(just a little about)

# Monte Carlo Generation of Multijet Processes

M. E. Peskin
SLAC ATLAS Forum
February 2007

As you folks know, I am interested in the question of analyzing generic signatures of new physics associated with heavy particles and missing energy.

John Conley, Tommer Wizansky, and I are working on a tool for this study. This is a progress report. It contains some nontrivial concepts - `well-known' in the literature - that I hope you find interesting.

We wish to identify the events characteristic of supersymmetry or similar models with dark matter and electroweak symmetry breaking. To do this, we must make the descent into the Standard Model background. Here are some of the levels that must be crossed:

| | |
|---|---|
| $\sigma_{tot}$ | 100 mb |
| jets w. $p_T > 100$ | 1 $\mu$b |
| Drell-Yan | 100 nb |
| $t\bar{t}$ | 800 pb |
| | |
| SUSY ($M < 1$ TeV) | 1-10 pb |

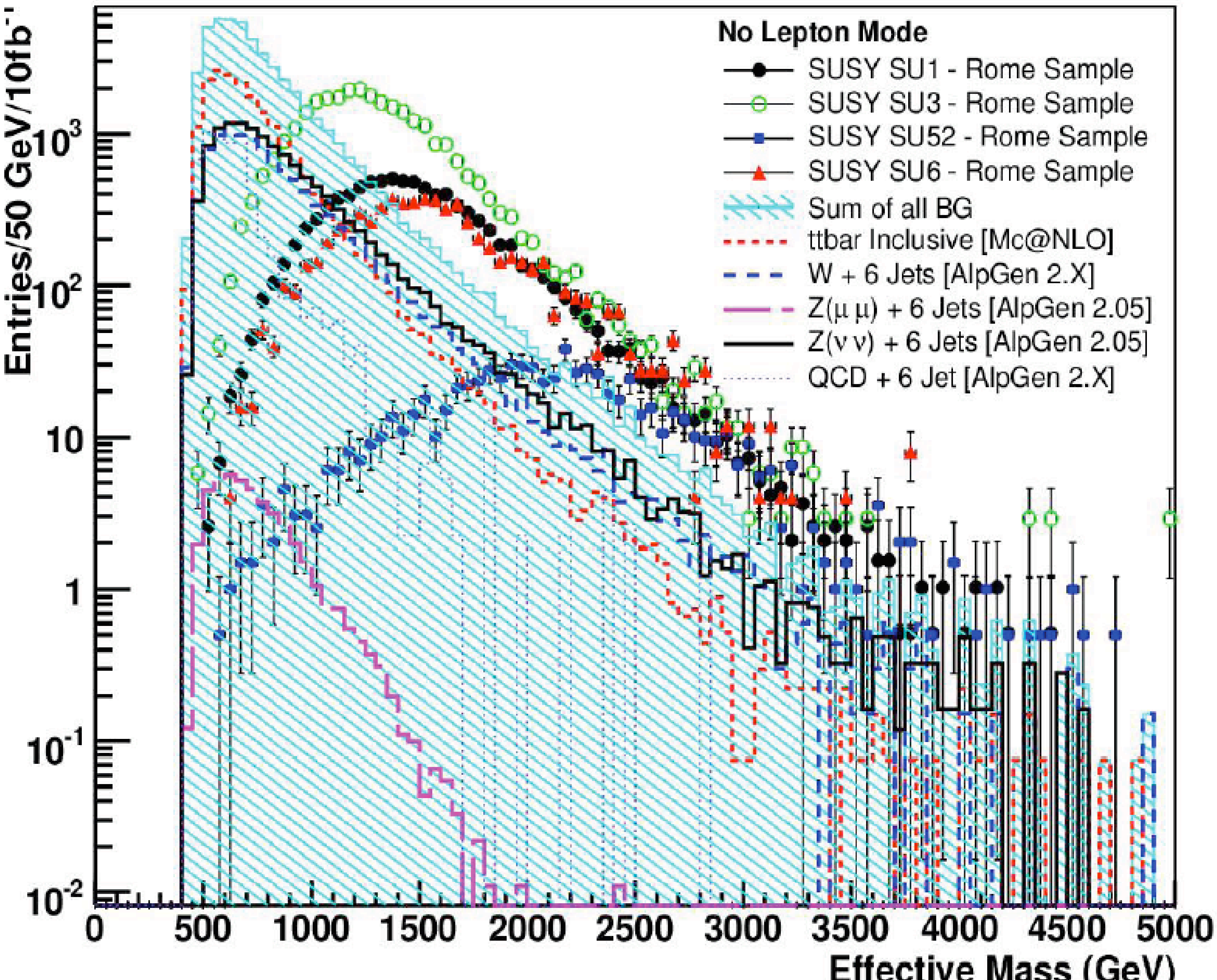
In studies for ATLAS and CMS, the dominant backgrounds to new physics do not come from QCD multijet events with jets faking missing energy. Rather, they come from heavy particle production within the Standard Model, production of $W, Z, t\bar{t}$ plus jets.

These reactions already offer missing energy, leptons, and hadronic activity. They populate the region of large HT associated with new physics to the extent the additional jets are radiated along with the heavy particles.

It is nontrivial to demonstrate correctly that pure QCD processes are not important. In particular, this requires the ability to assign low missing energy significance to special cases.

Missing energy from b jets is another source of exceptions. Jay Wacker is studying this now.

Here is a recent quantitative evaluation by Sanjay Padhi, using ALPGEN and the ATLAS full simulation code




Meff distribution subject to

$\not{E}_T > 100$

4 jets, 2 w.

