Tutorial 13: The future of the energy frontier

Dr. M Flowerdew

July 2, 2014

In this tutorial, we take a brief look into some of the future projects and developments that will define the high-energy frontier in the future. These large, complex projects take decades to develop and build, due in part to the technical difficulties, but also due to the cost, and the need for complex international negotiations to ensure the success of these projects. There is still some uncertainty over exactly which projects will ultimately be built, the final decisions of what to build depend in part on the LHC Higgs boson measurements and searches for BSM physics in the next few years. We will briefly discuss some of the more well-developed proposals, and along the way review some basic accelerator physics.

1 The LHC upgrades

The full exploitation of the potential of the LHC is a central part of the current European strategy for particle physics. Over the next decade, a series of upgrades are planned, culminating in the so-called *HL-LHC*[1], due to start in about 2023. The first long shutdown is currently in progress. The primary objective here is to complete the installation of safety systems on the LHC magnets that will allow them to accelerate protons up to the design energy of $6.5-7$ TeV, instead of $3.5-4$ TeV as in Run 1. This higher

[1]High-luminosity LHC

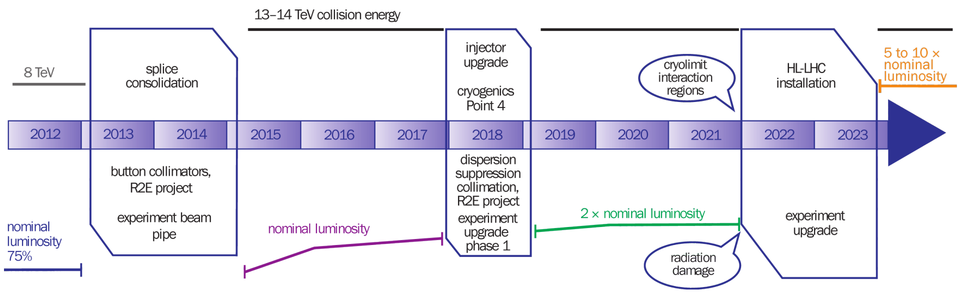

Figure 1: Planned timeline of the LHC and HL-LHC upgrades.

collision energy will eventually allow the physics of the TeV regime to be fully explored by ATLAS and CMS.

The second round of upgrades, expected to occur in 2018, will further upgrade the LHC magnet systems, to allow the instantaneous luminosity to be increased to double the nominal value of 10^{34} cm^{-2}s^{-1}. By the end of Run 3, as much as 300 fb^{-1} of data may be collected, compared to 20 fb^{-1} today.

Finally, in 2022-23, the HL-LHC will be installed. The main purpose here is to increase the instantaneous luminosity as much as possible, potentially up to 10 times the nominal value. This is needed because the statistical precision of LHC measurements (and searches) improve only as $1/\sqrt{L}$, if L is the integrated luminosity. So, to halve the statistical uncertainty after Run 3 would require a four-fold increase in L, or an additional run time of nine years, as Run 3 is planned to be three years long. With the HL-LHC, this increase will be achieved within about seven months. The ultimate target for the HL-LHC is to collect a few ab^{-1} of data by about 2030 (1 ab^{-1} = 1000 fb^{-1}).

During all of these shutdown periods, the LHC detectors are being repaired and upgraded, to cope with the increased hit rates and radiation damage of the upcoming run. The most extreme changes will occur for the HL-LHC, where, for example, ATLAS will need to install a completely new, all-silicon, inner detector.

Now, we will look at some parts of the physics case for these upgrades.

1.1 Prospects for LHC measurements and searches

The LHC upgrades are intended to provide a vast increase in the numbers of event candidates, with an important cost: *pile-up*. This refers to the number of pp collisions that occur simultaneously in each bunch-crossing, which could reach about 140 for the HL-LHC, compared to an average of about 20 in 2012. This has a number of effects, making identification of the primary vertex more difficult for example, but perhaps the most important effect is on the trigger rate, which will increase dramatically. It is possible to improve the algorithms used to select triggered events, and to increase the output bandwidth, however it is certain that some thresholds (usually in p_{T}) will need to be raised in each new run. These factors mean that not all LHC measurements will be improved by the upgrade plans. In particular, many precision measurements of SM phenomena are already systematically limited, and will not benefit from the increased statistics. Examples include most measurements of the top quarks, W and Z boson production and the W boson mass. In addition, searches for new particles with low masses (e.g. $\lesssim 200-300$ GeV) will become increasingly difficult as the trigger thresholds increase.

