Data-driven analysis methods for the measurement of reconstructed jets in heavy ion collisions at RHIC and LHC

Gabriel de Barros[1]
[1]*Universidade de São Paulo*

Bo Fenton-Olsen[2], Peter Jacobs[2,3] and Mateusz Ploskon[2]
[2] *Lawrence Berkeley National Laboratory*
[3]*CERN*

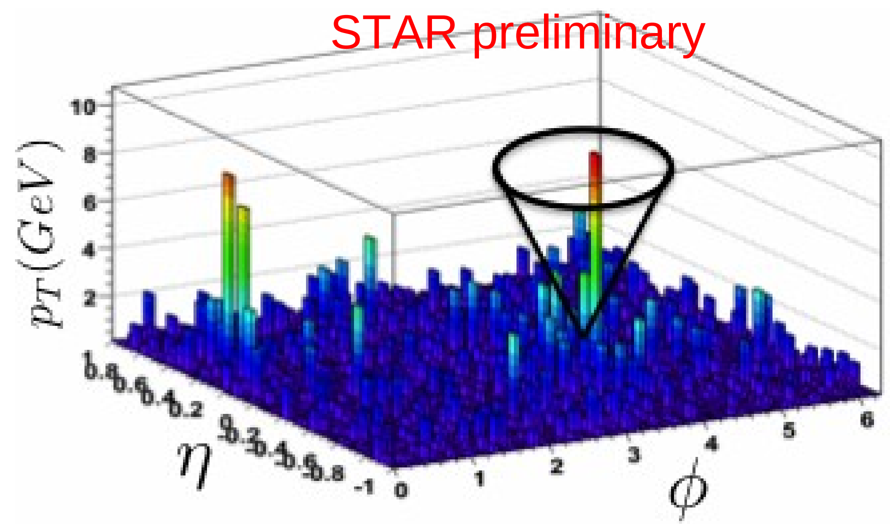
Jet reconstruction in heavy ion collisions

Very challenging task due to large underlying event

Jets are easy to identify but difficult to measure accurately due to large local background fluctuations

Origin of fluctuations: combinatoric “jets”

- random recombination of hadrons from soft background and multiple overlapping true jets
- Experimental noise, no underlying physical distribution



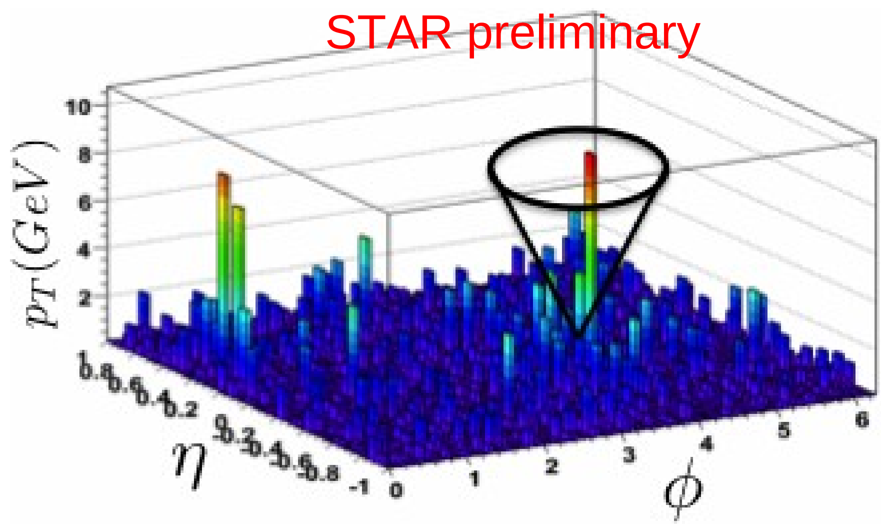
Jet reconstruction in heavy ion collisions

Very challenging task due to large underlying event

Jets are easy to identify but difficult to measure accurately due to large local background fluctuations

Origin of fluctuations: combinatoric “jets”

- random recombination of hadrons from soft background and multiple overlapping true jets
- Experimental noise, no underlying physical distribution



Experimental approaches thus far:

I. Reduce background fluctuations by suppressing low p_T contribution

Explicitly: cut on hadron p_T or calorimeter cell energy

Implicitly: high B-field, hadronic calorimetry

→ Jet quenching: possibly biased jet population

II. Restrict analysis to very high p_T jets; MC assessment of remaining background systematics

→ Limited applicability (high p_T only); dependence on fragmentation model

HI Jet Reconstruction: a re-assessment

Even in principle:

- Cannot discriminate most hadrons as “jet” or “background” on event-by-event basis
- Cannot know local background density on jet-by-jet basis

Jet quenching has meaning, and can be measured, only on an ensemble-averaged basis

HI Jet Reconstruction: a re-assessment

Even in principle:

- Cannot discriminate most hadrons as “jet” or “background” on event-by-event basis
- Cannot know local background density on jet-by-jet basis

Jet quenching has meaning, and can be measured, only on an ensemble-averaged basis

Therefore, the crucial experimental limitation:

- ***is not*** the large background relative to signal
- ***is*** the precision with which we can know the background fluctuations and correct for their effects on an ensemble-averaged basis

Well-defined observables: inclusive cross section, semi-inclusive coincidence rates,...

HI Jet Reconstruction: a re-assessment

Even in principle:

- Cannot discriminate most hadrons as “jet” or “background” on event-by-event basis
- Cannot know local background density on jet-by-jet basis

Jet quenching has meaning, and can be measured, only on an ensemble-averaged basis

Therefore, the crucial experimental limitation:

- ***is not*** the large background relative to signal
- ***is*** the precision with which we can know the background fluctuations and correct for their effects on an ensemble-averaged basis

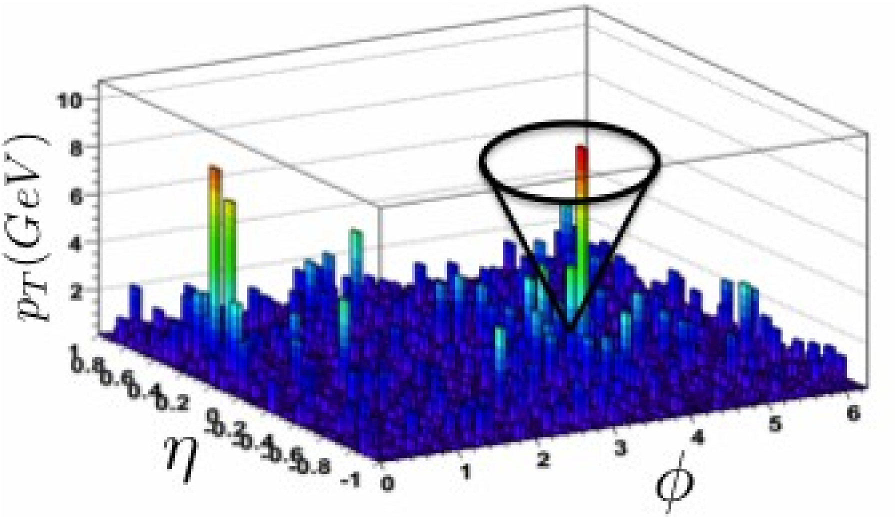
Well-defined observables: inclusive cross section, semi-inclusive coincidence rates,...

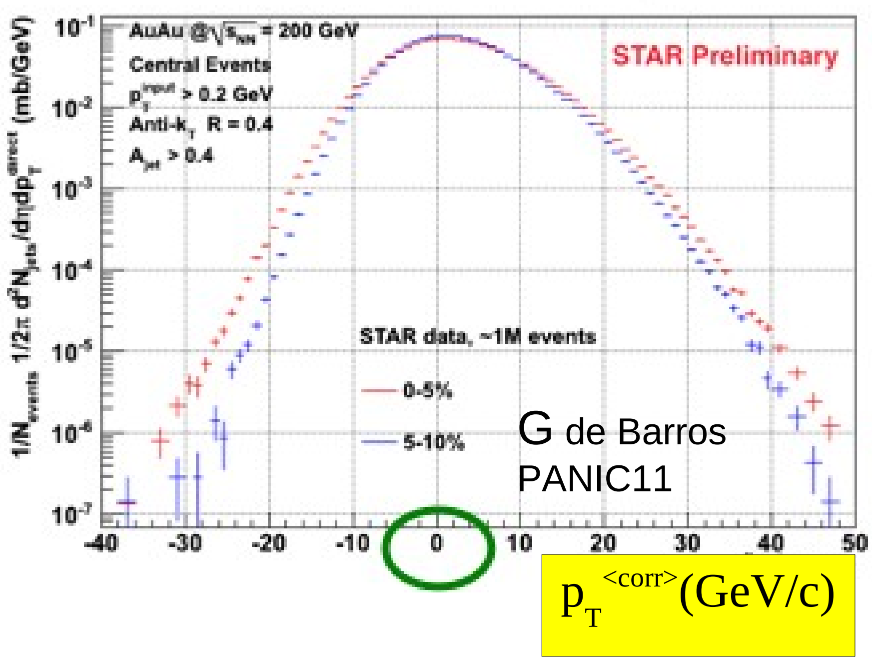
We seek new HI Jet analysis methods that are systematically well-controlled over a very broad energy range, at both RHIC and LHC