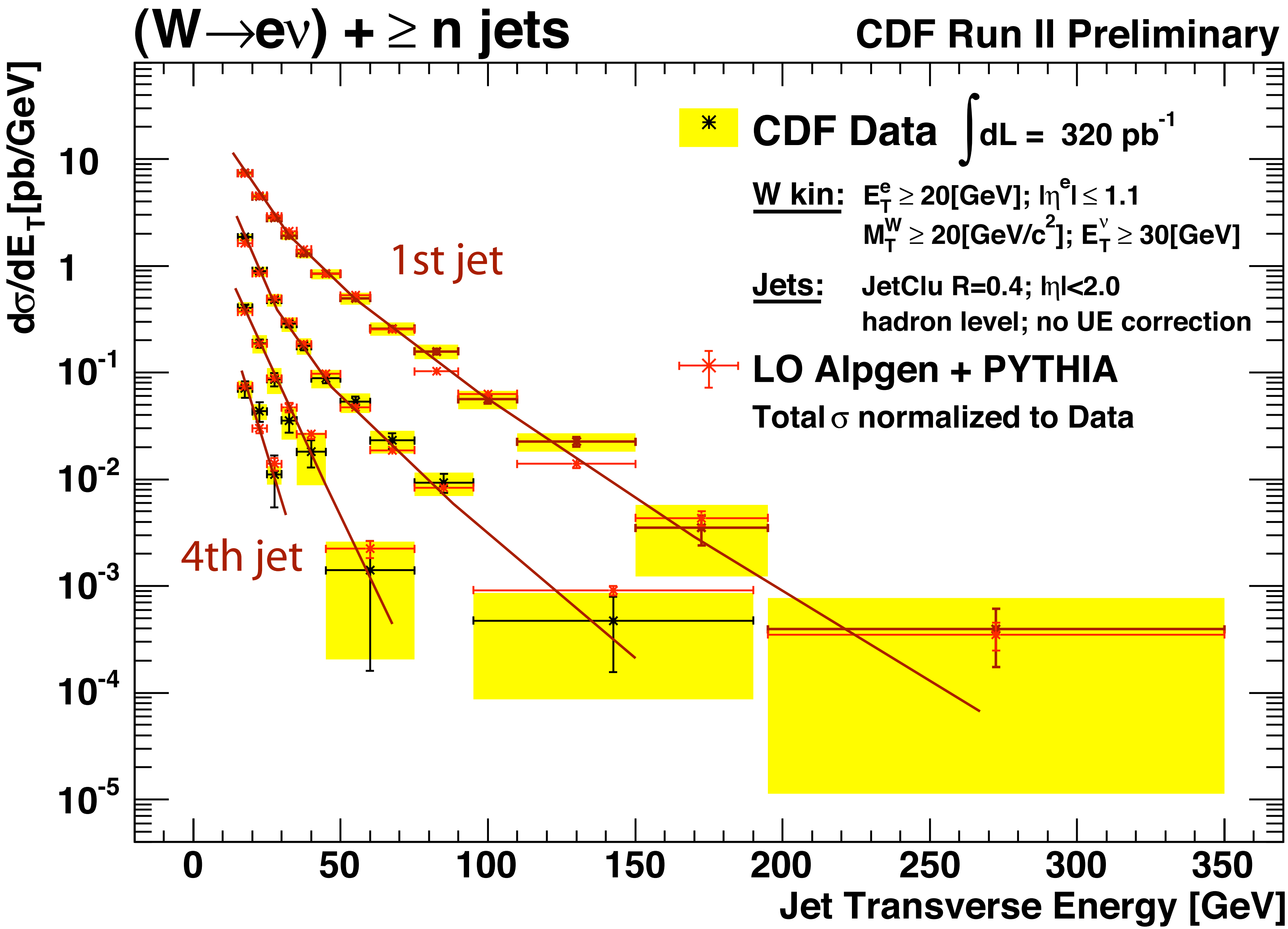
$E_T > 100$

I personally would like to concentrate on $W, Z, t\bar{t}$ + jets processes. For these, there is a program:

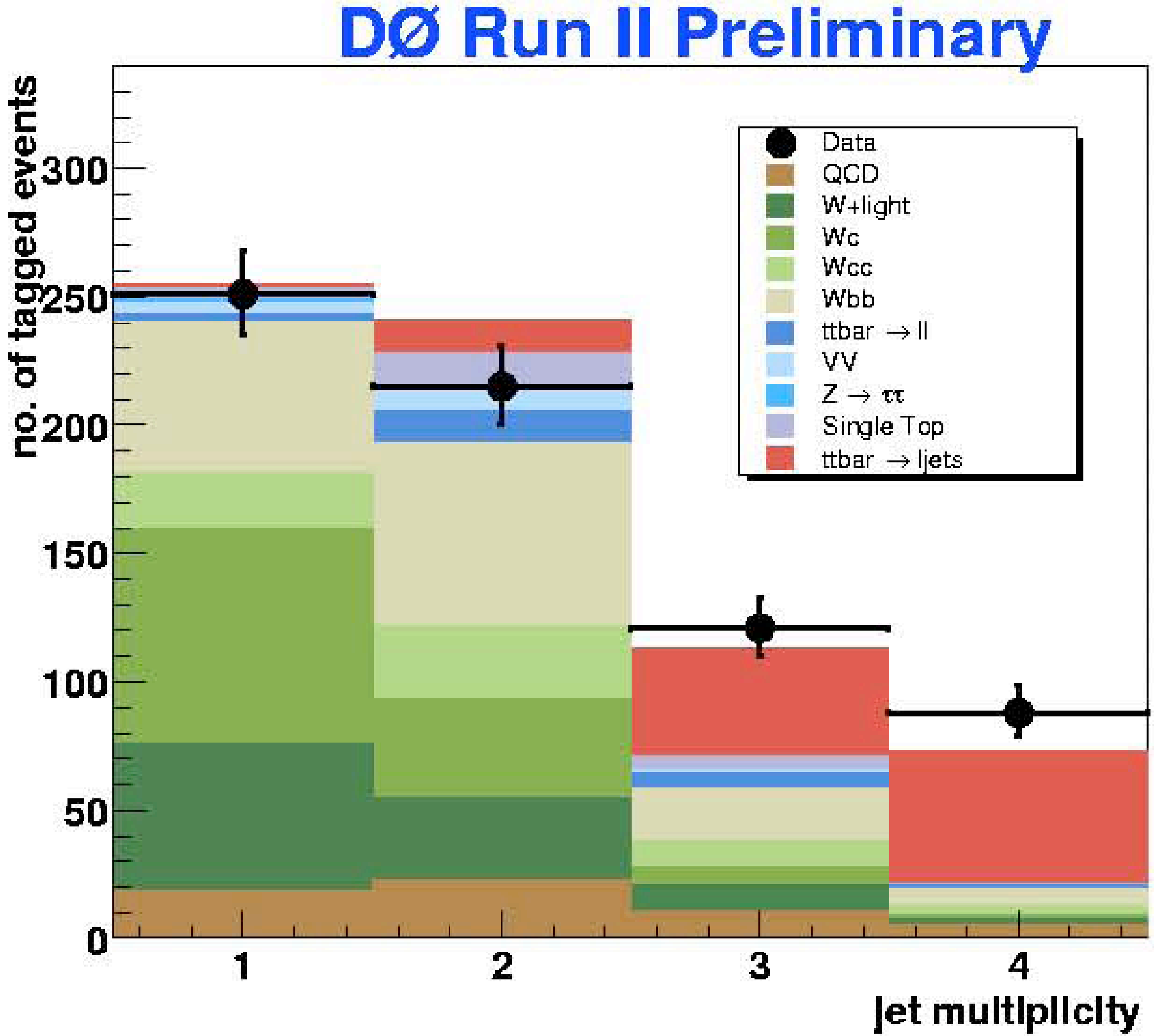
Formulate a model of QCD radiation associated with $W, Z, t\bar{t}$ .

Find data samples in which this model can be validated by or fit to data. Theses should include cases with hard gluon radiation.

Extrapolate into regions with many jets, many leptons, or large HT.
QCD radiation falls off systematically as we go into these regions, and, at the same time, these are the regions where signals are expected.



(W→eν) + ≥ n jets
CDF Run II Preliminary
dσ/dE_T [pb/GeV]
CDF Data ∫dL = 320 pb^-1
W kin: E_T^e ≥ 20[GeV]; |η^e| ≤ 1.1
M_T^W ≥ 20[GeV/c^2]; E_T^ν ≥ 30[GeV]
Jets: JetClu R=0.4; |η|<2.0
hadron level; no UE correction
LO Alpgen + PYTHIA
Total σ normalized to Data
1st jet
4th jet
Jet Transverse Energy [GeV]

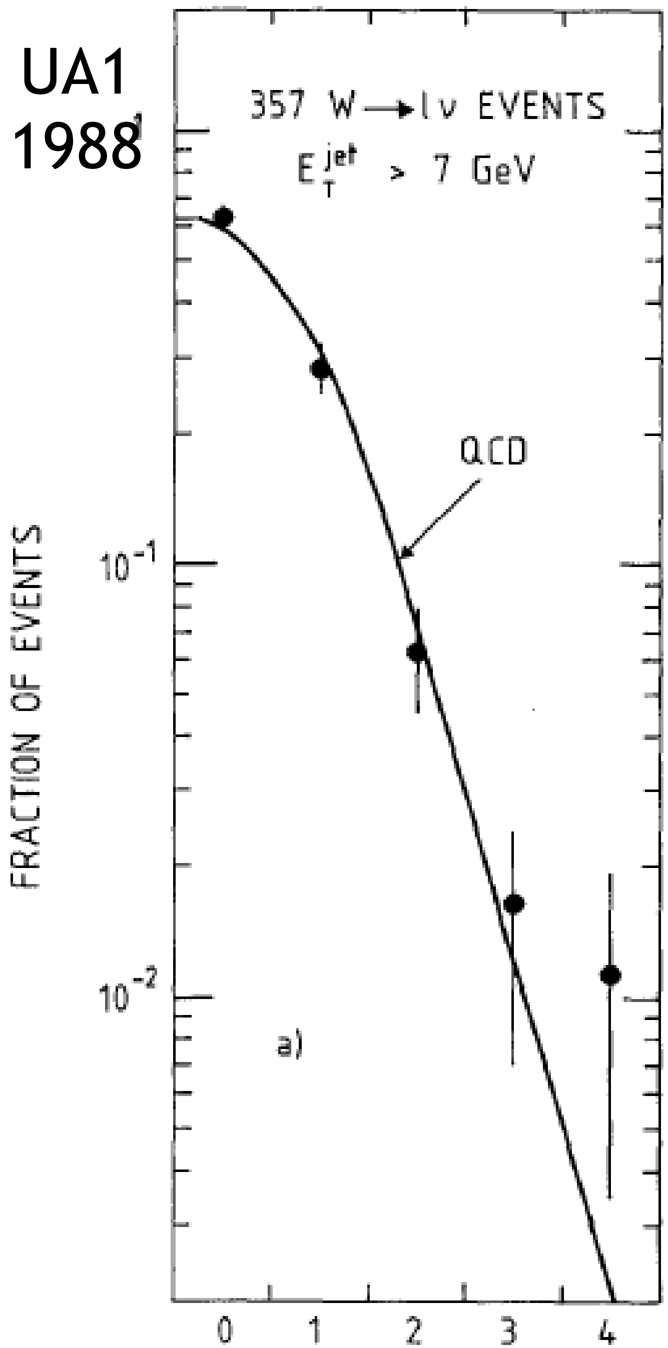

DØ Run II Preliminary
no. of tagged events
0
50
100
150
200
250
300
1
2
3
4
jet multiplicity
Data
QCD
W+light
Wc
Wcc
Wbb
ttbar → ll
VV
Z → ττ
Single Top
ttbar → ljets

