Finding an Upper Limit in the Presence of Unknown Background

Fermilab Workshop on Confidence Limits

S. Yellin, March 28, 2000

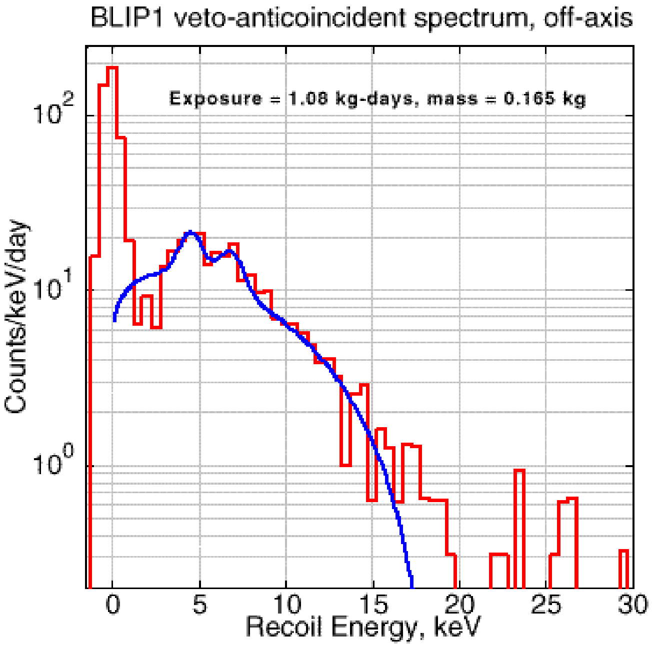
Problem: What if you don't understand your background, but know it's present because its distribution is different from that of the signal? How should you find the upper limit of the signal?

Solution: Look where the signal+background is observed to be especially small and use only that region to set an upper limit. But isn't that cheating? Maybe signal+background is small there because you've selected a region with a downward statistical fluctuation in the signal.

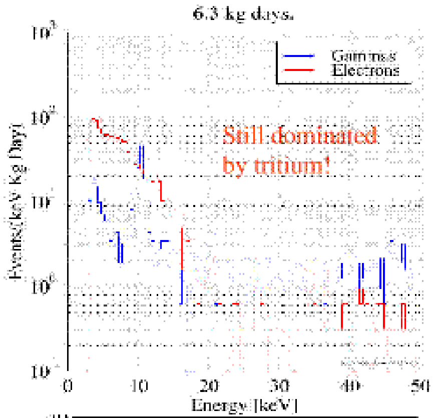
Plan: I'll explain the "**Maximal Gap**" method, then generalize to the "**Optimum Interval**" technique. Finally, I'll show results from testing and using the methods.

