A New Method for Resolving Combinatorial Ambiguities at Hadron Colliders

Felix Yu

UC Irvine

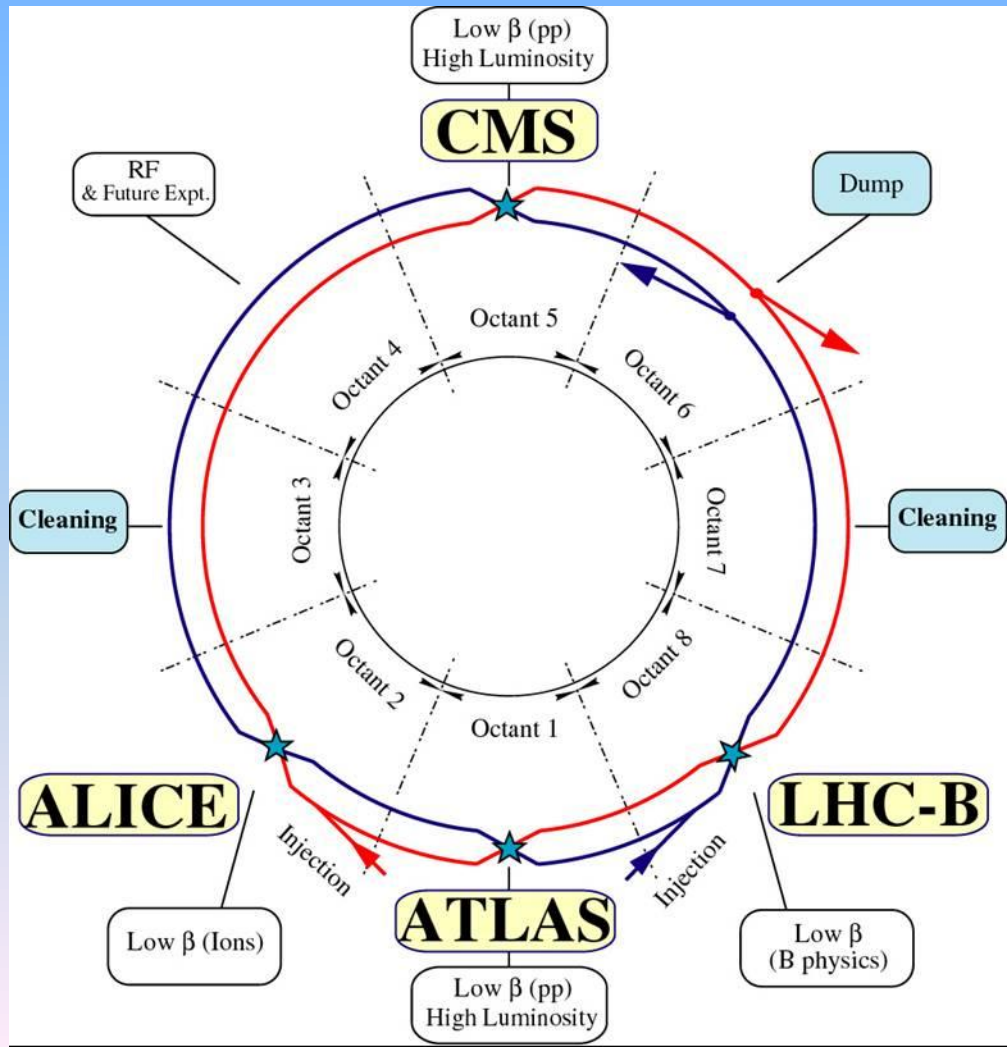
arXiv: 1009.2751 with A. Rajaraman

LHC Theory Initiative Fellows Meeting, Brookhaven National Lab

October 8, 2010

The LHC Era is here!

- The current 7 TeV run will finish in 2011 with ≈1 fb^{-1} (Marc-Andre Pleier's talk)
- Potential for new physics (cf. Workshop on Topologies for Early LHC Searches, SLAC, Sept. 22-26, 2010, Mariangela Lisanti's talk)
- One hope is for the direct production of dark matter



Motivation

- Collider signals of SUSY, UED
  - Cascade decay chains ending in LSP, LKP
  - Z_2 symmetry ensures the LSP, LKP is stable
    - If neutral, is usually a good dark matter candidate
  - Z_2 symmetry ensures the kinematics of the event are not readily measurable
- Many kinematic methods have been developed to aid determination of the LSP, LKP mass

Motivation

- Kinematic edges
  - Thresholds and maximums of invariant mass distributions provide algebraic expressions relating underlying cascade decay chain masses
- Polynomial method
  - Solve momentum conservation equations
  - Use non-linear constraints to solve for LSP, LKP momenta
- m_{T2}, m_{CT2} method
  - m_{T2} Kink
  - Sub-system m_{T2}
  - m_{T2} as a discovery variable

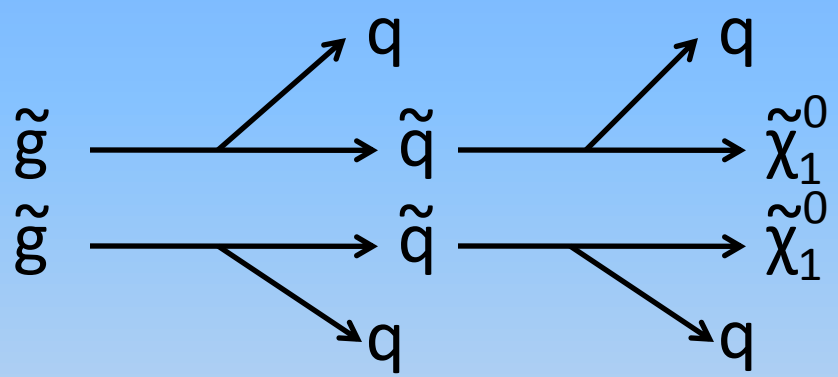
Baer, Chen, Paige, Tata, Hinchliffe
Allanach, Lester, Parker, White
Barr, Gripaios, Nojiri, Cheng, Gunion, Han, Marandella, McElrath, …

Motivation

- Need the LSP, LKP mass to determine the mass scale
  - Differences in masses are easy
- Usual procedure to solve for LSP, LKP mass:
  1. Use cuts to isolate some collection of objects (*e.g.* 4 jets inclusive + 0 leptons + MET)
  2. Hypothesize an underlying decay chain topology (*e.g.* pair-produced gluinos decaying to neutralinos via squarks)
  3. Assign objects to decay chains (ordered, if possible)
  4. Apply your favorite mass reconstruction technique

Motivation

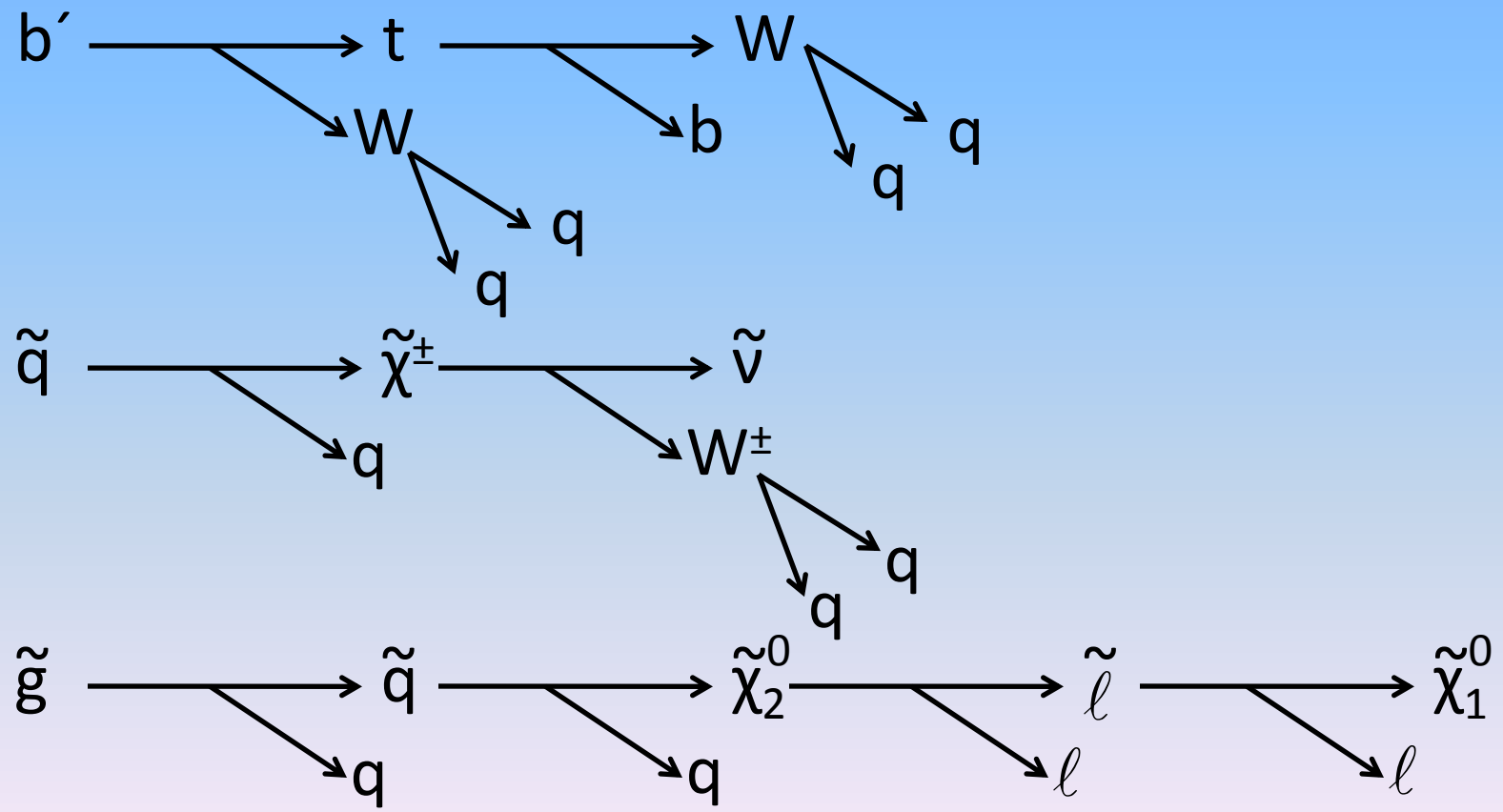
- Decay chain assignments must deal with combinatorial ambiguities



- For pair-produced gluinos decaying to LSPs via squarks, we have 4 quarks that can be grouped into 3 pair-pair combinations
  - Important note: this combinatorial ambiguity is present even if dealing with only signal events

Motivation

- Other possibilities: in busy cascade decay chains, it is important to associate particles correctly
- Usually, there are additional tools to eliminate wrong combinations



Outline

- Motivation
- Our Simple Models
- Brief Review of Basic Kinematics
- The well-studied hemisphere method
- The new p_T v. M method
- Comparison
- Conclusions

Models

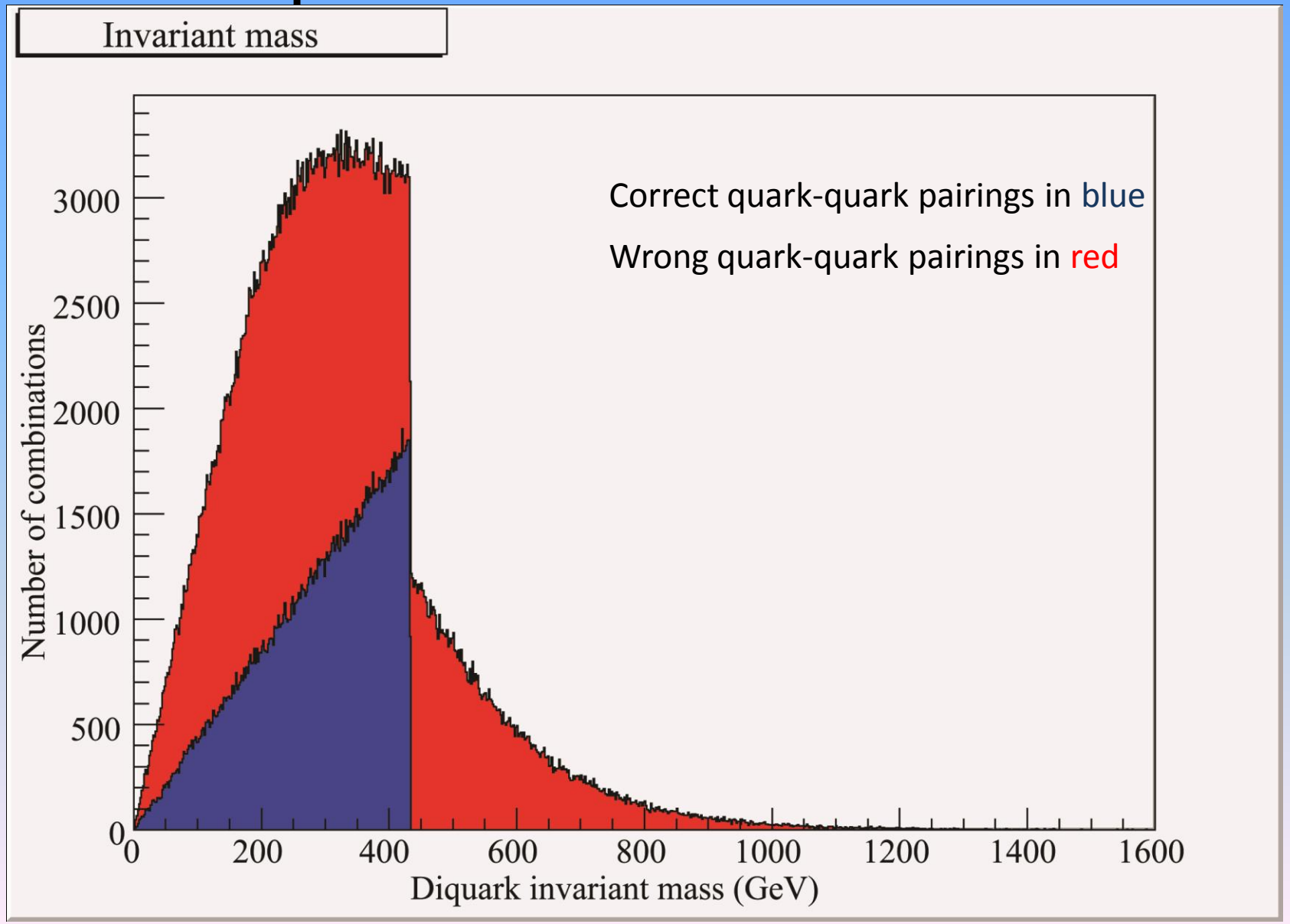
- Consider gluino pair-production in a 7 TeV LHC (and 14 TeV) LHC, both decay identically via on-shell or off-shell squarks to LSP neutralino