- Fully data-driven: no modeling of backgrounds
- To measure jet quenching: fragmentation biases must be minimal, and transparent

  → To do this: utilize STAR and ALICE capabilities to measure individually almost all jet constituents over a wide p_T range

HI jet reconstruction: FastJet





Jet defined operationally: output of reconstruction algorithm (not necessarily interpretable perturbatively)

FastJet: collect all jets in acceptance i=1,...N

Event-wise estimate of background density:

$$\rho = \text{median}\left\{\frac{p_T^{\text{jet, i}}}{A_{\text{jet, i}}}\right\}$$

Spectrum corrected event-wise for median background density:

$$p_{T,i}^{\langle corr \rangle} = p_{T,i} - \rho \cdot A_i$$

ρ is median: ~half the jet population has $p_T^{<corr>} < 0$

- contains crucial information about background

Toy Model Event Generator

Develop Toy Model Event Generator that

- is simple (and therefore transparent to interpret)
- captures the essential features and complexity of jet reconstruction in the HI background
- approximates experimental conditions at RHIC and LHC

→ apply model to explore algorithms

Toy Model Event Generator

Develop Toy Model Event Generator that

- is simple (and therefore transparent to interpret)
- captures the essential features and complexity of jet reconstruction in the HI background
- approximates experimental conditions at RHIC and LHC

→ apply model to explore algorithms

- Generic "particles": primary only, no decays, no detector effects
- Generic acceptance (both RHIC and LHC): $|\eta|<1.0$, full azimuth

- Soft component: Boltzmann distribution
  - RHIC: $<p_T>$=500 MeV
  - LHC: $<p_T>$=700 MeV
- Hard jet component:
  - Distribution: $T_{AA} * d\sigma_{jet}/dp_T$ (from p+p measurement or PYTHIA)
  - Various fragmentation models (to separate unfolding and fragmentation biases):
    1. *None: jets are modeled as single high p_T particles*
    2. *PYTHIA (vacuum) fragmentation, etc.*

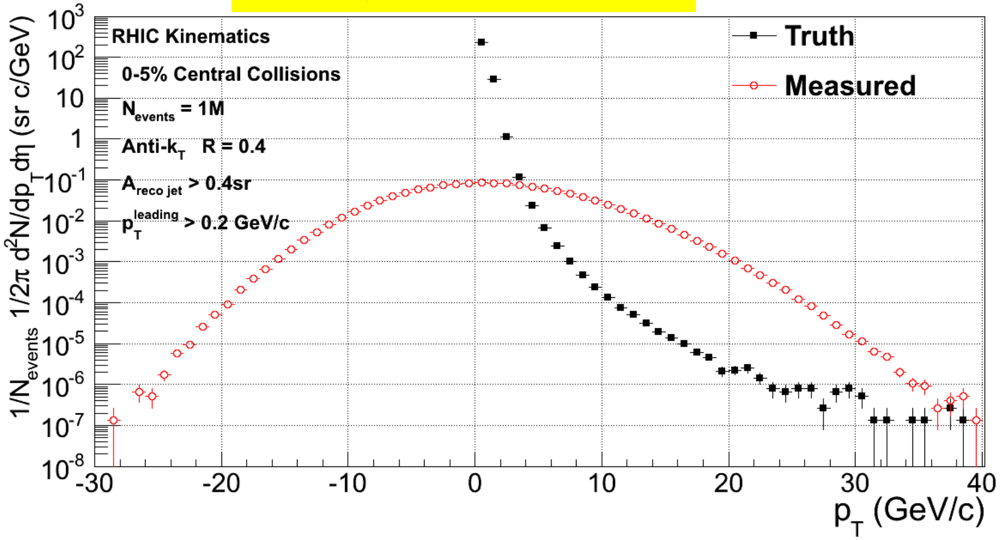
- *"Central" collisions: total multiplicity (charged+neutral) in acceptance*
  - *RHIC, Au+Au 0-5%: Mtot=2000*
  - *LHC, Pb+Pb, 0-5%: Mtot=4800*

- throw millions of such events and analyze like data, with anti-k_T, R=0.4

Toy model: inclusive jet spectrum

Dramatic broadening due to HI event background

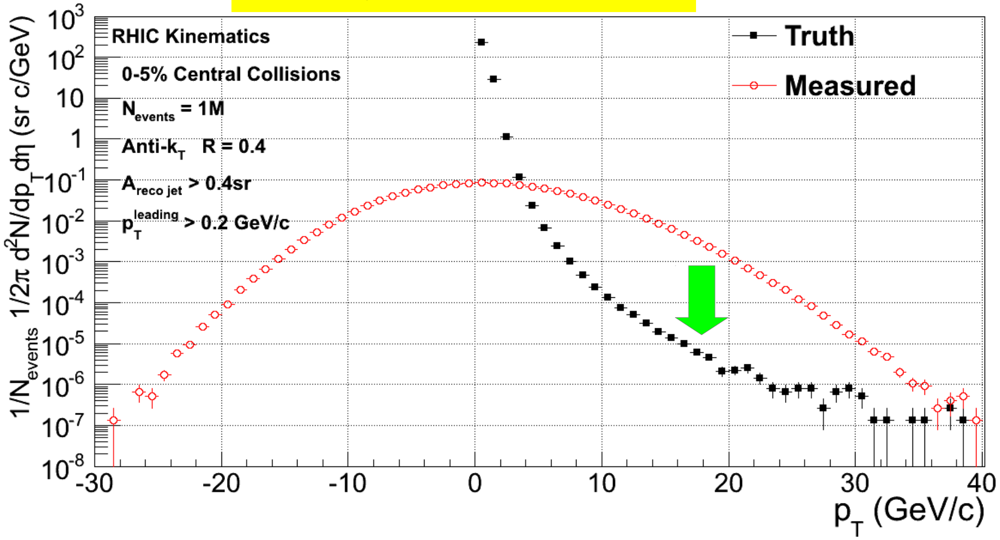
RHIC, central Au+Au



Toy model: inclusive jet spectrum

Dramatic broadening due to HI event background

RHIC, central Au+Au



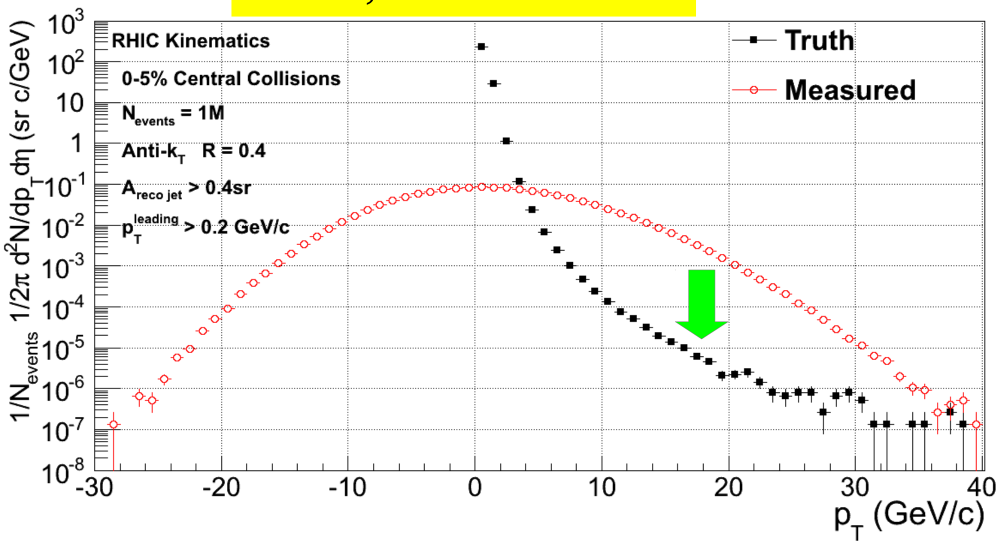
Toy model: inclusive jet spectrum

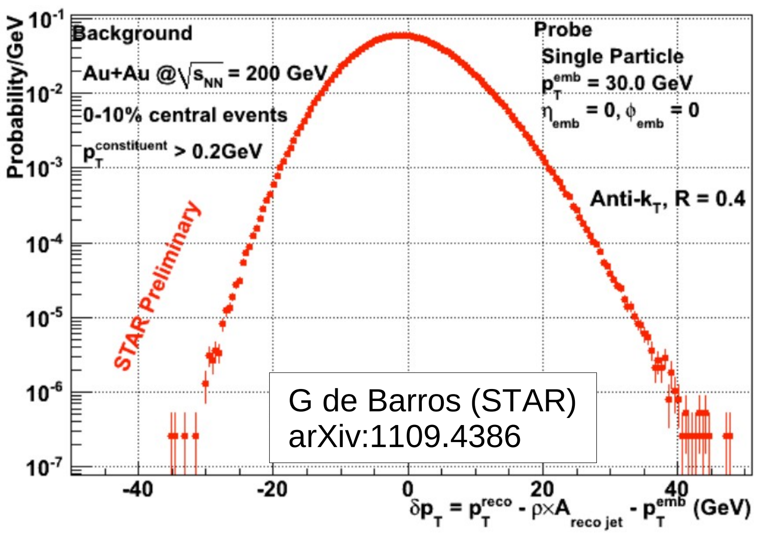
Dramatic broadening due to HI event background