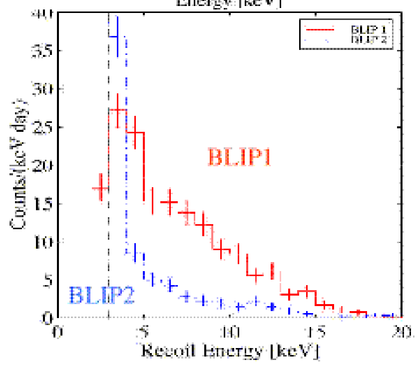
Tritium

Run 15: BLIP1 low-energy spectrum dominated by tritium



Run 18: after etch

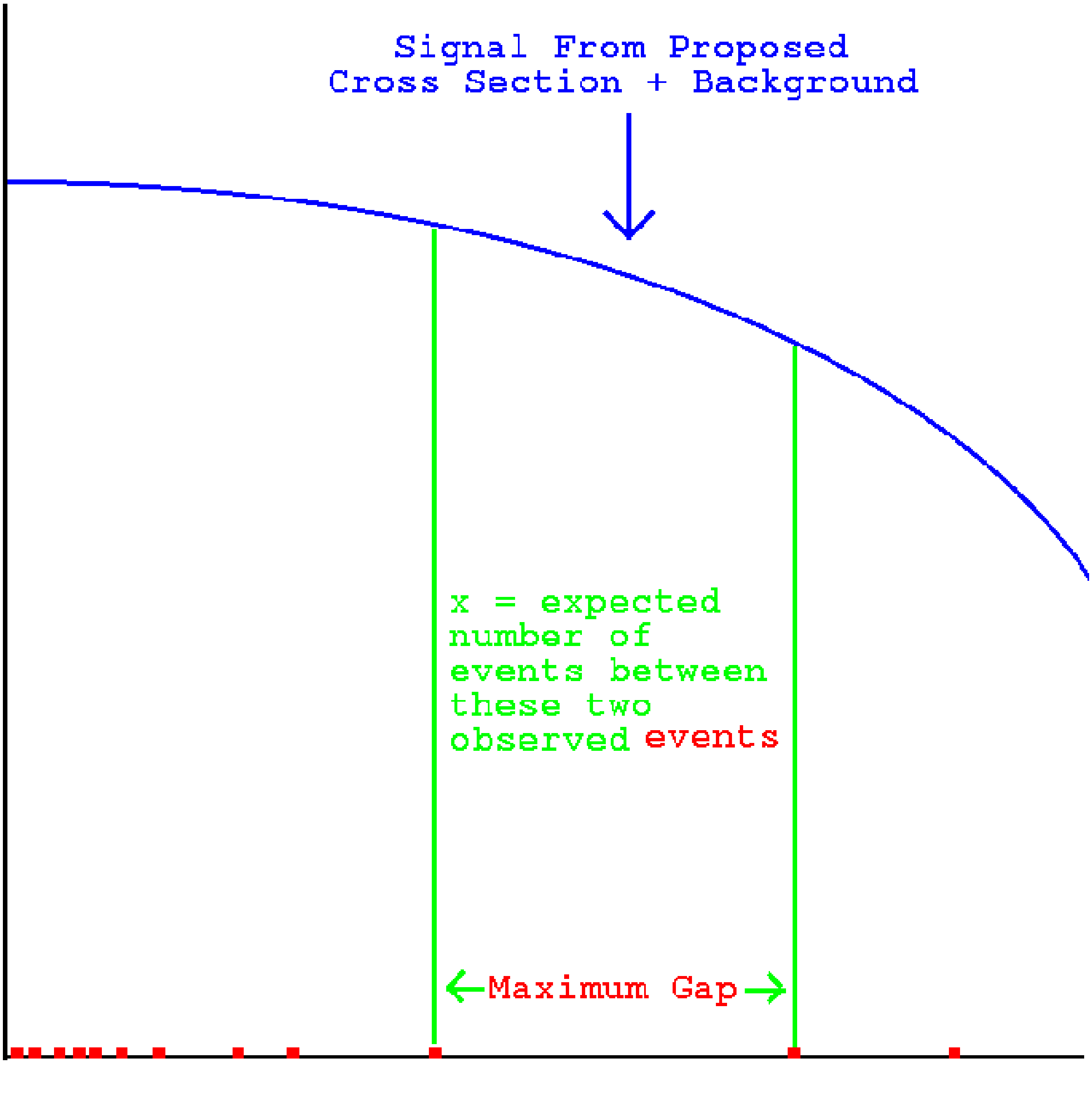




Resisted etch or external to detector - current run to answer

Maximum Gap Method

Exclude a proposed cross section as too large
if there's an "x" that's too large.



Confidence Level By Which Large x Excludes a Signal

$$C_0(x,\mu)=\sum_{k=0}^{m}\frac{(kx-\mu)^k e^{-kx}}{k!}\left(1+\frac{k}{\mu-kx}\right)$$

x: Maximum expected number of events in an empty range.
μ: Total number of events expected in the entire range.
m: Greatest integer $\leq \mu/x$.

For a 90% confidence level upper limit, increase the proposed signal until C_0 reaches 0.90.

Features:

1. Independent of the true signal and background, the true signal will be accepted with at least 90% probability.

2. The upper limit is not very sensitive to the boundaries of the entire range; extending a boundary into a region with high background will not affect the location or size of the maximum gap.

3. The answer doesn't depend on how data are binned because the data are not binned.

4. The limit is parameter independent; it doesn't change if a one-one change of variables is made.

5. The same formula works for any signal shape; the only dependence on the assumed signal is through μ.

Maximum Gap → Optimum Interval

Instead of looking for the maximum range with zero events, look for the maximum range with n events for some $n \neq 0$ and find the confidence level by which the cross section is rejected:

$$C_n(x, \mu),$$

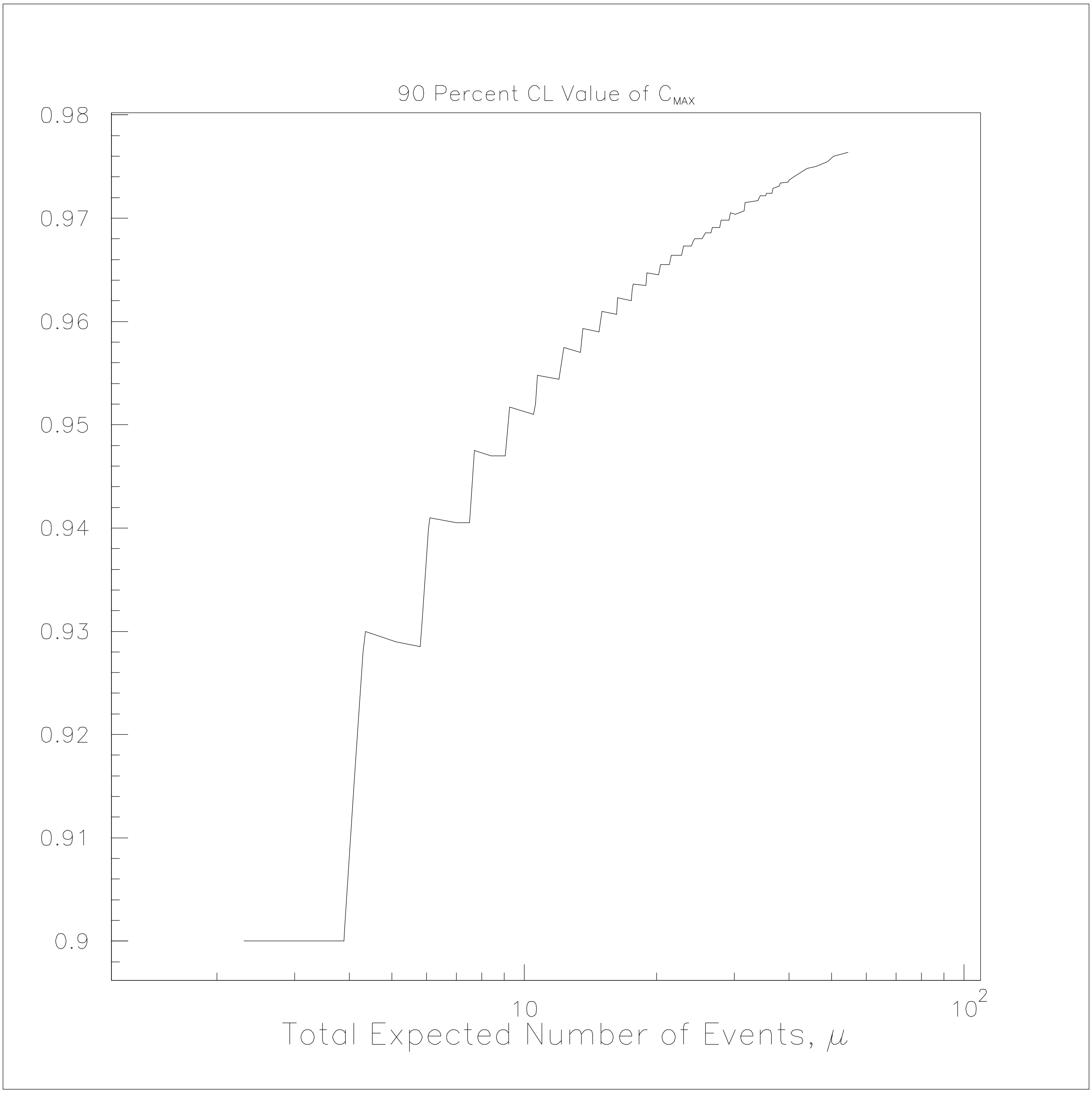
x: Maximum expected number of events in a range with n events.
μ: Total number of events expected in the entire range.

For a 90% confidence level upper limit, increase the proposed cross section until C_n reaches 0.90.