The main beneficiaries of the luminosity upgrades are measurements

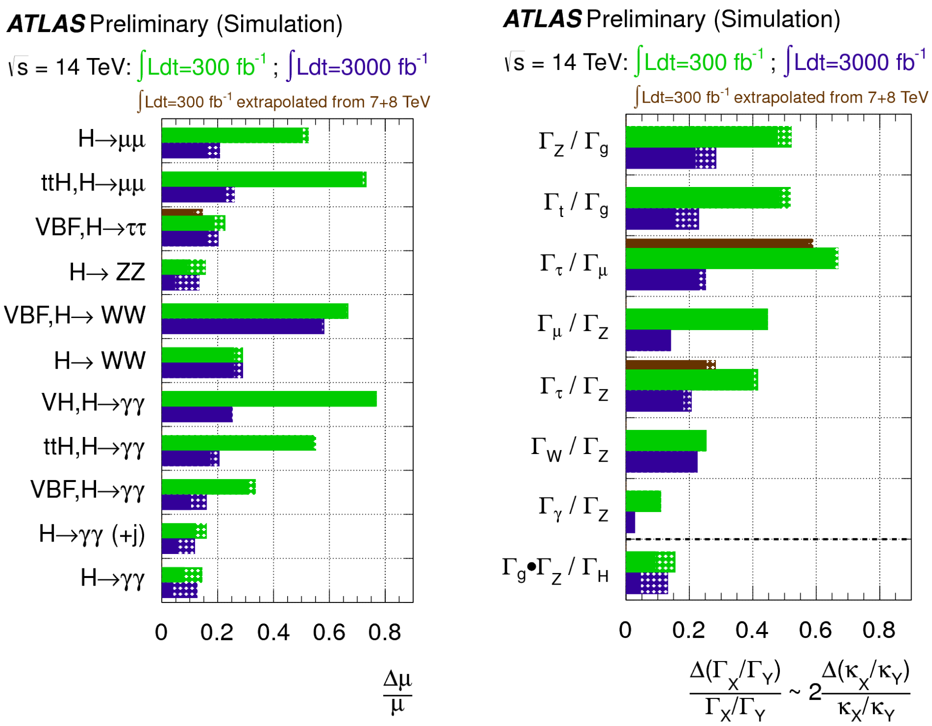

Figure 2: Expected uncertainties of Higgs boson measurements at the LHC and HL-LHC (300 fb^{-1} and 3000 fb^{-1}). Left: signal strength μ, right: various ratios of coupling constants. For comparison, some current uncertainties on μ with 20 fb^{-1} are: $\sim 20\%$ for $h \to \gamma\gamma$, $\sim 25\%$ for $h \to ZZ$ and $\sim 30\%$ for $h \to \tau\tau$ and $h \to WW$ (see tutorial 8).

and searches that are currently statistically limited. This includes nearly all studies of the Higgs boson. As made clear in tutorial 8, the decay modes of the Higgs boson that are easiest to detect tend to have small branching fractions, and therefore benefit greatly from more data. In addition, the increase in energy compared to Run 1 will increase the production cross-sections for associated modes (Vh and $t\bar{t}h$), where the associated object(s) can be used in the trigger, allowing more difficult decay modes such as $h \to b\bar{b}$ to be better explored.

Figure 2 shows the anticipated precision of Higgs boson measurements in about 2022 (300 fb^{-1}) and after the HL-LHC (3000 fb^{-1}). In addition to significant improvements in the precision of the discovery channels (see the figure caption), entirely new channels appear that have not been convincingly observed in current data. These include associated production with the decay $h \to \gamma\gamma$, and the anticipated observation of the decay $h \to \mu^+\mu^-$. In most cases, a precision of around 20% can ultimately be reached. These channels will allow the SM Higgs couplings to be over-constrained, and ra-