Measure ensemble-averaged distribution of fluctuations via embedding known probes into real events

$$\delta p_T \equiv p_T^{<corr>} - p_T^{embed} = p_T^{jet} - \rho \cdot A_{jet} - p_T^{embed}$$

RHIC, central Au+Au





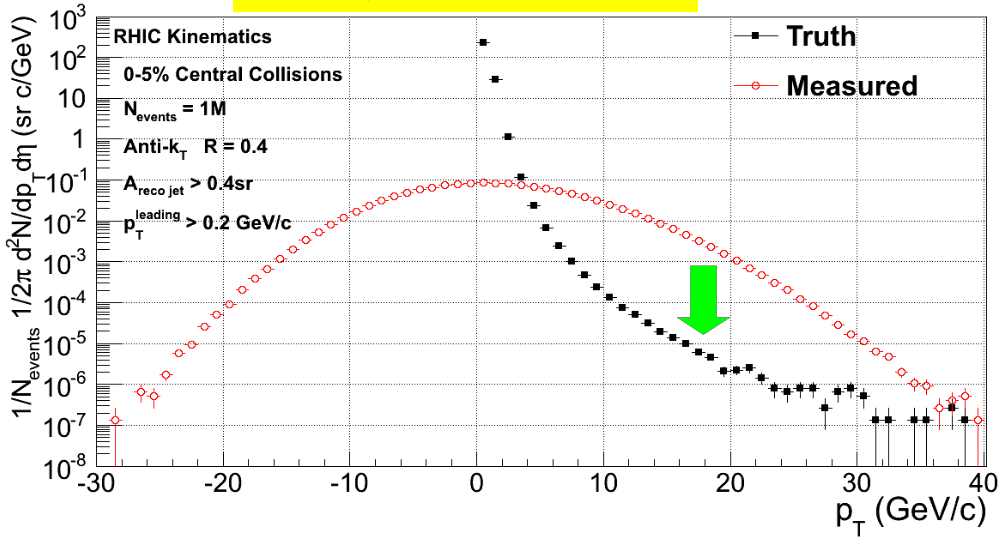
Toy model: inclusive jet spectrum

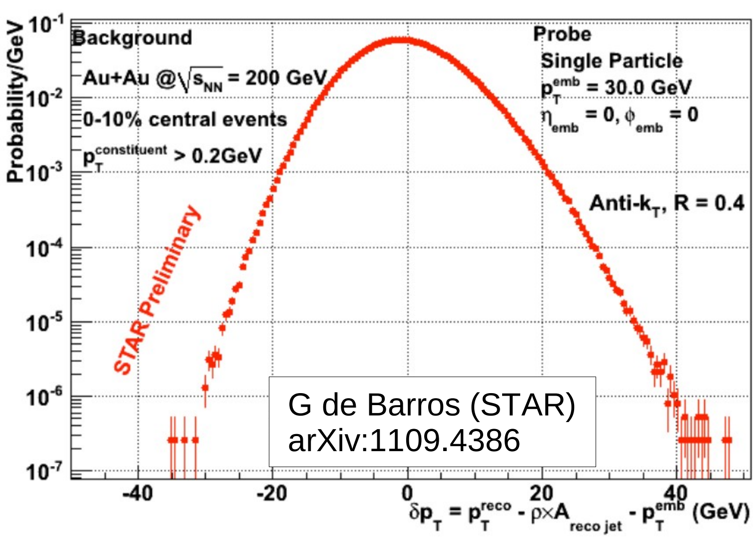
Dramatic broadening due to HI event background

Measure ensemble-averaged distribution of fluctuations via embedding known probes into real events

$$\delta p_T \equiv p_T^{<corr>} - p_T^{embed} = p_T^{jet} - \rho \cdot A_{jet} - p_T^{embed}$$

RHIC, central Au+Au





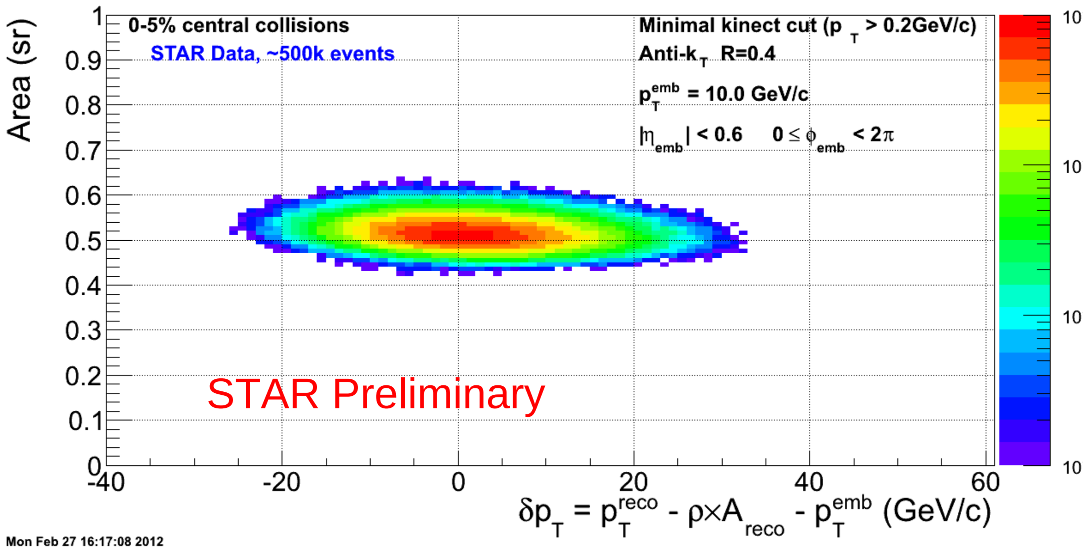
Crucial feature for unbiased measurement: δp_T distribution is independent of fragmentation pattern of embedded jet (G de Barros (STAR) arXiv:1109.4386)

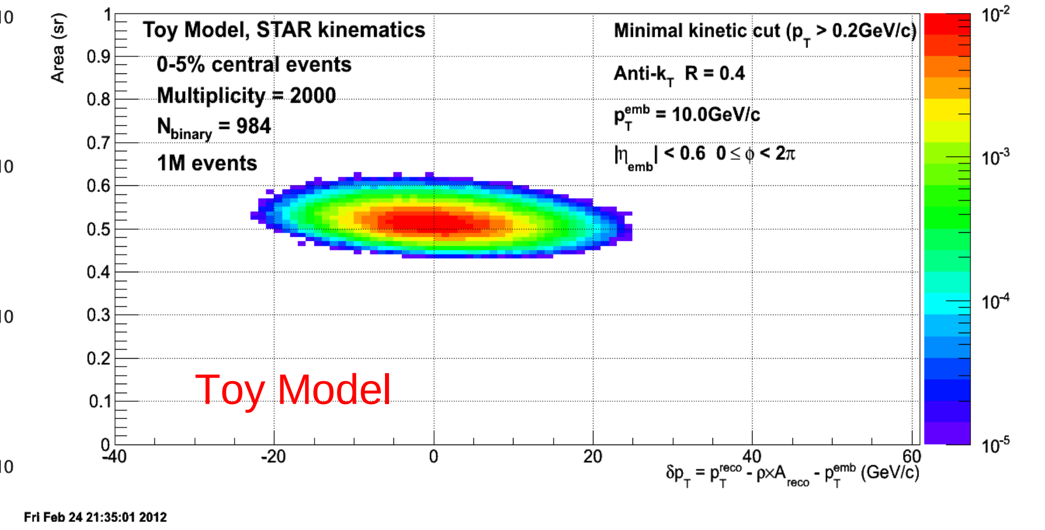
Correction for background fluctuations via "unfolding"

