Physics – Uspekhi **55** (10) 965 – 976 (2012) ©2012 Uspekhi Fizicheskikh Nauk, Russian Academy of Sciences

REVIEWS OF TOPICAL PROBLEMS PACS numbers: **29.20.–c**, 29.20.Ej, **29.27.–a**, 41.75. Jv, 52.27.Lw, 52.38.Kd

High energy particle colliders: past 20 years, next 20 years and beyond

V D Shiltsev

DOI: 10.3367/UFNe.0182.201210d.1033

Contents

Abstract. Particle colliders for high-energy physics have been in the forefront of scientific discoveries for more than half a century. The accelerator technology of the colliders has progressed immensely, while the beam energy, luminosity, facility size, and cost have grown by several orders of magnitude. The method of colliding beams has not fully exhausted its potential but has slowed down considerably in its progress. This paper briefly reviews the colliding beam method and the history of colliders, discusses the development of the method over the last two decades in detail, and examines near-term collider projects that are currently under development. The paper concludes with an attempt to look beyond the current horizon and to find what paradigm changes are necessary for breakthroughs in the field.

1. Introduction: tosays's colliders

1.1 Method

Particle accelerators have been widely used for physics research since the early 20th century and have greatly progressed both scientifically and technologically since then. To gain an insight into the physics of elementary particles, one accelerates them to very high kinetic energy, collides them with other particles, and detects products of the reactions transform the particles are transformed into other particles. It is estimated that in the post-1938 era, accelerator science has influenced almost 1/3 of physicists and physics studies and on average contributed to Nobel Prize-winning research in physics every 2.9 years [1]. Colliding beam facilities that produce high-energy collisions (interactions) between particles of approximately oppositely directed beams have paved the way for progress since the 1960s.

V D Shiltsev. Accelerator Physics Center,
Fermi National Accelerator Laboratory (Fermilab), FNAL,
PO Box 500, MS221, Batavia, IL, 60510, USA
Tel. +1 (630) 840 52 41. Fax +1 (630) 840 60 39

Received 11 April 2012
Uspekhi Fizicheskikh Nauk **182** (10) 1033 – 1046 (2012)
DOI: 10.3367/UFNr.0182.201210d.1033
Translated by V D Shiltsev; edited by A M Semikhatov

The center-of-mass (CM) energy $E_{\rm cm}$ for a head-on collision of two particles with masses m_1 and m_2 and energies E_1 and E_2 is

$$E_{\rm cm} = \left[2E_1E_2 + (m_1^2 + m_2^2)\,c^4 + 2\sqrt{E_1^2 - m_1^2c^4}\,\sqrt{E_2^2 - m_2^2c^4}\,\right]^{1/2}. \qquad (1)$$

For many decades, the only arrangement of accelerator experiments was a fixed-target setup, where a beam of particles accelerated by a particle accelerator hit a stationary target set in the path of the beam. In this case, as follows from Eqn (1) for high-energy accelerators with $E_2 \gg m_2c^2$, the CM energy is $E_{\rm cm} \approx \sqrt{2E_2m_1c^2}$. For example, $E_2 \approx 1000$ GeV protons hitting $E_1 \approx 1$ GeV stationary protons can produce reactions with the energy of about 45 GeV. A more effective colliding beam setup, in which two beams of particles are accelerated and directed against each other, has much higher CM energy $E_{\rm cm} \approx 2\sqrt{E_1E_2}$.

In the case of two equal masses of particles (e.g., protons and protons, or protons and antiprotons) colliding with the same energy $E = 1000$ GeV, we obtain $E_{\rm cm} = 2E = 2000$ GeV. Such an obvious advantage led to the first practical proposals of colliding-beam storage rings in the late 1950s [2, 3].

Almost three dozen colliders have reached the operational stage between the late 1950s and now. Schematic drawings of several collider types are shown in Fig. 1. In storage ring configurations in Fig. 1a and 1b, particles of each beam circulate and repeatedly collide. This can be done in a single

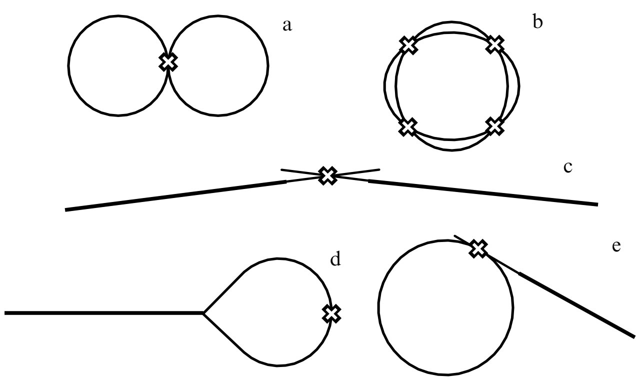

Figure 1. Schematics of particle collider types.

ring if the beams consist of the same-energy antiparticles. In linear colliders, first proposed in Ref. [4], the beams are accelerated in linear accelerators (linacs) and transported to the collision point: either in the simple two-linac configuration depicted in Fig. 1c, or with use of the same linac and two arcs, as in Fig. 1d. Another possible linac-ring configuration is shown in Fig. 1e.

1.2 Brief history of colliders, beam physics, and key technologies

The first colliding lepton beam facilities were built in the early 1960s almost simultaneously at three laboratories: e^-e^- AdA colliders at the Frascati Laboratory near Rome in Italy, the VEP-1 collider at the Novosibirsk Institute of Nuclear Physics (USSR), and the Princeton–Stanford Colliding Beam Experiment at Stanford (USA). Their CM energies were 1 GeV or less. Construction of the first hadron (proton–proton) collider, the Intersecting Storage Rings, began at CERN (Switzerland) in 1966, and in 1971 this collider was operational and eventually reached $E_{cm} = 63$ GeV. The first linear collider was the e^-e^+ SLAC Linear Collider (SLC) constructed at Stanford in the late 1980s. Detailed discussions of the history of colliders can be found, e.g., in [5, 6], while Refs [7–9] give an overview of the development of the colliding beam facilities in Novosibirsk. (Because this article is not intended to be a historical overview, only the most recent relevant references are provided below.)

The energy of colliders has been increasing over the years as is demonstrated in Fig. 2. There, the triangles represent the maximum CM energy and the start of operation for lepton (usually, e^+e^-) colliders and dots are for hadron (proton, antiproton, ion, proton–electron) colliders. We see that until the early 1990s, the CM energy on average increased by a factor of 10 every decade and, notably, the hadron colliders were 10–20 times more powerful. Since then, following the demands of high-energy physics, the paths of the colliders diverged: the Large Hadron Collider (LHC) was built at CERN to reach record high energies in particle reactions, while new e^+e^- colliders called "particle factories" have focused on detailed exploration of phenomena at much lower energies.

The exploration of rare particle physics events requires both appropriately high energy and a sufficiently high number of them. The event rate dN_{exp}/dt in a collider is the product of the interaction cross section σ_{int} and the factor L called the luminosity:

$$\frac{dN_{exp}}{dt} = L\sigma_{int} \,. \tag{2}$$

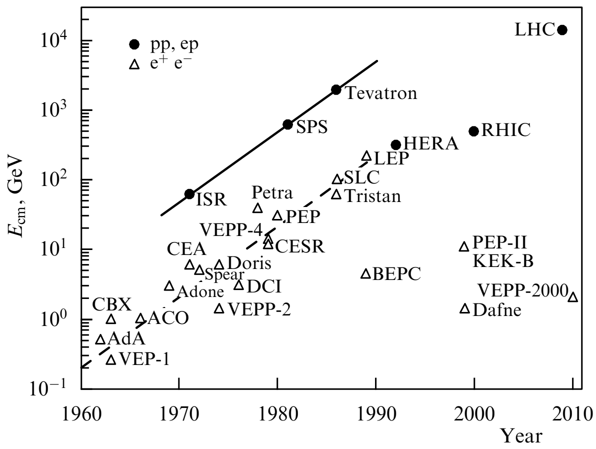

Figure 2. Colliders over the decades.

If two bunches containing N_1 and N_2 particles collide with a frequency f, the luminosity is

$$L = f\,\frac{N_1 N_2}{A} \,, \tag{3}$$

where A is the effective overlap area of the beams. In the simplest case of two bunches with identical Gaussian transverse beam profiles characterized by rms widths σ_x and σ_y, the overlap area is approximately $4\pi\sigma_x\sigma_y$ (we here omit any corrections due to the nonuniform longitudinal profile of the luminous region). The beam size can in turn be expressed in terms of the rms normalized transverse emittance $\varepsilon_{x,y}$ (which is an approximate adiabatic invariant of particle motion during acceleration) and the amplitude function $\beta_{x,y}$ (which is a beam optics quantity determined by an accelerator transverse, most often magnetic, focusing system):

$$\sigma^2_{x,y} = \frac{\varepsilon_{x,y}}{\gamma\beta_{x,y}} \,, \tag{4}$$

where $\gamma = E/mc^2$ is the relativistic Lorentz factor. Therefore, the basic equation for luminosity (3) can now be rewritten in terms of emittances and the amplitude β functions at the interaction point (which we denote by asterisks) as

$$L = f\gamma\,\frac{N_1 N_2}{4\pi\sqrt{\varepsilon_x\beta^*_x\varepsilon_y\beta^*_y}} \,. \tag{5}$$

Therefore, to achieve high luminosity, we have to maximize the population of bunches with as low emittances as possible and to collide them at a high frequency at locations where the focusing beam optics provide the lowest values of the amplitude functions $\beta^*_{x,y}$. Increasing the beam energy and hence the factor γ in Eqn (5) is also helpful in general.