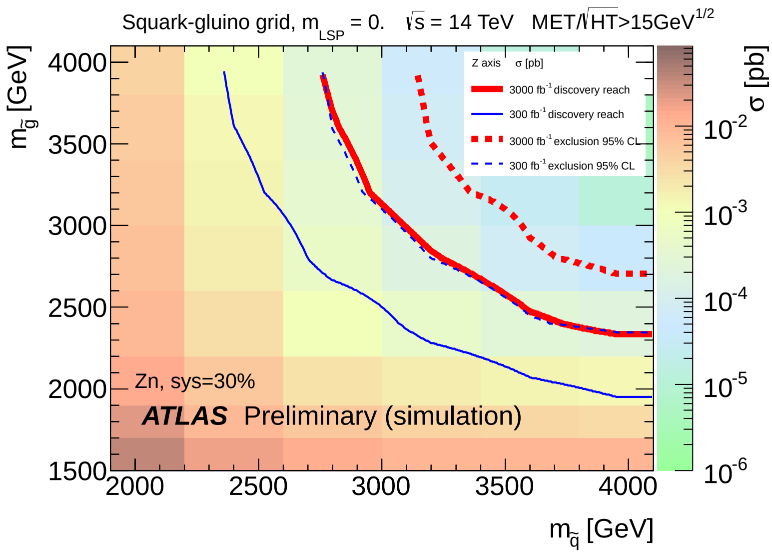

Figure 3: Expected mass reach in searches for squarks and gluinos at the LHC and HL-LHC (300 fb^{-1} and 3000 fb^{-1}). The solid lines show the 5σ discovery range, while the dotted lines show the exclusion reach at 95% confidence level. The current limits with 20 fb^{-1} are $m(\tilde{g}, \tilde{q}) \gtrsim 1400$ GeV (see tutorial 10).

tios of couplings to be extracted, as shown in Figure 2 (right). However, the problems with extracting the absolute couplings discussed in tutorial 9 still apply.

Searches for high-mass particles will also benefit greatly from the LHC upgrades. Again, the increase in beam energy for Run 2 plays a major role too. To see why this is the case, recall from tutorial 5 that the average value of Björken x required to produce a particle of mass m with a beam energy of E_b is approximately $m/(2E_b)$. The parton density functions fall rapidly at high values of x, and so an increase in E_b can have a dramatic effect on the production cross-section of massive particles. For example, the production rate for gluino pairs with a mass of 2.5 TeV increases by a factor of more than 2500 when $\sqrt{s}$ increases from 8 TeV to 13 TeV. For the same reason, higher integrated luminosities allow for a greater reach in mass, even when $\sqrt{s}$ is fixed.

Using supersymmetry as an example, Figure 3 shows the projected sensitivity to squark and gluino production, where these particles decay into jets and a neutralino. The current result in this channel was seen in tutorial 10. With 300 fb^{-1} of data, squarks and gluinos with masses of up to 2 TeV can potentially be discovered with 5σ significance, beyond the current exclusion of ~ 1440 GeV. If nothing is observed, the HL-LHC will have an exclusion reach of about 3 TeV. Even for top squark production alone, masses of up

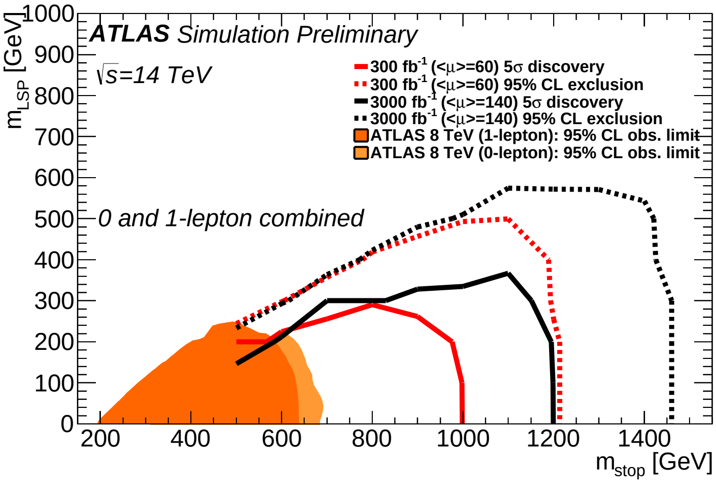

Figure 4: Similar to Figure 3, but for the top squark pair production, with $\tilde{t} \to t\tilde{\chi}_1^0$. The currently excluded region is shown in the lower left corner.

to 1.4 TeV could be excluded at 95% CL with the HL-LHC (Figure 4). Together, these results fully explore the mass range that is usually considered "natural" for a supersymmetric theory.

2 The ILC

To improve measurements of the Higgs boson[2] beyond the precision shown in Figure 2, a different kind of collider is needed. The precision of any measurement made at a hadron collider is fundamentally limited due to the uncertainty of the initial state, the large background from QCD processes and pile-up. In addition, if any non-SM processes are discovered at the LHC, the same features mean that it will be difficult to fully understand their nature at the LHC. Thus, there is a strong argument to produce a lepton collider to further study this particle, where the initial state is precisely known. The most well-developed concept for a lepton collider is the International Linear Collider, or ILC. However, most of what follows is relevant for any lepton collider, including the alternatives listed in Section 2.5.

The proposed layout of the ILC is shown in Figure 5. After the electrons and positrons are produced, they enter two damping rings, where they circulate with an energy of 5 GeV. The purpose of these is to reduce the emittance of the beams, ready for use in the collider (see Section 2.2). Once prepared, the beams enter the main linacs, which are each currently projected to be about 15 km long. On the return jouneys, the leptons are

[2]and anything else that the LHC may discover.