Response matrix: δp_T distribution

Comparison of Toy Model vs Data

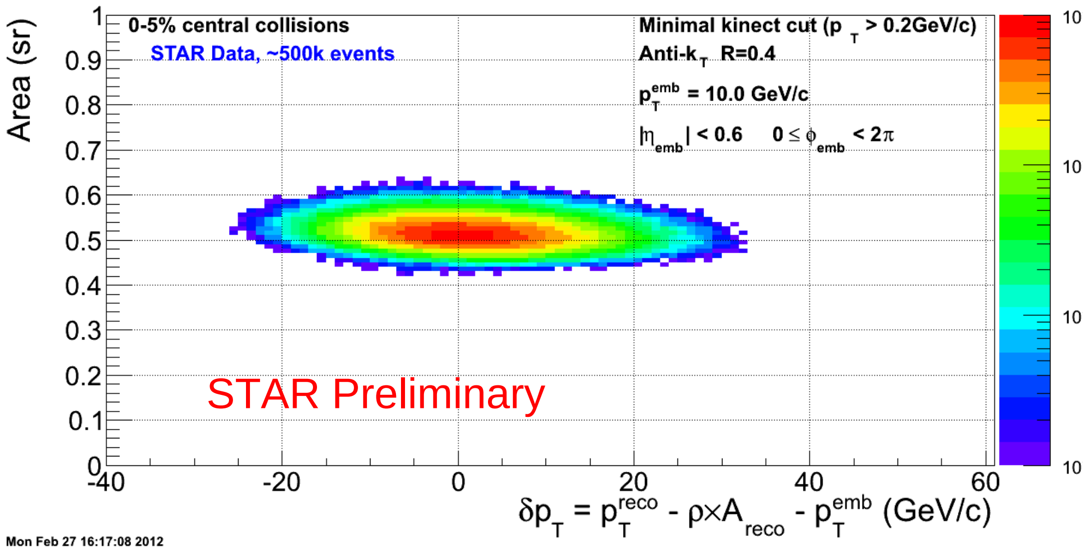
$$\delta p_T \equiv p_T^{<corr>} - p_T^{embed} = p_T^{jet} - \rho \cdot A_{jet} - p_T^{embed}$$

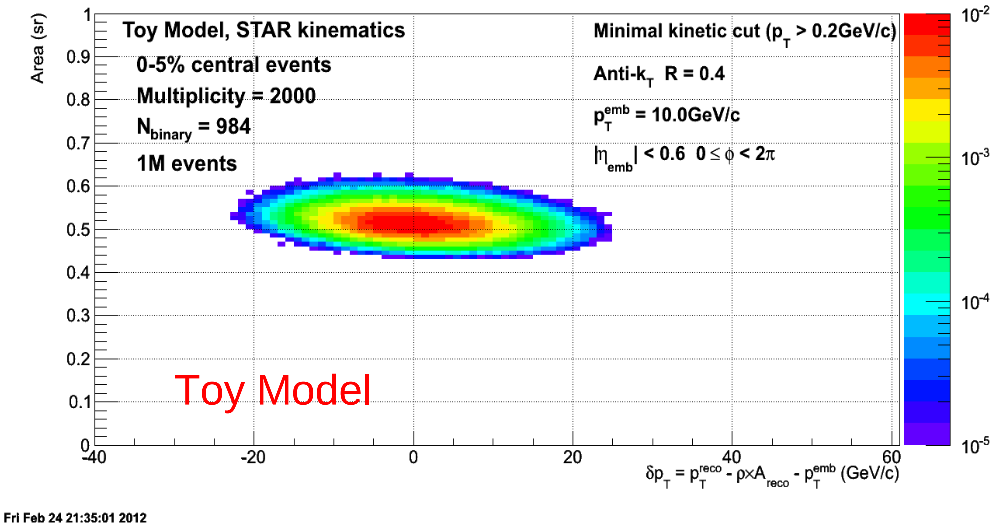


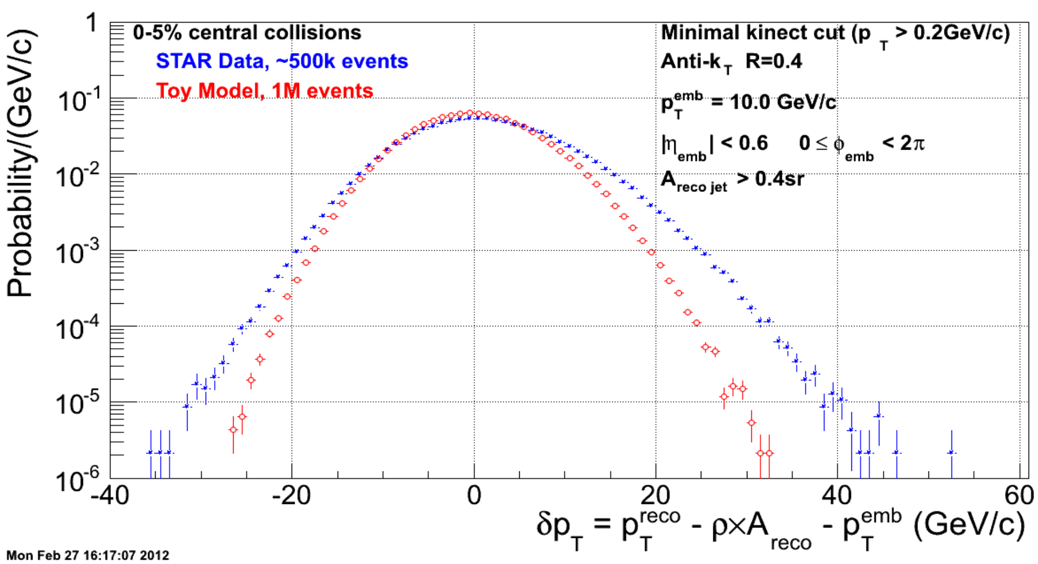


Comparison of Toy Model vs Data

$$\delta p_T \equiv p_T^{<corr>} - p_T^{embed} = p_T^{jet} - \rho \cdot A_{jet} - p_T^{embed}$$







Qualitatively similar, differing in detail

- e.g. Toy Model missing flow effects, will account for most of the difference

Current implementation sufficient for generic exploration of problem

- Flow complicates interpretation, leave out for now

Unfolding and Fragmentation Biases

Unfolding: it is essential to limit sensitivity to statistical noise (via regularization)
- Regularization imposes bias (smoothness) in exchange for reduced variance

- Various techniques: we use Iterative Bayesian
  - ➔ regularization via number of iterations

Unfolding and Fragmentation Biases

Unfolding: it is essential to limit sensitivity to statistical noise (via regularization)

- Regularization imposes bias (smoothness) in exchange for reduced variance

- Various techniques: we use Iterative Bayesian
  - → regularization via number of iterations

Heavy ion jet reconstruction: two essential biases in the problem

- Fragmentation bias
- Unfolding bias (~correction for background fluctuations)

These play against each other:

- Try to lower unfolding bias: suppress background fluctuations via momentum cut on jet constituents

- But this induces fragmentation bias

Jet quenching measurements: essential to minimize fragmentation bias

Unfolding and Fragmentation Biases

Unfolding: it is essential to limit sensitivity to statistical noise (via regularization)

- Regularization imposes bias (smoothness) in exchange for reduced variance

- Various techniques: we use Iterative Bayesian
  - → regularization via number of iterations

Heavy ion jet reconstruction: two essential biases in the problem

- Fragmentation bias
- Unfolding bias (~correction for background fluctuations)

These play against each other:

- Try to lower unfolding bias: suppress background fluctuations via momentum cut on jet constituents