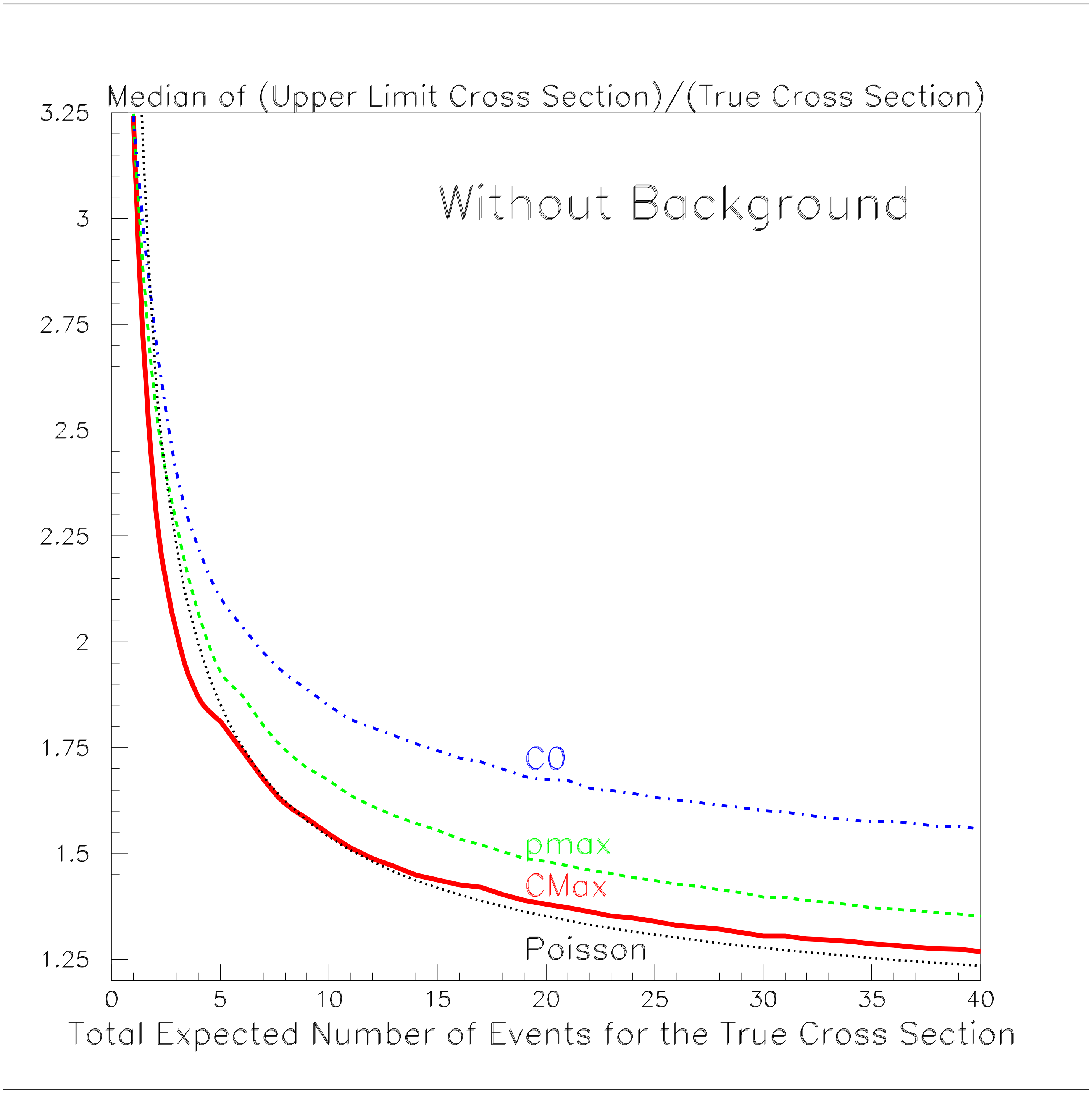
Features: Same as for the maximal gap method using C_0.
But Which n should be used? **Answer:** The one that gives the lowest upper limit. But don't cheat!

