Today Lecture

• Monte Carlo Event Generators
Early Days Analyses

- The self-evident discoveries of the early days are over
  - Not that discoveries were easy back then, but the analysis challenges were different
- A typical modern analysis looks for a small number of signal events over piles of similar-looking background events
  - Precision understanding of both the expected signal and background event properties is critical

A plot from a PRL announcing the discovery of what we now call J/ψ.
Modern Analyses

- Can miss a real signal or misidentify a background for signal due to small mistakes in the knowledge of:
  - Topology and detailed physics properties of signal and background events
  - Response of the detector in measuring properties of signal or (one of the) background events
- The example on the right is what most people think is a result of some unaccounted systematic effect by the CDF experiment
  - E.g. either a relatively small shift in the measured energy of jets or in the modeling of the properties of W+jets events could cause something like the observed discrepancy

A plot of invariant mass of two jets in $W+2\text{jets}$ topology from CDF
Top plot: is data minus background taken from the bottom plot
• Further discoveries in particle physics are likely to be made by finding small “deviations” from the “expected”
  • Need to know the “expected” to the precision smaller than the size of the “deviation”
    • Or you won’t know if it’s a deviation you are seeing or just a lack of your knowledge in the “expected”
    • Can easily mean having to know the expected to a fraction of a percent level
      • Both the rate and the shape

• Processes that contribute to the “expected” at hadron colliders are usually QCD processes
  • Just because the x-sections are large
  • Even a W or photon production relies heavily on QCD calculations (see diagrams)
    • As we know, those are not easy (all sorts of divergences, some things are hard to calculate)
Comparisons with the “Expected”

- If one uses elaborate selections using many variables in filtering events, correlations are important
  - E.g. a plot of the cross-section versus a single variable that theorists can give us will not work
  - Experimentalists need to know event properties versus (or cross-section as a function of) “all variables”
- The easiest way to pass this kind of information is to teach experimentalists to “simulate” events of interest using an “event generator”
  - Written by theorists, so they can continue improving it, e.g. include higher order corrections
  - Used by experimentalists to learn how events produced in various processes look like in designing their analyses
  - Turned out to be a very productive model as experimentalists kept finding deficiencies in describing various processes, reporting them back to theorists who in turn have been learning something new and also improving these event generators
Why Monte Carlo

• In reality what we are calculating is an integral that includes a convolution of:
  • PDFs
  • Hard scattering cross-section (high $Q^2$) and form-factors/contributions accounting for soft (low $Q^2$) emissions (initial state radiation for emissions before hard scattering and final state radiation / fragmentation for “after”) 
  • Hadronization

• This integral is very complex and has peculiar integration borders
  • Monte Carlo is very suitable for this case as it is easy, fast and one can code as much complexity into the function being integrated
Event Generators

• There are many, some are getting less used, new ones keep coming up

• Some of the old and very popular “multi-purpose” event generators:
  • Pythia, Herwig
    • Multi-purpose you set the type of process and they do “everything” for you:
      • Properly use parton distribution functions (if you use hadrons in initial state)
      • Calculate the hard scattering diagram using perturbative QCD (or electroweak)
      • Add non-perturbative contributions (e.g. initial state radiation)
      • Take “final state partons” and emulate fragmentation process to turn these partons into hadrons
      • Apply necessary boosts etc.

• The payback for multi-purposeness is occasional imperfect description of some observables (more on this later)

• Generally one needs to know the limitations of the generator you use
Simulation’s paradigm

Basic strategy

Divide event into stages, separated by different scales.

- **Signal/background:**
  - Exact matrix elements.

- **QCD-Bremsstrahlung:**
  - Parton showers (also in initial state).

- **Multiple interactions:**
  - Beyond factorization: Modeling.