Figure 3 demonstrates the impressive progress in luminosities of colliding beam facilities since the invention of the method. Again, the triangles are for lepton colliders and full circles are for hadron colliders. We can see that over the last 50 years, the performance of the colliders has improved by more than 6 orders of magnitude and reached record high values over 10^{34} cm^{-2} s^{-1}. At such a luminosity, we can expect, for example, to produce 100 events over one year of operation (about 10^7 s) if the reaction cross section is 1 femtobarn (fb $= 10^{-39}$ cm^2).

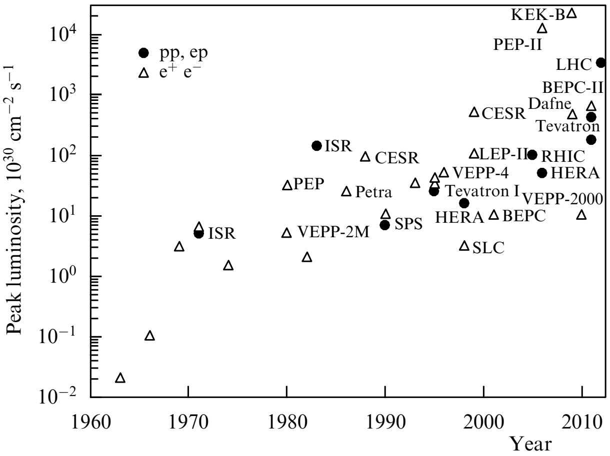

Figure 3. Peak luminosities of particle colliders.

Needless to say, such great progress in both the energy and the luminosity of colliders has come from numerous advances in accelerator science and technology. The format of this article does not allow going through all the advances in detail — the interested reader can refer for details, for instance, to the comprehensive handbook [6] — but it is worth listing at least the major ones, because they are relevant to our subsequent discussion on future colliders. References provided below are given to only relatively recent developments of the last two decades.

The maximum energy of colliders is determined by practical considerations, of which the first is the size of the facility. For a linear collider, the beam energy is the average accelerating gradient G times the length of the linac l:

$$E = eGl\,. \tag{6}$$

In the (only) linear collider SLC, the average gradient in a 3-km-long linac, which was powered at a frequency of 2.856 GHz ($\lambda_{\mathrm{RF}} = 10.5$ cm wavelength in normal-conducting copper structures), did eventually reach some 21 MV m^{-1}. Notably, although the length of the linac did not change, the beam energy approximately tripled over 30 years of operation because of upgrades, such as quadrupling the number of RF power sources (pulsed klystrons), increasing the peak output power of the klystrons from 35 MW to 65 MW, and further increasing the power by compressing the length of the RF pulses.

In circular colliders, the maximum momentum p and energy E of ultra-relativistic particles is determined by the radius of the ring R and the average magnetic field B of bending magnets:

$$pc = eBR \quad \text{or} \quad E\,[\mathrm{GeV}] = 0.3\,B\,[\mathrm{T}]\,R\,[\mathrm{m}]\,. \tag{7}$$

Again, evolution of energy was driven by practical considerations: for example, the maximum field of normal-conducting magnets of about 2 T at some moment was not sufficient for the energy demands because of the required longer accelerator tunnels and increasing magnet power consumption. Development of superconducting (SC) magnets (Fig. 4), which employ high electric current carrying NbTi wires cooled by liquid helium below 5 K opened the way to higher fields and record high energy hadron colliders [10]. The latest of them, the 14 TeV c.m. LHC at CERN, uses 8.3 T double bore magnets in a 26.7 km circumference tunnel.

To remain superconducting, such magnets need to operate within very strict limits on the power deposited into the low-temperature components (vacuum pipes, cold iron, SC cable, etc.), typically of the order of 1 W m^{-1} or less, which makes them of no practical use in high-energy lepton accelerators because relativistic electrons and positrons quickly lose energy due to the synchrotron radiation:

$$\delta E = \frac{4\pi}{3}\,\frac{e^2\gamma^4}{R} = 88.5\left[\frac{\mathrm{keV}}{\mathrm{turn}}\right]\frac{E^4\,[\mathrm{GeV}]}{R\,[\mathrm{m}]}\ . \tag{8}$$

The total power radiated into the walls becomes prohibitively high: e.g., 22 MW or about 800 W/m even in the largest-radius e^+e^- collider, LEP (in the same tunnel that is now occupied by LHC), at a beam energy of 105 GeV and relatively low average beam current of 4 mA. Besides the need to replenish electron beam power by the accelerating RF cavities, the synchrotron radiation leads to significant heating and outgassing of the beam vacuum pipe. On the other hand, the attainment of sufficiently long lifetimes of continuously circulating beams requires gas pressures of 1–10 nTorr or less. This technological challenge has been successfully resolved in modern (lower-energy) colliding ''factories''

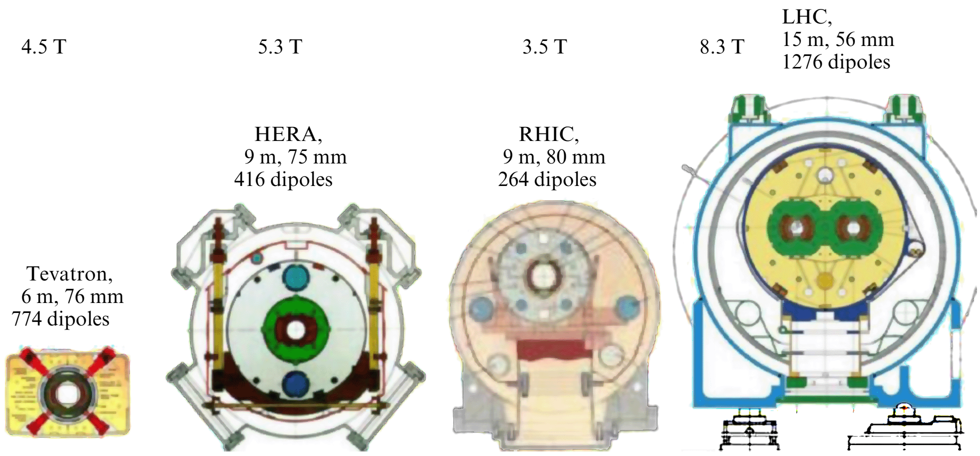

Figure 4. Superconducting dipole magnets for high energy hadron colliders: Tevatron (NbTi, warm-iron, small He plant, 4.5 K), HERA (NbTi, Al collar, cold iron), RHIC (simple and economical design), and LHC (2 K super fluid He, double bore). Courtesy of A Zlobin.

operating with multi-Ampère beams. Radiation of protons (ions) is smaller by a significant factor, $(\gamma_p/\gamma_e)^4 = (m_e/m_p)^4$ [see Eqn (8)], but can still be of a certain concern in very high-energy, high-current SC accelerators, for instance, the LHC.

From the very beginning of colliders, it was understood that their effective operation required the stability of particle motion in the rings. It is desired for particles to exhibit bound transverse oscillations in the fields of focusing magnets, as in the solution of Hill's equation

$$x(s) = A\sqrt{\beta(s)}\,\cos\left[\psi(s) + \delta\right], \tag{9}$$

where the phase $\psi(s)$ advances around the ring and determines the so-called tune, or the number of such betatron oscillations per turn:

$$\nu = \frac{1}{2\pi}\oint \mathrm{d}\psi(s)\,. \tag{10}$$

Imperfections of the guiding magnetic fields can lead to multi-turn instabilities due to linear or nonlinear resonances and cause particles to leave the acceptable aperture. At first, this was avoided by a careful choice of the tune(s) — which is usually set away from integer or rational values, $\nu \neq n/m$ — and with careful construction of the magnets. The magnetic field quality is given by the multipole coefficients in the expansion

$$B_x + \mathrm{i}B_y = B_0\sum_{n=0}(b_n + \mathrm{i}a_n)\left(\frac{x+\mathrm{i}y}{R_0}\right)^n, \tag{11}$$

where R_0 is the reference radius, the pole number is $2(n+1)$, a_n and b_n are the normal (skew) multipole coefficients, and b_0 is unity. Over the years, collider magnet builders have perfected magnet designs to the level of around 10^{-4} in the undesired multipoles: this was particularly challenging for SC magnets where such quality required very tight tolerances of less than a few dozen microns for the placements and position stability of the conductor coils under enormous magnetic forces [10]. Even more demanding were the field quality tolerances for the special, very strong magnets now routinely used in modern colliders to over-compress beams at the interaction points and arrange minimal possible beta-functions at the collision points $\beta^*_{x,y}$; per Eqn (5), such magnets were of great help in achieving higher luminosities.