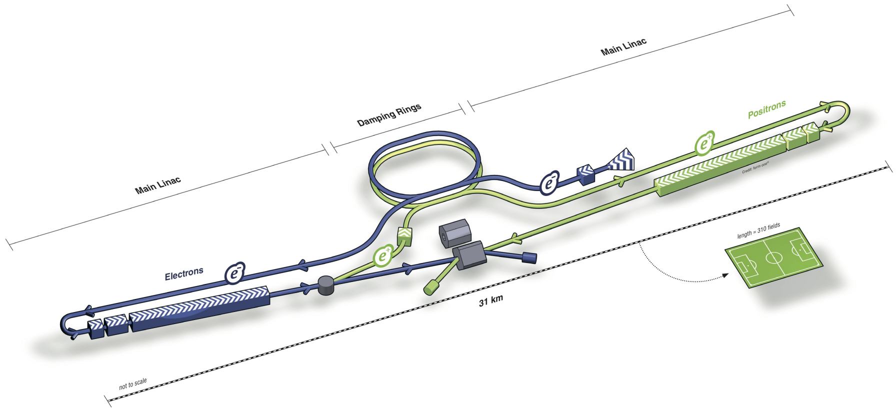

Figure 5: Proposed layout of the ILC, from their Technical Design Report.

accelerated, ready for collision at the interaction point, after which they are dumped. The positrons are created from the interactions in a Ti-alloy of $\lesssim 30$ MeV gamma rays, which in turn are created by passing the high-energy electron beam through a helical undulator before reaching the interaction point.

We will now use the ILC as an example to discuss some principles of accelerator physics.

2.1 Emittance and focussing

A key parameter of any particle beam is its emittance. This describes the range of deviations from the ideal path that the particles take. Usually, the properties of the beam are very different in the horizontal and vertical planes, and therefore the emittances in these two directions are considered separately. We shall call these directions x and y, respectively, with z being along the beam propagation direction. This defines a set of coordinates that have different orientations along the ideal beam line, but they are more convenient to use in this context than absolute coordinates.

If we consider just one direction, say x, then the emittance ϵ_x is the volume of phase space occupied by some specified fraction of the beam. The phase space is parameterised by x and $x' = \mathrm{d}x/\mathrm{d}s \approx \frac{1}{c}\mathrm{d}x/\mathrm{d}t$, where s parameterises the distance along the ideal beam line. Due to Liouville's theorem, this volume is conserved as it propagates through the accelerator.

At the point of production, the position and direction of the particles are essentially uncorrelated, meaning that the volume defining the emittance is an ellipse (as in Figure 6 (left)). Even though individual particles have complicated paths through the accelerator, we can understand the propagation of the beam as a whole by considering just this envelope. This is because particles in the beam that start inside the envelope cannot cross it – two

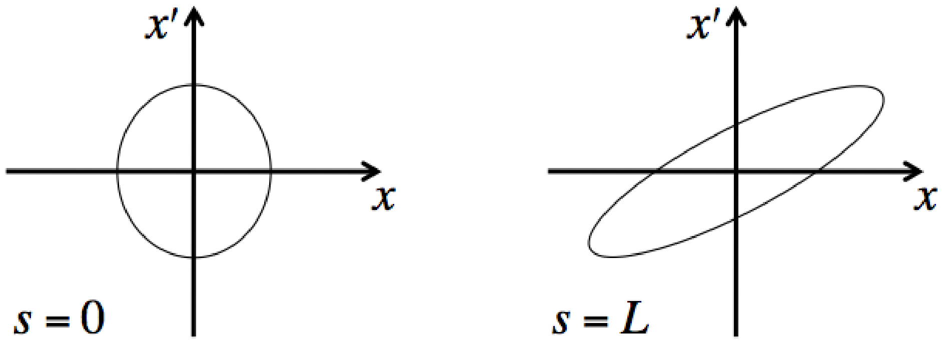

Figure 6: Sketch of a beam envelope in $x - x'$ space for a beam that freely propagates over a distance L. It is assumed that x and x' are uncorrelated at $s = 0$, for simplicity.

particles with the same position and velocity will experience the same force, and their future paths must be identical.

We begin by considering the free propagation of the particles within this initial ellipse. The regions of the ellipse with positive x' will migrate to higher values of x over time, and regions with negative x' will migrate to negative values of x. This is illustrated in Figure 6 (right). This is an intuitive result: the beam spreads out in x over time due to the initial spread in x velocities.