	Gluino Mass	Squark Mass	Neutralino Mass	Kinematic Edge
Model A	600 GeV	400 GeV	100 GeV	433 GeV
Model B	600 GeV	800 GeV	100 GeV	500 GeV

- Generate 100,000 events using MadGraph/MadEvent 4.4.26, decay using BRIDGE
  - No ISR/FSR, hadronization (see talks by Jonathan Walsh, Stefan Hoche, David Krohn, Duff Neill for how great a simplification this is)
  - Only consider parton level

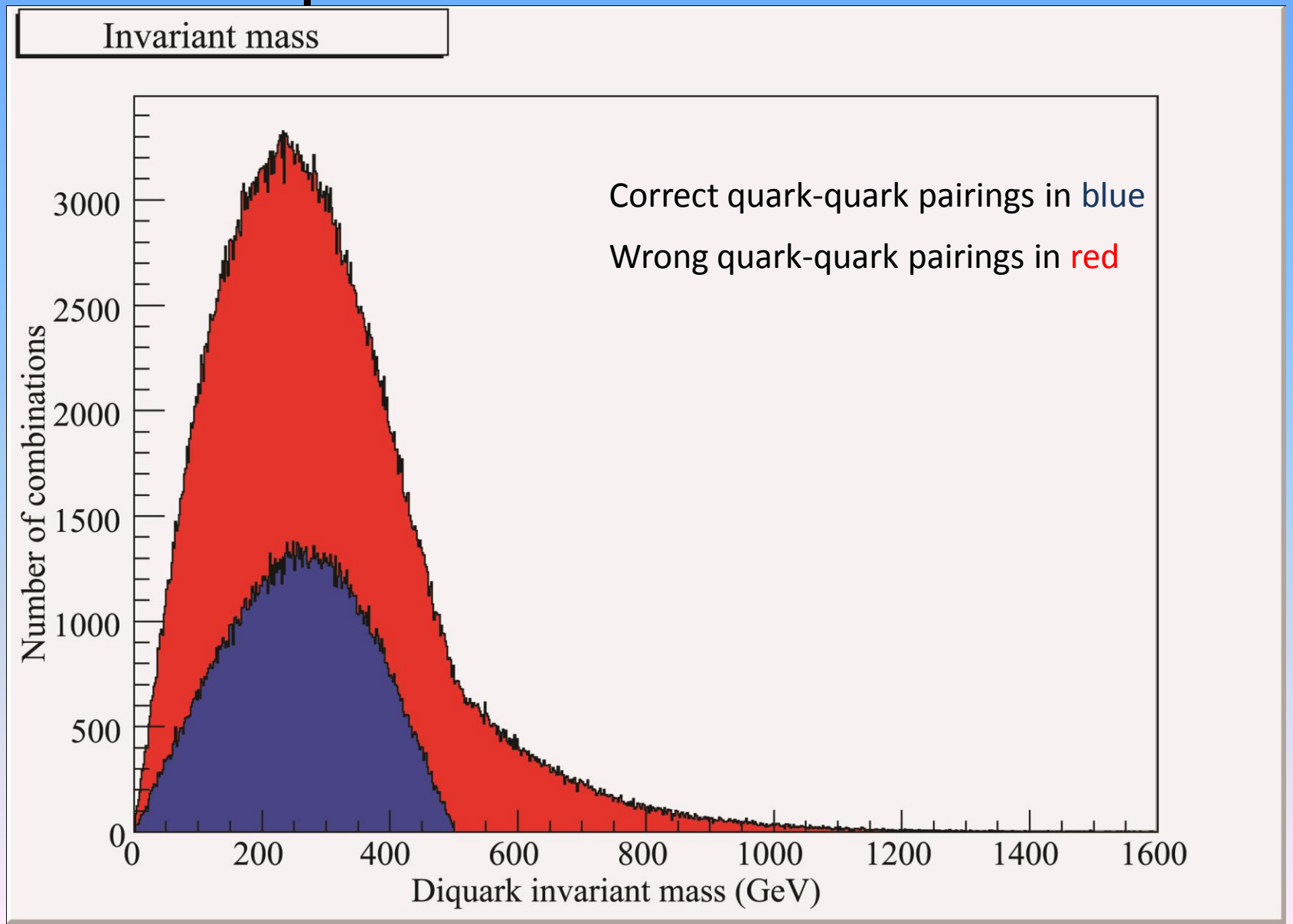
Kinematic edge in invariant mass

On-shell squark scenario



Kinematic edge in invariant mass

Off-shell squark scenario



The Hemisphere Method

- Basic idea: divide an event into hemispheres where each decay chain falls entirely into separate hemispheres
- Two steps
  - Step 1: Choose 2 seeds
    - These are the central axes for the hemisphere clusters
  - Step 2: Cluster remaining objects with the given seeds
    - Figure of merit is minimum pdR: roughly, a momentum-weighted angular separation

      $pdR \equiv (|\Delta p|\ \Delta R)$, where $\Delta R \equiv \sqrt{[(\Delta\phi)^2 + (\Delta\eta)^2]}$

The Hemisphere Method

- Our implementation

1. Choose highest p_T object as seed 1

2. **PDR1**. Of the remaining objects, choose the maximum pdR object as seed 2

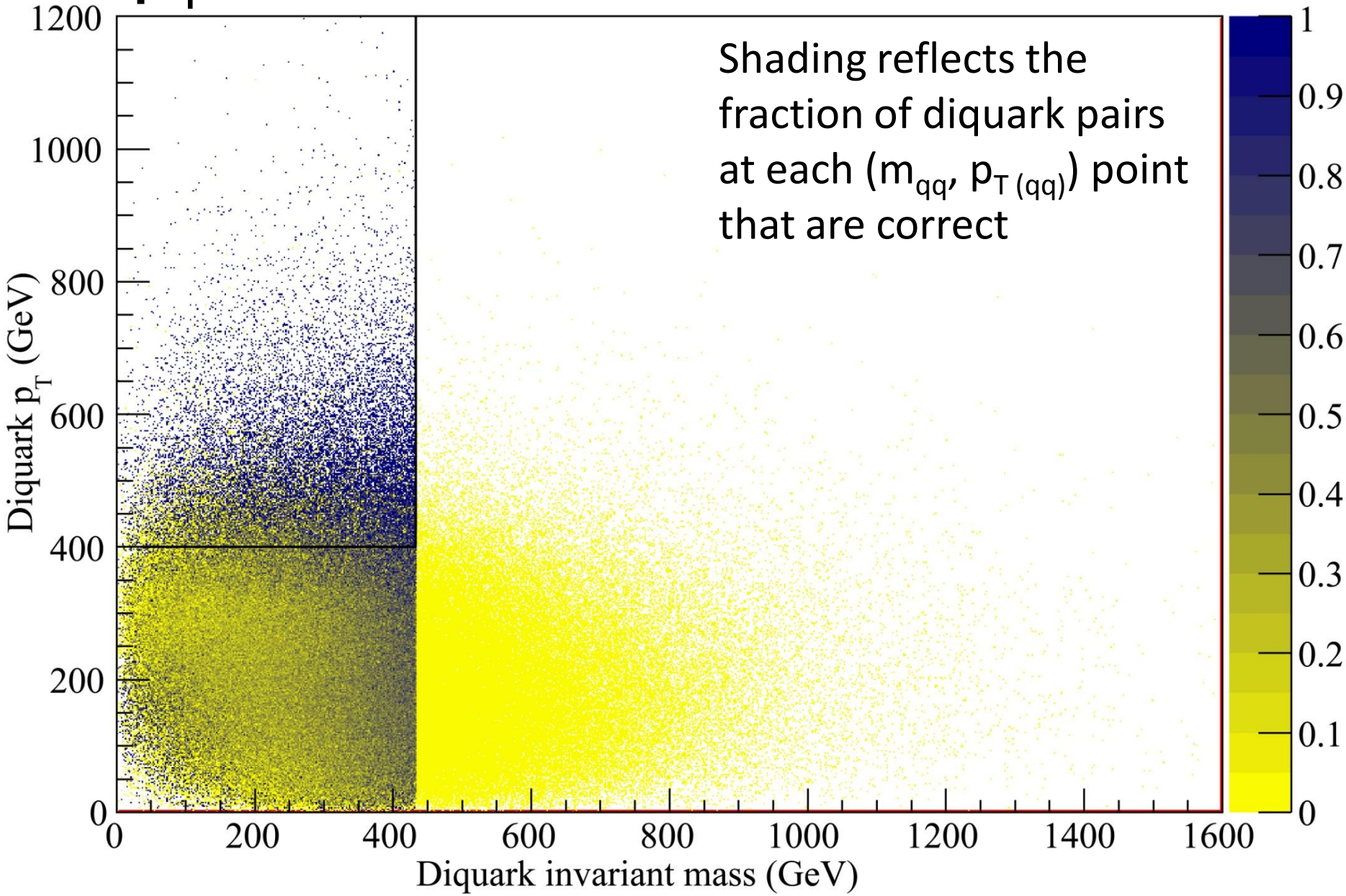
   PDR2. Of the remaining objects, choose the maximum invariant mass object as seed 2

3. For each remaining objects, calculate pdR w.r.t seed 1 and w.r.t. seed 2. Cluster the object with whichever seed has the smaller pdR, *i.e.* the closer seed in momentum-weighted angle space.

The New p_T v. M Method

- Plot p_T v. M for each qq pair of the event
  - Each event has 6 unique qq pairs; 2 pairs are correct, 4 pairs are wrong
  - The 6 qq pairs can be grouped into 3 unique pair-pair combinations, one of which is correct
  - We want to isolate the correct pair-pair combination
- Observe excesses at high invariant mass (wrong diquark pairs) and high p_T (correct diquark pairs)

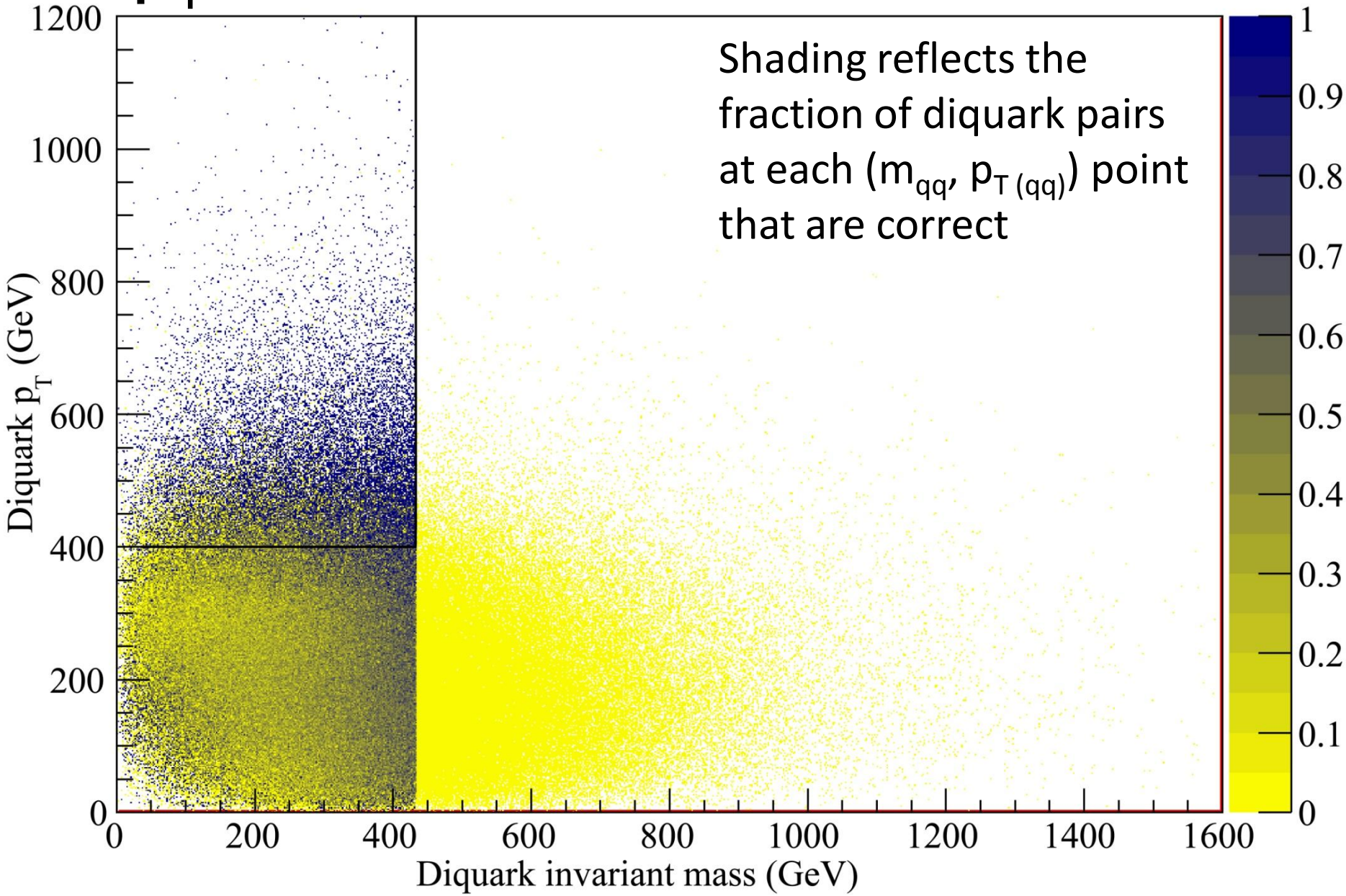
The p_T v. M method – Model A 7 TeV



The p_T v. M Implementation

- Plot p_T v. M for each qq pair of the event
  - If possible, observe the invariant mass edge
  - For the (wrong) pairs with invariant masses larger than the edge value, gradually increase the p_T cut such that the survival rate of pairs drops below 5%
  - Extrapolate this cut to the upper left region with high p_T and low invariant mass
    - This region will characteristically have high purity, *i.e.* be dominated by correct diquark pairs
    - Only use pair-pair combinations where both diquark pairs lie in the boxed region