Unfortunately, the growing demand for luminosity led to nonlinearity that was orders of magnitude stronger due to the very nature of the colliders, namely, due to electric and magnetic forces of the opposite bunch at the interaction points. The nonlinearity due to collisions is characterized by a dimensionless parameter called the ''beam–beam parameter'' or the ''beam–beam tune shift'':

$$\xi_{x,y} = \frac{Nr_c}{2\pi\gamma(\sigma^*_x + \sigma^*_y)}\,\frac{\beta^*_{x,y}}{\sigma^*_{x,y}}\,, \tag{12}$$

(here, all the parameters are for the opposite bunch, and $r_c = e^2/mc^2$ is the classical radius of the particles), which can reach a value ranging from $\xi \sim 0.03$ in hadron colliders to $\xi \sim 0.13$ in lepton colliders. Beyond this limit, a variety of so-called ''beam–beam effects'' have usually led to intolerable operational conditions due to beam losses, beam size blowups, etc.

It follows from Eqns (3) and (5) that the path to higher luminosity via higher beam intensity and smaller beam size almost automatically leads to higher beam–beam parameters. To bypass the above-mentioned ''beam–beam limit,'' several methods have been implemented over the decades, including (a) making a careful choice of working tunes away from the most detrimental resonances; (b) operation with very flat bunches [wide in the horizontal plane and narrow in the vertical one, see Eqn (12)]; (c) breaking beams into a significant number of bunches separated at all but a few interaction points, such that the single bunch intensity N is reduced; more recently, (d) compensation of the beam–beam effects using electron lenses [11]; (e) reduction of the strength of the beam–beam resonance in the ''round beams'' scheme with strongly coupled vertical and horizontal motion [12]; and (f) using the so-called ''crab-waist'' collision method that beneficially modifies the geometry of the colliding bunch profiles only at the interaction points [13]. Despite all the inventions, the beam–beam effects are still considered to be setting a limit, not fully resolved, on the performance of most colliders.

There are two ways to achieve high luminosity within the beam–beam limit: either to increase the beam current $I = efN$ or to decrease the beam emittance ε. In addition to the synchrotron radiation power, which increases linearly with I, the major limitations on the current come from so-called coherent beam instabilities and from the demands of the radiation protection from inevitable particle losses. The instabilities are caused by beam interaction with the electromagnetic fields induced by the beam itself in the vacuum chambers and RF cavities, or caused by unstable clouds of secondary particles, like electrons or ions, which are formed around the circulating beams [14, 15]. Beam-based transverse and longitudinal feedback systems and electron/ion clearing (either by a weak magnetic or electric field or by modulation of the primary beam current profile, making the secondaries unstable) are now in routine use to avoid coherent instabilities. Incoherent particle losses may have a variety of causes: unstable single-particle motion, diffusion due to scattering on residual vacuum molecules or on the other particles within the bunch, high-frequency noises in the guiding magnetic field or the EM fields of RF cavities, ground motion and artificial vibrations, etc. To avoid damage or excessive irradiation of the accelerator components, allowing them to be accessed and maintained if necessary in the tunnel, sophisticated collimation systems are utilized, which usually employ a series of targets (which scatter the halo particles) and absorbers (which intercept the particles in dedicated locations) [16], or use more sophisticated techniques like collimation by bent crystals [17] or by hollow electron beams [18].

As regards the quest for smaller beam emittances (phase space area), the paths of the lepton and hadron colliders were different. In the electron machines, where any transverse errors are naturally damped by the synchrotron radiation, the emittance is determined by quantum fluctuations of the radiation, which mostly excite horizontal oscillations; therefore, the challenge is to design focusing beam optics to minimize the effect in the horizontal plane and avoid its coupling to the vertical plane, ideally below 1% [19]. Hadron colliders cannot enjoy fast damping due to the synchrotron radiation, at least for energies less than 10 TeV, and therefore the only ways to progress for them were either to generate low-emittance (high-brightness) beams in the sources or to arrange beam ''cooling'' (phase space reduction, usually at the low- or

Table 1. Past, present, and possible future colliders; hadron colliders are in bold, lepton colliders in italics, facilities under construction or in decisive design and planning stage are in parenthesis (...).

	Early 1990s	Early 2010s	2030s
Europe	**H*E*RA**, (**LHC**) *LEP*, (*Dafne*)	**LHC** (*Super-B*, **HL-LHC**, **LH*e*C**, ***E*NC**)	**HE-LHC** *CLIC*?
Russia	*VEPP-2*, *VEPP-4* (**UNK**, *VLEPP*)	*VEPP-2000*, *VEPP-4M* (**NICA**, *Tau–Charm Factory*)	**NICA**? *Higgs Factory*?
United States	**Tevatron**, (**SSC**) *SLC*, *CESR*, (*PEP-II*)	**RHIC** (***e*RHIC**, ***EL*IC**)	*Muon Collider*? *PWLA/DLA*?
Asia	*Tristan*, *BEPC* (*KEK-B*)	*BEPC* (*Super-KEKB*)	*ILC*? *Higgs Factory*?
Total	9 (7)	5 (9)	1 + ?

medium-energy accelerators in the injector chain), using either "stochastic cooling" or "electron cooling" methods, and the latter has been recently exemplified in a relativistic system [20].

Operation of colliders with progressively smaller and smaller beams brought up many issues relevant to alignment of magnets, vibrations, and long-term tunnel stability [21]. Radiation backgrounds in high-energy physics detectors necessitated a careful design of the accelerator–detector interface in high-luminosity colliders. High-energy physics demands for polarized beam collisions and very precise CM energy calibration $(\delta E/E) \sim 10^{-5}$ have been largely satisfied by, correspondingly, the development of polarized particle sources married with sophisticated methods to maintain beam polarization along the acceleration chain [22] and by novel method of "resonant depolarization" [19, 23].

To conclude this section, we note that colliders have had 50 glorious years, as not only many important particle discoveries were made in them, but they also initiated a wide range of innovation in accelerator physics and technology, which resulted in a 100-fold increase in energy (for both hadron and lepton colliding facilities) and a 10^4-10^6-fold increase in the luminosity. At the same time, it is obvious that the progress in the maximum CM energy has drastically slowed down since the early 1990s (and lepton colliders even went backwards in energy; see Fig. 2). Moreover, the number of facilities in operation has dropped from 9 to 5, as indicated in Table 1, which lists all operational colliders as of the early 1990s and now (early 2010s) and accounts for the projects under construction or under serious consideration at the time (in parenthesis).

1.3 Past 20 years: Achievements and problems solved

Tectonic changes in the field happened in the early 1990s, when two flagship projects were terminated — the 6 TeV CM proton–proton complex UNK [24] in Protvino, Russia, in 1991 and the 40 TeV CM proton–proton Superconducting Super Collider in Texas, USA, in 1993 [25]. Our current landscape shows the end of the Tevatron era (the 26-year-long, ~ 2TeV CM energy proton–antiproton collider run ended in September 2011) and is dominated by the LHC at CERN. The Tevatron, LEP, and HERA established the Standard Model (SM) of particle physics. The next generation of colliders is expected to explore it at deeper levels and to eventually lead to the exploration of the smallest dimensions beyond the current SM. In the following sections, we outline possible colliders for the next 20 years and take a look at the issues and options for the even more distant future.

2. Next 20 years: Physics, technologies, and facilities

The future of colliders is ultimately driven by the demands of particle physics, but should stay within the limits of the available technologies and financial resources. All the projects currently under construction or at the design stage (see Table 1) satisfy those three requirements and therefore have good prospects of becoming operational and delivering results in the next 20 years. Schematically, they can be categorized by the area of promising physics as follows:

2.1 LHC upgrades and lower-energy colliders

2.1.1 Energy frontier. The LHC luminosity upgrade project HL-LHC [26] will employ novel SC magnet technology based on Nb_3Sn superconductors for tighter focusing at the interaction points and quintuple the performance of the energy frontier machine by the mid-2020s to 5×10^{34} cm^{-2} s^{-1}, with luminosity leveling at 14 TeV CM energy in proton–proton collisions, and will enable obtaining about 250 fb^{-1} of integrated luminosity per year, with the ultimate goal of 3000 fb^{-1} for both ATLAS and CMS experiments.