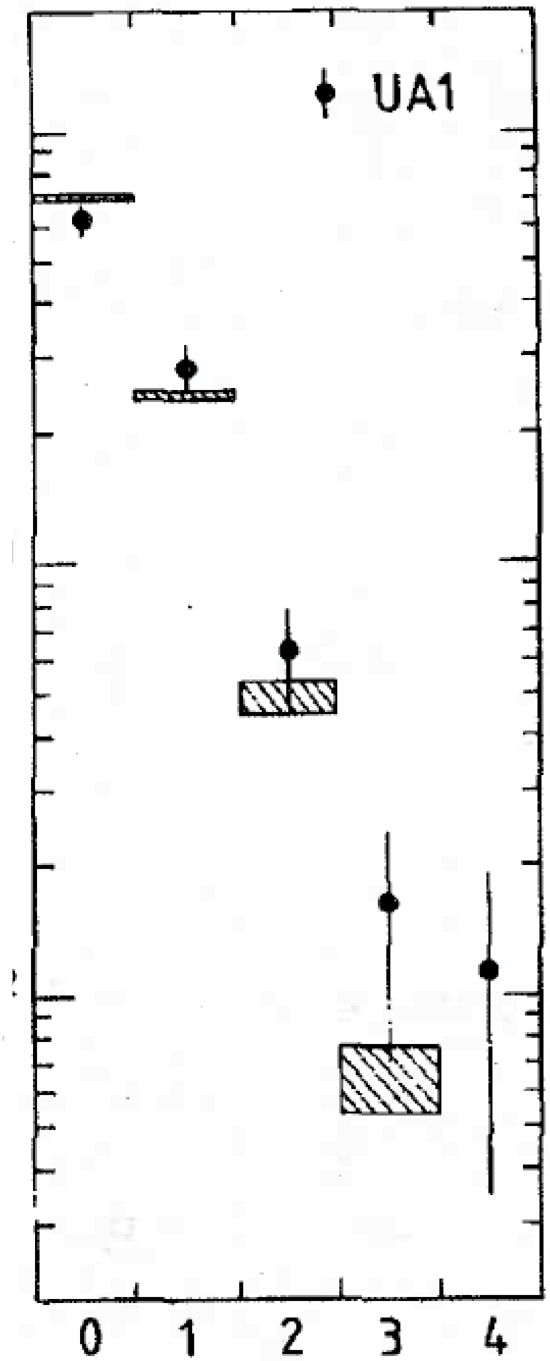
This kind of analysis is a standard part of the Tevatron culture.
But still, it deserves a name.

I like to call it a staircase.

# Berends-Giele staircase

UA1
1988




number of jets

Berends,
Giele,Kuijf,
Kleiss, Stirling
1989

Multilepton signatures of SUSY exhibit similar staircases, and for these also the backgrounds are supposed to be dominated by processes with heavy SM particles with additional jets.

Here we should refer to the Baer-Tata staircase.