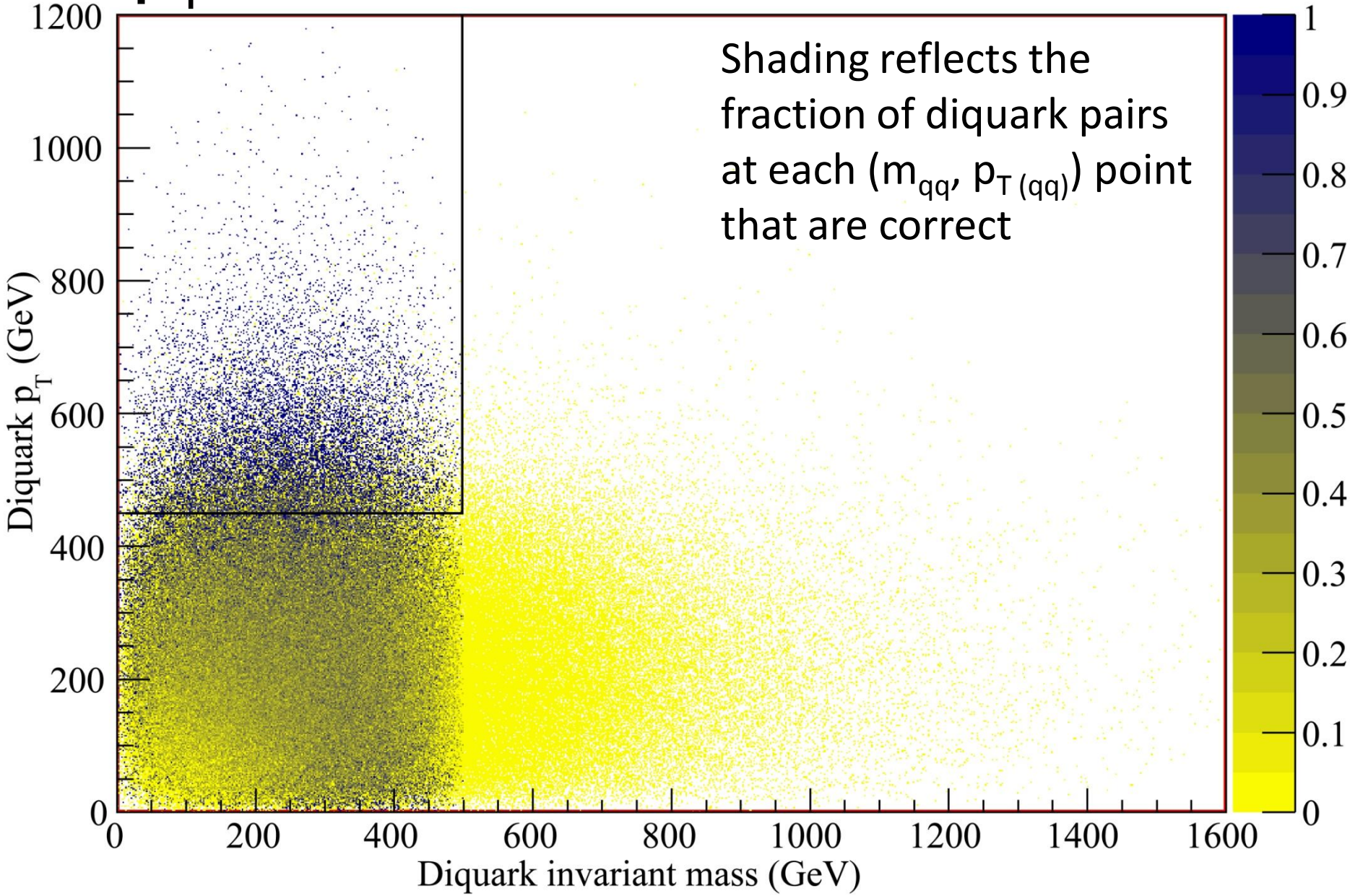
The p_T v. M method – Model A 7 TeV



Survival Probability

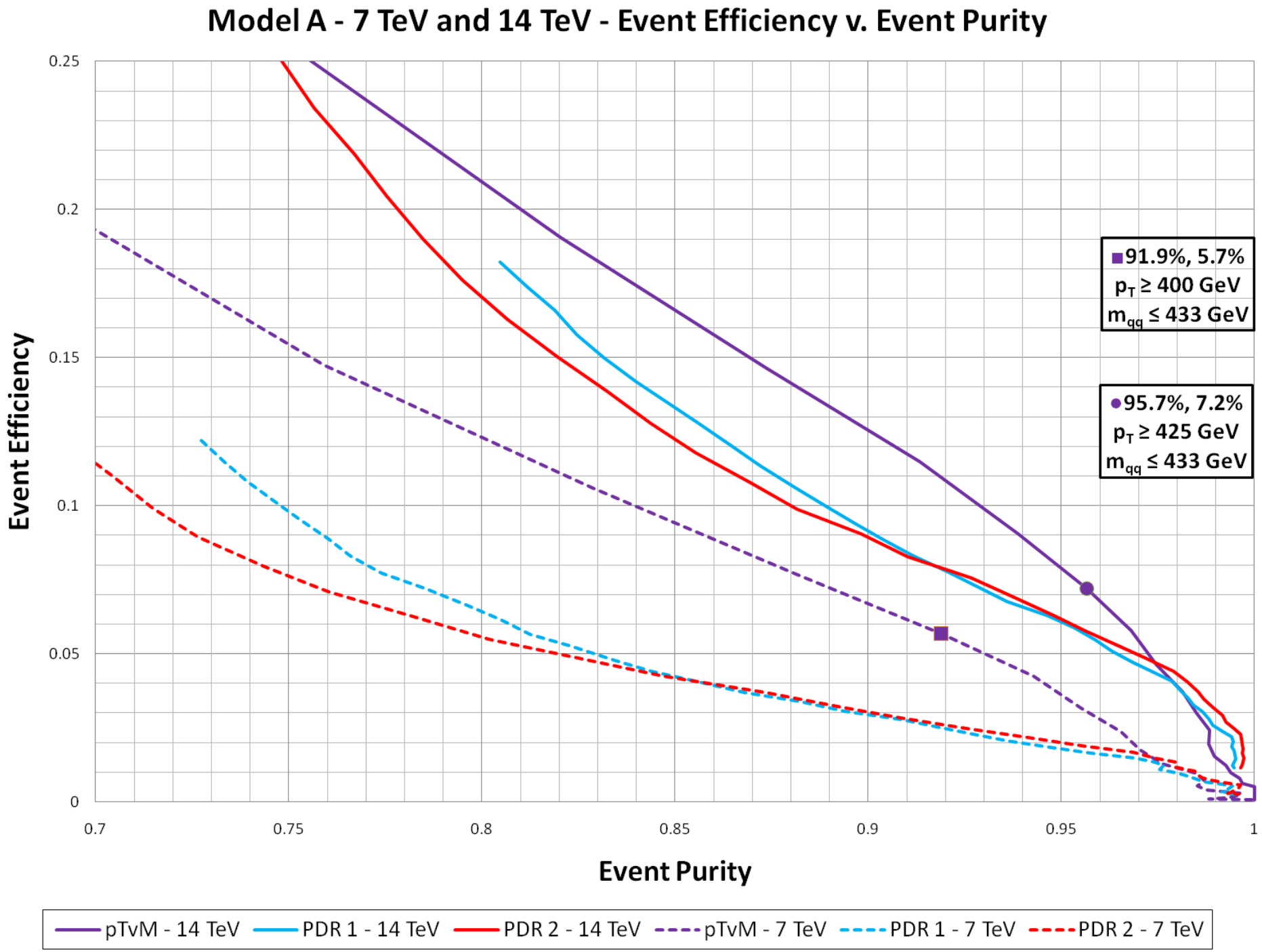
Cut on min p_T	Survival %
0	100.000%
25	98.525%
50	94.402%
75	88.347%
100	80.857%
125	72.162%
150	62.948%
175	53.796%
200	44.963%
225	36.548%
250	29.014%
275	22.481%
300	16.960%
325	12.398%
350	8.982%
375	6.275%
400	**4.378%**
425	3.007%
450	2.077%
475	1.427%
500	0.994%

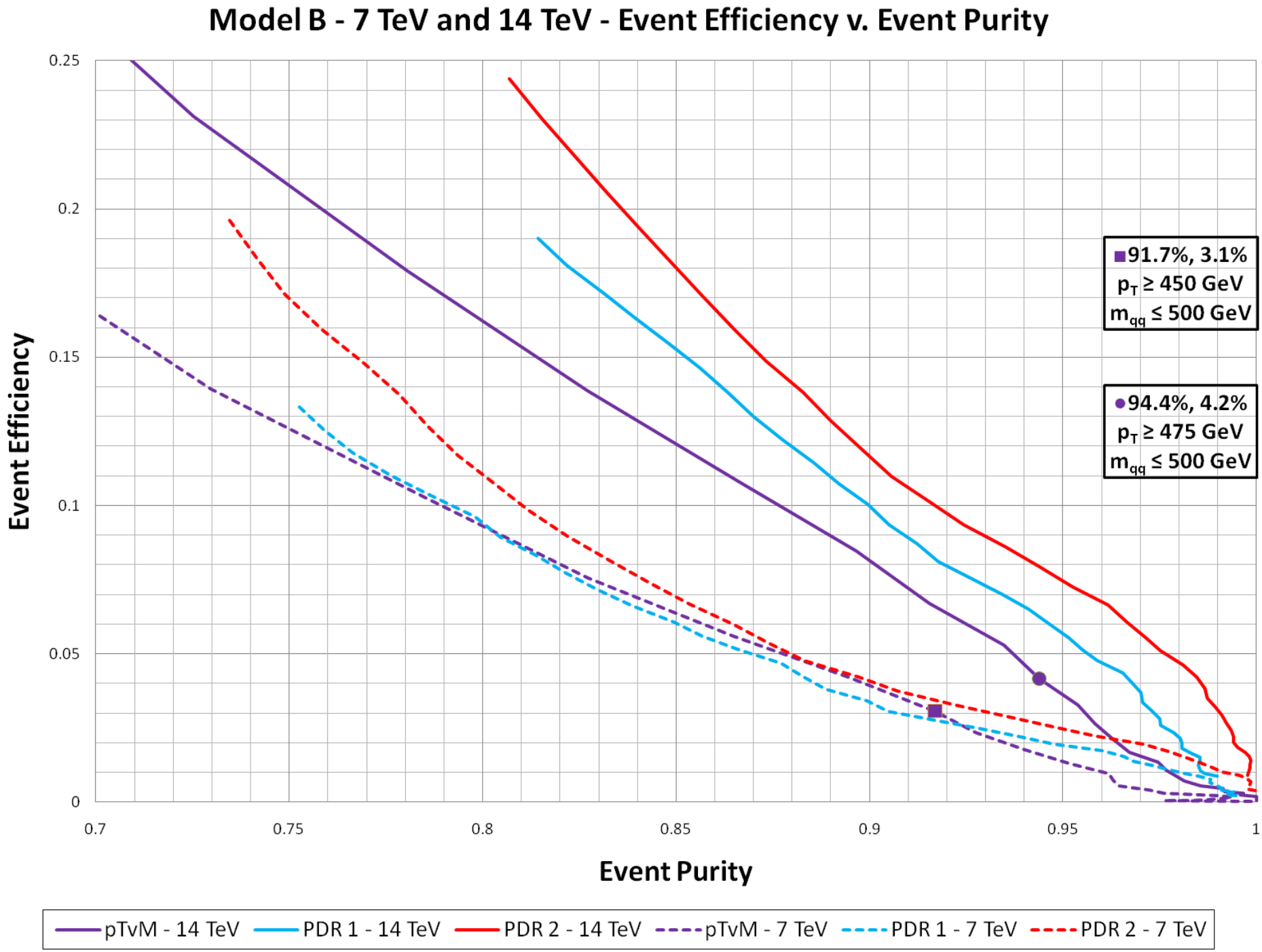
The p_T v. M method – Model B 7 TeV



Comparison between the hemisphere method and the p_T v. M method

- Will use event efficiency v. event sample purity
  - Event efficiency is the percentage of events that pass cuts
  - Event purity is the percentage of remaining events that are fully correctly assigned
- Variable cut
  - Hemisphere method: Vary the object-seed pdR difference
    - Example: For a given object, its pdR w.r.t seed 1 is 500 GeV, and its pdR w.r.t seed 2 is 505 GeV. By the pdR measure, it could equally well be clustered with seed 1 or 2. By imposing an increasing pdR difference cut (a minimum difference in pdR between an object and the two seeds), we can gradually eliminate these ambiguous assignment scenarios.
  - p_T v. M method: Vary the survival probability in the (wrong) diquark pairs region at high invariant mass
    - Equivalently, require a higher p_T cut