2.1.2 Low-energy hadron collisions. Investigation of the mixed phase of quark–gluon matter and polarization phenomena at relatively low hadron energies has recently become of significant interest for the high-energy physics community, and it is the main goal of the Nuclotron-based Ion Collider fAcility (NICA) currently under development at JINR (Dubna, Russia) [27]. NICA will allow the study of ion–ion (Au^{+79}) and ion–proton collisions in the energy range 1–4.5 GeV/amu with an average luminosity of 10^{27} cm^{-2} s^{-1} and also polarized proton–proton (5–12.6 GeV/amu) and deuteron–deuteron (2–5.8 GeV/amu) collisions; in that mode, luminosities up to 10^{31} cm^{-2} s^{-1} are foreseen. The plans indicate the start of operation and initial physics results later in this decade.

2.1.3 Electron–hadron collisions. Deep inelastic electron–nucleon scattering is the focus of a new electron-hadron collider project, the LHeC [28], in which polarized electrons of 60 GeV to possibly 140 GeV collide with LHC protons of 7000 GeV with a design luminosity of about 10^{33} cm^{-2} s^{-1}. This would exceed the integrated luminosity collected at the previous HERA ep collider at DESY by two orders of magnitude in a kinematic range 20 times wider in the transferred momentum Q^2. A similar approach of reusing an existing beam facility and adding an accelerator for another species is taken in two collider projects in the US: the eRHIC at BNL [27] and the Electron–Ion Collider (ELIC) at JLab [28]. The eRHIC design is based on one of the existing RHIC (Relativistic Heavy Ion Collider) hadron rings, which can accelerate a polarized nuclear beam to 100 GeV/nucleon and polarized protons up to 250 GeV, and a new 20–30 GeV multi-pass energy-recovery linac (ERL) to accelerate polarized electrons; the luminosity varies from 10^{33} cm^{-2} s^{-1} to 10^{34} cm^{-2} s^{-1}, depending on the energy and species. The ELIC proposal reuses the CEBAF 3–7 GeV polarized electron accelerator and requires the construction of a 30 to 150 GeV storage ring for ions (p, d, He^3, and Li, and unpolarized light to medium ion species). To attain very high luminosities in the ELIC, from 5×10^{33} cm^{-2} s^{-1} to 10^{35} cm^{-2} s^{-1}, an ERL-based continuous electron cooling facility is anticipated to provide low emittance and simultaneously very short ion bunches. Although with a lower CM energy than LHeC, both projects in the US have the advantage of colliding both electron and ion species with polarized spins. It is believed that only one of the two can be supported and constructed.

Complementary physics programs can be realized at the proposed electron–nucleon collider ENC at the upcoming Facility for Antiproton and Ion Research (FAIR) at GSI Darmstadt (Germany) by utilizing the 15 GeV antiproton high-energy storage ring HESR for polarized proton and deuteron beams, with the addition of a 3.3 GeV storage ring for polarized electrons [29]. This will enable electron–nucleon collisions up to a CM energy of 14 GeV with peak luminosities in the range $10^{32} - 10^{33}$ cm^{-2} s^{-1}.

2.1.4 Electron–positron factories. In the late 1990s–early 2000s, two asymmetric-energy e$^+$e$^-$ B-factories, the KEKB collider for the Belle experiment at KEK and the PEPII collider for the BaBar experiment at SLAC, achieved tremendous success in confirmation of the Standard Model (SM) in the quark flavor sector and indicated that the Kobayashi–Maskawa mechanism is the dominant source of observed CP violation in nature. Despite that, two fundamental questions remain unanswered in the flavor sector of quarks and leptons: (a) it is not clear why the SM includes too many parameters and (b) there is still a serious problem with the matter–antimatter asymmetry in the universe. To extend the reach of physics beyond two B-factories, much higher (by a factor of about 40) luminosity Super-B factories are being either set up or considered for construction: one in Italy [32] and the other in Japan [33]. Both are asymmetric-energy e$^+$e$^-$ colliders with beam energies of about 4 GeV and 7 GeV and with a design luminosity approaching 10^{36} cm^{-2} s^{-1}, which is to be achieved via somewhat higher beam currents and very small beta-functions at the interaction points $\beta_y^* \approx 0.3$ mm, made possible by employment of the abovementioned 'crab waist' setup. Physics experiments at the Super-KEKB in Japan are expected in 2015, after which the first results are expected. Ultimately, the Belle-II detector should collect 40 times more B-meson samples per second than its predecessor — roughly 800 B–B pairs per second — and accumulate an integrated luminosity of 50 ab^{-1} = 50,000 fb^{-1} by 2021.

Many similar technical solutions, e.g., the 'crab waist,' will also be employed in the TauCharm factory project in Novosibirsk (Russia) [34], which calls for a CM collision energy variable from 3 GeV to 4.5 GeV (from J/ψ resonances to charm baryons), the luminosity in excess of 10^{35} cm^{-2} s^{-1}, and longitudinal polarization of at least one (electron) beam. If we project to the very end of the next 20 years, the landscape of collider physics is much less certain; there are several directions to advance and the choice among the options will be based upon the results from the LHC. The relevant results are expected to be available starting in 2012–2013 (e.g., the anticipated discovery of the Higgs boson), but they might easily slip well into the 2020s. We discuss five possibilities for an after-LHC collider of the 2030s.

2.1.5 Higher-energy LHC. One of the most feasible opportunities is an energy upgrade of the LHC to 33 TeV CM proton–proton collisions [35]. The HE-LHC in the existing LHC tunnel will require 20 T dipole magnets, which are currently thought to be possible via combination of the NbTi, Nb_3Sn, and HTS (high-temperature superconductor) SC magnet technology. Such a collider could follow the HL-LHC and start operation in the early 2030s. Despite the presumed feasibility of the machine, its energy reach is limited to ~ 2.5 times the LHC energy and it is not fully clear yet whether such a (relatively) small energy advance will justify its construction.

2.1.6 Higgs Factory. The Higgs boson, with the mass $m_{\rm H}$ of about 125 GeV, has been recently discovered at the LHC, and the detailed studies and precise measurements of this unique spin-0 elementary particle might be of enough significant interest to justify construction of an e$^+$e$^-$ collider — a dedicated "Higgs factory." The maximum cross section, and arguably the optimal CM energy for studies of a number of Higgs boson properties, is $E_{\rm cm} \sim m_{\rm H} + (110 \pm 10)$ GeV $\sim$ 240 GeV, and several opportunities for the facility are now under discussion, including one based on the ILC-type linear collider (see below), as well as several ring–ring versions [36]. The biggest challenge for the latter is the requirement to replenish energy loss of electrons and positrons due to synchrotron radiation of the order of 10 GeV per turn, even in 20 km or longer tunnels [see Eqn (8)], which implies extensive use of high-gradient SC RF accelerating cavities. Other challenges to attainment of the required luminosity 10^{34} cm^{-2} s^{-1} (equivalent to 20,000 events per year under the assumption of an e$^+$e$^- \to$ HZ cross section of about 200 fb $= 2 \times 10^{-37}$ cm^2) are the significant electric power consumption of the order of 100 MW, needed for continuous acceleration of an approximately 10 mA beam current and the need for very small beam emittances and very large momentum acceptance of the ring to accommodate the energy losses at the interaction points (see the discussion on the beamstrahlung effect below). The cost-saving option of the Higgs factory in an existing tunnel, e.g., the 26.7 km long LHC tunnel or 21 km long UNK tunnel, looks particularly attractive.

An alternative way for production of the Higgs bosons is the reaction $\mu^+\mu^- \to$ H (the so-called s-channel reaction), which has an advantageously large cross section for muons, $(m_\mu/m_{\rm e})^2 \approx 40{,}000$ times higher than for electrons, and (another advantage) needs a $\mu^+\mu^-$ collider with CM energy $E_{\rm cm} \sim m_{\rm H}$, lower by a factor of two. The third advantage of that scheme is the significantly smaller CM energy spread: $\delta E_{\rm cm}/E_{\rm cm} \sim 0.01 - 0.003\,\%$ (compared to $\approx 0.2\%$ for the e$^+$e$^-$ factories), which allows much better study of the outstandingly narrow-width Higgs particle decays [37]. Production of $\sim 4{,}000$ events per year will require the

Table 2. Comparison of lepton collider alternatives.

Collider	ILC	CLIC	MC
CM energy, TeV	0.5	3	1.5 – 4
CM energy spread, rms	$\sim 2\%$	$> 5\%$	$\sim 0.1\%$
Luminosity, $cm^{-2}\,s^{-1}$	2×10^{34}	$2\times10^{34*}$	$(1-4)\times10^{34}$
Feasibility report	2007	2012	2014 – 2016
Technical design	2013	2016	$\sim$ 2020
Total number of elements	38,000	260,000	10,000
Overall 'high-tech' length, km	36	$\sim$ 60	14 – 20
Wall plug power, MW	230	580	120 – 200

* Peak luminosity within a 1% c.m. energy spread.

luminosity of 10^{31} $cm^{-2}\,s^{-1}$ at least, which seems to be very challenging because of the short muon lifetime and difficulty of muon production (see a discussion on high-energy muon colliders below).