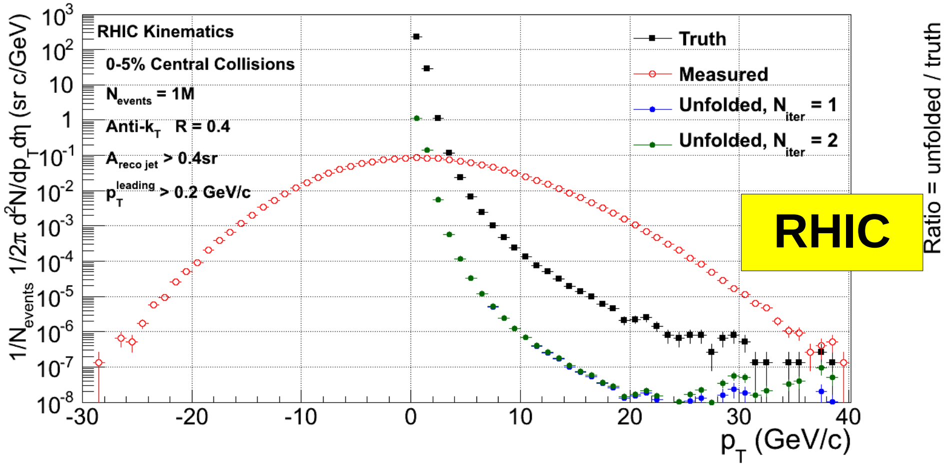
- But this induces fragmentation bias

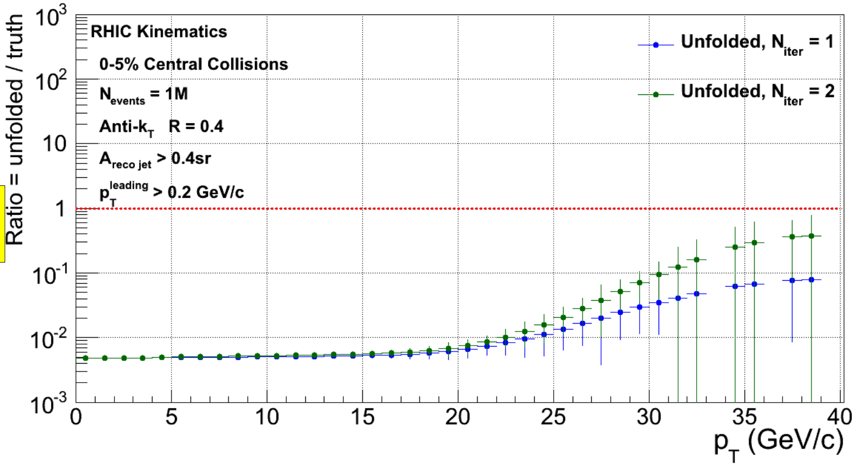
Jet quenching measurements: essential to minimize fragmentation bias

Toy Model study: isolate Unfolding Bias effects by choice of Single Particle fragmentation

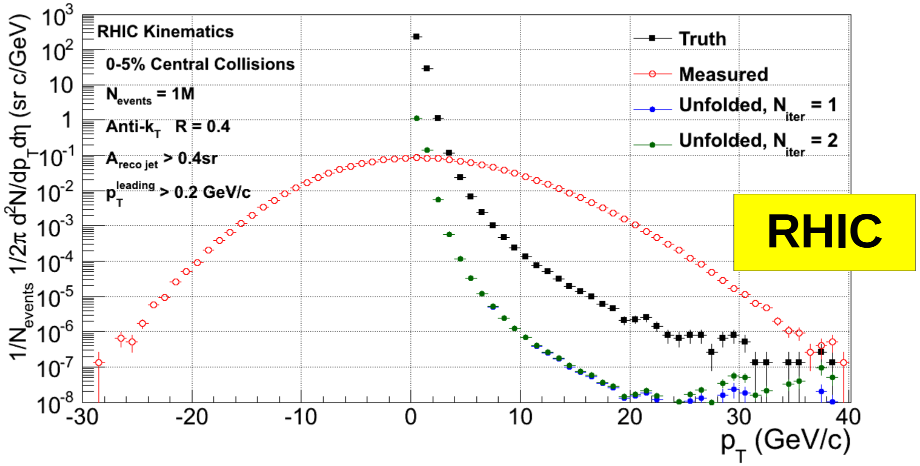
- Then assess fragmentation bias by more physical choice of jet fragmentation