Model A - 7 TeV and 14 TeV - Event Efficiency v. Event Purity
Event Efficiency
Event Purity
0.25
0.2
0.15
0.1
0.05
0
0.7
0.75
0.8
0.85
0.9
0.95
1
91.9%, 5.7%
$p_T \geq 400$ GeV
$m_{qq} \leq 433$ GeV
95.7%, 7.2%
$p_T \geq 425$ GeV
$m_{qq} \leq 433$ GeV
pTvM - 14 TeV
PDR 1 - 14 TeV
PDR 2 - 14 TeV
pTvM - 7 TeV
PDR 1 - 7 TeV
PDR 2 - 7 TeV


Model B - 7 TeV and 14 TeV - Event Efficiency v. Event Purity
Event Efficiency
Event Purity
0.25
0.2
0.15
0.1
0.05
0
0.7
0.75
0.8
0.85
0.9
0.95
1
91.7%, 3.1%
$p_T \geq 450$ GeV
$m_{qq} \leq 500$ GeV
94.4%, 4.2%
$p_T \geq 475$ GeV
$m_{qq} \leq 500$ GeV
pTvM - 14 TeV
PDR 1 - 14 TeV
PDR 2 - 14 TeV
pTvM - 7 TeV
PDR 1 - 7 TeV
PDR 2 - 7 TeV

Discussion

- On-shell decay chains
  - p_T v. M is significantly better than the hemisphere method in retaining more event efficiency for a given event purity
- Off-shell decay chains
  - p_T v. M is marginally to moderately worse than the hemisphere method
    - Possibly because of the flexibility in choosing second hemisphere axis

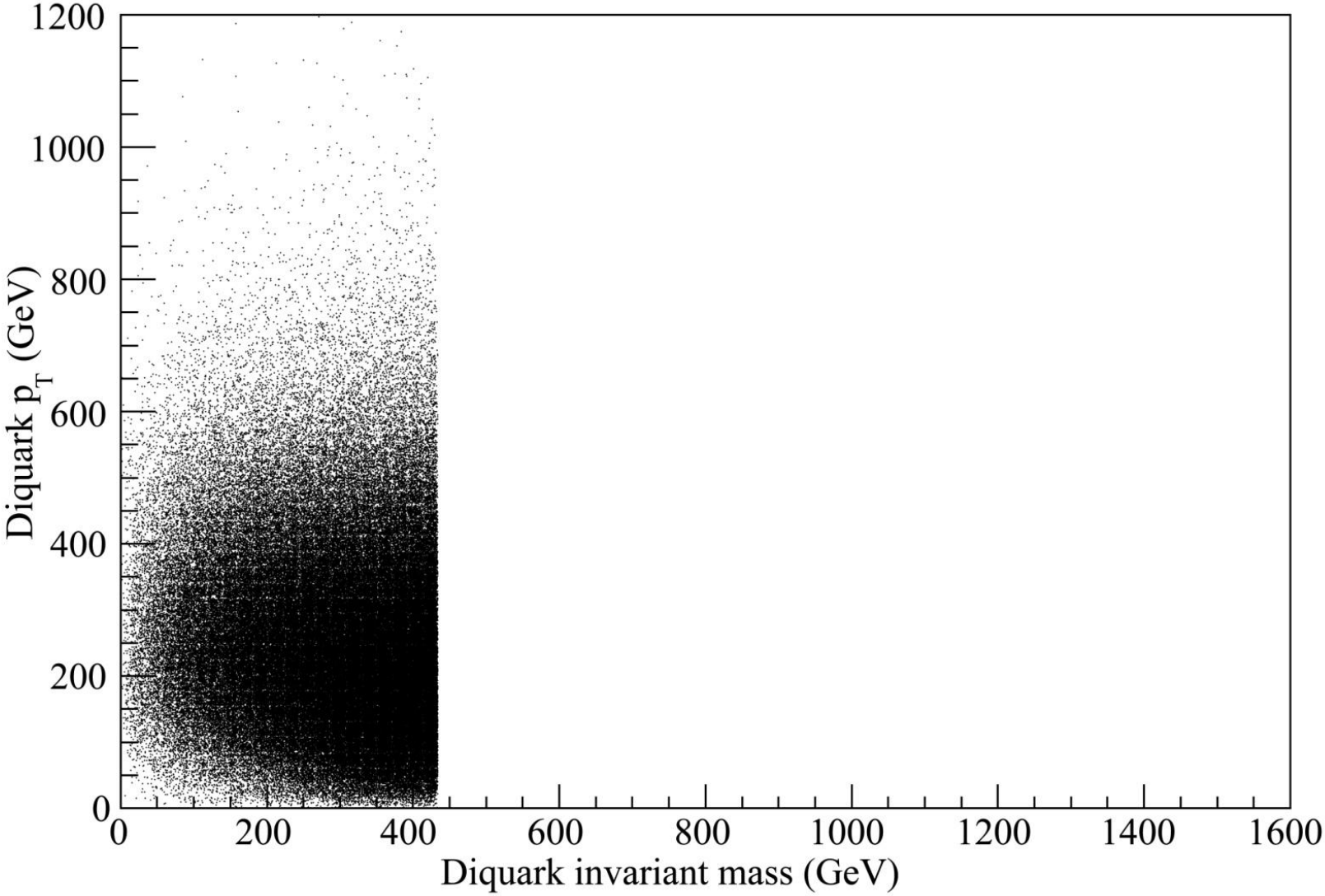
Future Work

- Apply p_T v. M at reconstruction level
  - SPS1a including ISR/FSR, detector simulation
- Perform shape analysis of p_T v. M
  - Optimize the p_T v. M cut
- Reorganize p_T v. M to be an event-by-event variable
- Perform a detailed study of p_T v. M and the hemisphere method in off-shell cases
  - Should use both in parallel since it is not known a priori whether the decay chain is on-shell or off-shell

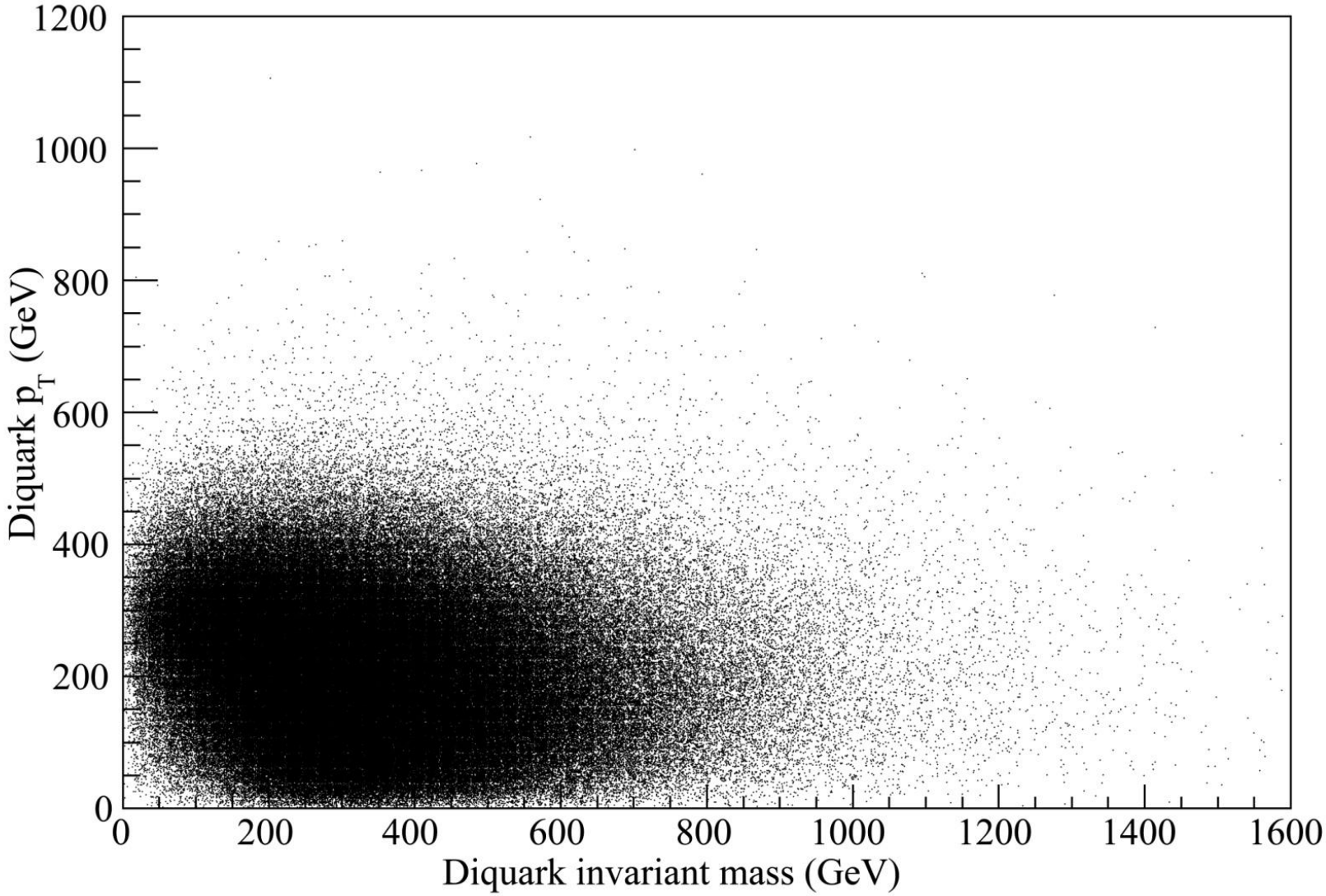
Conclusions

- Distinguishing combinatorial ambiguities is important for new physics searches at the LHC
- The p_T v. M method is better than the hemisphere method for on-shell decay chains in delivering high purity event samples
  - The hemisphere method is better suited for off-shell decay chains
- The p_T v. M method is easy to implement and flexible
  - A “robustness” study in a simulated collider environment is underway

The p_T v. M method – Model A 7 TeV – **Correct** Diquark Pairs Only



The p_T v. M method – Model A 7 TeV – **Wrong** Diquark Pairs Only



Kinematic edge in invariant mass

- An oft-used feature of cascade decay chain kinematics is the invariant mass edge
- On-shell squark – characteristic triangular shape

$$m_{qq} = m_{qq}\big|_{edge}\sqrt{\frac{1+\cos\theta^*}{2}}$$

$$m_{qq}\big|_{\text{edge}} = \sqrt{\frac{(m_{\tilde{g}}^2 - m_{\tilde{q}}^2)(m_{\tilde{q}}^2 - m_{\tilde{\chi}_1^0}^2)}{m_{\tilde{q}}^2}}$$

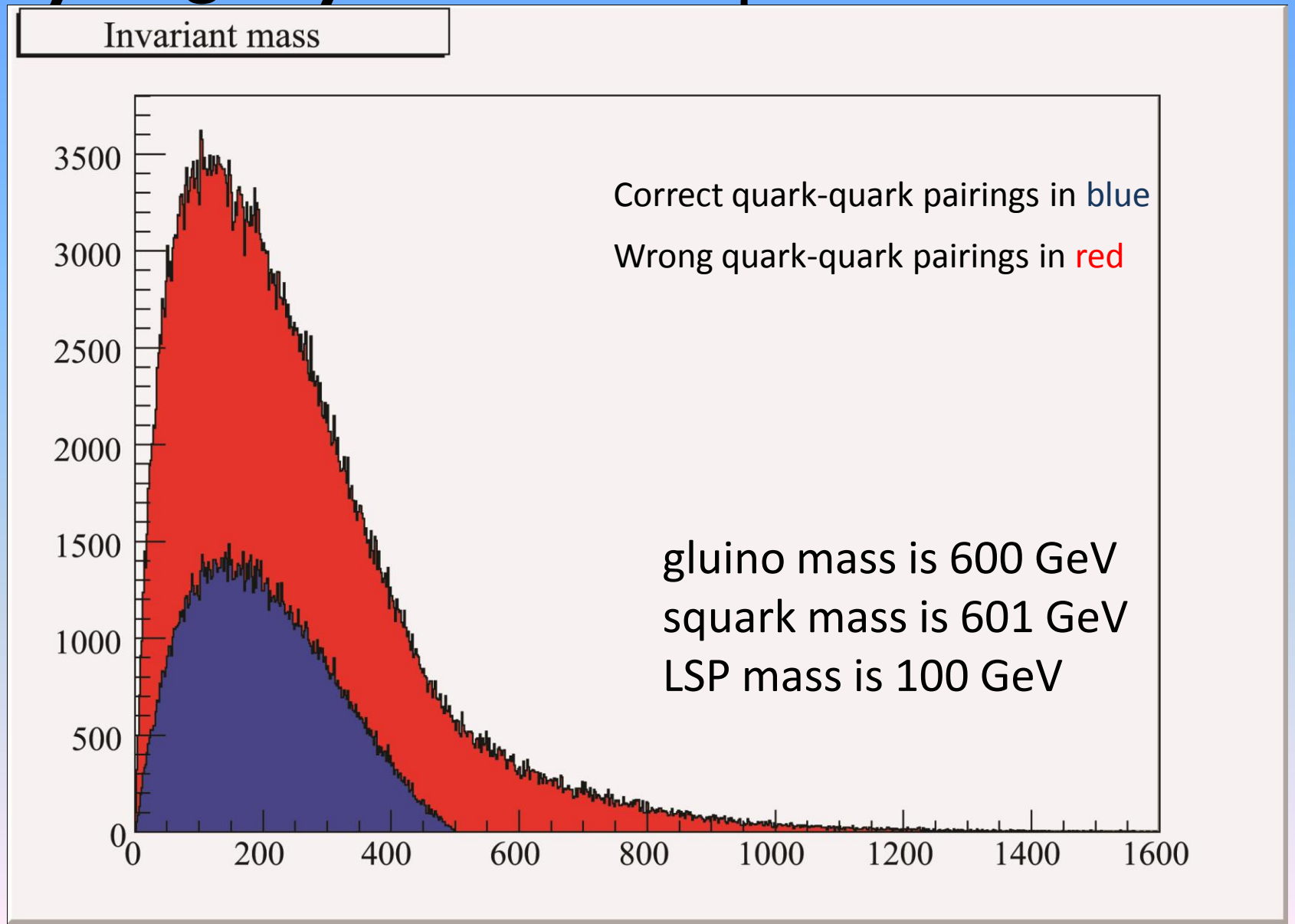
Θ^* is the angle of the second quark in the gluino rest frame

