnetic system cannot concentrate the final beam spot beyond certain fundamental limits set by the brightness of the original source and the ability of the accelerating system to maintain it.

The second limit is more complex. The interaction between one beam and another produces several effects. If beams of opposite charge collide, one pinches the other, usually increasing its density; but if that pinching action becomes too severe, the beam blows up! In addition, the extremely high electric and magnetic fields that arise in the process cause the particles to radiate; the energy thereby lost diversifies the energy of the different particles in the bunch, which makes it less suitable for experiments.

And there is an additional feature that aggravates the problem. As the energy of colliders increases, the cross sections of the interesting reactions decrease as the square of the energy. Therefore the luminosity—and therefore the density of the interacting bunches—must increase sharply with energy. Thus all the problems cited above will become even more severe.

As a result of all these factors, a linear collider is not really linear in all respects; in particular, the brightness of the beam must increase as a high power of its energy. This fact is difficult to express as a simple cost-scaling law. It suffices to say that all these effects eventually lead to a very serious limit on electron-positron linear colliders. Where this limit actually lies remains in dispute. At this time an upper bound of several TeV per beam is a reasonable estimate. We can hope that human ingenuity will come to the rescue again—as it has many times before when older technologies appeared to approach their limits.

THIS DISCUSSION of linear electron-positron colliders is a part of a larger question: "How big can accelerators and colliders, be they for electrons and positrons or for protons, become?" As indicated, the costs of electron-positron linear colliders may be linear for awhile, but then costs increase more sharply because of new physical phenomena. The situation is similar for proton colliders. The cost estimates for the largest proton collider now under construction—CERN's Large Hadron Collider—and for the late lamented SSC are roughly proportional to energy. But this will not remain so if one tries to build machines much larger than the SSC, such as the speculative Eloisatron, which has been discussed by certain European visionaries. At the energy under consideration there, 100 TeV per beam, synchrotron radiation becomes important even for protons and looms as an important cost component. Indeed, physical limits will cause the costs eventually to rise more steeply with energy than linearly for all kinds of machines now under study.

But before that happens the question arises: "To what extent is society willing to support tools for particle physics even if the growth of costs with energy is 'only' linear?" The demise of the SSC has not been a good omen in this regard. Hopefully we can do better in the future.