- **Hadronization:**
  - Non-perturbative QCD: Modeling.
Hard Scattering

- Hard scattering in general purpose generators is usually calculated at lower orders (i.e. LO)
  - Otherwise there are some substantial complexities in separating soft and hard parts which can yield incorrect results
    - The way you separate hard scatter and soft emissions is unique to a specific order, if you use a single standard algorithm to describe initial and final state radiation, the way you calculate hard scatter should be compatible with that
  - Generators would use a known LO calculation and the initial event weight is essentially the cross-section
    - One sometimes needs to be a bit careful with selecting appropriate scales, but that’s pretty much it

- Examples:
  - qq-bar -> qq-bar, qq-bar->Z, qq->Z+q etc.
Final State Evolution

• Once final partons are available, apply forward evolution to make Final State Radiation (FSR), use Next-to-Leading-Log (or MLLA) to re-sum multiple emissions
  • Here you will create softer secondary jets and create parton showers
  • Implementation details slightly differ (angular ordering, \( kT \) ordering, by hand/by construction etc.) but the results are very similar

• Hadronization models:
  • Strings in Pythia: create a color string, which is then “broken” to make real hadrons
  • Clusters in Herwig: combine nearby partons into colorless clusters, turn them into final state hadrons
    • Both include various “by hand” tunings based on a lot of comparisons with data, so that by now they reproduce data very well

• At the end you have all final particles turned into hadrons
  • But we haven’t yet connected the incoming partons in the hard scatter with the initial hadrons
The two initially alternative approaches seem to have converged to something similar

- Still “imperfect science” (can’t hope QCD to just work there as the process is way too soft), but works very well in practice
Initial State Radiation

• Here they usually use “backward evolution” (using the same Sudakov form-factors approach, just backwards) to create emissions that happened before the hard scatter
  • May seem strange, but doing this in forward direction re-connecting PDFs and the hard scattering would be very inefficient
  • The expression one would write for the probability of multiple successive emissions is easy to reverse computationally, so it’s just a trick that is a matter of convenience
• This stage creates additional (typically soft) emissions that could have happened off the incoming partons
• Next you walk back to the partons before these emissions, these you would consider to come from inside the hadrons you have been colliding
  • So these will fall under what you call PDFs
Parton Distribution Functions

- PDFs have to be in-sync with the order in which the cross-section is calculated
  - So one would usually use one of the LO PDF sets
    - CTEQ, MRST, CT10 etc.
  - Mixing orders is like either removing some types of emissions altogether or double counting
- Convolute everything we have constructed so far with the PDFs for the incoming hadrons
  - Each step adds a new multiplicative weight to the “event”
  - One could keep track of these weights, but an easier way is to “throw” a coin every time and go with the probability
    - Once you generate many trials, the sum of all your “events” will be the integral you are calculating
Underlying Event

• Hadron-hadron collisions always have the “main process” and a “shadow event”
  • It is definitely associated with color “bleaching”, and this “shadow” part tends to be universal (changes little with the properties of the “main collision”)
• Two approaches: “add-on” soft component (fairly empirical, no good explanation why) or multiple interactions (other partons within the same hadron interact in addition to the ones forming hard scattering process)
  • In both cases a lot of additional tuning to make data and simulation agree
Running Pythia or Herwig

• Easy! You set input parameters in the card file using one of the allowed options (from the manual):
  • Process name (like pp->Z+jets), beam energies, which PDF set you want to use (or it will pick some default)
• And run (usually you will want to make a pre-determined number of events like 1,000 or 100,000 etc.)
• The result is a file that contains events of the type (e.g. W production) you requested
  • In each event some specific “outcome” (momentum of the Z boson, it’s direction, what it decayed into etc.) has happened according to the weights that were calculated, so this is just like it happens in real life where outcomes follow QM probabilities
  • If you know how to read these files, you can learn exactly what happened in each “event” and can even build a “feynman-like” diagram tracing which particles underwent decay, which emitted new particles etc.
    • With Pythia one will get somewhat stuck in connecting hadrons and final partons as they are separated by the hadronization “string” which makes mother-daughter relationship a little ambiguous
    • Herwig with its clustering model allows a more direct tracing of mother-daughter relationships
Next Time

- Some of the “specialized” generators
  - MadGraph, AlpGen
- Detector simulation methods
  - GEANT

- This week lab: we will run Pythia and decode the output