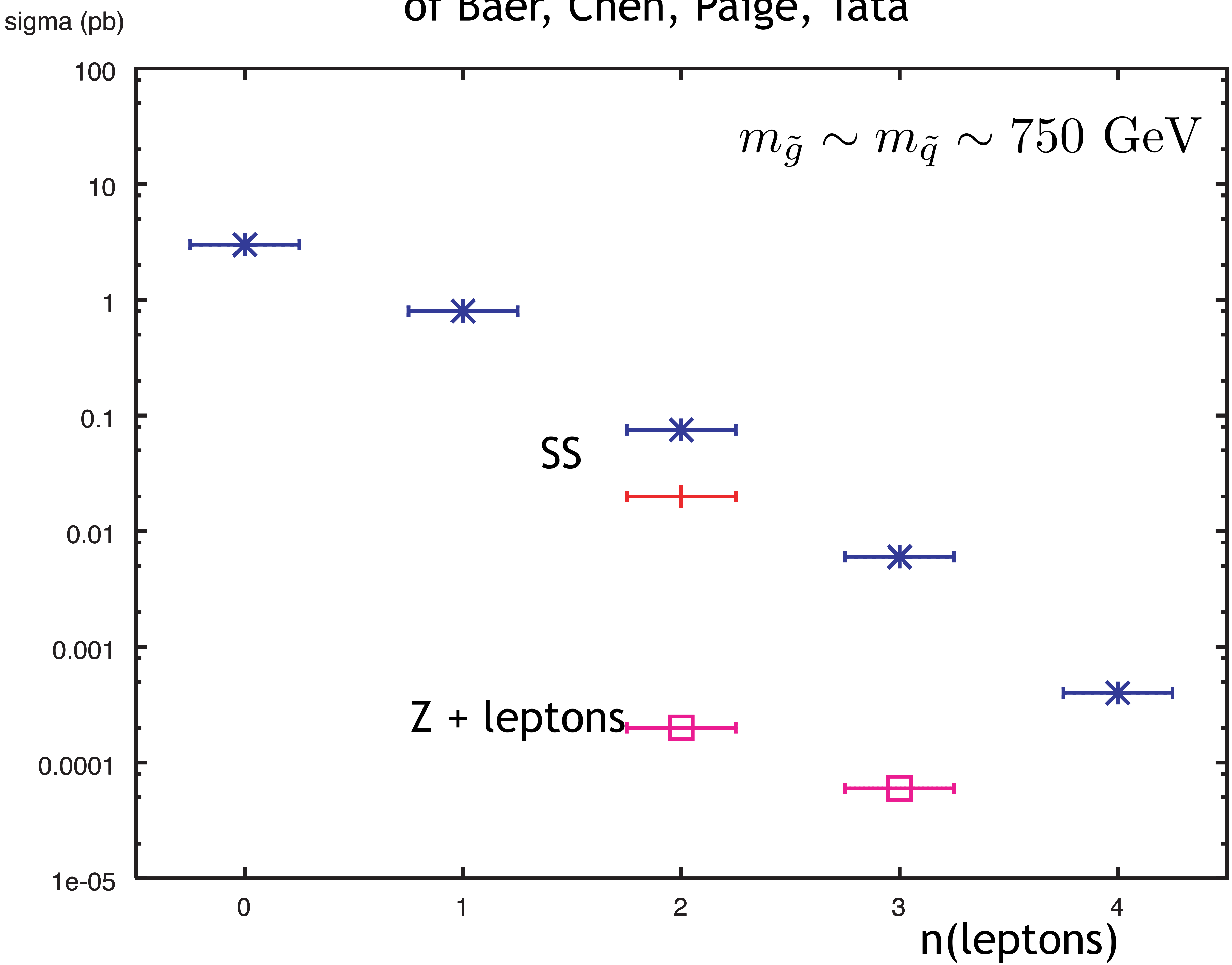
It is well appreciated that SUSY models predict 2, 3, 4 - lepton events in a steadily decreasing progression.

The Standard Model also produces such events, from multiple heavy-quark decays and jets faking leptons.

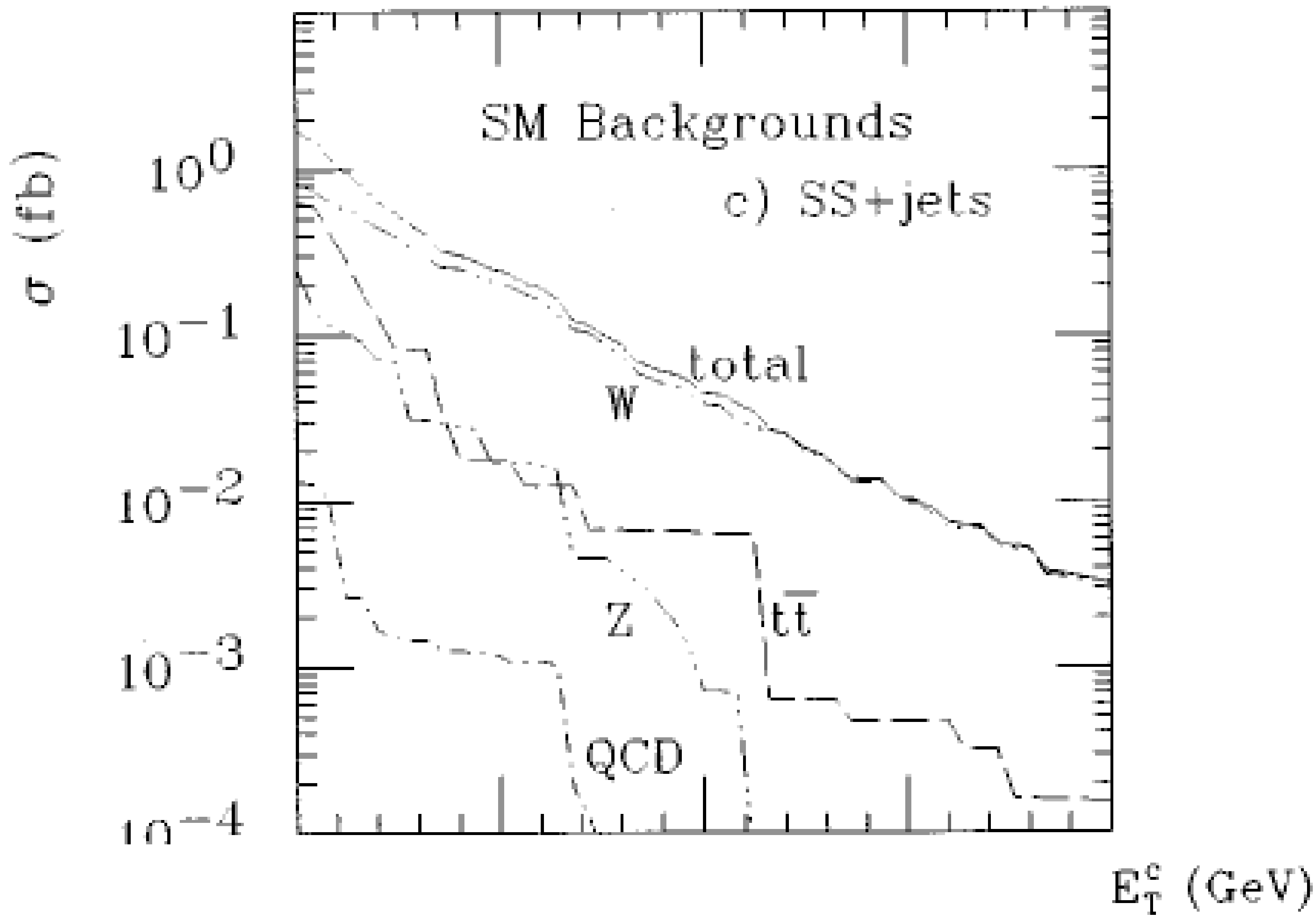
Fortunately, these come from the same W, Z, $t\bar{t}$ + jets processes that we have already been discussing.