- Off-shell squark – small number of events near edge

$$m_{qq}\big|_{\text{edge}} = m_{\tilde{g}} - m_{\tilde{\chi}_1^0}$$

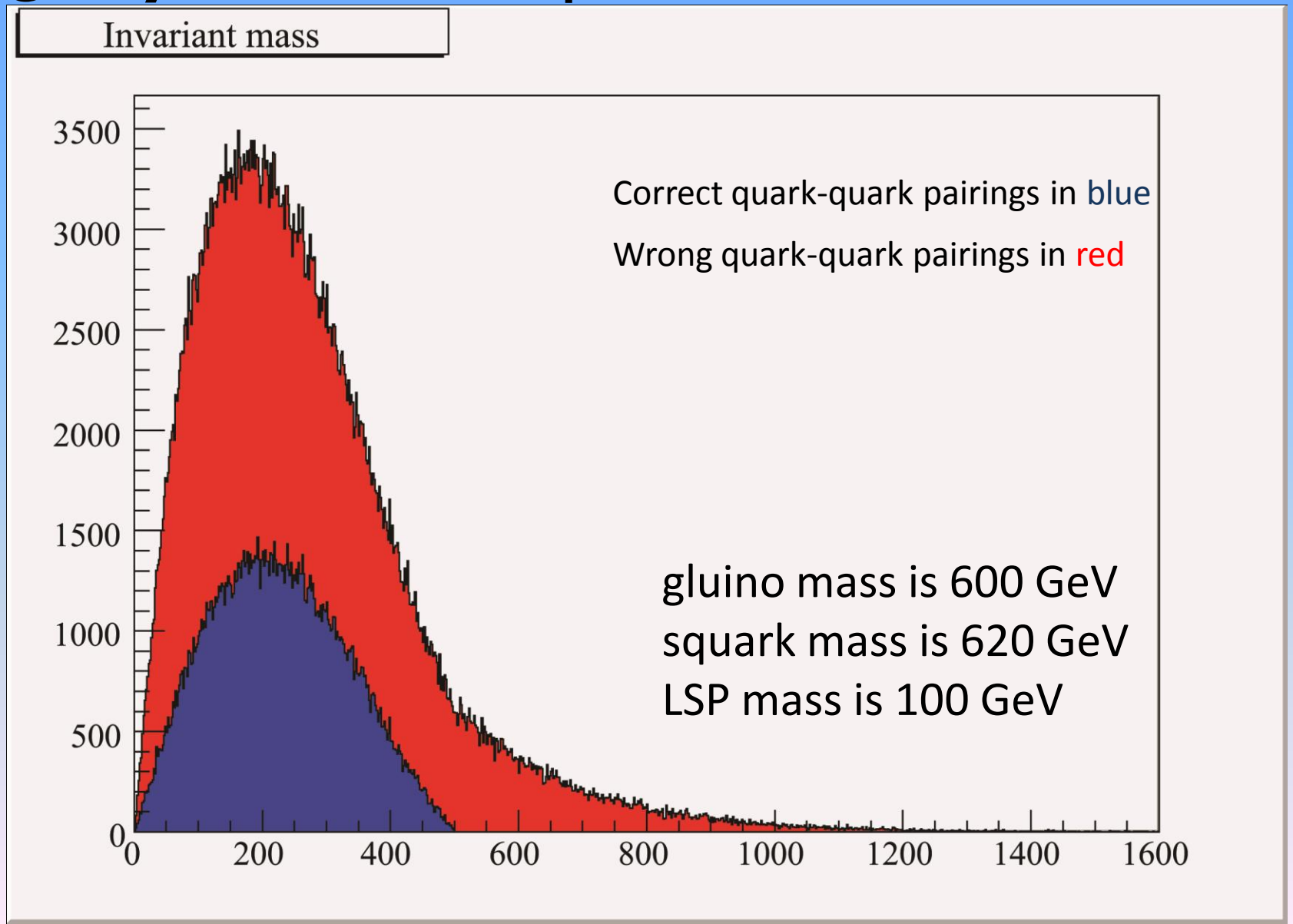
Kinematic edge in invariant mass

Very slightly off-shell squark scenario

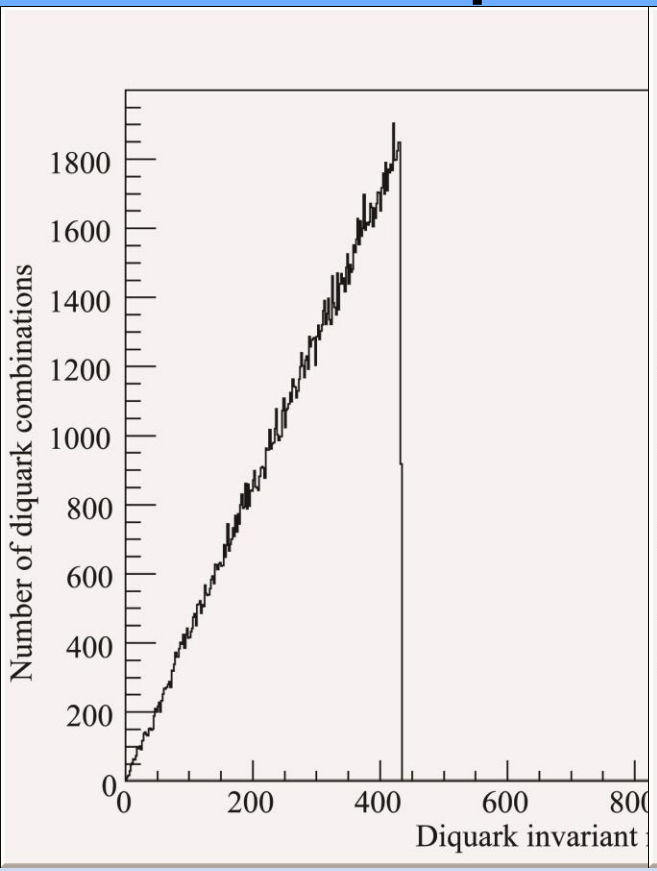


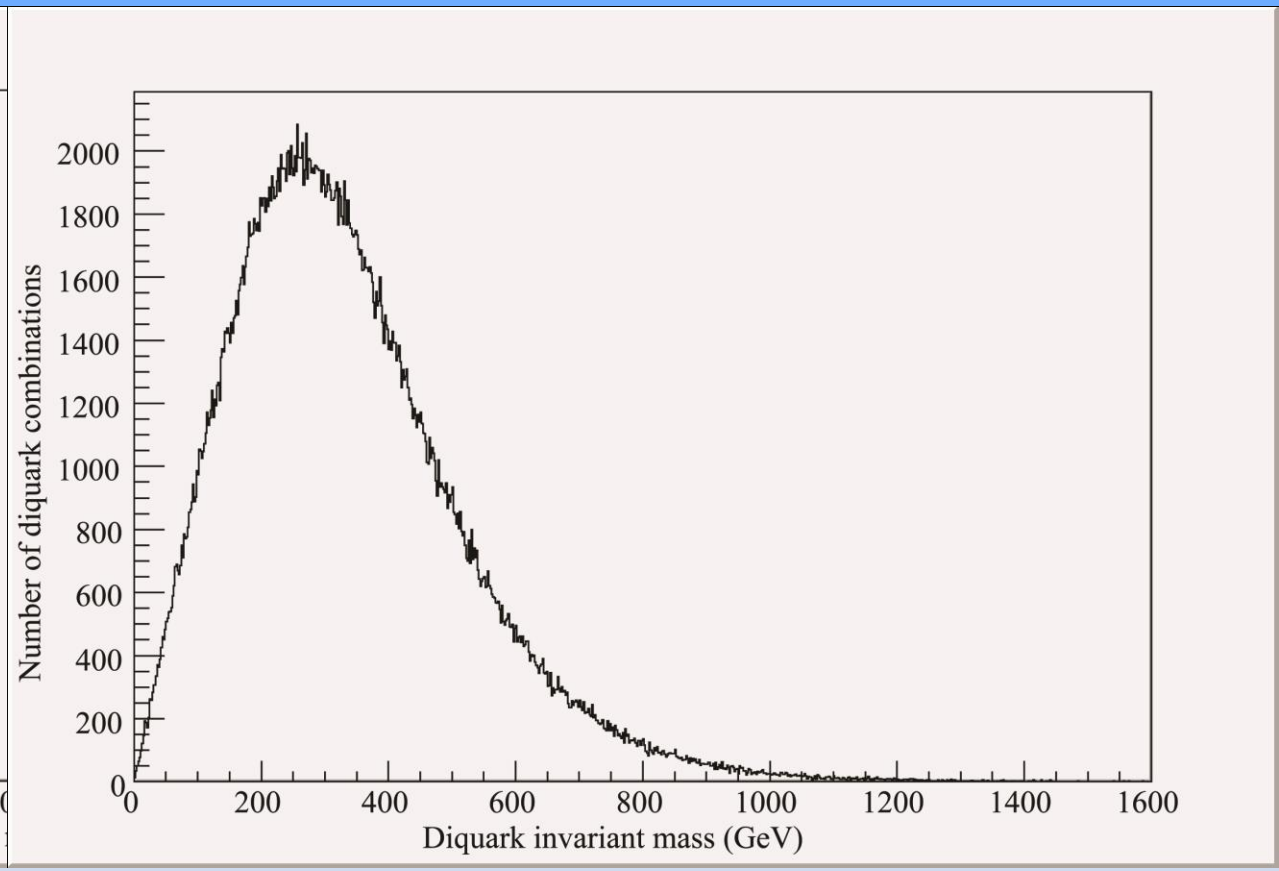
Kinematic edge in invariant mass

Slightly off-shell squark scenario

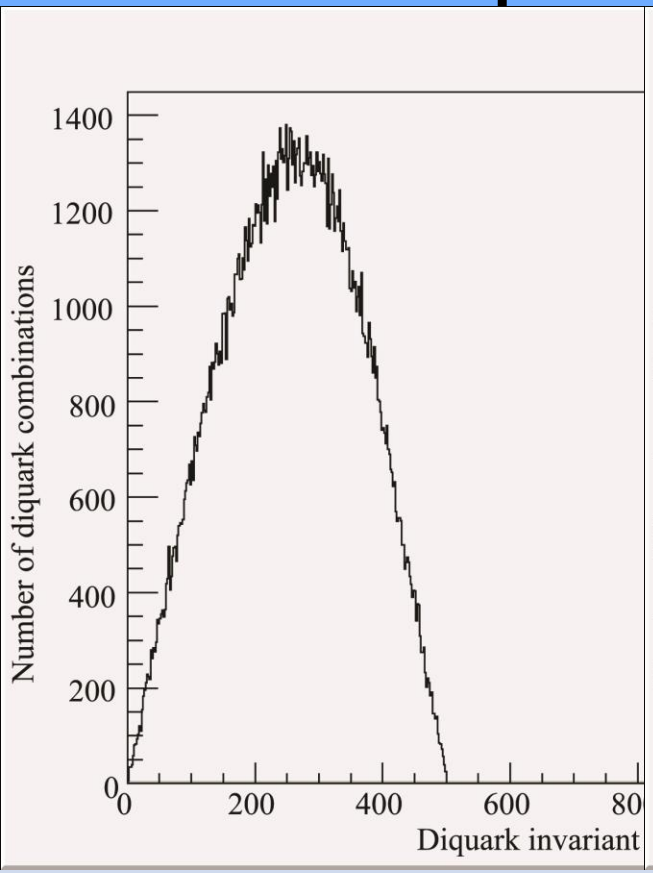


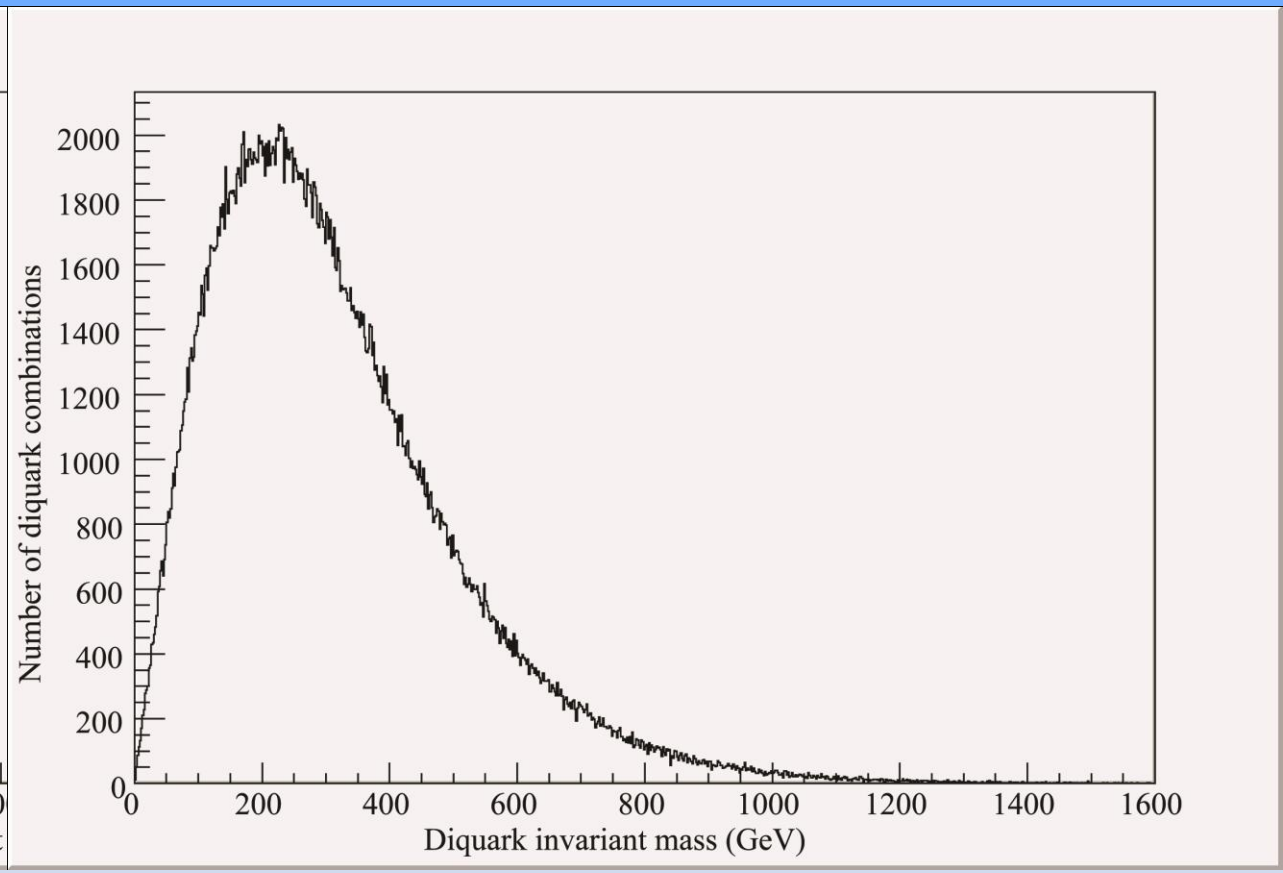
Kinematic edge in invariant mass
On-shell squark scenario





