CDF/PHYS/MONTECARLO/PUBLIC/5938
April 27, 2002

The Merits of Stand-Alone Monte Carlo Generation

Henry Frisch
Enrico Fermi Institute, University of Chicago

Talk at the CDF Monte Carlo Workshop, April 25, 2002

Abstract

CDF tradition has been to generate Monte Carlo (MC) events in the same job in which detector simulation and reconstruction are performed. I argue here that an alternative structure, a public library of stable validated datasets at the generator level, is more cost-effective in terms of collaboration manpower. Other benefits are a higher level of validation of the physics, access to more generators, access to wider resources, and better code maintenance and documentation.

1 Benefits of Separating Generators from Detector-Dependent Code

1. There has been an explosion of new tools and techniques:
   (a) Leading Order (LO) generators:
      i. Powerful symbolic generators such as COMPHEP, GRACE
      ii. Fast matrix element generators for many particles such as MADGRAPH
   (b) NLO 1-Loop automated calculators on the horizon (exist for e^+e^-)
   (c) Major collaborations/efforts underway in Europe, ...
   (d) Distributed computing efforts (GRID, SAM,...)
2. Simple and Robust Interfaces-
   (a) Parton-level generator to Shower Monte Carlo- Les Houches #1
   (b) Shower Monte Carlo to detector simulation- Les Houches #2, STDHEP (Fermilab supported standard- Mrenna now involved).
3. **Add** manpower, money, and resources

4. Theorists work closely with experimenters when the generator is theirs

5. Validation is possible by many interested parties, not just CDF, and especially interested theorists.

6. Those generators that have survived for many years are exceptionally well-documented, and have proven themselves responsive to requests from the users (e.g., Pythia, Herwig, COMPHEP, ...).

7. These codes are maintained by the authors- this represents a large investment of manpower that has expertise and continuity, and which would be very expensive to reproduce, even if possible.

8. Lastly, separating out the generators means that it is easy to run the identical events over-and-over through a detector simulation, comparing different versions of the simulation as it evolves. Alternatively, one can run the identical events through two *different* detector simulations (for example, a parametric simulation and a GEANT-based simulation), and compare on an event-by-event level as well as on an average level.

2 Planning for the Evolution of Theoretical Tools

1. We will be doing precision comparisons in *many* channels.

2. We need several generators for each process for validation.

3. The field is moving fast- we need to keep up with new generators, new versions, new advances, with little time lag

4. There are more subtleties as we get to more complex processes; we will need better communication and closer working with the authors of the Monte Carlo programs.

3 Questions for CDF to Answer

1. How many FTE's are needed to keep current, and to test new developments, tools, programs, etc.?

2. Do we have the resources? We need to define the goals and the scope of the MC needs of the Physics groups versus time for the next several years.

3. What are the benefits vs costs of incorporating *all* generators inside the CDF framework?

4 Proposal for CDF

So, to make sure that we can meet the needs of the Physics groups over the next several years, I think we should do the following:

1. **We make a (simple!) estimate of the MC tasks (both lists of processes$\times$ generators and perhaps deeper issues, such as NLO needs, tuning of shower MC's, and and resources, and manpower needed for them[1].)**

2. **We move to a hybrid scheme, in which we still have Pythia and Herwig available embedded inside CDF code, but also to have the ability to use stand-alone generators easily with CDF detector simulations.**

5 MCDB Status

The Monte Carlo Data Base (MCDB) is supported by the Fermilab Computer Division, with Steve Mrenna being the point person. The CD has devoted money, time, and resources to it, and it is being strongly supported. Although we have just begun, it has gone well so far; the present status is as follows (the main hangup at the moment is with the CDF end- getting generated files through the farms is the bottleneck at present).
The structure of the MCDB is shown in Figure 1.