Electroweak backgrounds, e.g. $pp \to W^+W^+ \to \ell^+\ell^- + \text{jets}$ are at the fb level.

signal cross sections from one of the models
of Baer, Chen, Paige, Tata



dissection of the Same Sign dilepton background from this paper



To model staircases, we need a simulation code that accurately produces many QCD jets with correct transverse momentum distributions.   This is the territory of ALPGEN and MADGRAPH, and the higher-loop versions of these that Lance Dixon envisions.

These codes are like highly tuned BMW's.

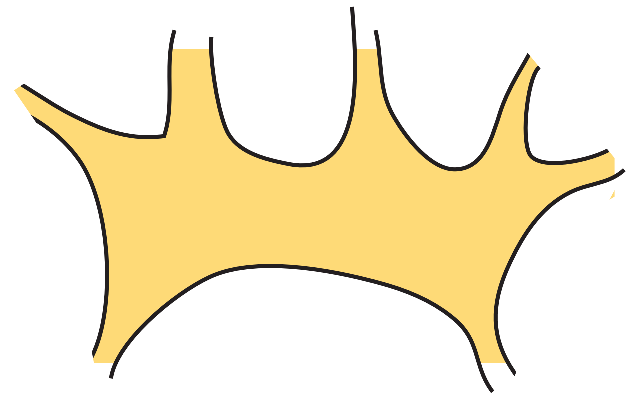
But it may be useful to have a codes that is built more like a \`62 Chevy, one that lets you poke under the hood and adjust the carburetor by hand.

It is possible to get quite far with a few simple - but quite nontrivial - tricks. Let's now discuss them.

In the rest of this talk, I will discuss methods for generation of $gg \to ng$ events.

The techniques I will discuss do generalize to the processes of interest with heavy particles. The strategy, following Ellis, Kleiss, and Stirling, is to include the decays of the heavy particles to massless fermions. This makes the calculations easier and, as a bonus, generates the final states with the correct spin correlations.

The first trick is to use the large-N limit of QCD. Ignoring color factors proportional to $1/N_c^2 \sim$ 10%, a multigluon diagram is represented by the color flow



We can treat the gluons as ordered around the diagram and therefore distinguishable. We can ignore the interference of different color structures. (A part of this approximation is already used in the treatment of color in PYTHIA and HERWIG.)

The diagrams contributing to a given color structure have only the singularity structure

$$1/(k_1 \cdot k_2)(k_2 \cdot k_3)(k_3 \cdot k_4) \cdots (k_n \cdot k_1)$$

which is much less complicated than it could be.

I will only try to work to the tree level. I am assuming, then, that it is possible to build a reasonable model by multiplying the tree cross sections by K-factors. With a simple enough description, we can try to fit these K-factors to experiment.

Of course, each tree amplitude still requires a large number of Feynman diagrams.

There are easier to discuss with the use of spinor products:

$$\langle 12\rangle = \overline{u}_L(1)u_R(2) \qquad [12] = \overline{u}_R(1)u_L(2)$$

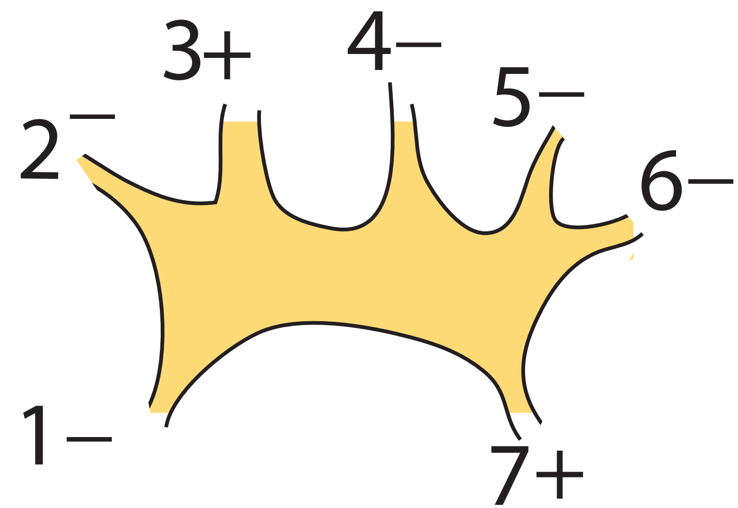
build from the spinors of massless 4-vectors. These obey

$$|\langle 12\rangle|^2 = |[12]|^2 = 2k_1 \cdot k_2$$

Every initial and final gluon has a definite helicity + or -. From now on, I will label all helicities as outgoing (- in -> + out).

Certain amplitudes are very simple after all diagrams are summed. Notate, for example

$$i\mathcal{M}(1^-, 2^-, 3^+, 4^-, 5^-, 6^-, 7^+) =$$




Then tree amplitudes with zero or one + or - vanish!

Parke and Taylor found that the amplitudes with two - take a very simple form.  For example,

$$i\mathcal{M}(1^+,2^-,3^+,4^-,5^+,6^+,7^+) = ig^5 \frac{\langle 24\rangle^4}{\langle 12\rangle\langle 23\rangle\langle 34\rangle\cdots\langle 71\rangle}$$

These are MHV amplitues.  The amplitudes with two + (anti-MHV) take a similar form in terms of square brackets.