Kinematic edge in invariant mass
Off-shell squark scenario





The Hemisphere Method

- Our cuts

  Cut 1. The highest p_T object must have $p_T \geq 200$ GeV

  PDR1 Cut 2. The minimum pdR between seed 1 and seed 2 must be 1800 GeV

  PDR2 Cut 2. The invariant mass of seed 1 and seed 2 must be larger than the kinematic edge value

  Cut 3. Discard all singlet-triplet events

  Cut 4. The maximum seed-object invariant mass must be less than or equal to the kinematic edge value

The Hemisphere Method

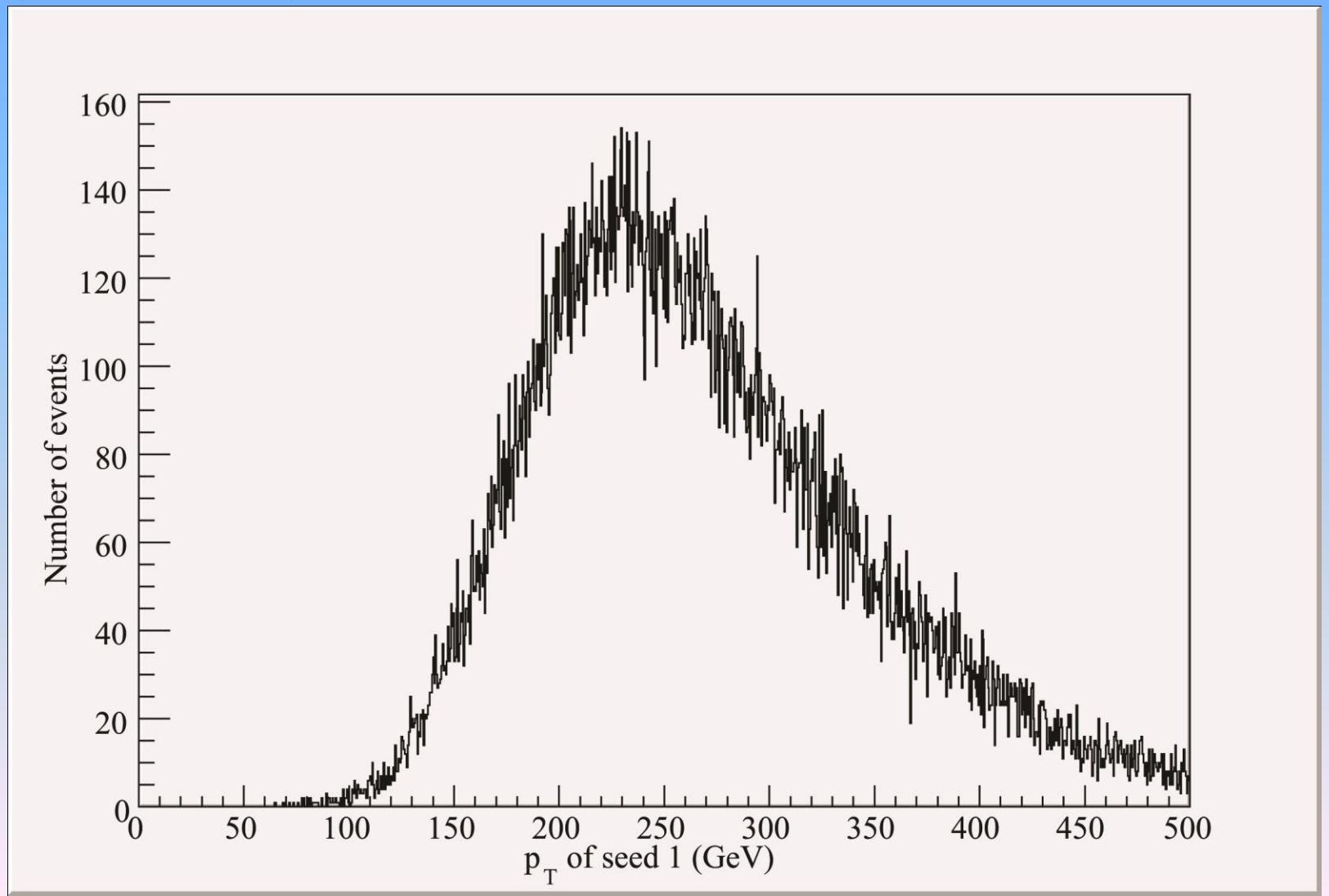
- Our cuts do not include realistic detector cuts (η-acceptance, minimum p_T, isolation requirements)

PDR 1 Cut Performance	Cut 1	Cuts 1-2	Cuts 1-3	Cuts 1-4
Model A – 7 TeV	78.8%	25.2%	12.4%	12.2%
Model A – 14 TeV	81.7%	35.8%	18.5%	18.2%
Model B – 7 TeV	81.8%	27.1%	13.4%	13.3%
Model B – 14 TeV	83.9%	37.5%	19.2%	18.7%

PDR 2 Cut Performance	Cut 1	Cuts 1-2	Cuts 1-3	Cuts 1-4
Model A – 7 TeV	78.8%	51.4%	26.1%	25.7%
Model A – 14 TeV	81.7%	58.1%	30.5%	30.1%
Model B – 7 TeV	81.8%	38.5%	19.6%	19.6%
Model B – 14 TeV	83.9%	46.1%	24.4%	24.4%

The Hemisphere Method – Cut 1

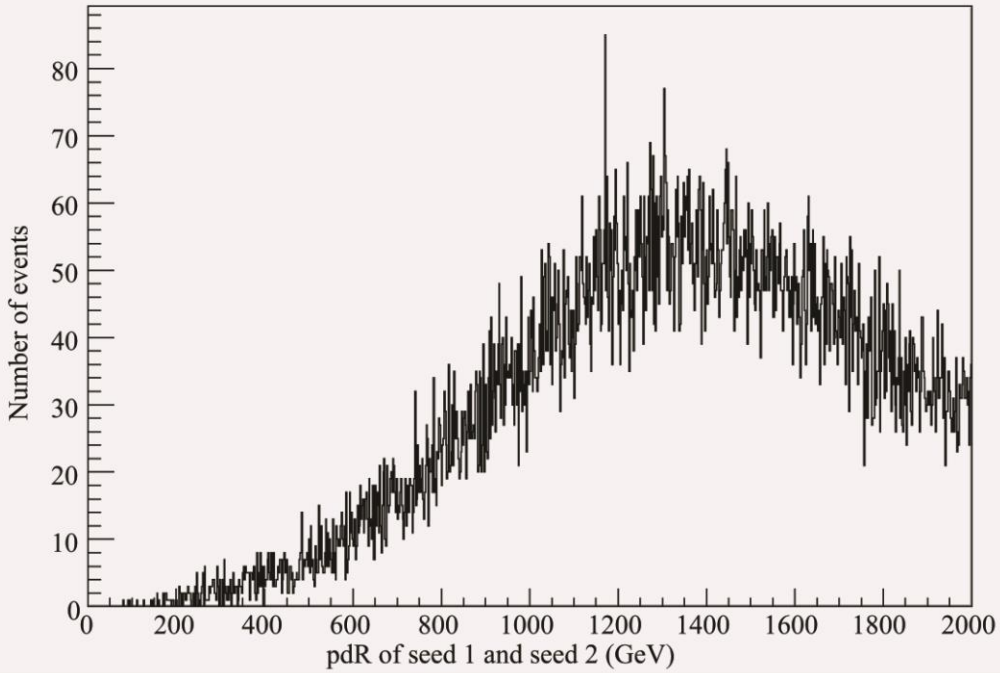
p_T (of seed 1) ≥ 200 GeV



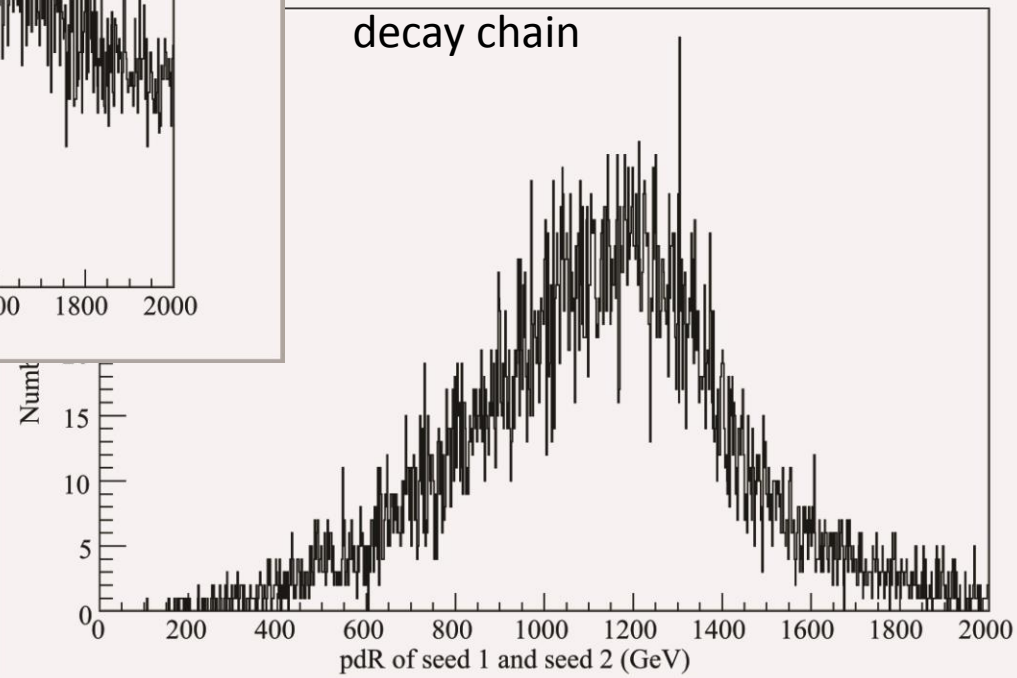
The Hemisphere Method – PDR 1 Cut 2

pdR (of seed 1 and seed 2) ≥ 1800 GeV

Seed 1 and seed 2 on opposite decay chains



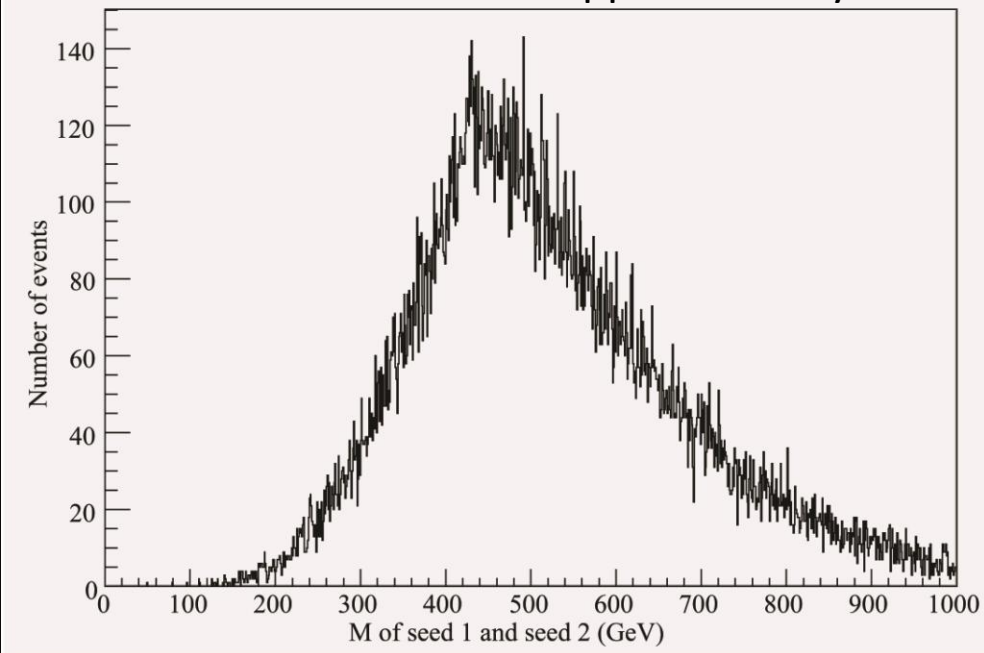
Seed 1 and seed 2 on the same decay chain