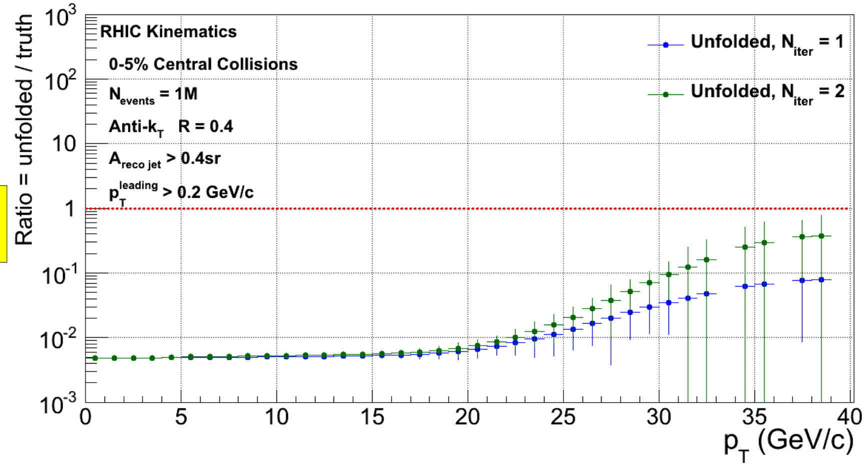
Unfolding of inclusive spectrum



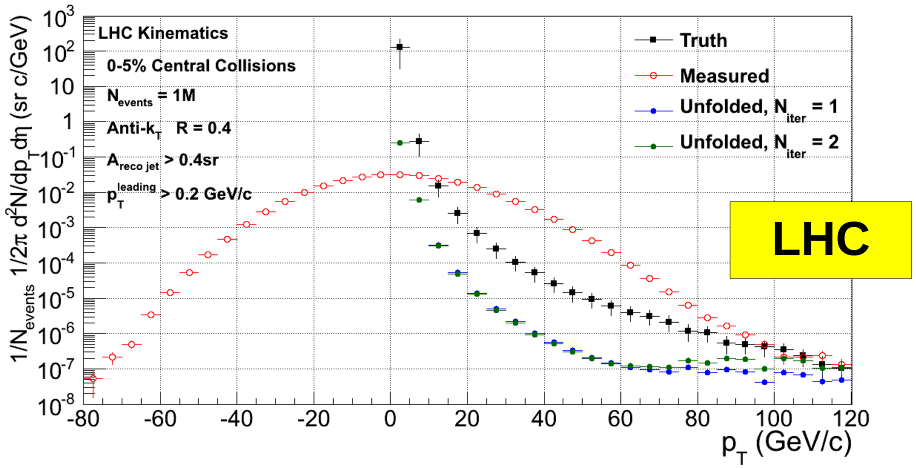


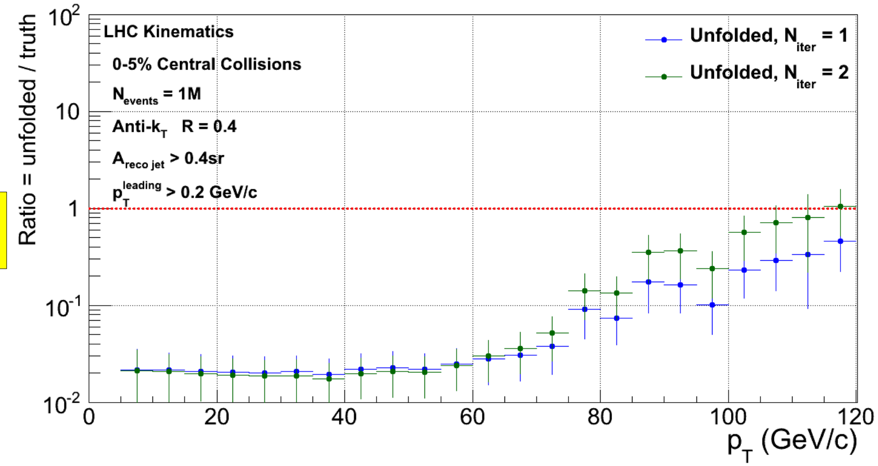
Unfolding of inclusive spectrum



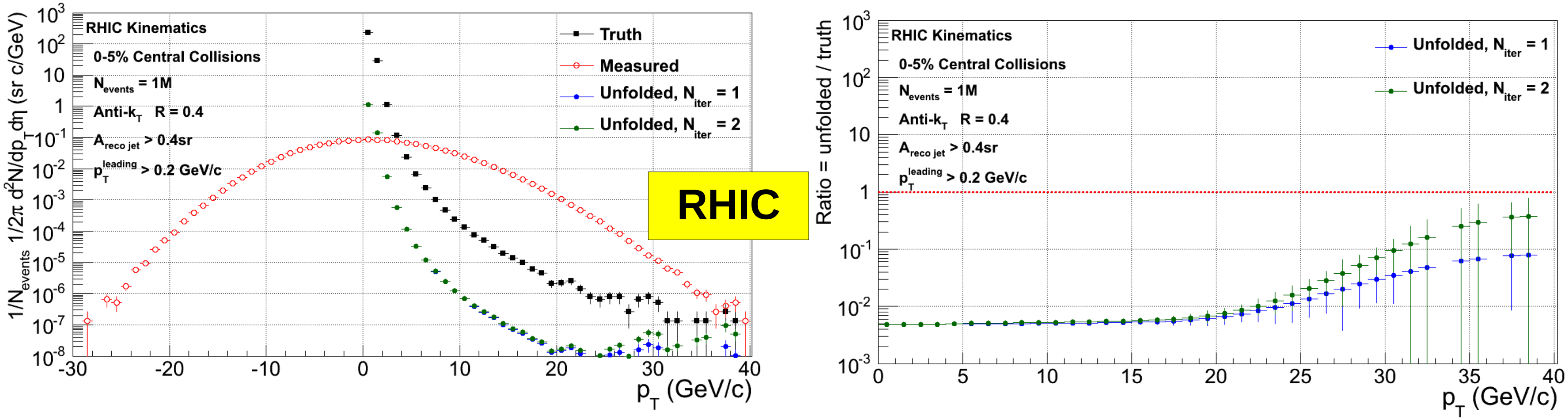


Unfolding does not converge to Truth → highly biased (and wrong) solution

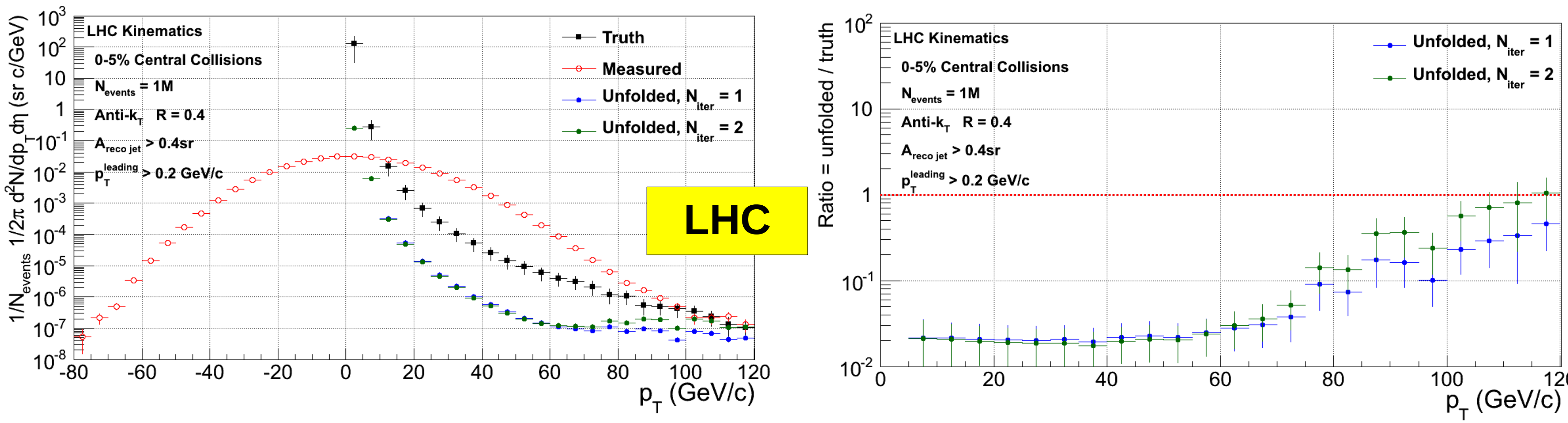




Unfolding of inclusive spectrum



Unfolding does not converge to Truth → highly biased (and wrong) solution



What went wrong? Answer: overwhelming population of combinatoric (noise) "jets"
no underlying physical distribution, problem not well-posed

Solution: eliminate combinatoric jet population ***before*** unfolding
what's left? hard jet population with p_T smeared by background fluctuations:
→ well-posed unfolding problem

Unfolding of inclusive spectrum: suppression of combinatoric jet population

The only available tool for inclusives: impose a fragmentation bias

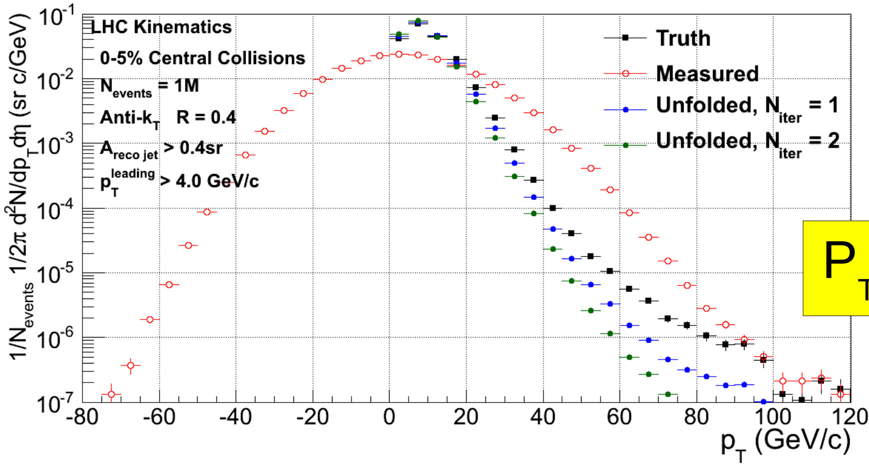
Reasonable choice:

- require accepted jet candidates to have at least one hadron constituent with $p_T > p_T^{threshold}$ ("$p_T^{leading}$" cut)

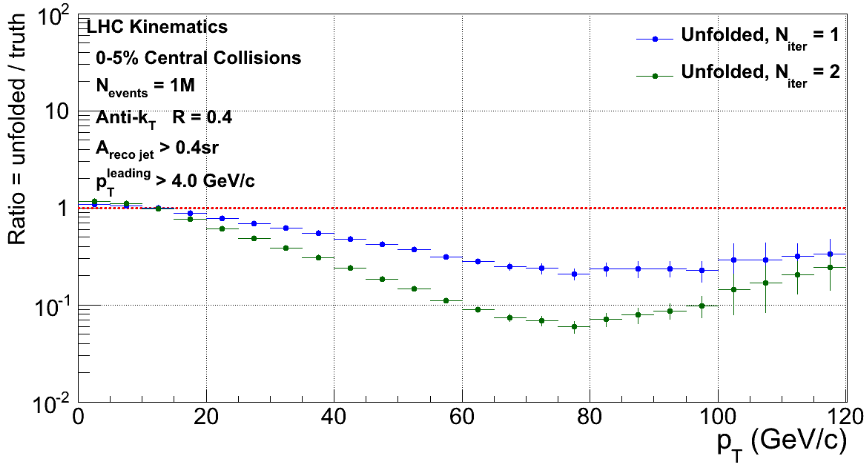
Bias is not zero, but is transparent and may be relatively mild (depends on p_T)

- Jet candidates can still have much of their radiation carried by very soft hadrons

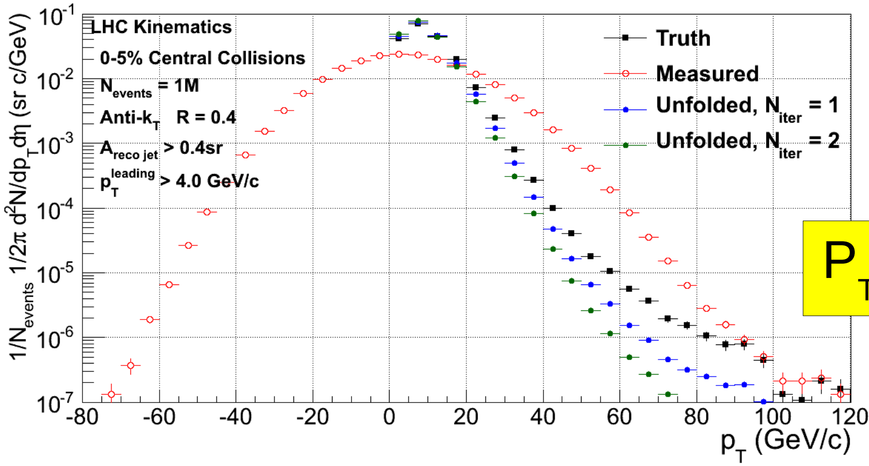
Inclusive spectrum with $p_T^{leading}$ bias: LHC