2.2 Post-LHC energy frontier lepton colliders: ILC, CLIC, and Muon Collider

2.2.1 Energy Frontier Lepton Collider. It is presently widely believed that a multi-TeV lepton collider will be needed to follow the LHC discoveries. A physics program that could be pursued by a new lepton collider with sufficient luminosity would include understanding the mechanism behind mass generation and electroweak symmetry breaking; searching for, and possibly discovering, supersymmetric particles; and hunting for signs of extra space–time dimensions and quantum gravity. By the beginning of the 2020s, the results obtained from the LHC will be expected to more precisely establish the desired lepton collider energy. The most viable options currently under consideration are the e^+e^- linear colliders ILC (International Linear Collider) [38] and CLIC (Compact Linear Collider) [39] or $\mu^+\mu^-$ Muon Collider [40]. Each of these options has its own advantages, challenges, and issues [39, 41].

2.2.2 ILC and CLIC. The biggest challenge for the linear e^+e^- colliders is to accelerate particles to the design energy within a reasonable facility footprint and with the maximum possible power conversion from the "wall-plug" to the beams. The ILC employs pulsed 1.3 GHz SC RF cavities with an average accelerating gradient of 33.5 MV m^{-1}, the total length of the 0.5 TeV CM energy facility is about 31 km, and the design power efficiency (beam power/total AC power) is about 8%. The CLIC setup is based on two-beam acceleration in 12 GHz normal conducting RF structures with an average gradient of 100 MV m^{-1}, the total length of the main tunnel of the 3 TeV CM collider is 48 km, and the overall beam power efficiency is $\sim 5\%$. Both projects have in principle demonstrated the technical feasibility of their key acceleration technologies. Both have very tight requirements on the beam emittance generated in several-km-long injection rings, the emittance preservation in the main linacs where the beams are subject to minuscule transverse kicks due to vibrations and other dynamic misalignments, and the need for ultimate-precision beam position monitors to stabilize beams trajectories on every shot using fast beam-based feedback systems. The stability tolerances are even tighter for the elements of several-km-long "final focus" systems — accelerator beamlines to focus beams to unprecedented beam sizes $\sigma_y^*/\sigma_x^* = 6$ nm/640 nm in the ILC and $\sigma_y^*/\sigma_x^* = 0.9$ nm/450 nm in the CLIC. Another challenge that is not easy to circumvent is the CM energy spread induced by beamstrahlung (the energy loss caused by radiation of gamma quanta by the incoming electron due to its interaction with the EM field of the electron (positron) bunch moving in the opposite direction) at the very moment of collision of short bunches with the rms length $\sigma_z = (50-300)$ μm, which for the parameters of interest can be approximated as

$$\frac{\delta E}{E} \propto \frac{\gamma N^2 r_e^3}{\sigma_x^2 \sigma_z} \tag{13}$$

and can be several percent, even reaching 10% (see Table 2). The induced radiation generates undesirable background in the detectors, makes handling the beams after the collision more difficult and, most importantly, sets limits on the energy resolution of the narrow resonances, such as in the Higgs- and expected Z'-boson decay reactions.

2.2.3. Muon collider. Muons, which can be regarded as heavy electrons, are essentially free of all synchrotron radiation related effects, which are proportional to the fourth power of the Lorentz factor γ^4 and are therefore $(m_\mu/m_e)^4 = 207^4 \approx 2\times10^9$ times smaller. Hence, a multi-TeV $\mu^+\mu^-$ collider [41] can be circular and therefore have a compact geometry that will fit on existing accelerator sites (see Fig. 5 for a possible schematic of a muon collider, MC, facility on the 6×7 km^2 FNAL site). The collider has a potentially higher energy reach than linear e^+e^- colliders, the CM energy spread in a 1.5-4 TeV $\mu^+\mu^-$ collider can be as small as 0.1%, it requires less AC power, and it operates with significantly smaller number of elements, requiring high reliability and individual control for effective operation (see Table 2). An additional attraction of a MC is its possible synergy with the Neutrino Factory concept [42] because the beam generation and injection systems of that facility and of a MC are similar (perhaps identical) [43]. As mentioned above, due to the higher mass of the muon and superb energy resolution, a Higgs factory based on low(er)-energy $\mu^+\mu^-$ collisions is also very attractive.

The biggest challenges of a MC come, first, from the very short lifetime of the muon: $\tau_0 = 2$ μs is just long enough to allow acceleration to high energy before the muon decays into an electron, a muon-type neutrino, and an electron-type antineutrino ($\mu^- \to e^-\nu_\mu\bar{\nu}_e$), and, second, from the methods of muon production as tertiary particles in the reactions $pN \to \pi + \ldots \to \mu + \ldots$, such that the beams of muons are generated with very large emittances. A high-energy, 1–5 TeV CM high-luminosity, 10^{34} $cm^{-2}\,s^{-1}$ muon collider would require a reduction by a factor of 10^6 of the 6-dimensional muon beam phase space volume (muon cooling). Although

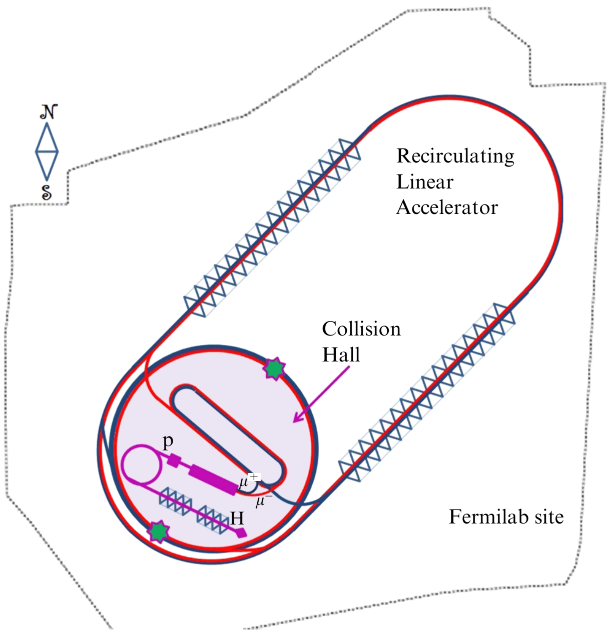

Figure 5. Schematic for a 4 TeV Muon Collider at the FNAL site.

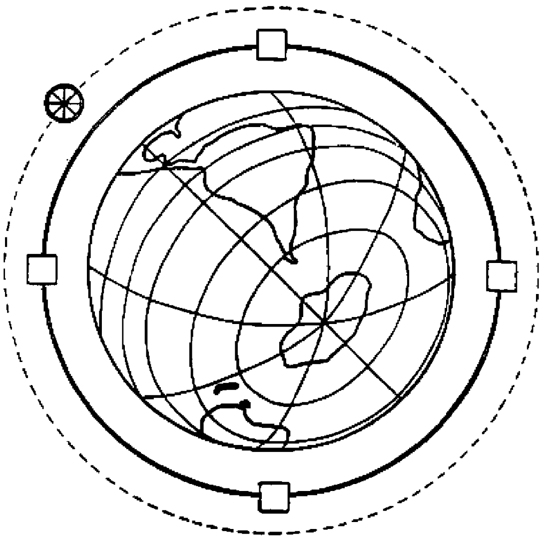

Figure 6. Enrico Fermi's "ultimate accelerator" encircling Earth.

significant progress has been achieved over the past decade in developing the concepts and technologies needed to produce, capture, and accelerate muon beams with high intensities of the order of 10^{21} muons/year, the feasibility of the high-luminosity multi-TeV muon collider can be ascertained only after demonstration of the fast ionization cooling of muons and resolution of the related issue of the breakdown of normal conducting RF cavities in strong magnetic fields. The latter is expected to be addressed in 2014–16, while a convincing demonstration of the 6D cooling might take another 4 to 6 years [41].

3. Beyond 2030: New methods and paradigm shift

Our forecast for the future of colliders beyond 2030 requires several assumptions: on the available resources, on the desired science reach, and on the possible ways to the goal. We start with the money.

3.1 Possible development of colliders in the resource-limited world

Currently, the world's particle physics research budget is some \$3B. That is about 1.2% of the world's total spending on basic science research (about \$250B). For comparison, global R&D funding of some \$1,400B is about 2% of the world's GDP $\sim$ \$70,000B (all numbers are taken from [44]). We argue that the era of exponential expansion for most of about 200 sciences (see, e.g., [45]) is gone, and for particle physics, the new era of much more modest growth, or even a slowdown, in financial support began in the 1990's, as we can also conclude from Fig. 2 above. Although most of today's scientific leaders in scientific communities, governments, and universities came of age during the previous "golden era" and have not yet fully adjusted to new realities, we base our predictions on the assumption that the field will stay approximately within the current financial limits, while still vigorously exploring new methods and directions. This immediately cuts a number of unrealistic possibilities, such as Enrico Fermi's ultimate accelerator or "globaltron" [46] (see Fig. 6), whose cost would exceed \$20,000B even under a modest estimate of \$0.5B per kilometer of a high-tech accelerator.