Notice that the squares of these give just the denominator seen earlier for color-ordered amplitudes.

What about all of the other amplitudes that are not MHV. Recently, Britto, Cachazo, Feng, and Witten found a remarkable recursive formula that allows one to evaluate these.

$$\mathcal{M}(1,\cdots,n) = \sum_{splits} \mathcal{M}(a\cdots\hat{i}\cdots b)\frac{1}{s_{a\cdots b}}\mathcal{M}(b+1\cdots\hat{j}\cdots a-1)$$

where the amplitudes on the right are evaluated at complex (analytically continued) but still lightlike, on-shell momenta.

This allows the n-point amplitudes to be recursively evaluated in terms of amplitudes with fewer legs. We can stop when we reach MHV. At 5 points all amplitudes are MHV or anti-MHV.

By defining a C++ bispinor class, it is possible to write a fairly simple yet efficient code. It computes

2 -> 6 gluon amplitudes at 800/sec

2 -> 10 gluon amplitudes at 100/sec

This is almost, not quite, as fast as one would like for event generation. With some improvements, it may be possible to write a parton shower Monte Carlo without approximating the amplitudes for successive parton emissions.

To make this useful, we also need to generate n-particle phase space efficiently.

In the 1980's, Ellis, Kleiss and Stirling wrote a simple algorithm (RAMBO) for n-body massless phase space: Generate n vectors, each drawn from the distribution

$$\int \frac{d^3k}{2k}\; e^{-|k|}$$

Boost to the CM frame, and rescale the total energy to $\sqrt{s}$ .

This generates a flat distribution over n-body massless phase space.

More recently, Draggiotis, van Hameren, and Kleiss described a similar algorithm SARGE that generates n-body massless phase space weighted with the denominator of QCD color-ordered amplitudes.

Start with $p_1 = (1, 0, 0, 1) \quad p_2 = (1, 0, 0, -1)$

Choose three variable with the distribution $\frac{d\xi_+}{\xi_+}\frac{d\xi_-}{\xi_-}d\phi$

Let $k^0 = (\xi_+ + \xi_-) \quad k^3 = (\xi_+ - \xi_1)$

so that $k$ is lightlike with azimuthal angle $\phi$ around the axis of the p's. Then the measure above is proportional to

$$\frac{d^3k}{2k}\frac{p_1 \cdot p_2}{(p_1 \cdot k)(k \cdot p_2)}$$

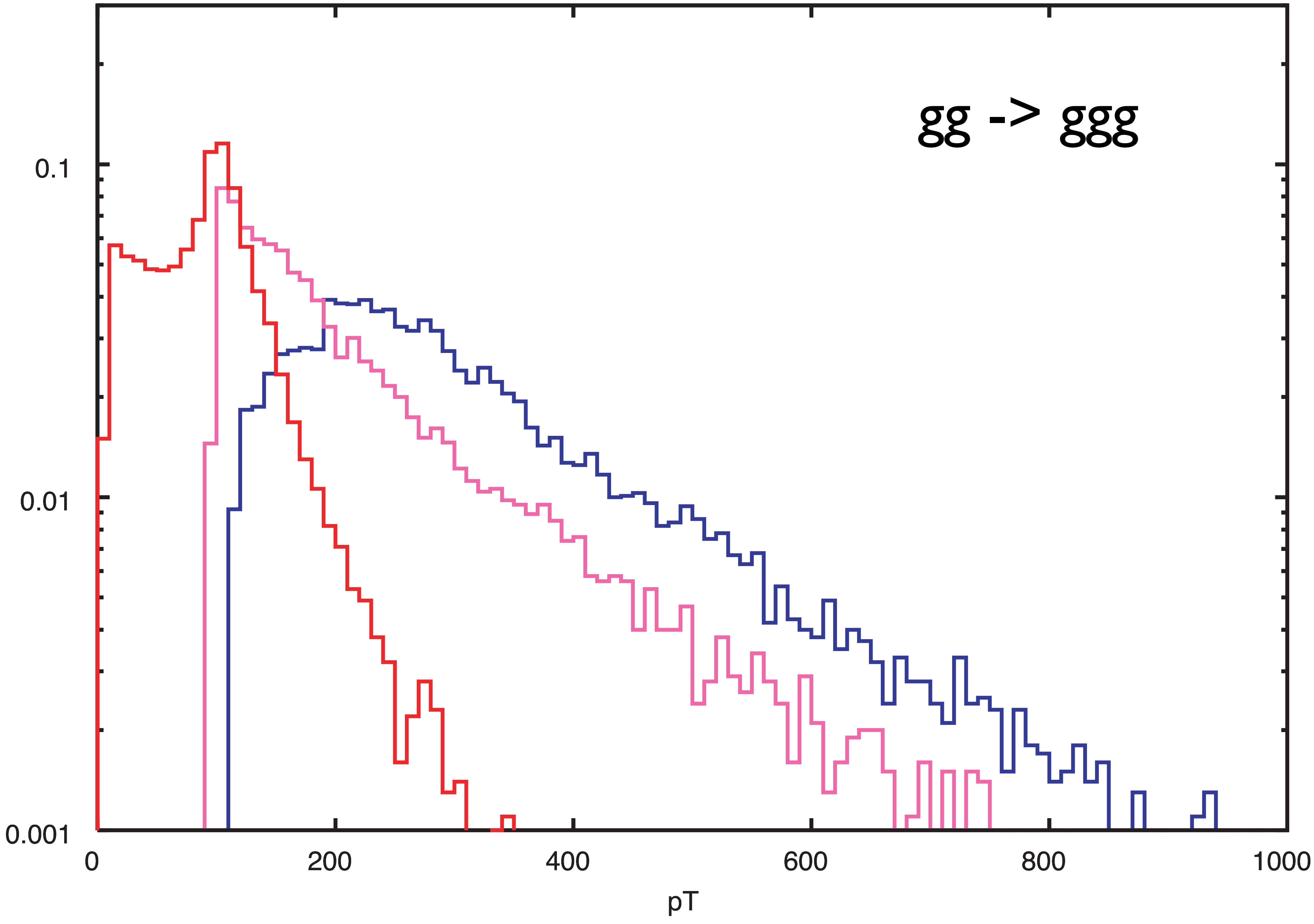
The final result is Lorentz-invariant and scale-invariant, so we can now go to the frame of $(k, p_2)$ and repeat.