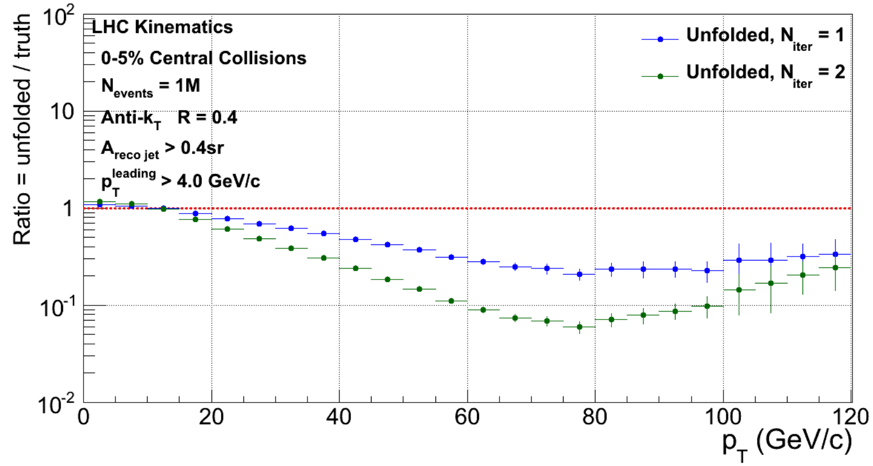


$P_T^{threshold}$=4GeV

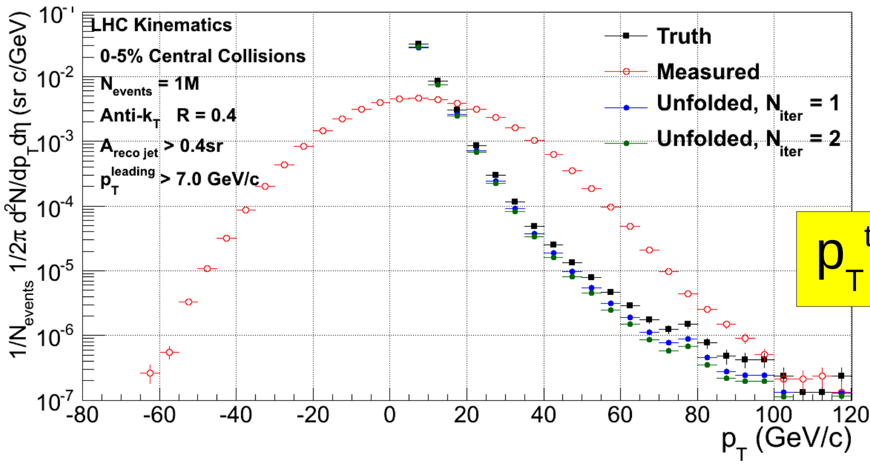


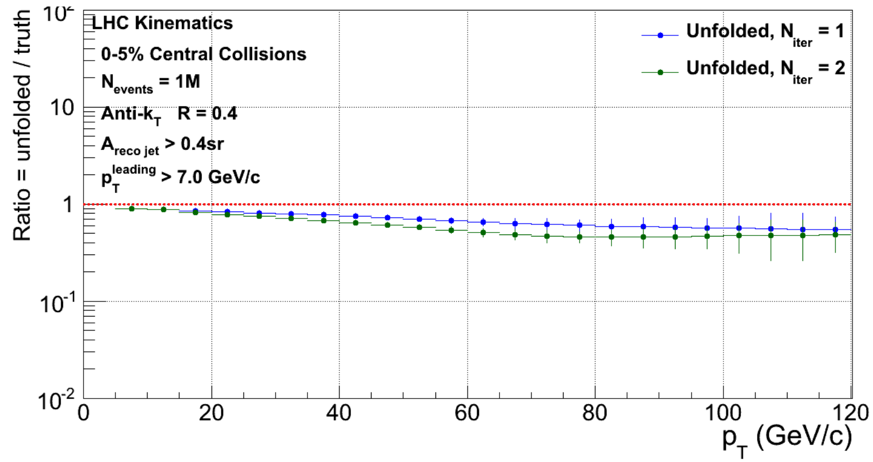
Inclusive spectrum with $p_T^{leading}$ bias: LHC





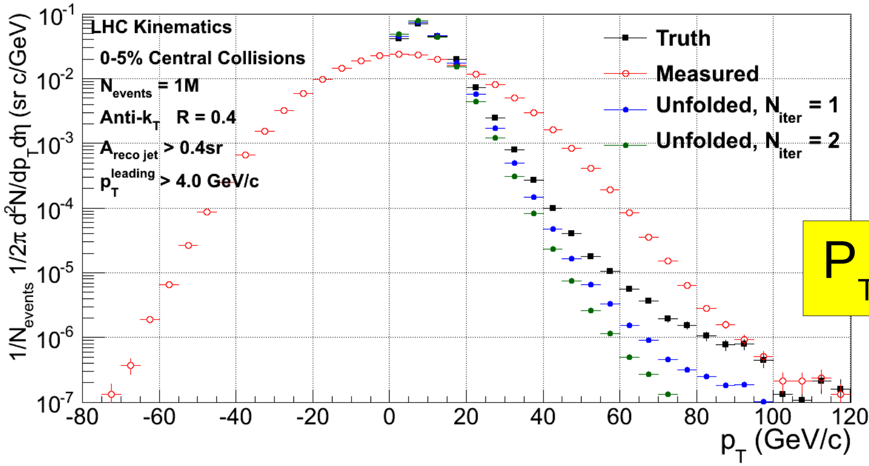
$P_T^{threshold}$=4GeV

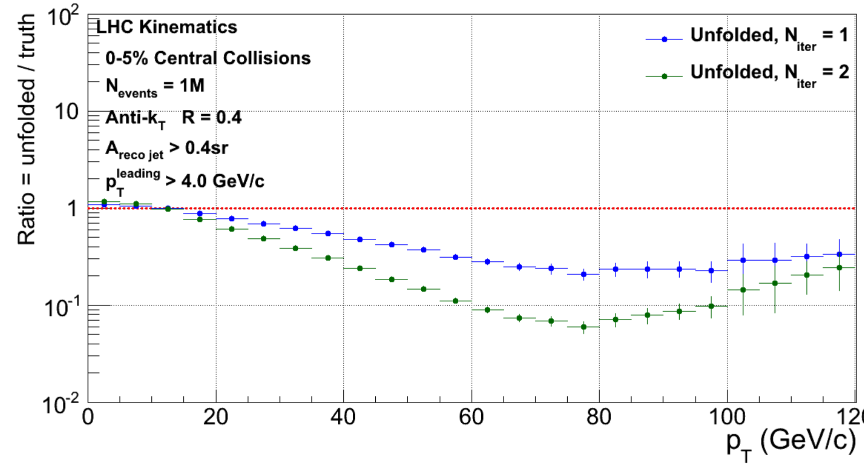




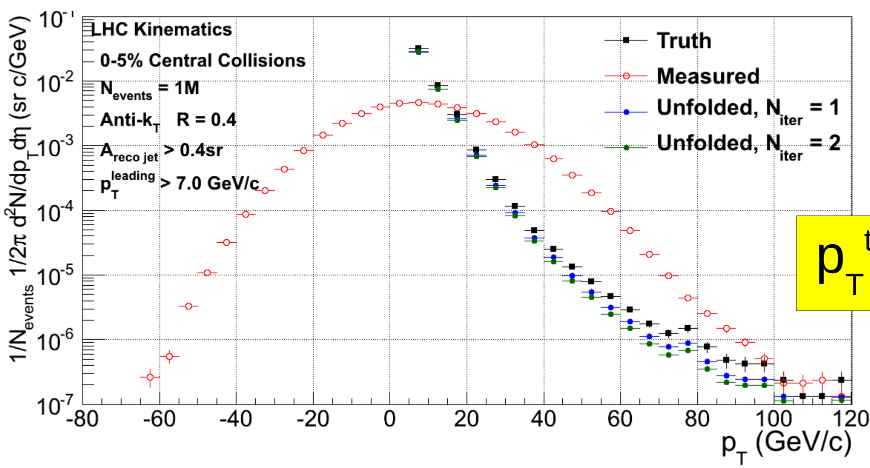
$p_T^{threshold}$=7GeV

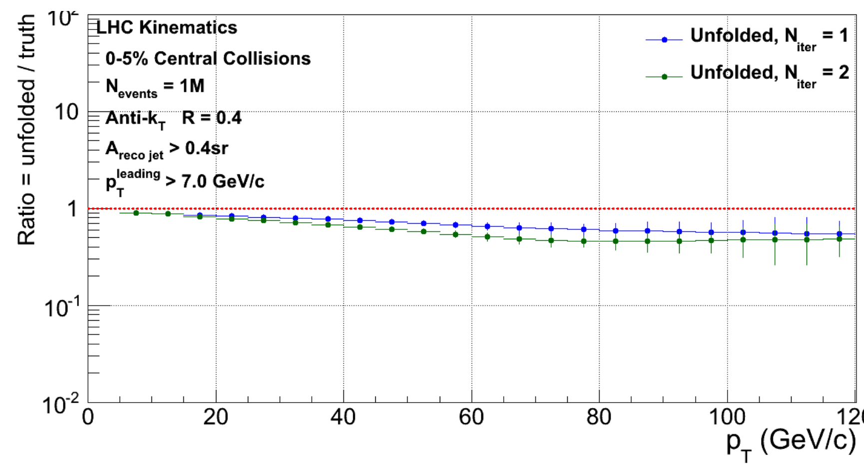
Inclusive spectrum with $p_T^{leading}$ bias: LHC



