

1 A theorist's view of collider phenomenology

The principal questions are: For any given model of new physics,

- What are the discovery signatures?
- What is the discovery reach of collider X?
- How do we distinguish this model from the rest?

In order to answer these questions, we follow the recipe.

2 Recipe

2.1 Define the theory

In order to define the theoretical framework, we need to specify:

- The new particles and their quantum numbers (spin, charge, color, etc.)
- Their interactions *at the relevant energy scale*, i.e. the Lagrangian describing the new physics, with parameters given at the collider energy scale. Often the new physics input parameters are given at higher energies, e.g. the GUT scale or the Planck scale, in which case we need to RGE evolve them down to the electroweak scale. But sometimes we need to RGE evolve *up*, as in the case of the fermion Yukawa couplings. The set of new physics input parameters forms the “parameter space” of the new physics model.
- We also need to identify potentially interesting regions of the parameter space. For example, certain regions may already be ruled out from direct searches, from indirect constraints (precision electroweak data, flavor-changing neutral current processes) or simply because of theory bias, e.g. they do not have a good dark matter candidate or are fine-tuned.
- Given the Lagrangian, we can derive the Feynman rules for the new physics model and do some preliminary computations of relevant particle properties, for example the mass spectrum of the new particles (at tree-level or one-loop, as appropriate) and their widths (lifetimes).

2.2 Production cross-sections of new particles

Next we have to compute the production cross-sections of the new particles, as they determine the expected total number of events in which new particles are present:

$$N_{total} = \sigma_S L, \tag{1}$$

where σ_S is the cross-section for a particular final state, and L is the total integrated luminosity of the data collected in the collider experiment. For example, during the Tevatron Run I each experiment collected about 110 pb^{-1} , while in Run II there is hope for something like 8 fb^{-1} . In the initial stages of the LHC the data taking rate will be 10 fb^{-1} per year.

If N_{total} is too small, we may already have to give up at this step. How small is small? That depends on the type of collider, as well as the size of the backgrounds (see below Sec. 2.6 and 2.7).

2.3 Branching fractions of new particles

Having figured out how many new particles were made, we still have to decide on the specific method to discover them, i.e. on the discovery signature. This is because the new particles will typically have several possible decay modes and we have to consider each one separately. A Z -boson, for example, can decay to

- two quarks (jets) with a branching fraction $\sim 70\%$;
- two neutrinos (missing energy) ($BR \sim 20\%$);
- e^+e^- pair ($BR \sim 3.33\%$);
- $\mu^+\mu^-$ pair ($BR \sim 3.33\%$);
- $\tau^+\tau^-$ pair ($BR \sim 3.33\%$).

Each one of these decays will lead to a characteristic Z -boson signature. Apriori we do not know which one is best for discovery, we can only determine this later, once we have considered the relevant backgrounds (see Secs. 2.6 and 2.7). Another example is the W -boson which can decay to

- two jets ($BR \sim 2/3$);
- an electron plus missing energy ($BR \sim 1/9$);
- a muon plus missing energy ($BR \sim 1/9$);
- a τ plus missing energy ($BR \sim 1/9$).

A third example is top pair production at the Tevatron. Each top quark decays to a W boson and a b -quark, so a top pair gives $W^+W^-b\bar{b}$. Depending on the subsequent decays of the two W 's, we have a ‘‘dilepton’’, a ‘‘semileptonic’’, and a ‘‘hadronic’’ top sample.

In order to get the total number of events in which a particular signature is realized, we have to multiply (1) by the respective branching fractions. Consider, for example, Higgs boson production via the process $f\bar{f} \rightarrow Zh$, which is relevant at both LEP, Tevatron and LHC (Fig. 1). Choosing the decays $Z \rightarrow \ell^+\ell^-$ and $h \rightarrow b\bar{b}$, we have a signature of 2 opposite sign leptons (2OSL), and 2 b -jets (2B)

$$f\bar{f} \rightarrow Zh \rightarrow \ell^+\ell^-b\bar{b}. \quad (2)$$

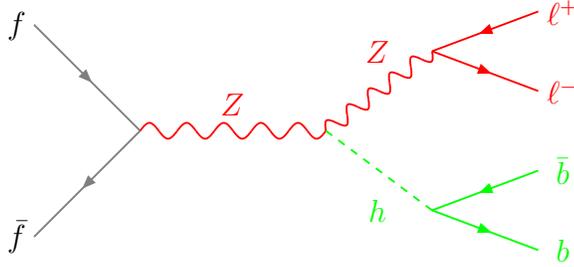


Figure 1: $f\bar{f} \rightarrow Zh$ production diagram.