This effect can of course be reversed by focussing. If a magnetic field is arranged such that its magnitude in the y direction varies as $-gx$, where g is a constant, then the force experienced in the x direction by a particle with velocity $v \approx (0, 0, c)$ will be $-gcx$, i.e. a restoring force. This field acts like a lens with a focal length of p/egl, where p is the beam energy, and l is the length of the focussing magnet, reducing x' for parts of the ellipse with positive x, and vice versa for negative x. Figure 7 shows an example where the focussing exactly compensates for the increased beam width, effectively reversing the sign of x' with respect to the beam before focussing. After propagating for another distance L, the original shape of the beam from Figure 6 (left) can be recovered. Thus, with repeated focussing, the overall beam size in x can be maintained.

The field required to focus in the x direction can be achieved with a quadrupole magnet. However, the full quadrupole field in the $x - y$ plane has components $B = (-gy, -gx, 0)$, giving a total force on the particle of $F = v \times B = (-gvx, gvy, 0)$. This will defocus the beam in the y direction, increasing its divergence. Fortunately, it is possible to achieve focussing in both directions by alternating quadrupole magnets that focus in x and y. To see why this works, recall that the focal length f of two lenses with focal

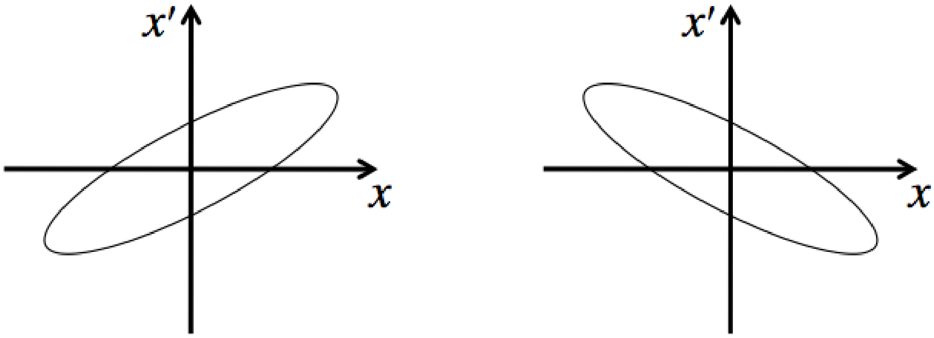

Figure 7: Sketch of a beam envelope in $x - x'$ space before (left) and after (right) focussing in the x direction. It is assumed that the focussing is perfectly tuned to the beam, i.e. that $x' \to -x'$ for all particles in the beam.

lengths f_1 and f_2 and separated by a distance d is given by

$$\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2} - \frac{d}{f_1 f_2}. \tag{1}$$

If f_1 and f_2 have opposite sign, then f can always be made positive (and therefore focussing) as long as d is sufficiently large. In particular, if $f_2 = -f_1$, then f is always positive.

The instantaneous luminosity of a collider ultimately depends upon the physical size of the beam at the interaction point. To achieve best results, this should be as small as possible, and this can be achieved by placing strongly focussing magnets close to the interaction point. To see why this is the case, consider what would happen in Figure 7 with a stronger focussing magnet. The emittance ellipse after the magnet would stretch up and down to more extreme values of x'. Due to Liouville's theorem, the width of the ellipse would shrink, and after a short propagation time the beam size in x would be much smaller than it started. This is where the interaction point should be located. However, high values of x' correspond to particles travelling at large angles with respect to the ideal beam line, and so the beam will quickly diverge again, requiring more focussing to avoid losses from collisions with the beam pipe wall. Therefore, the final focussing magnets should be as close to the interaction point as possible, to increase the maximum tolerable beam divergence. In addition, it is desirable to have the overall emittance as low as possible, which allows for a smaller beam size for a given maximal divergence. This is achieved through beam cooling.

2.2 Beam cooling

The restrictions of Liouville's theorem only apply to a closed system that does not exchange energy with its surroundings. If either of these assumptions are broken, then it is possible to alter the emittance of the beam. Even the act of accelerating the beam (see Section 2.3) reduces the transverse emittance as defined, because p_z increases while p_x and p_y remain the same. Therefore, $x' = p_x/p_z$ is reduced, and similarly for p_y. However, cooling is usually understood to mean a reduction in the normalised emittance $\gamma\epsilon_{x,y}$.

One of the simplest ways to reduce the (normalised) emittance of a beam is to wait until it is spatially extended (as in Figure 6) and insert a beam stop restricting its width. Due to the correlation between x and x' in that case, this will also reduce a substantial fraction of the particles that contribute most to the beam divergence. This technique is most useful in the early stages of beam production, where the particle energy is relatively small.