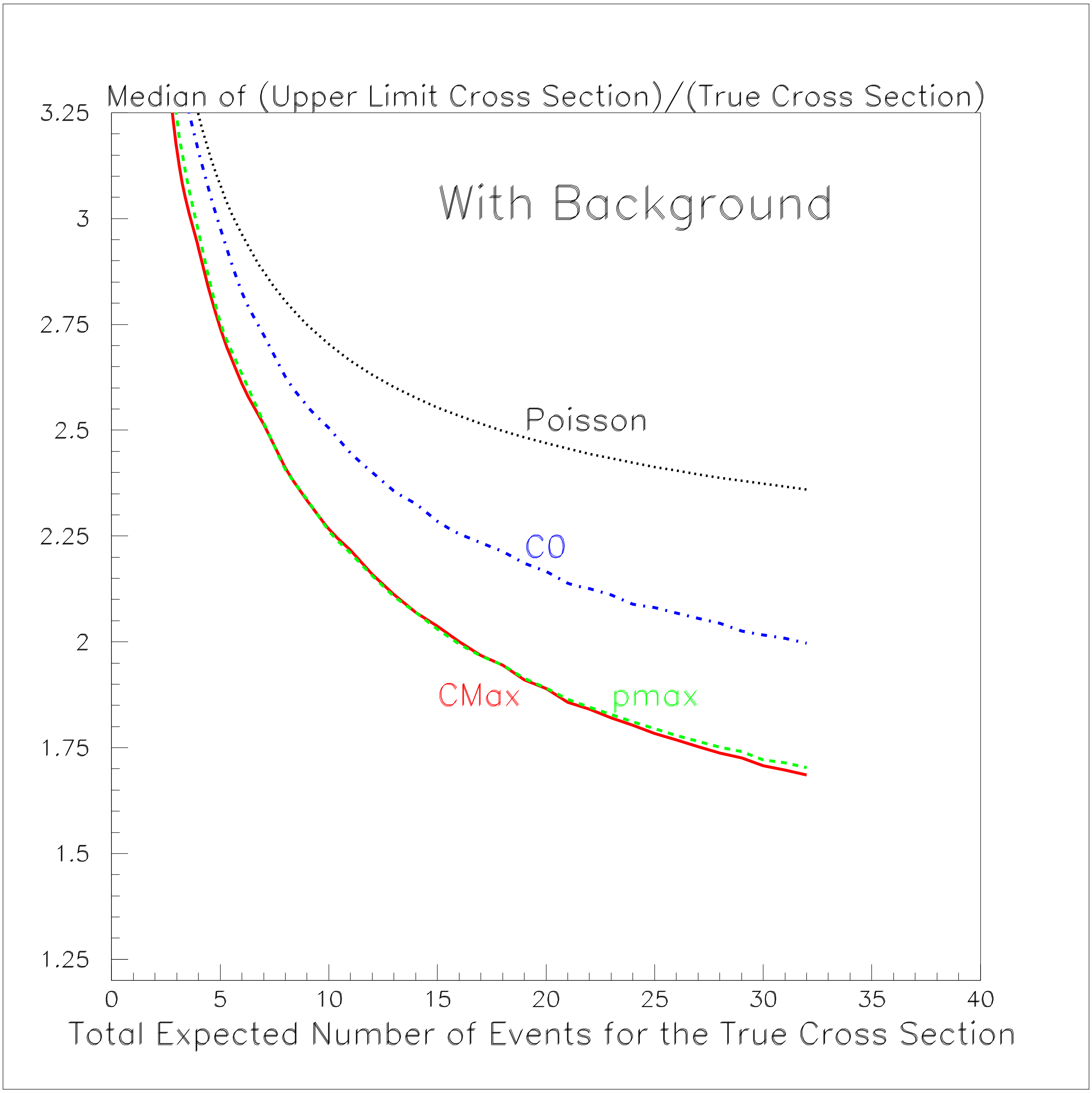
For a given assumed signal size, call C_{Max} the value of the biggest C_n – the C_n that rejects the signal with the most confidence. I.e., over all possible intervals within the entire experimental range, use the "optimum" one, as measured with C_n. Then find how high the maximum C_n must be to permit rejection to, say, a 90% confidence level.

Features: Same as for the maximal gap method. But so far, the required functions can only be evaluated by Monte Carlo. They all have been tabulated for a limited range of μ.

