Vertexing status and plans

Sho Uemura

SLAC

Overview

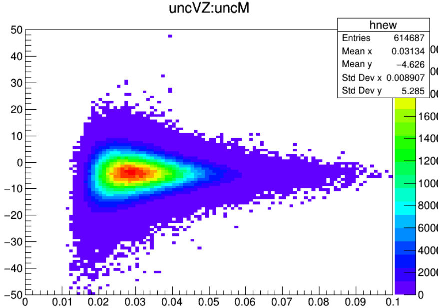
- How should the analysis work?
- What's been going on
- What I'm worried about

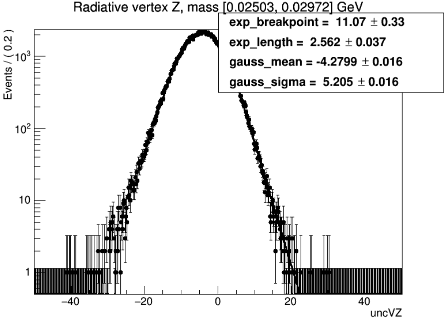
The analysis

- We want to set exclusion regions in mass vs. ϵ^2 for visibly decaying A’
  - Find limits on production cross-section (relative to radiative) as a function of mass and $c\tau$, then exclude based on compatibility with the cross-section predicted by ϵ^2
- Find significance and a preferred region for any observed signal
- Anything else?

Limit setting

- Inputs:
  - Vertex Z vs. mass
  - Efficiency vs. Z at each A' mass (from MC)
  - A' vertex distributions (production×decay×efficiency) as a function of mass and coupling
- First step: take a resolution-limited mass slice (2.5 σ_m wide), get a 1-D data set
- Can take sideband slices, fit those and interpolate to get a background model for the slice
  - Gaussian with exponential tail seems to work
  - What if that's not all there is to the tail?
- Then what?



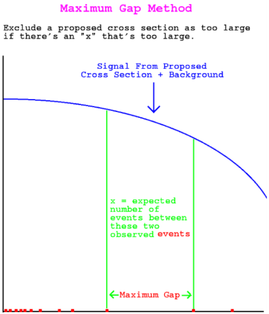


Cut-and-count

- Define a zcut such that we expect b events above the cut, count events above the cut, get a confidence interval for the signal rate past the zcut: makes no assumptions about signal shape
  - Feldman-Cousins intervals?
  - What's optimal b? 0.5? Does the answer change if there are irreducible backgrounds?
- For each $c\tau$, this sets an interval for the production cross-section
- Check for consistency with the appropriate ϵ^2

Optimum interval (Yellin, arXiv:physics/0203002)

- Sets a limit on A' production cross-section for a given $c\tau$
- Assume a production cross-section, then the expected trident distribution (background plus A's) dN/dz is known: do a change of variables from z to N, so the expected distribution is uniform in N and has total length μ (= total expected events)
- For every interval between two events (of width x expected events and containing n observed events), compute the probability $C_n(x, \mu)$ given the expected distribution that all intervals containing n events are narrower than this one; take the interval with largest C to be the "optimum interval" that most strongly rejects the proposed cross-section; that value of C is called C_{max}
- You pay a statistical penalty for picking the best interval and the best n; if you want a 90% confidence interval on the cross-section, the C_{max} threshold for excluding a cross-section is larger than 0.9
- If there's an unexpected background (or one where we just don't know the spatial distribution), the optimum interval automatically avoids it

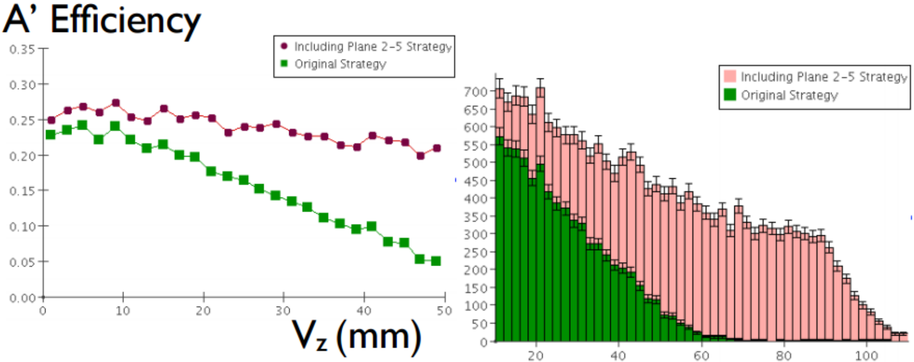


We found a signal . . . now what?

- Do a profile likelihood ratio to get a local p-value for a nonzero signal at given mass and $c\tau$
- Also get a p-value for consistency with the predicted production cross-section
- Significance: toys should work (since any signal strength we find is likely to be weak — we're not going to find a 5σ signal)
  - The "upcrossings" method of estimating the look-elsewhere effect for strong signals is specific to 1D searches, but can be generalized to 2D (arXiv:1105.4355)
- Look at esum distribution? Can we compare goodness of fit with full-diagram and A'?