The primary method for cooling high-energy beams is in a damping ring. For a circular collider, the collider ring itself can act as a damping ring, while for the ILC the damping ring would be separate from the main linac, as shown in Figure 5. As the particles circulate around the ring, they lose energy to synchrotron radiation. This reduces all components of the particles' momenta, while acceleration to maintain a constant beam energy only increases p_z. Thus, over time, the beam divergence decreases. This damping is more efficient in the y (vertical) direction than x, as the radiation is emitted mainly in the horizontal plane.

The energy lost per particle per revolution due to synchrotron radiation scales as γ^4/r, where r is the radius of curvature, assuming $\beta \approx 1$. Thus, electron and positron beams can be cooled very effectively, even with low beam energies (5 GeV in the proposed ILC damping ring). It is this energy loss that ultimately limits the beam energy that a circular e^+e^- collider can sustain. The damping time for (anti)protons is much longer for the same accelerator parameters, although at the LHC the high beam energy makes these losses non-negligible. However, *stochastic cooling* is often used in hadron colliders, to accelerate the cooling process. This uses readings taken of the beam in one part of the ring to correct the beam profile in another part of the ring, which is possible because the straight-line distance between the two points is shorter than the path taken by the beam. It is best if corrections can be applied to parts of a bunch, rather than the whole bunch, and so typically the beam is stretched in z before the corrections are applied (and then compressed again after correction). Over time, the average deviations from the ideal beam line can be reduced, cooling the beam.

Figure 8: Example RF cavity proposed for the ILC acceleration stage.

2.3 RF cavity acceleration

After damping, it is planned to accelerate the ILC beams in the main linacs using radio-frequency (RF) cavities like the one illustrated in Figure 8. As a particle bunch passes through each cell in the cavity, it is accelerated by the oscillating electric field, which temporarily points in the correct direction to accelerate the bunch.

Exercise: Why can a static electric field not be used to accelerate particles to high energies?

The precise shape of the ILC resonator is highly optimised to produce the best field properties for accelerating the ILC particle bunches. We will instead consider a simpler variation called a pill-box cavity, shown in Figure 9. In this case, the resonant volume is a simple cylinder. This works as an accelerator because there are solutions of Maxwell's equations where the electric field points purely along z, while the magnetic field is circular in ϕ, as shown. The boundary conditions at the cavity walls mean that the electric field must vanish at $\rho = a$, where ρ is the radius variable in cylindrical coordinates. In this case, both electric and magnetic fields are described by Bessel functions, and for the lowest frequency mode the maximum electric field strength is found along the axis of the cylinder. The resonant frequency is fixed by the cavity's radius, and is $f = 2.405c/2\pi a$. This in turn determines the ideal distance between adjacent cavities, $d = \pi a/2.405$.[3] such that successive cavities accelerate the particles constructively.

For efficient particle acceleration with minimal energy consumption, it is beneficial to have a high quality resonator. Resistance in the cavity walls leads to losses, and for this reason the planned ILC cavities will be made of superconducting materials. In this way, field gradients of about 30 MV m^{-1} can be achieved.

Perhaps surprisingly, best results occur if the bunches arrive just before the maximum of the oscillation in the electric field, rather than at the maximum. This can be understood using Figure 10. Point S is defined as the ideal time of arrival for a bunch to be accelerated, where it will just reach the next cavity at the same point in its oscillation. A particle travelling slightly faster than the average will arrive early, perhaps at point P. In

[3] Assuming $v = c$.

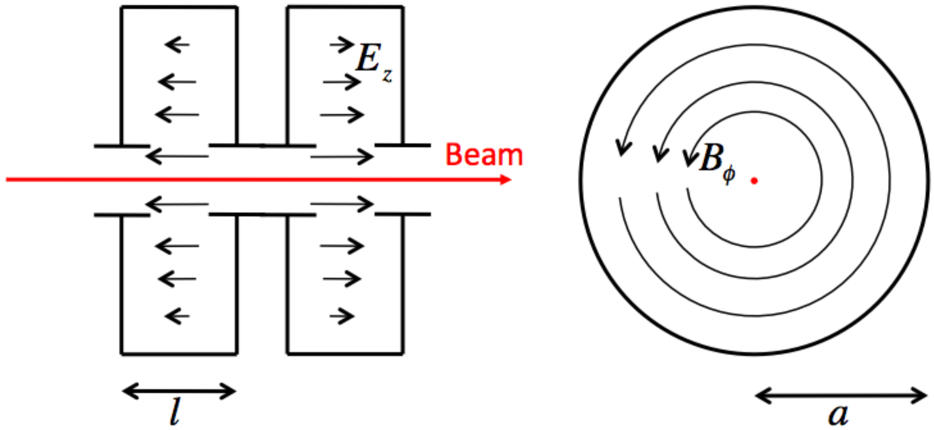

Figure 9: Schematic of a simple "pill box" RF accelerator cavity. Projections parallel and perpendicular to the beam are shown, together with arrows indicating the directions of the electric and magnetic field inside the cavity. With the electric field in this configuration, the particle bunch should be in the right-hand cavity.