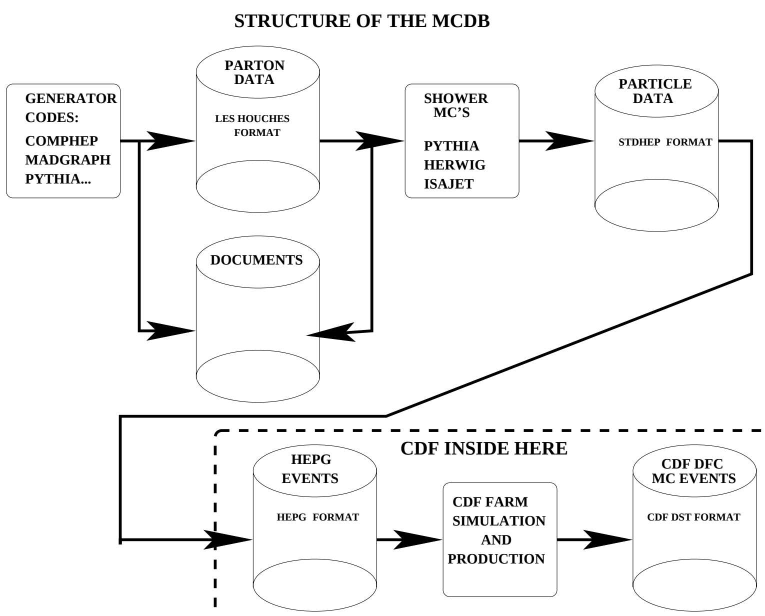

Figure 1: The structure of the Monte Carlo DataBase data flow. The data from the parton generator MC are stored in the Les Houches Accord format, and is human readable (ASCII text). The data from the Shower Monte Carlo are in the Fermilab-supported STDHEP format.

1. We have 1.2 TBytes on STKEN as an initial working space.
2. Eduard Boos came to Chicago for a week, and got the initial installation of COMPHEP working and an initial documentation set up.
3. Lev Dudko came to Fermilab for 9 weeks; COMPHEP now has batch mode, etc.
4. Lev has created a structure for easy documentation; it is easy to use, and is being used[2].
5. Tim Stelzer has Madgraph in very good shape. He and Steve are in the process of getting it so that it's publicly available.
6. We have an initial 'wishlist' of processes that have been requested by the CDF and D0 high-Et physics groups; the generation of these is in progress[3].

6 Short-term Wish List

Most of these relate to COMPHEP, as that's furthest along.

1. Fill in the present 'wishlist' of processes.
2. Interface TAOLA (Belyaev is has a working version)
3. MADGRAPH interface (Stelzer and Mrenna have a working version)
4. Get a Les Houches interface for Herwig
5. NLO Generators- MCFM, Wbbgen, etc. (needs more thought too)
6. Automate the path of STDHEP$\rightarrow$HEPG$\rightarrow$CDF Farm so that one button does it (Carron, Wurthwein, Litvinov, Miro)
7. Outsource generation to external farms (e.g. UC, Liverpool, ... has GRID and/or SAM overtones)

7 Conclusions

We have at least 6 years of precision measurements ahead of us, in which we will work at a level of detail far beyond what we ever dreamed of when we began the Collider. We will need MC tools that are much more powerful than we have at present, even though the present ones are already more powerful than those available in Run I. We need to plan the structures so that we can exploit new tools rapidly, simply, and robustly. This will require changes from our Run I modes of operation. Lastly, we are looking for a few good like-minded folks who are interested in specific signatures and who would like to see them generated in the fashion described in the body of this talk.

8 Acknowledgments

The people who have been working on this the most are: Eduard Boos, Lev Dudko, Sasha Belyaev, Steve Mrenna, Tim Stelzer, Sebastian Carron;+ Elena Gerstein (HEPG interface), Stan Thompson (HEPG interface); Fermilab CD. Boos and Dudko in particular are among the authors of COMPHEP, as is Belyaev, and have been very responsive to requests to modify the interface and usage of COMPHEP to make it easy to use in a hadron-collider environment. Steve Mrenna is the key person at Fermilab, and is exceptionally knowledgeable as well. Tim is a major player with Madgraph. All five of these are theorists with a deep understanding of the issues, and as such are an invaluable resource. In addition, they are very interested in Tevatron collider physics, and thus are great fun to talk with.

References

[1] CDF5821 gives a list of the Monte Carlo datasets requested initially by the top di lepton group. This list includes the different background channels, as well as indicates which generators will be needed. This is a start at an example list of needs for a specific analysis; by taking the logical OR of several of these each Physics group can make simple initial estimate of its MC Precision analyses, such as the W mass, some of the B-physics measurements, possibly even the top mass, or some of the QCD specialized analyses, may have special needs. These could be folded in too.

[2] See http://www-d0.fnal.gov/∼dudko/mcdb/, which is the top-level page of the documentation. This will be moved to a more public page when the MCDB is a little further along.

[3] For the current status of requested samples,
see http://www-d0.fnal.gov/∼dudko/mcdb/7/18_1.html.