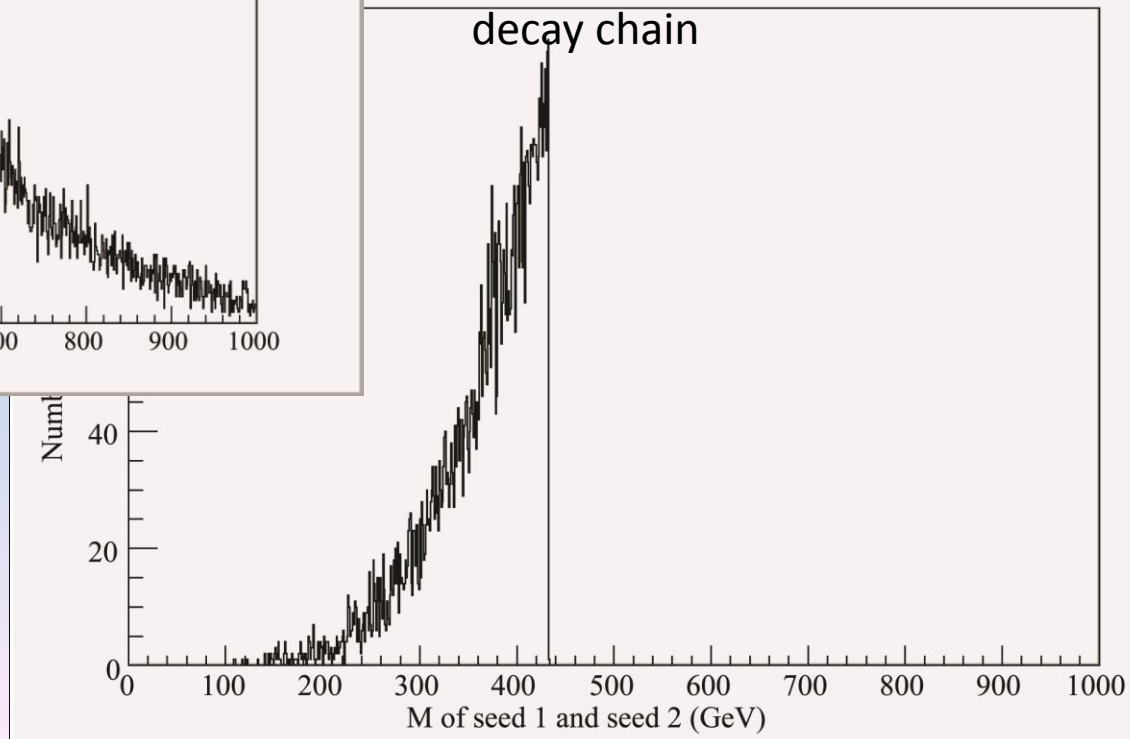
The Hemisphere Method – PDR 2 Cut 2

M (of seed 1 and seed 2) > M_{edge}

Seed 1 and seed 2 on opposite decay chains

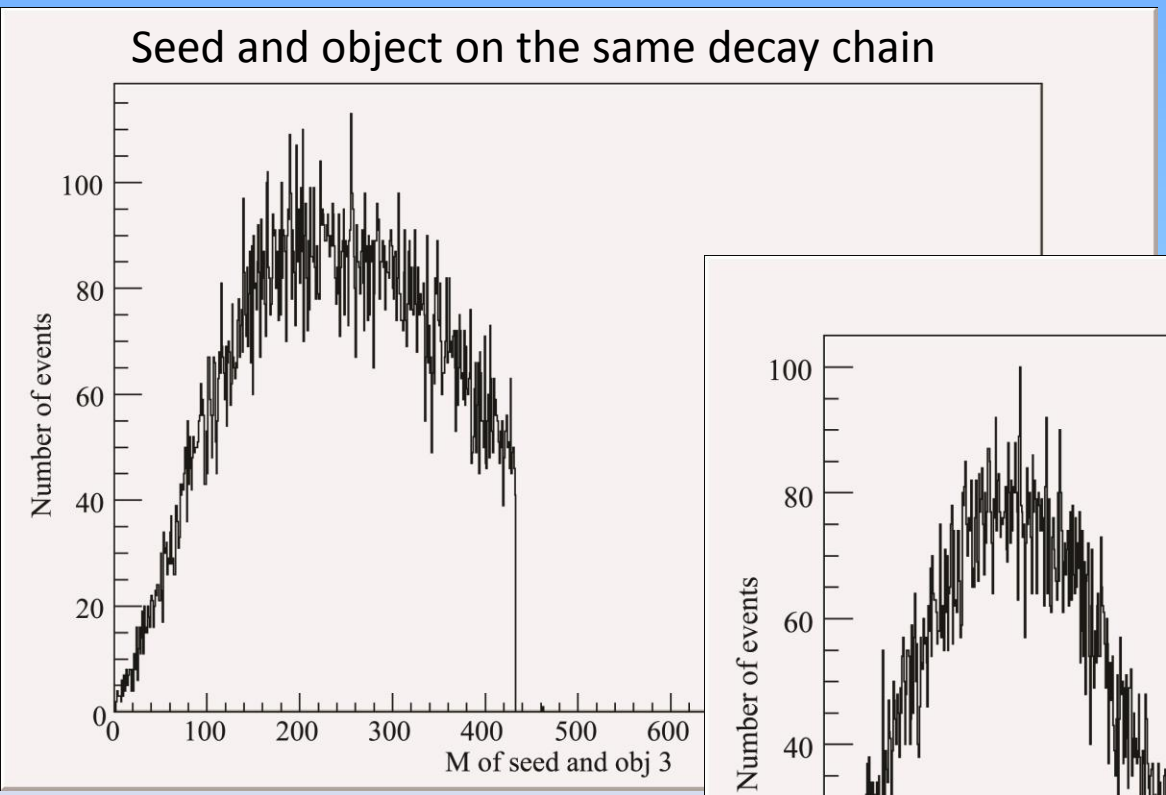


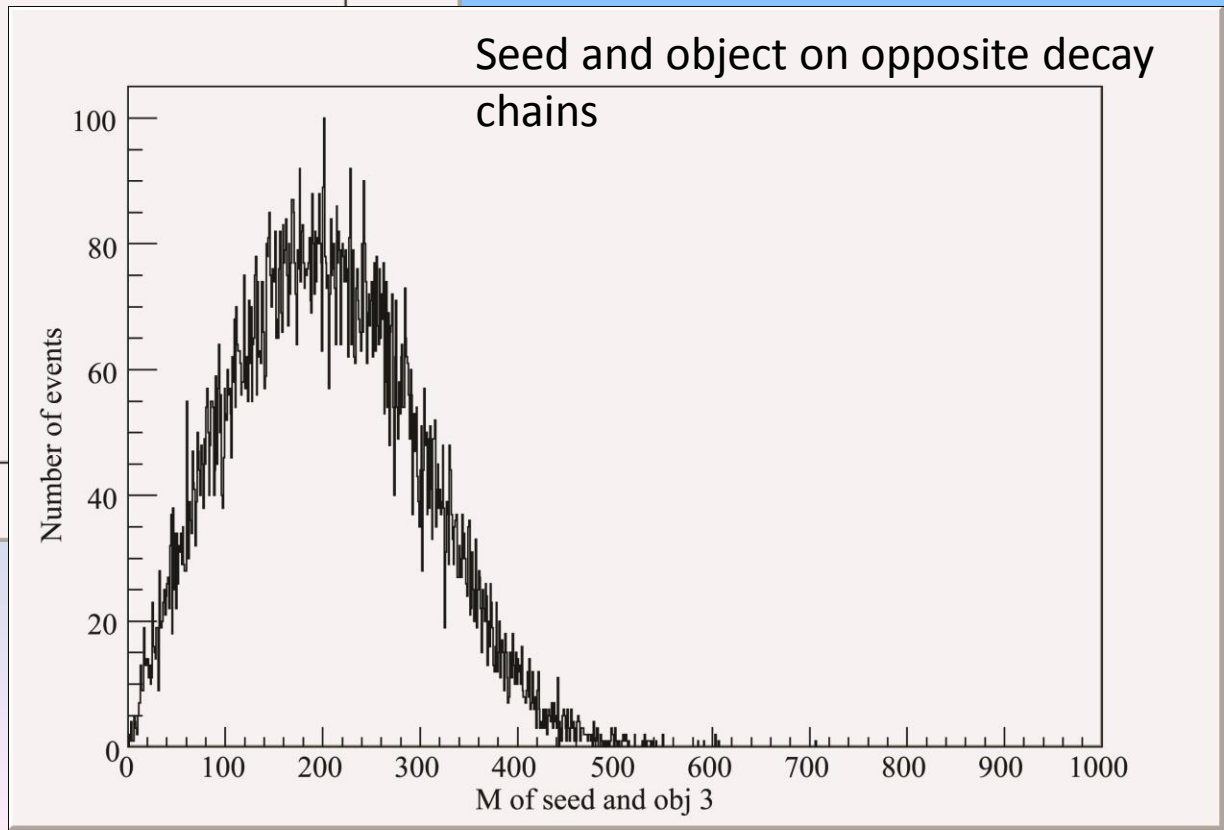
Seed 1 and seed 2 on the same decay chain