The colliding beam facilities discussed in Section 2 can be distributed into several logarithmically broad categories, depending on the cost of construction (Table 3).

From what we know now, category I and II colliders can be built by a single country, with a relatively modest international contribution; category III requires significant cooperation at least within one of the big regions (Europe, Asia, America); category-IV machines will arguably need the close cooperation of two or all three regions; and category V colliders must be truly international. Following our basic assumption (that the particle physics budget will not grow much beyond $\sim$\$3B in current prices), it is hard to imagine that any of the projects in category V are possible, because none of the possible scenarios — "do nothing but build for 3–10 years," or "carve some 1/3 of the budget for construction and finish it in 10–30 years," or at an even slower pace — seem realistic, especially for an energy reach only comparable to the LHC. Category-III machines appear totally feasible financially, but they only add to the exploration of the phenomena discovered by the LHC; thus, there will always be an alternative: "why don't we just run LHC longer?" We are therefore left with category-IV colliders, which appear to be on the border of financial affordability. There is a problem with them, however: they have limited energy reach, equivalent to some 5 TeV CM energy in parton collisions (that is the muon collider energy, and the HE-LHC CM energy divided by an equivalent number of partons in a

Table 3.

Category	Cost, billions of dollars	Facility
I	$\leqslant 0.3$	NICA, ENC
II	0.3 – 1	Super-B factories, c – τ factory, eRHIC, ELIC
III	1 – 3	Higg factory, HL-LHC
IV	3 – 10	HE-LHC, LHeC, MC, Higgs factory–ILC
V	10 – 30	ILC, CLIC

proton: $33/(6-10) \approx 3-5$ TeV). It is not certain whether there will be a demanding physics case for such CM energies to justify the required significant investment.

3.2 Future technologies: Acceleration in microstructures, in plasma and in crystals

Below, we offer some options for the physics reach of colliders in the distant future within an approximate budget of \$10B in current prices. The cost models of modern colliders are quite complicated and do not fit the scope of this article, but we can safely assume that future facility of such cost should not exceed a few dozen km in length and simultaneously require less than a few dozen MW of beam power (or correspondingly stay under ~ 100 MW of AC wall power consumption). Can such colliders reach energies orders of magnitude beyond current ones, namely, 100–1000 TeV?

Achieving the energies of interest within the given footprint, requires fast acceleration. At present, three main opportunities for high-energy colliders are being actively discussed, which can be schematically classified by the type of media used for acceleration: solid state structures, plasma, and crystals. (Below, we give references the most recent reviews.)

3.2.1 Acceleration in dielectric structures. Most of the present accelerators use RF fields ($f_{\mathrm{RF}} < 10-30$ GHz) in resonant normal-conducting or SC structures powered by conventional RF sources and are, in general, limited to gradients of ~ 100 MV m^{-1} due to surface breakdown phenomena. A combination of direct beam excitation (or wakefields radiated by a short intense "driving" bunch of electrons propagating in a high-impedance environment) and hard dielectric materials for structure fabrication (quartz, diamond, garnets etc., which are characterized by lower power losses and higher breakdown gradients than metals) allow the respective accelerating gradients ~ 100 MeV m^{-1} and ~ 1 GeV m^{-1} in microwave 100 GHz and 1 THz dielectric structures [47]. Conceptually, a dielectric wakefield accelerator (DWA)-based linear collider would consist of a large number of $\sim$ 100-GeV modules (stages) with the gradient ~ 0.3 GeV m^{-1}, each driven by a separate ~ 1 GeV high-intensity electron beam. Even without going into the difficulties associated with staging, cost, and power considerations, it is difficult to imagine that more than a 3 TeV CM energy DWA facility can fit within a 10 km site.

A further increase in the gradient to 1–3 GeV m^{-1} is thought to be possible in μm scale dielectric structures driven by lasers operating in the optical or near-infrared mode [48]. In various options, either external fiber lasers are coupled to the structures or semiconductor lasers can be integrated on the same slab right next to the microcells they power. The advantage of such an approach is that laser power sources can operate at very high repetition rates of 10–100 MHz, which helps reach higher luminosity [see Eqn (5)], but again, a 10 TeV CM energy collider will require 12 km of total linac length [48]. We note that staging sequential accelerating modules made of smaller-size structures (microns vs millimeters) will require proportionally tighter synchronization, alignment, and mechanical stability tolerances to keep beam trajectories and emittances under control.

3.2.2 Acceleration in plasma. In the past decade, plasma-wakefield acceleration (PWA) methods have much attention because of the promise to sustain extremely large acceleration gradients. Electric fields due to charge separation in dense plasma are of the order of

$$E_0 = \frac{m_{\mathrm{e}} c \omega_{\mathrm{p}}}{e} \approx 100 \left[\frac{\mathrm{GeV}}{\mathrm{m}}\right] \sqrt{n_0 [10^{18}\ \mathrm{cm}^{-3}]}\,, \qquad (14)$$

where $\omega_{\mathrm{p}} = (4\pi e^2 n_0/m_{\mathrm{e}})^{1/2} = 2\pi c/\lambda_{\mathrm{p}}$ is the electron plasma frequency, n_0 is the ambient electron number density, and $\lambda_{\mathrm{p}} \approx 30\,[\mu\mathrm{m}]\,(n_0\,[10^{18}\ \mathrm{cm}^{-3}])^{1/2}$ is the plasma wavelength that sets the characteristic scale length of the wakefield. For example, the generation of PWA gradients of the order of 30–100 GeV m^{-1} at plasma densities $n_0 = 10^{17}-10^{18}$ cm^{-3} have already been demonstrated in small-scale (a few cm to a meter) experiments [49].

There are two ways to separate electrons and ions in the plasma: by lasers and by external beams. In laser–plasma accelerators, a longitudinal accelerating electric field is generated by the ponderomotive force of an ultraintense laser pulse with a duration of the order of the plasma period. This force, proportional to the gradient of the laser intensity, pushes the plasma electrons out of the laser beam path, separating them from the less mobile ions. This creates a traveling longitudinal electric field in the wake of the laser pulse, with a phase velocity close to the speed of light, i.e., as is needed for accelerating relativistic particles. The structure of the wakefield has a broad phase space where negatively charged particles can be both accelerated and focused. Depletion of the laser energy determines the energy gain and the length of a single acceleration stage, after which a new laser pulse must be coupled into plasma for further acceleration. Practical considerations indicate that this stage-to-stage distance is of the order of 1 m, and minimization of the total linac length requires operation at relatively low densities $n_0 \approx 10^{17}$ cm^{-3}, energy gain per stage 10 GeV, and average gradient ~ 5 TeV km^{-1} [50].

An alternative LWA concept involves the passage of an ultra-relativistic electron bunch through a stationary plasma (either preformed by ionizing a gas by a laser or through field ionization by a Coulomb field of the relativistic electron bunch itself). The plasma electrons are repelled by the bunch. Generating large-amplitude wakefields requires short high-density electron bunches compressed in all three spatial dimensions to sizes smaller than λ_{p}, for example, bunches of around 10^{10} electrons a few micrometers in spot size and ~ 10 μm long. The wake produces a high-gradient accelerating field, Eqn (14), and transverse focusing for a negatively charged witness bunch behind the drive bunch [51].

The figure of merit in wakefield accelerators (plasma or dielectric) is the transformer ratio R_{T} = (maximum accelerating field behind the drive bunch)/(the maximum decelerating field inside the drive bunch), which is limited to ~ 2 for finite-length longitudinally symmetric drive bunches and can reach somewhat higher values only for specially prepared triangular bunch current shapes. A proposed beam-PWA collider design [52] consists of a conventional 25 GeV electron drive beam accelerator, which produces drive bunches distributed in counter-propagating directions to many (tens to hundreds, depending on the final energy) one-meter-long plasma cells for both the electron and positron arms of the collider. Although each cell provides a 25 GeV energy gain ($R_{\mathrm{T}} \approx 1$), an average geometric accelerating gradient reaches only 0.25 TeV km^{-1} because of approximately 100-meter-long cell-to-cell gaps needed to bring a fresh drive beam to the next cell from a single source of the high-intensity 25 GeV electron bunches. As in other setups, serious issues associated with staging and transfers from one cell to another (alignment, synchronization, etc) are anticipated. The number of stages can be significantly reduced if higher-energy drive beams are available, e.g., some 0.6 TeV energy gain might be possible in a single ~ 400 m plasma cell driven by 1 TeV protons if such high-energy proton bunches can be compressed longitudinally to under a bunch length of 1 mm and kept tightly focused transversely [53]. Without going into practical considerations and just projecting such gradients further, we can imagine an ultimate ~ 10 TeV collider within a 10 km footprint.