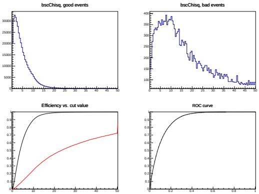
Events without L1 hits

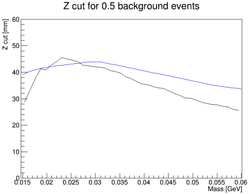
- We need these to get acceptance past $Z = 50$ mm or so; on the other hand no-L1 events have significantly worse mass and vertex resolution, and tails that extend an additional 10-20 mm
- I think we need to split the data set
- Can't trust that L1 efficiency is the same in data and MC: how do you split A' production between data sets?
  - Use total trident rate (all tracking strategies, double-counting excluded) in both data sets?
- Need to combine L1 and no-L1 datasets: doable for both cut-and-count and optimum interval methods

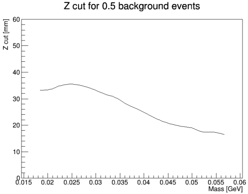


Setting cuts

- Vertex tails mostly come from L1 scatters
- Using data to test cuts: reject high-Z fakes (“background”), keep low-Z tridents (“signal”)
- Need to check with MC to make sure these cuts don’t reject high-Z A’, and maybe find more variables to cut on
- Should check mass dependence of cuts
- Eventually (soon) we will need to use the analysis to tune the cuts for best sensitivity





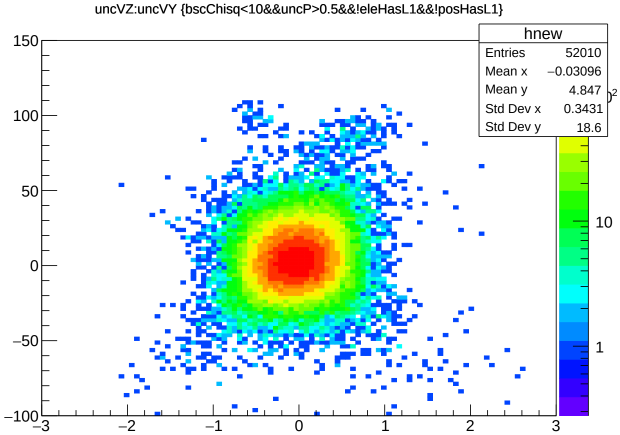


Software, MC, other stuff

- A' MC: Brad has made displaced vertex A'
  - Only making $c\tau = 1$ mm (for now?), since I just want to get good coverage in Z for getting efficiencies
- Big sample of trident MC: it'd be nice to have an MC sample of the same size as the full data set (current MC is about 1/5 the lumi of the unblinded 10%)
  - May be the only way we're going to see (and be prepared to cut out) a small background, O(10) events
  - But we'll never know if it has everything
- Unblinding $Z < 0$: no reason not to
  - May tell us something about vertex tails
  - But maybe there are too many differences

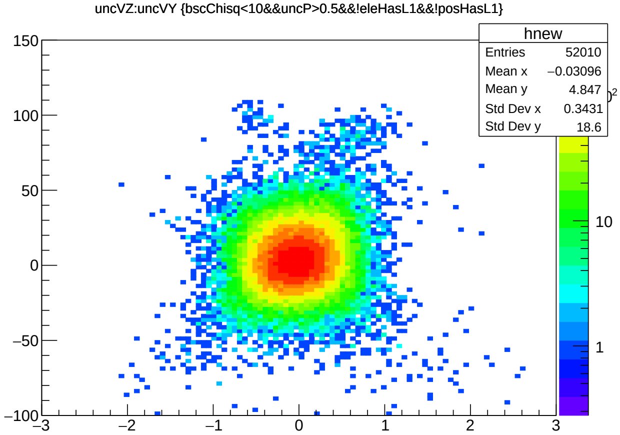
What I'm worried about

- Background from L1
- What if there's background not accounted for in our model?
- The yield is too low (lower than expected, too low to set limits)



Known sources of tridents at large Z

- Beam-gas
  - Expected to be a few events at most; should get a real estimate from Takashi
- Beam halo/scattered beam from target hitting L1
  - We see tridents being made in inactive L1 silicon: beam halo electrons, or electrons scattered by the target
  - This will be O(2000) tridents in the full data set, but concentrated near Z=100 mm: hopefully we can reject these without throwing out too much of our A' space

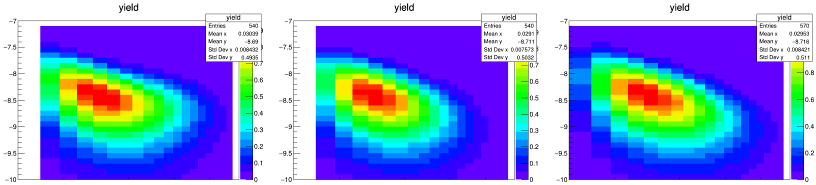


What if we have an unexpected background?

- Suppose we have a background at large Z, slowly varying with mass
  - Real vertices, like beam-gas tridents
  - Fake vertices, like some kind of subexponential vertex tail
- It's not obvious now in 10% of the data, so if we have something like this it's going to be small — O(50) events, O(5) events in a mass slice
- Count events past zcut, fit to a low-order polynomial? Could this improve our limit?
  - Doesn't work very well for edges of the mass distribution (similar problem to bumphunt)
  - Seems dangerous, too easy to really screw things up
  - I'd prefer to take the hit to our limit, but use this when calculating significance (we can quote signal significance over known low-Z backgrounds and over low-Z+high-Z backgrounds)

Yield

- Yield (A' past zcut) $\log \epsilon^2$ vs. mass (15-60 MeV)
- Assuming A' efficiency constant to $z = \infty$ (obviously not true), radiative fraction 15%
- Left to right:
  - Real data (L1 required, reasonable cuts), zcut derived from data
  - Real data (L1 required, reasonable cuts), zcut 35 mm
  - Loose data (L1 not required, loose bumphunt-style cuts), zcut 35 mm
- Not good enough! That's way under 2σ even with very optimistic assumptions



Why is the yield so low?

- Matt G. estimated 6σ significance at 3 PAC days; we got 1.9
- The trident yield in data is lower than MC, but less than a factor of 2
- We should figure out what's wrong