The number of signal events with this signature will be

$$N_{signature} = \sigma_S L B(Z \rightarrow \ell^+\ell^-) B(h \rightarrow b\bar{b}). \quad (3)$$

At this step, we may also be forced to give up, if $N_{signature}$ is too small. Often there are cases where the production cross-section σ_S is large, and N_{total} is significant, but the branching fractions are very small and lead to uninteresting values for $N_{signature}$.

2.4 Detector acceptance

Having produced enough events $N_{signature}$ of a given signature, we now need to ask, how often did the detector correctly recognize these events as such, i.e. how often did our particles pass particle ID cuts. We can define an ‘‘acceptance’’

$$\epsilon_a \equiv \frac{N_{reconstructed}}{N_{signature}} \quad (4)$$

which tells us what fraction of the original $N_{signature}$ events were correctly identified by the detector. In our example of Zh production in (3), we will require 2OSL and 2B in the detector. Accounting for the detector effects, we therefore have

$$N_{reconstructed} = \sigma_S L B(Z \rightarrow \ell^+\ell^-) B(h \rightarrow b\bar{b}) \epsilon_a. \quad (5)$$

ϵ_a is typically a number of order 0.1 – 0.5, i.e. the detectors perform rather well.

2.5 Cuts

At this point we know how many signal events with the signature (2) are expected to be *seen* in the detector. Unfortunately, the actual observed number of events for this signature will also include events due to all possible backgrounds (a few examples are shown in Fig. 2). How do we separate signal events due to Fig. 1 from background events due to Fig. 2 and the like? This is where cuts come in. Those are conditions which are satisfied for the majority of the signal events, and false for the majority of background events, or at least that is the hope.

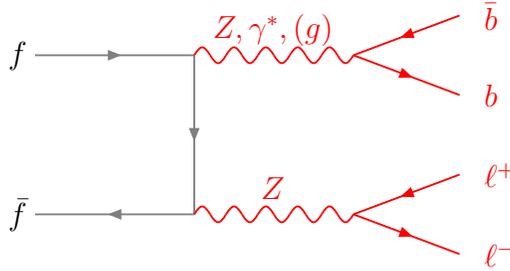


Figure 2: An example of a background to the signature (2).

In our case, the two b -jets from the Higgs decay will have an invariant mass close to m_h while the b 's due to Z , γ^* or g exchange will generally have lower invariant mass.

We can define the “efficiency” of the cuts for the signal as

$$\epsilon_c \equiv \frac{N_S}{N_{reconstructed}}, \quad (6)$$

where N_S is the number of signal events which ultimately pass the cuts and enter the final plots. It is clear that

$$N_S = \sigma_S L B(Z \rightarrow \ell^+ \ell^-) B(h \rightarrow b\bar{b}) \epsilon_a \epsilon_c. \quad (7)$$

Sometimes what is called an “efficiency” is the product of the efficiency defined in (6) and the acceptance:

$$\epsilon_a \epsilon_c = \frac{N_S}{N_{signature}}.$$

The definition is usually clear from the context.

2.6 Backgrounds

It is clear from the discussion above that the proper choice of cuts cannot be done without simulating all relevant backgrounds. So in fact already at the previous steps one would have to simulate the backgrounds, run them through the detector simulation and see the effect of the cuts. If the desired degree of separation of signal from background is not reached, one has to go back and redesign the cuts, possibly even tweak the identification requirements in Sec. 2.4.

Once the final cuts are selected, one obtains for the total number of background events after cuts an expression similar to (7)

$$N_B = \sum_i \sigma_B^i L B^i \epsilon_a^i \epsilon_c^i, \quad (8)$$

where the sum runs over all relevant background processes and B^i is a shorthand for the product of relevant branching fractions.

2.7 Reach

The reach analysis involves comparing signal to background. The significance S is

$$S = \frac{N_S}{\sqrt{N_B}} \quad (9)$$

and 3σ (5σ) implies $S = 3$ ($S = 5$). A high significance is not necessarily enough for a good reach, as we may not have enough statistics, so often in addition to (9) one also requires

$$N_S > 5. \quad (10)$$

There are three types of variables at play in (9):

- The significance S .
- The total integrated luminosity L (which depends on the particular collider).
- The new physics parameters (they enter through N_S).

Most often the results for the reach are presented versus a model parameter, and for a fixed L , as in Fig. 3, or for a fixed significance, as in Fig. 4.