3.2.3 Acceleration in crystal channels. The density of charge carriers (conduction electrons) in solids, $n_0 \approx 10^{22}$–10^{23} cm^{-3}, is significantly higher than what was considered above in plasma, and correspondingly, longitudinal fields of up to 100 GeV cm^{-1} or 10 TeV m^{-1} are possible [see Eqn (14)]. The new effects at higher densities are due to intense energy radiation in high fields and increased scattering rates, which result in fast pitch-angle diffusion over distances of $l_\mathrm{d}\,[\mathrm{m}] \sim E\,[\mathrm{TeV}]$. This leads to particles escaping from the driving field; it was therefore suggested to accelerate particles in solids along major crystallographic directions, which provide a channeling effect in combination with low emittance determined by an Ångström-scale aperture of atomic "tubes." Channeling with nanotubes is also being discussed. Positively charged particles are channeled more robustly because they are repelled from ions and hence experience weaker scattering. Radiation emission due to betatron oscillations between atomic planes is thought to be the major source of energy dissipation, and the maximum beam energies are limited to about 0.3 TeV for positrons, 10^4 TeV for muons, and 10^6 TeV for protons [54]. X-ray lasers can efficiently excite solid plasma and accelerate particles inside a crystal-channel waveguide, although ultimate acceleration gradients of 10 TeV m^{-1} might require relativistic intensities, exceeding those currently conceivable for X-rays [55]. Moreover, only disposable crystal accelerators, e.g., in the form of fibers or films, are possible at such high externally excited fields, which would exceed the ionization thresholds and destroy the periodic atomic structure of the crystal (and hence acceleration will occur only in a short time before full dissociation of the lattice). For laser and plasma fields of about 1 GV cm^{-1} = 0.1 TV m^{-1} or less, reusable crystal accelerators can probably be built that can survive multiple pulses [56]. Side injection of powerful X-ray pulses into continuous fiber 0.1–10 km long allows avoiding the multiple staging issues intrinsic to other methods and reaching energies of 10–1000 TeV, if imperfections like crystal dislocations are kept under control and unintended crystal curvatures are less than the inverse "critical" radius $R_\mathrm{c}\,[\mathrm{m}] \sim 2E\,[\mathrm{TeV}]$, such that the channeling conditions remain [17].

We note that due to the imposed facility footprint limit, limited bending fields available, and problems with synchrotron radiation losses, circular colliders do not seem conceivable for ultra-high CM energies 10–100 times the LHC energy, and we with necessity come to a linear configuration, as in Fig. 1c. Based on the purely energy gain arguments, heavier particles are preferred in all novel acceleration methods because they radiate less (once again, we refer to the synchrotron radiation energy loss formula (8)) and it can be argued that acceleration of electrons and positrons beyond ~ 1–3 TeV is impractical. Also, even at that energy, the CM energy spread due to bremsstrahlung of electrons at the interaction point in Eqn (13) becomes prohibitively large at any other practical beam parameters, and special measures will need to be taken, e.g., conversion of electrons and positrons in high-energy γ-quanta and $\gamma-\gamma$ collisions [57]. Among the remaining options of heavier particles, protons seem to be the only choice for CM energies beyond 10^4 TeV = 10 PeV; for the next 2–3 decades, when we do not envision energies of practical realization greater than 1 PeV, muons are a much more attractive option because (i) they are point-like particles and, contrary to protons, do not carry an intrinsic energy spread of elementary constituents and (ii) they do not have issues associated with nuclear interactions with accelerating media of plasma or solids. The fact that muons are not stable particles and decay as $\mathrm{d}N/\mathrm{d}t = -N/\gamma\tau_0$ becomes irrelevant in the fast acceleration setups, for which the "surviving" beam fraction at the final energy E,

$$\frac{N}{N_0} \approx \left(\frac{m_\mu c^2}{E}\right)^{\kappa}, \tag{15}$$

is very close to 1 for the exponent $\kappa = (m_\mu c/\tau_0 G) \ll 1/\ln(E/m_\mu c^2)$ or, conversely, the average accelerating gradient $G \gg 3$ MeV m^{-1}, a condition that easily holds for any setup considered above (see Table 4). We note that at $G \gg 0.3$ TeV m^{-1}, accelerating the tau leptons can be discussed although production of needed quantities of τ particles is questionable.

3.3 Luminosity limits

Following Ref. [56], we explore the possible luminosity reach of ultimate-energy colliders with $E_\mathrm{cm} > 100$ TeV. Given that the acceleration of heavy particles in solid media/crystals is the technology of choice, it seems reasonable to limit the minimal overlap area of the colliding beams to the crystal lattice cell size $A \sim 1$ Å^2 = 10^{-16} cm^2 and to assume that the

Table 4. Options for future particle colliders.

Collider type	Dielectric based	Plasma based	Crystal channeling
Accelerating media	Microstructures	Ionized plasma	Solid crystals
Energy source: option 1 option 2	Optical laser e^- bunch	e^- bunch Optical laser	X-ray laser
Preferred particles	Any stable	e^-, μ^-	μ^+, p^+
Max accelerating gradient, GeV m^{-1}	1 – 3	30 – 100	100 – 10^4
CM energy reach in 10 km	3 – 10	3 – 50	10^3 – 10^5
Number of stages/10 km: option 1 option 2	10^5 – 10^6 10^4 – 10^5	~ 100 10^3 – 10^4	~ 1

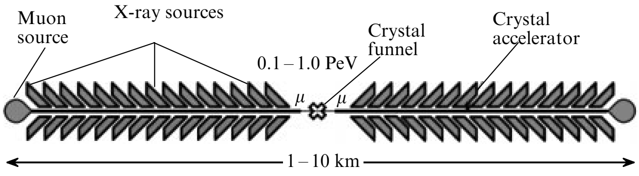

Figure 7. Concept of a linear X-ray crystal muon collider.

crystals of each collider arm will be aligned channel to channel. The other factor in luminosity formula (3) is the number of particles N, which cannot be made arbitrarily high due to the beam loading effect. Effective acceleration with the transfer ratio $R_T \sim 1$ is possible only if the number of particles in the beam does not exceed the number of particles in the plasma volume excited by an external source. Such a volume is about $100\lambda_p$ (longitudinal extent before the excitation decays) $\times\lambda_p^2$, which results in $N_0 \sim 10^3$ particles per individual bunch. Of course, exciting many parallel atomic channels n_{ch} will proportionally increase the luminosity $L = fN^2/A = f \times 10^{16} \times 10^6 n_{ch}$ [cm^{-2} s^{-1}], which can reach 10^{30} cm^{-2} s^{-1}, for example, at $f = 10^6$ Hz and $n_{ch} \sim 100$. Making the product fn_{ch} greater than 10^8 Hz can be very costly because the total beam power $P = fn_{ch}NE$ will exceed 16 MW per beam, which we consider a practical limit from our collider cost considerations (see above). It might be beneficial instead to attempt to combine (focus) all the channeling beams into one using some kind of *crystal funnel* (Fig. 7) and, thus, gain a factor of n_{ch} in the luminosity. Overall, in the power-limited scenario, the luminosity scales at very high energies as

$$L\,[\mathrm{sm}^{-2}\,\mathrm{s}^{-1}] \approx 4 \times 10^{33\text{–}35}\,\frac{P^2\,[\mathrm{MW}]}{E^2\,[\mathrm{TeV}]\,fn_{ch}[10^8\,\mathrm{Hz}]}\,. \qquad (16)$$

The performance of ultimate-energy colliders cannot increase with energy: either it is independent of it (if total beam power is small) or it decreases as $\sim 1/E^2$. This fact, if not overturned by some future invention, indicates the need for a paradigm shift in high-energy particle physics, because so far the performance goals of new energy frontier facilities scaled as $L \propto E^2$, reflecting the fact that many important cross sections decrease as $\sigma_{int} \propto 1/E_{cm}^2$. Seemingly, the physics reach of colliders of the future will be limited to either resonances (phenomena with unusually high production rates at certain energies, indicating new particles) or other high-cross-section reactions. Still, even scaling as in Eqn (16) is much better than what other sources of ultra-high-energy particles can offer. For example, the number of cosmic ray events decreases with the CM energy $E_{cm} \approx (2Emc^2)^{1/2}$ even faster, approximately as $1/E_{cm}^7$, and at the highest energies $E_{cm} \sim 500$ TeV or $E \sim 10^8$ TeV, the event rates are extremely low, a few per week even for the largest-area observatories.