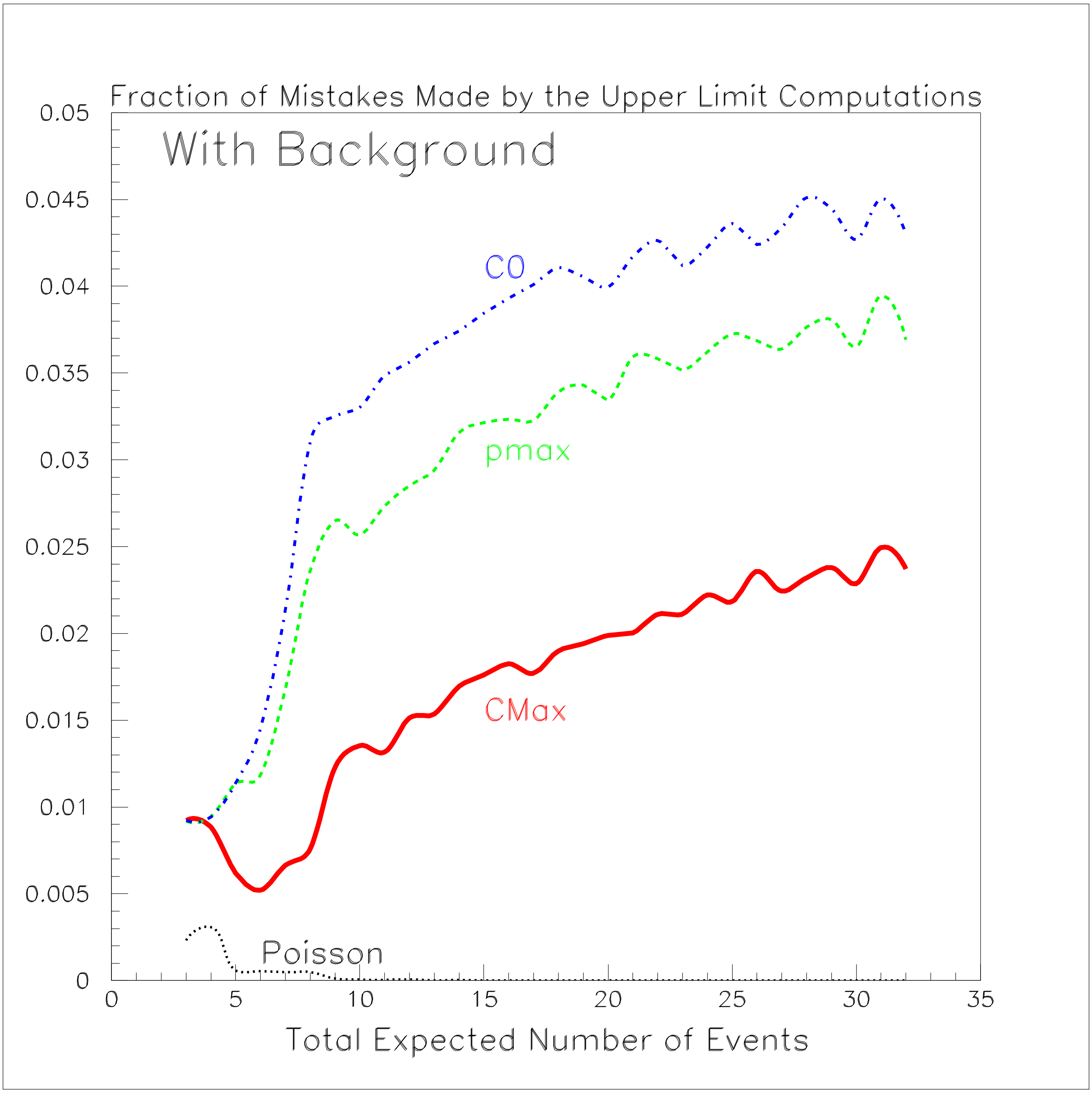


90 Percent CL Value of C_{MAX}
0.98
0.97
0.96
0.95
0.94
0.93
0.92
0.91
0.9
10
10^2
Total Expected Number of Events, μ


Median of (Upper Limit Cross Section)/(True Cross Section)
Without Background
C0
pmax
CMax
Poisson
3.25
3
2.75
2.5
2.25
2
1.75
1.5
1.25
0
5
10
15
20
25
30
35
40
Total Expected Number of Events for the True Cross Section