The Hemisphere Method – PDR 1 Cut 4

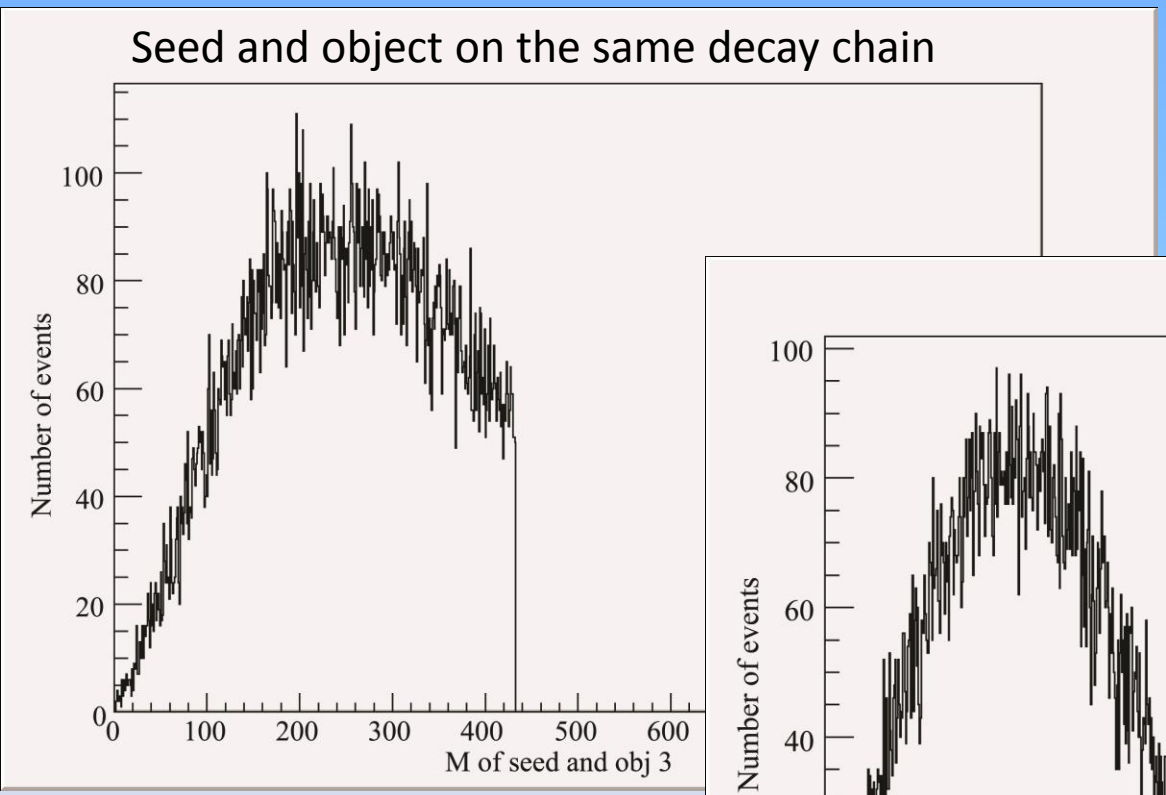
M (of seed and object) ≤ M_{edge}

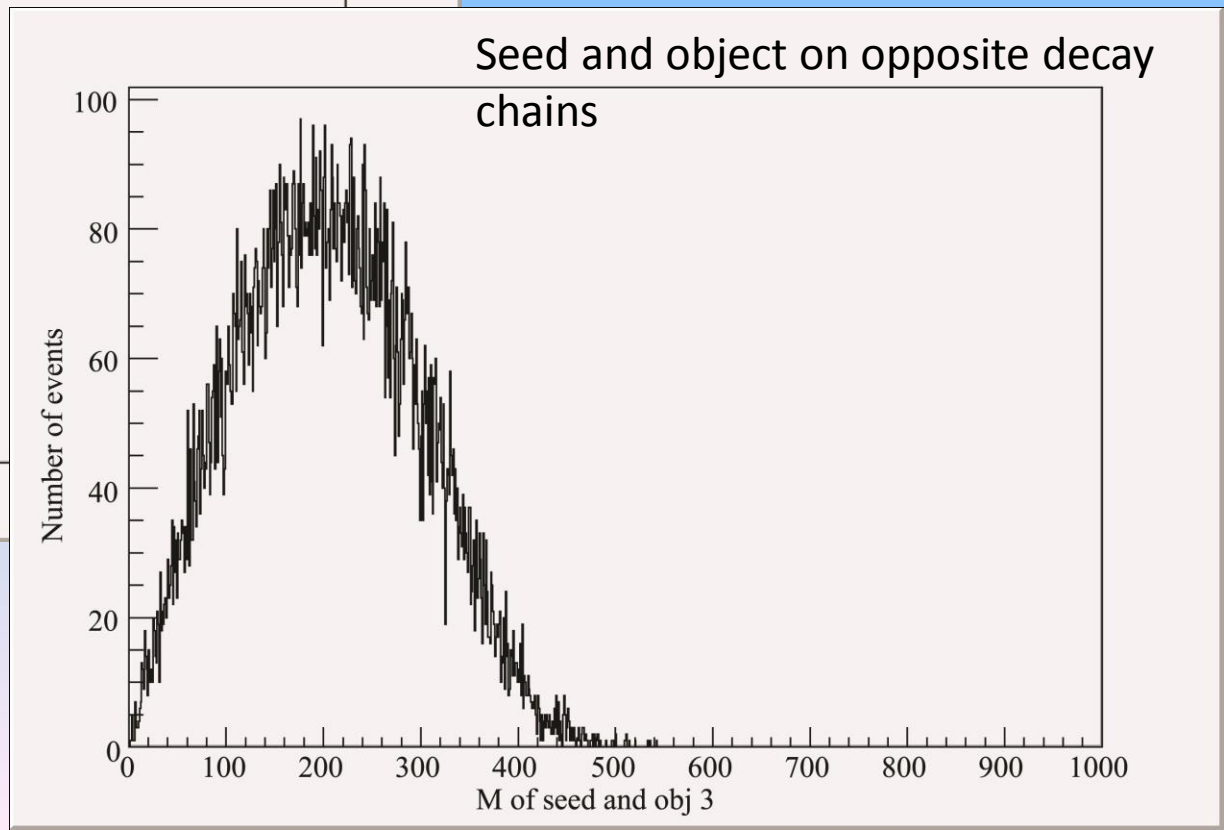




The Hemisphere Method – PDR 2 Cut 4

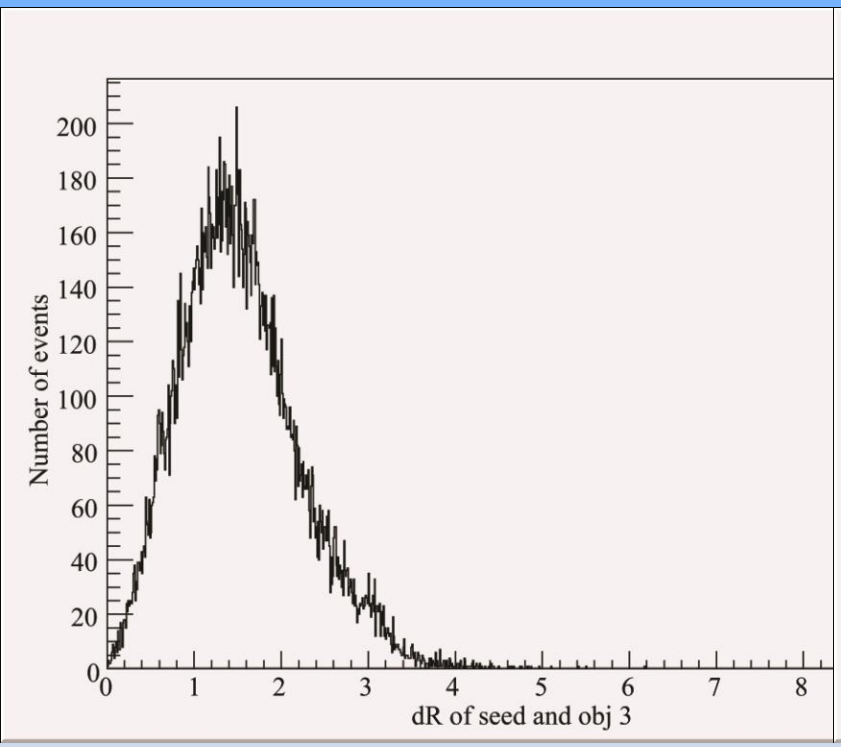
M (of seed and object) ≤ M_{edge}

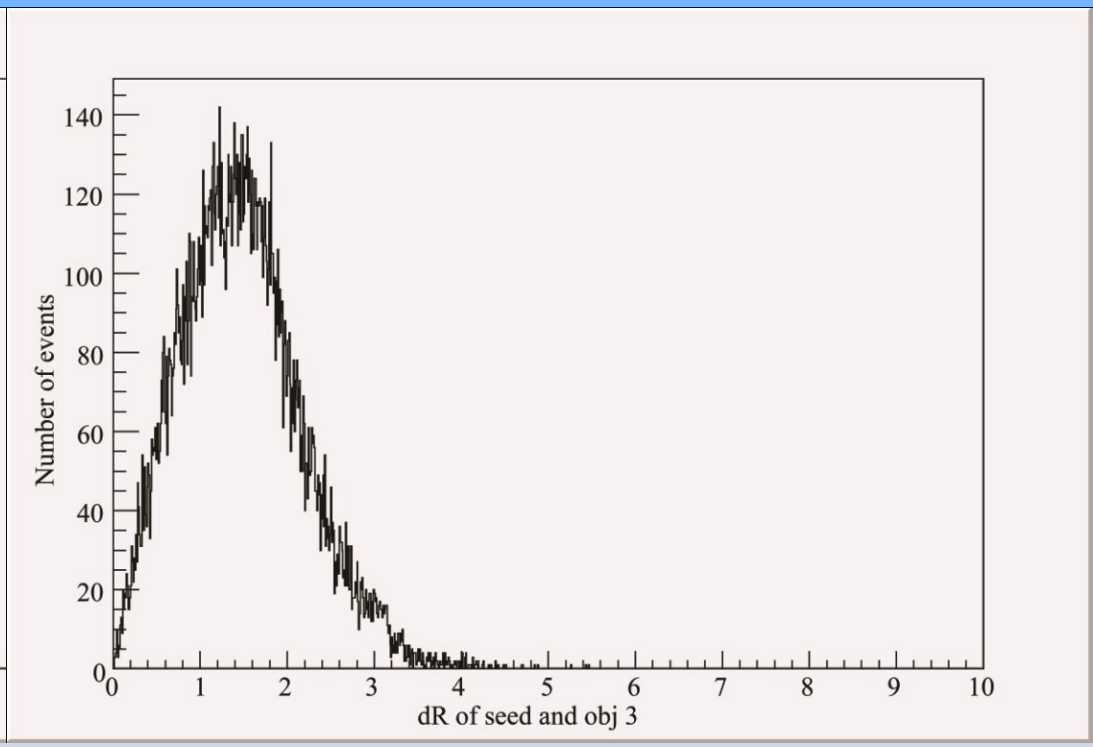




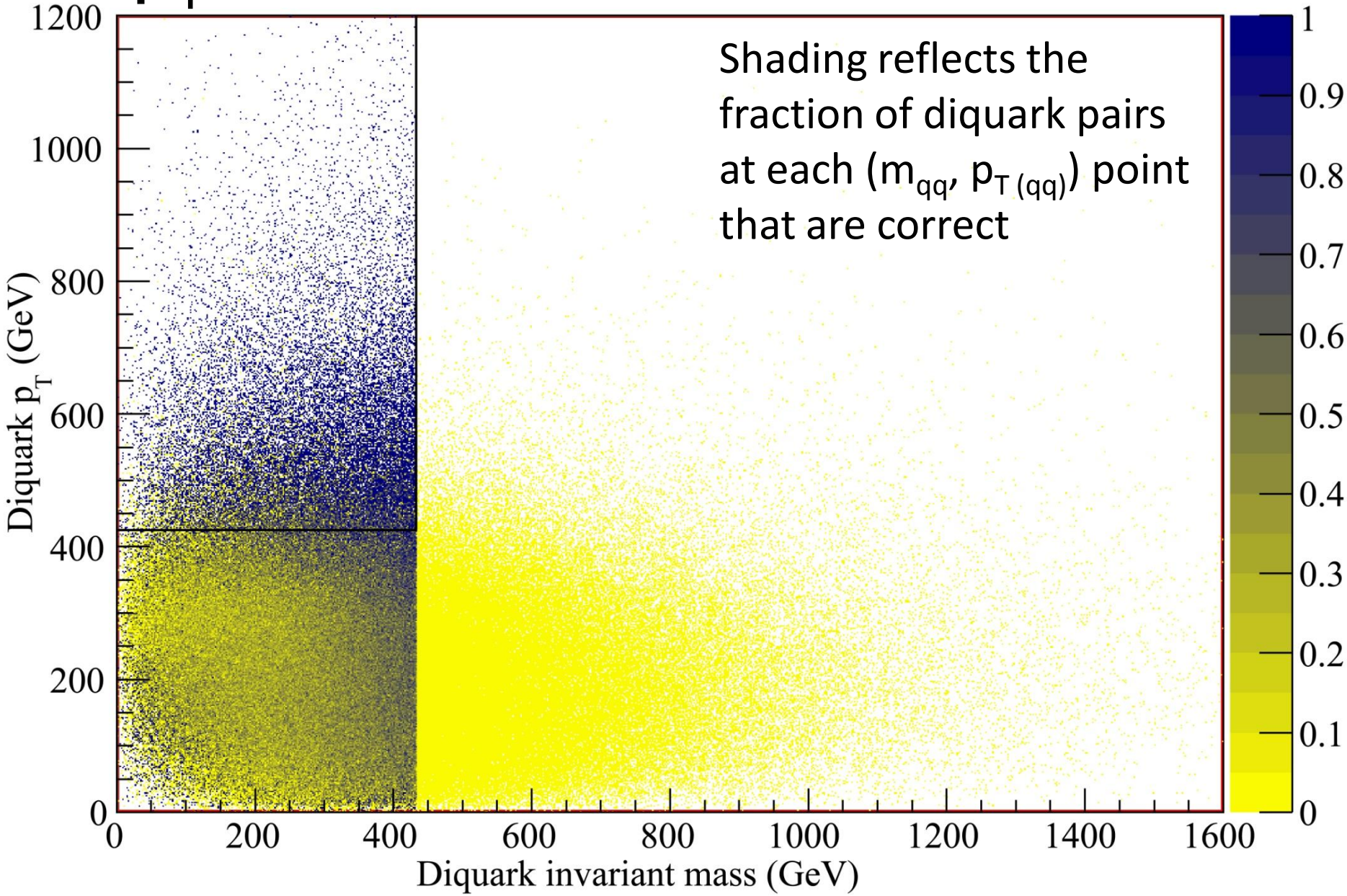
Alternative cut: maximum dR cut

- Constraining dR differences does not work

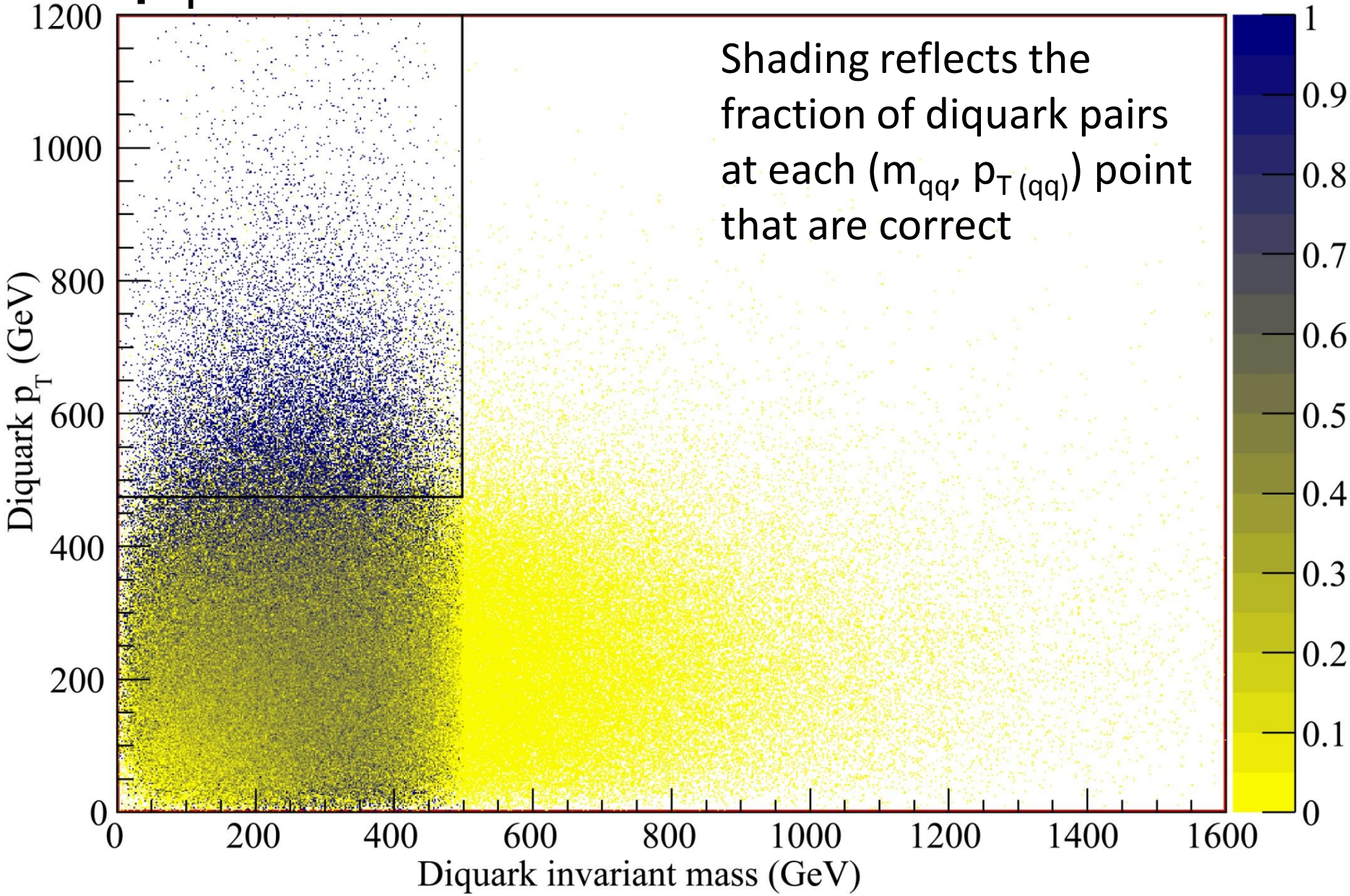




The p_T v. M method – Model A 14 TeV



The p_T v. M method – Model B 14 TeV



Discussion

- p_T v. M is more flexible
  - No distinction between choosing seeds and clustering
  - If seed 1 and seed 2 are incorrect, the hemisphere method fails
    - Requires strict cuts to ensure seed 1 and seed 2 are from different decay chains
- p_T v. M is readily generalized
  - Multi-jets, leptons, complicated decay chains including W and Z bosons