4. Conclusions

The colliding beam method has been a smashing success so far: almost 3 dozen colliders have been built over the past half-century, and CM energies of about 10 TeV have been achieved. At the same time, the pace of the energy progress has greatly slowed due to increasing size, complexity, and cost of the facilities and, as a result, the number of colliders currently in operation is about half of what we had 20 years ago. The prospects of facilities for the next 20 years are not very clear today; they will depend on discoveries made at current machines, first and foremost, the LHC.

It seems that economic realities will impose severe constraints on any collider beyond 2030 to be built under about \$10B at current prices, within a footprint of some 10 km, and with a total electric power consumption of tens to 100 MW. As discussed in the previous section, there are possibilities — which even currently conceivable methods can offer — of reaching ultra-high energies of the order of 100–1000 TeV within the abovementioned limits. The quest for energy will come at the price of the expected luminosities and will require at least three paradigm shifts: (1) development of a new technology based on ultrahigh acceleration gradients $\sim 0.1-10$ TeV m^{-1} in crystals, (2) acceleration of heavier particles, preferably muons, and (3) new approaches to physics research with luminosity limited to $\sim 10^{30\text{–}32}$ cm^{-2} s^{-1}.

As with any other shift in mainstream accelerator technology, the required switch to the acceleration of muons in linear crystal structures will take a decade or two for an R&D program to address several key issues:

a) development of economical high-intensity coherent X-ray sources, e.g., based on table-top, ~ GeV scale, electron accelerators [58, 59];

b) understanding the most effective mechanisms of coupling X-ray power to the excitation of a lattice — which can probably be studied even with the existing high-power coherent X-ray sources, like LCLS in the US or Spring-8 in Japan;

c) efficient production, injection, and manipulation of nm-size muon beams (this program can effectively gain momentum from current research on muon colliders);

d) methods of combination of multiple crystal channeling beams into one (experiments at the existing high-energy proton machines might provide important input here).

As in the past, as soon as the main technology issues are addressed, a great boost to new development can be given by a test facility, where new acceleration methods are used for exploration of interesting "low-energy" physics, e.g., a "table-top" factory of ω, ψ, τ, or even Z, W, or Higgs particles with decent luminosity and relatively low cost.

The author wishes to thank C Hill, A Valishev, E Eichten, and S Henderson for the helpful discussions.

References

1. Haussecker E F, Chao A W *Phys. Perspective* **13** 146 (2011)
2. Kerst D W et al. *Phys. Rev.* **102** 590 (1956)
3. O'Neill G K *Phys. Rev.* **102** 1418 (1956)
4. Tigner M *Nuovo Cimento* **37** 1228 (1965)
5. Edwards D A, Edwards H T *Rev. Accel. Sci. Tech.* **1** 99 (2008)
6. Chao A W, Tigner M (Eds) *Handbook of Accelerator Physics and Engineering* (River Edge, N.J.: World Scientific, 1999)
7. Skrinsky A, in *Proc. of the 1995 IEEE Particle Accelerator Conf. 1–5 May 1995, Dallas, Texas. 16th IEEE Particle Accelerator Conf. and Intern. Conf. on High-Energy Accelerators*, p. 14
8. Skrinskii A N *Usp. Fiz. Nauk* **138** 3 (1982) [*Sov. Phys. Usp.* **25** 639 (1982)]
9. Budker G I *Usp. Fiz. Nauk* **89** 533 (1966) [*Sov. Phys. Usp.* **9** 534 (1967)]
10. Tollestrup A, Todesco E *Rev. Accel. Sci. Tech.* **1** 185 (2008)
11. Shiltsev V et al. *Phys. Rev. Lett.* **99** 244801 (2007)
12. Danilov V et al., in *Proc. 1996 European Particle Accelerator Conf., 10–14 June 1996, Barcelona, Spain*, p. 1149

13. Raimondi P, Shatilov D, Zobov M, SIS Pubbl. LNF-07/003 (IR) (Roma: INFN Laboratori Nazionale di Frascati, 2007); physics/0702033
14. Ohmi K *Phys. Rev. Lett.* **75** 1526 (1995)
15. Raubenheimer T O, Zimmermann F *Phys. Rev. E* **52** 5487 (1995)
16. Mokhov N et al. *JINST* **6** T08005 (2011)
17. Biryukov V M, Chesnokov Yu A, Kotov V I *Crystal Channeling and Its Application at High-Energy Accelerators* (Berlin: Springer-Verlag, 1997)
18. Stancari G et al. *Phys. Rev. Lett.* **107** 084802 (2011)
19. Brandt D et al. *Rep. Prog. Phys.* **63** 939 (2000)
20. Nagaitsev S et al. *Phys. Rev. Lett.* **96** 044801 (2006)
21. Shiltsev V *Phys. Rev. ST Accel. Beams* **13** 094801 (2010)
22. Alekseev I et al. *Nucl. Instrum. Meth. Phys. Res. A* **499** 392 (2003)
23. Shatunov Yu, Skrinsky A *Particle World* **1** 35 (1989)
24. Yarba V, in *Proc. of the 1991 IEEE Particle Accelerator Conf., APS Beam Physics. 6 – 9 May 1991, San Francisco, CA., USA*, p. 2913
25. Wojcicki S *Rev. Accel. Sci. Tech.* **2** 265 (2009)
26. Rossi L, in *Proc. 2nd Intern. Particle Accelerator Conf., IPAC 2011, 4 – 9 September 2011, Spain*, p. 908
27. Trubnikov G et al., in *12th Russian Particle Accelerator Conf. RuPAC-2010, 27 September – 1 October 2010*, p. 14
28. Klein M, in *Proc. 2nd Intern. Particle Accelerator Conf., IPAC 2011, 4 – 9 September 2011, Spain*, p. 1948
29. Ptitsyn V et al., in *Proc. 2nd Intern. Particle Accelerator Conf., IPAC 2011, 4 – 9 September 2011, Spain*, p. 3726
30. Ahmed S et al., in *Proc. 2011 IEEE Particle Accelerator Conf., March 28 – April 1, 2011, New York, USA*, p. 2306
31. Lehrach A et al. *J. Phys. Conf. Ser.* **295** 012156 (2011)
32. Biagini M E et al. (SuperB Collab.), arXiv:1009.6178
33. Abe T et al., KEK Report 2010-1 (Eds Z Doležal, S Uno) (Tsukuba: KEK, 2010); arXiv:1011.0352
34. Levichev E *Phys. Part. Nucl. Lett.* **5** 554 (2008)
35. "The High Energy Large Hadron Collider", Preprint CERN-2011-003 (Geneva: CERN, 2011)
36. Blondel A, Zimmermann F, arXiv:1112:2518
37. Ankenbrandt C M et al. (Muon Collider Collab.) *Phys. Rev. ST Accel. Beams* **2** 081001 (1999)
38. "ILC Reference Design Report", ILC-Report-2007-001; http://www.linearcollider.org
39. Delahaye J P *Mod. Phys. Lett. A* **26** 2997 (2011)
40. Geer S *Annu. Rev. Nucl. Part. Sci.* **59** 347 (2009)
41. Shiltsev V *Mod. Phys. Lett. A* **25** 567 (2010)
42. Apollonio M et al. (The ISS Accelerator Working Group) *JINST* **4** P07001 (2009)
43. Poklonskiy A A, Neuffer D *Int. J. Mod. Phys. A* **24** 959 (2009)
44. "2012 Global R&D Funding Forecast" *Batelle R&D Magazine* 16 December (2011); http://www.rdmag.com/Featured-Articles/2011/12/2012-Global-RD-Funding-Forecast/
45. Shiltsev V *Mod. Phys. Lett. A* **26** 761 (2011)
46. Cronin J W (Ed.) *Fermi Remembered* (Chicago: Univ. of Chicago Press, 2004)
47. Gai W *AIP Conf. Proc.* **1086** 3 (2009)
48. Bermel P et al. *ICFA Beam Dynamics Newsletter* (56) 91 (2011)
49. Leemans W *ICFA Beam Dynamics Newsletter* (56) 10 (2011)
50. Schroeder C B et al. *Phys. Rev. ST Accel. Beams* **13** 1013014 (2010)
51. Muggli P *AIP Conf. Proc.* **1299** 52 (2010)
52. Seryi A *Nucl. Instrum. Meth. Phys. Res. A* **623** 23 (2010)
53. Caldwell A, Lotov K, Pukhov A, Simon F *Nature Phys.* **5** 363 (2009)
54. Dodin I Y, Fisch N J *Phys. Plasmas* **15** 103105 (2008)
55. Tajima T, Cavenago M *Phys. Rev. Lett.* **59** 1440 (1987)
56. Chen P, Noble R J *AIP Conf. Proc.* **398** 273 (1997)
57. Telnov V I, arXiv:0908.3136
58. Schlenvoigt H-P et al. *Nature Phys.* **4** 130 (2008)
59. Kneip S et al. *Nature Phys.* **6** 980 (2010)