this case, it experiences a lower electric field, and will be accelerated less in this cycle. Conversely, a late particle (at P') will be accelerated more than average. These lead to small oscillations around S as particles gain and lose energy in different RF cavities. Point U is a point of unstable equilibrium, which marks the divide between two successive stable equilibria. However, particles arriving later than U (but still within this cycle) will at some point be *de*celerated by the negative electric field in the other half of the cycle. If S were placed at the peak of the oscillation, this would happen to any particle arriving late, which is clearly undesirable.

2.4 Physics at the ILC

Due to the small Yukawa coupling of electrons, direct s-channel Higgs production is not detectable at e^+e^- colliders. Instead, for collision energies below $\sqrt{s} \sim 450$ GeV, the dominant Higgs production mode is associated production with a Z boson, as shown in Figure 11, with a maximum cross-section of about 300 fb at $\sqrt{s} \approx 250$ GeV. Above 450 GeV, $e^+e^- \to \nu_e\bar{\nu}_e h$ dominates, mediated by t-channel W boson exchange (and called WW fusion).

Exercise: Draw the leading order diagram for WW fusion.

This production mechanism gives a unique handle on the Yukawa coupling of the Z boson, irrespective of how the Higgs boson decays. Due to conser-

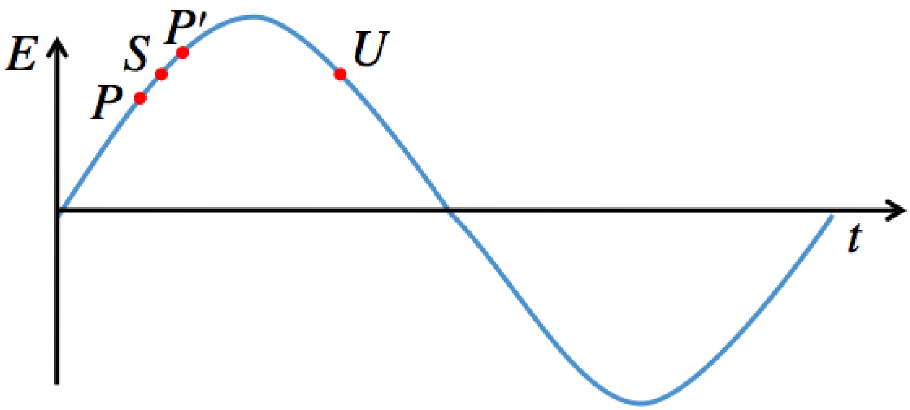

Figure 10: Sketch of the electric field at the centre of an RF cavity, as a function of time. The points indicate different times at which particles to be accelerated may pass through the cavity. S and U show stable and unstable equilibrium positions, respectively, while particles passing through at P and P' are both pushed towards S.

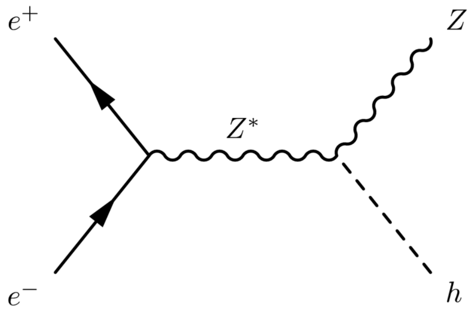

Figure 11: Diagram for Zh production at the ILC.

vation of momentum, we can write

$$p_{e^+} + p_{e^-} = \begin{pmatrix} \sqrt{s} \\ 0 \\ 0 \\ 0 \end{pmatrix} = p_Z + p_h. \tag{2}$$

If the Z boson decays leptonically, three out of the four momenta are known, allowing all four components of p_h to be precisely determined. The peak in the invariant mass distribution $\sqrt{p_h^2}$ can be used to measure the Zh production cross-section, which scales as g_Z^2 (using the notation from tutorial 8). In addition, number of detected $h \to ZZ^*$ decays, compared to the total Zh cross-section, will allow the ZZ branching fraction to be measured, which is proportional to $g_Z^2/\sum_i g_i^2$. Combining these two measurements allows the total Higgs decay width to be measured, as $\sigma(Zh)/\mathrm{BR}(h \to ZZ) \propto \sum_i g_i^2 \propto \Gamma_h$. With this information, it is possible to determine the absolute couplings involved in all Higgs boson decay modes, from the measurements of their rates (and also with LHC information on their ratios), in many cases to a precision of a few percent. With beam energies of up to 500 GeV, it will even be possible to do this for the top quark, via the $t\bar{t}h$ production channel.