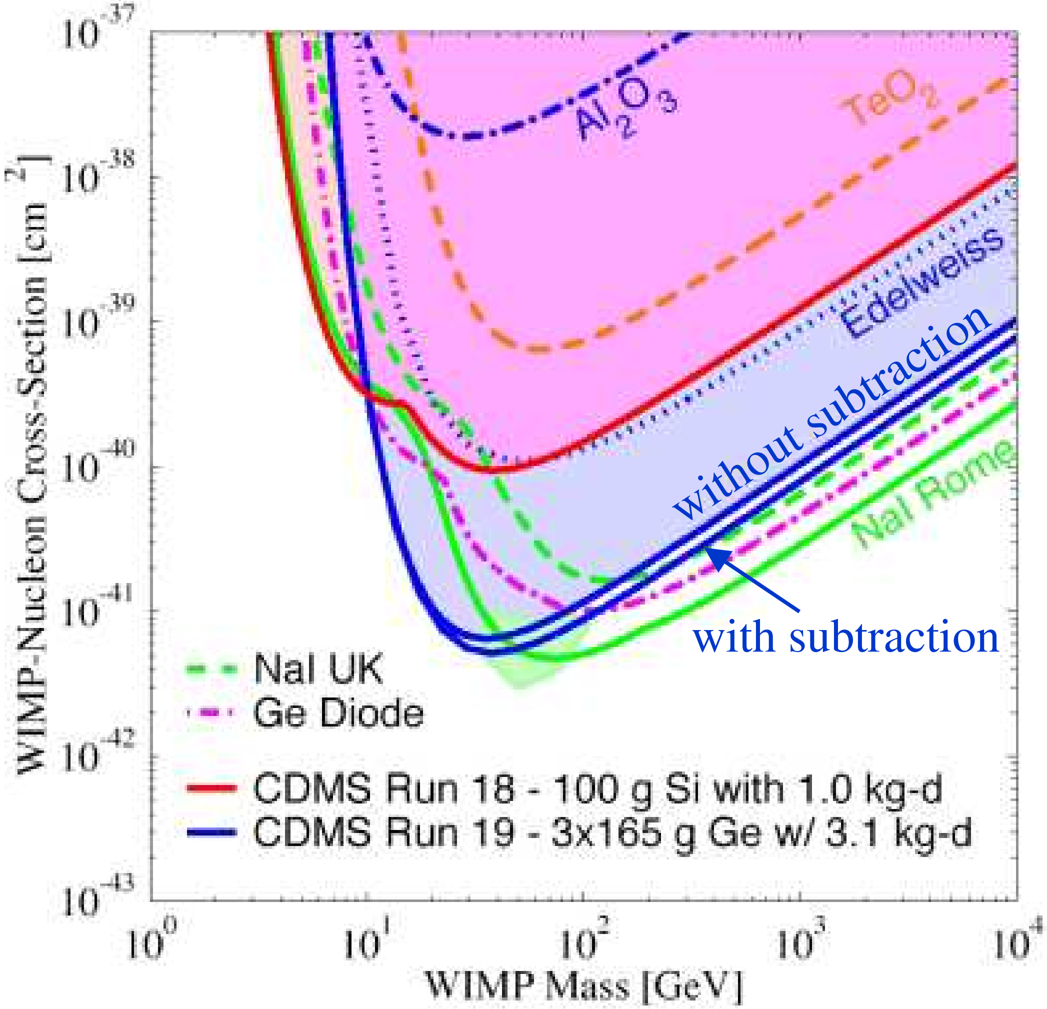
Median of (Upper Limit Cross Section)/(True Cross Section)
With Background
Poisson
C0
CMax
pmax
3.25
3
2.75
2.5
2.25
2
1.75
1.5
1.25
0
5
10
15
20
25
30
35
40
Total Expected Number of Events for the True Cross Section



Fraction of Mistakes Made by the Upper Limit Computations
With Background
C0
pmax
CMax
Poisson
0.05
0.045
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
0
5
10
15
20
25
30
35
Total Expected Number of Events

CDMS Results

Exclude half of DAMA claim region



Current run: x 4 in statistics => full exploration of DAMA claim

CDMSII deep underground: in final stage of review by DOE/NSF

Evaluation: When Did CDMS Find This Method Useful

During nearly all of the past lifetime of CDMS, there was no sign of a signal, but there was a background that we didn't know how to eliminate or subtract. Then the method was useful.

More recently, we've learned how to eliminate most of the background, and measure most of what was left. Now we use a version of the Unified Method (Feldman-Cousins).

We will eliminate most of the background we now measure, and the detector will become more sensitive. When that happens, the most important background may again be something we can't measure, and we may again need the Optimum Interval method, which is

1. **Safe:** Independent of signal and noise, a "90%" confidence region can be predicted to have at least a 90% probability of covering the truth, with a result that's not too sensitive to the experimental range used.
2. **Effective:** The Optimum interval method gets an especially low limit while making especially few mistakes.
3. **Cheap:** Once the software and Monte Carlo computed tables are available, very little cpu is needed.

For the software, tables, a draft of a paper on the subject, and the transparencies of this talk, see

http://slac.stanford.edu/~yellin/upperlimit.html