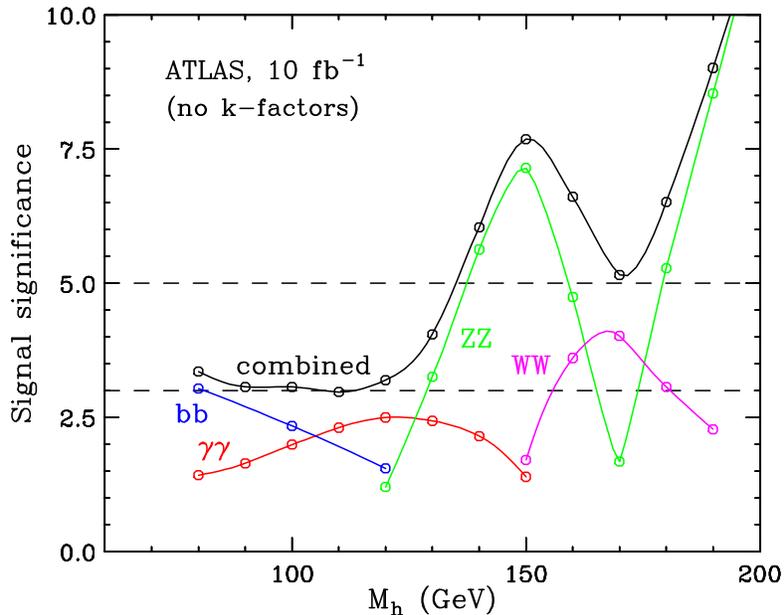


Figure 3: *LHC discovery potential for a light Standard Model Higgs boson. The curves are labelled by the Higgs decay mode used. Next-to-leading order corrections to both signal and background have been ignored.*

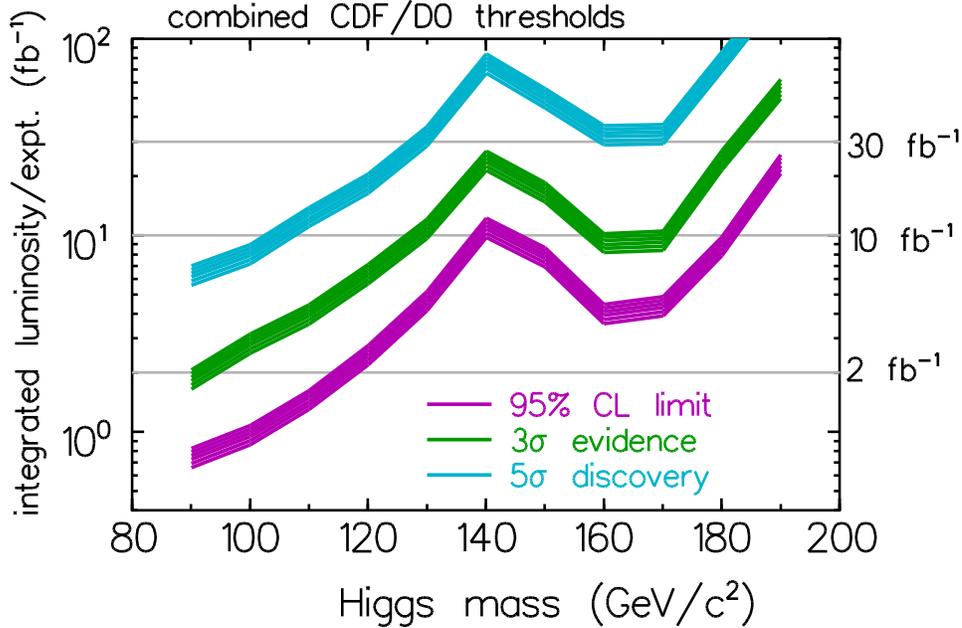


Figure 4: *Tevatron discovery potential for a Standard Model Higgs boson in run II. Shown are the integrated luminosities delivered per experiment required to exclude the Standard Model Higgs boson at 95% CL, observe it at the 3σ level or discover it at the 5σ level, as a function of the Higgs mass. In the low-mass Higgs region, below 140 GeV, the curves shown are the result of combining the Wh and Zh channels (where $h \rightarrow b\bar{b}$ and the W and Z decay leptonically), using the neural network selection, and the statistical power of both experiments. In the high-mass Higgs region, above 140 GeV, the curves shown are the result of combining various channels in which the Higgs boson decays to WW (where one W may be virtual). The lower edge of the bands is the calculated threshold; the bands extend upward from these nominal thresholds by 30% as an indication of the uncertainties in b -tagging efficiency, background rate, mass resolution, and other effects. (From the Tevatron Run II workshop.)*