Generating N-2 vectors in this way, we get the color-ordered QCD denominator for N particles.

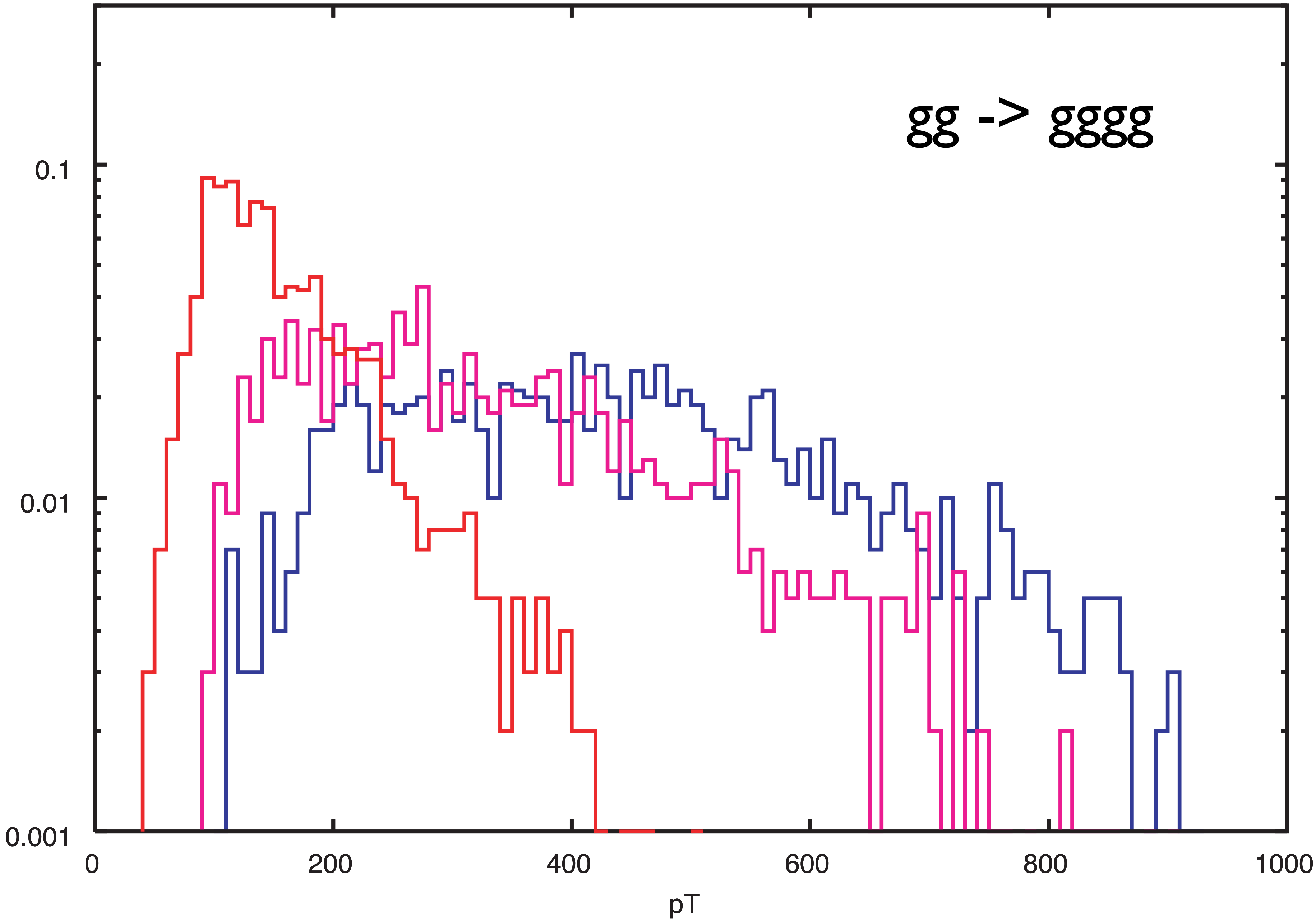
Including the initial state is more awkward, but we can orient the initial 4-vectors with respect to two of the final vectors.

To show you that this can actually reduced to practice, let me show you pT distributions for gg -> ggg, gggg generated with this technology. I am cheating by assuming a fixed gluon CM energy of 2 TeV.

The event generation speed is 30/sec for 3 g, 3/sec for 4g. This is not optimized - or maybe even correct; I generated these graphs in the past hour.



gg -> ggg
0.1
0.01
0.001
0
200
400
600
800
1000
pT


gg -> gggg
0.1
0.01
0.001
0
200
400
600
800
1000
pT

This is just the start of a project. I hope eventually to turn this into a useful simulation tool.