In addition, double-Higgs production will be accessible at the ILC. The cross-section is small, but unlike the LHC, it is possible to use the $b\bar{b}$ decay mode to aid detection. The Higgs bosons are again produced in association with Z bosons, and using this channel it should be possible to extract the self-coupling of the Higgs boson.

The clean environment of a lepton collider allows for precision measurements of the top quark, which has so far been produced only at hadron colliders. In particular, many ambiguities in top quark reconstruction (e.g. from pile-up jets) are absent, and this should allow the top quark mass to be measured with a precision of about 100 MeV (compared to 760 MeV from the latest Tevatron+LHC combination).

Finally, the ILC (or another lepton collider) could have an important role to play in non-SM physics. If the LHC discovers non-SM particles, and they are within reach of the ILC, they can be characterised more precisely than before. Using supersymmtry as an example, the ILC would be able to determine the precise masses and compositions of different SUSY particles within its reach. Also, with precise tracking (helped by a lack of pile-up), particle lifetimes could be measured down to 10^{-5} ns. The ILC could also discover particles that were missed at the LHC, such as charginos that are nearly degenerate with their associated neutralinos. In this case the soft pion from the decay $\tilde{\chi}_1^{\pm} \to \pi^{\pm}\tilde{\chi}_1^0$ may go unnoticed at the LHC, but could be reconstructed in a lepton collider.

2.5 Alternatives to the ILC

If new particles discovered at the LHC have masses higher than about 500 GeV, then the ILC will not be able to produce them directly. In this case, another proposal, the Compact Linear Collider (CLIC), could be used to explore higher energies, up to about 2.5 TeV. This would be achieved by using a high intensity drive beam to provide the RF power to accelerate the main particle beam, essentially working as a transformer to transfer power from one beam to the other. The field gradients produced in this way (for the main beam) would be much higher than in a conventional RF cavity ($\sim$ 100 MV m^{-1}), allowing higher beam energies to be produced. Ongoing work on using plasma wakefields to perform a similar function could ultimately produce accelerating fields of 50 GV m^{-1} or more, which could allow even higher energies to the explored.

Despite the apparent disadvantages, circular lepton colliders are also being discussed. With hindsight, we know now that LEP only narrowly missed out on seeing the Higgs boson (it had sensitivity up to 115 GeV). With a slightly increased beam energy, a circular e^+e^- collider could, in principle, produce a large sample of Higgs bosons to study. The limiting factor is the synchrotron radiation emitted by the electrons and positrons as they circulate around the ring, which scales as E^4. At its highest beam energies, LEP consumed several MW of power, and while there is a proposal to re-use the LEP/LHC ring for an e^+e^- collider (called LEP3), the synchrotron energy loss would severely limit the instantaneous luminosities that could be obtained[4].

Exercise: Estimate roughly by how much would the energy consumed by LEP3 would be larger than that of LEP2 at its highest energy, assuming that the beam current is unchanged.

Synchrotron radiation would also be reduced by increasing the radius r of the ring. CERN is currently investigating the feasibility of an $80-100$ km circumference collider ring. Due to its size, this would encircle Geneva, passing right under Lac Léman. This tunnel, like the LEP/LHC tunnel, could initially hold an e^+e^- collider, and then progress to hadron physics, colliding protons with energies of up to $\sqrt{s} \sim 80-100$ TeV.[5] The lepton collider option would be limited in its energy reach, up to about $\sqrt{s} \approx 400$ GeV, but the circular design would allow extremely large samples of Higgs bosons and top quarks to be collected.

Muons have a long enough lifetime that they can be accelerated to highly relativistic energies before they decay. This raises the possibility of building a muon collider. This would allow all of the same measurements as an e^+e^- collider (with the same beam energy and integrated luminosity), but in addition allow direct Higgs boson production via $\mu^+\mu^- \to h$. Also, as many extensions to the SM predict couplings and/or masses that depend on the generation number, colliding muons could give unique opportunities to discover particles that are produced less often in collisions of first-generation particles. The main challenge here is one of cooling – all of the processes described in Section 2.2 take time, whereas a muon beam must be cooled within a fraction of a second. Research is ongoing into ways of cooling a muon beam. For example, ionisation cooling works on the same principle as a damping ring, except that the energy loss is achieved by passing the beam through a block of absorber material, rather than through synchrotron radiation.

[4]In addition, it could not be built until the LHC ceases operations.

[5]There is also an *ep* proposal, like a higher energy version of the HERA collider.