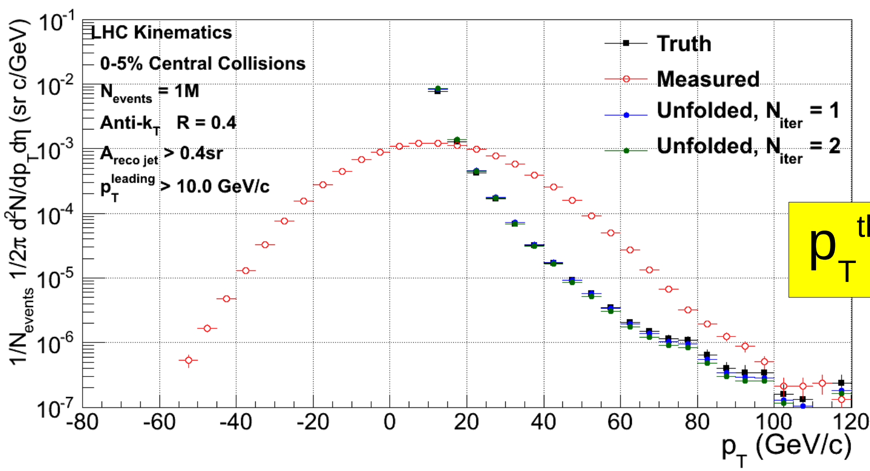


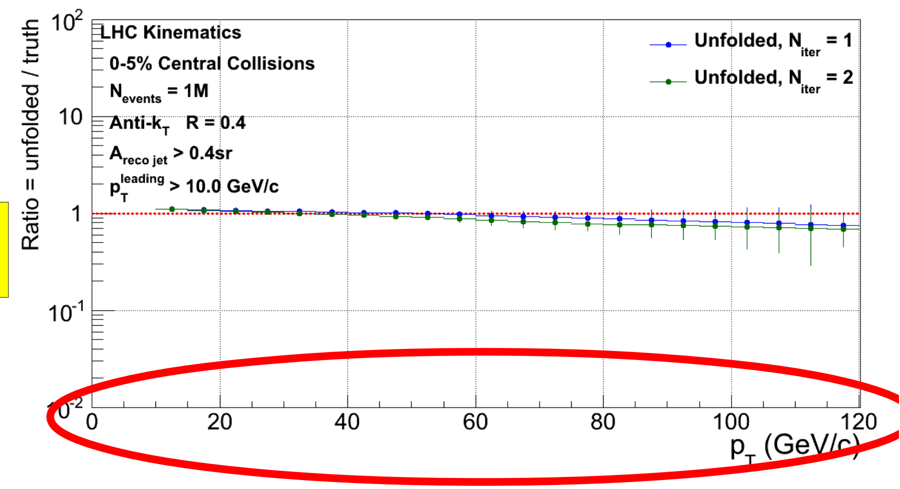
$P_T^{threshold}$=4GeV





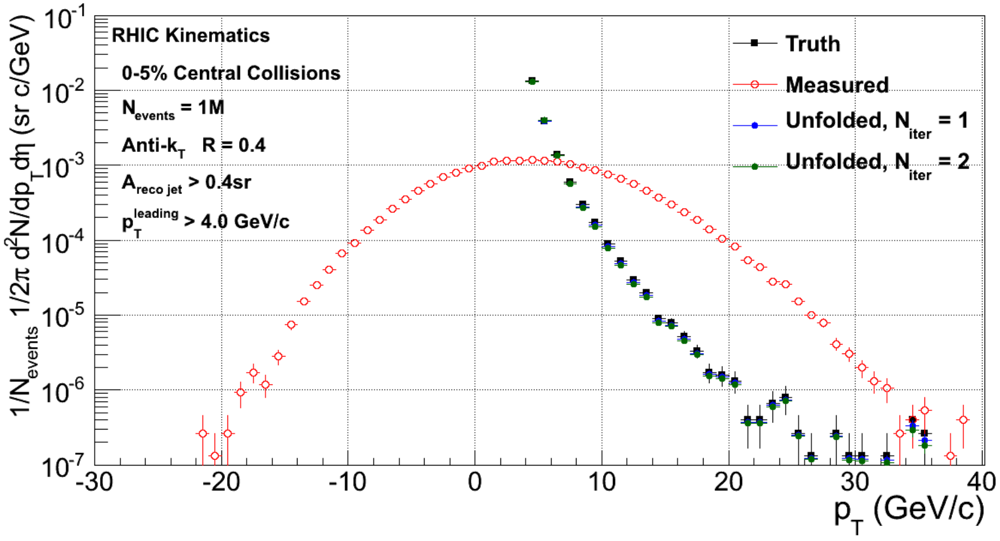
$p_T^{threshold}$=7GeV

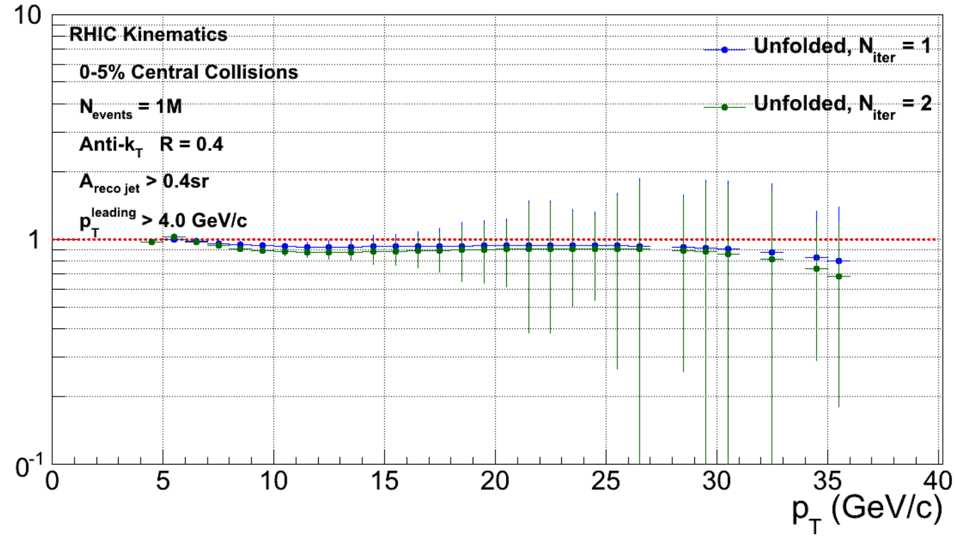




$p_T^{threshold}$=10GeV

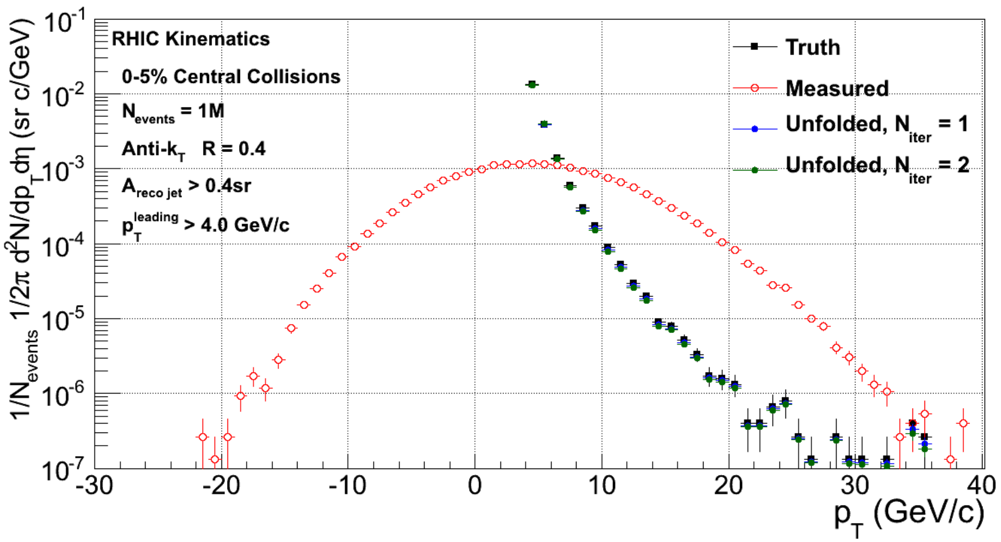
Inclusive spectrum with $p_T^{leading}$ bias: RHIC

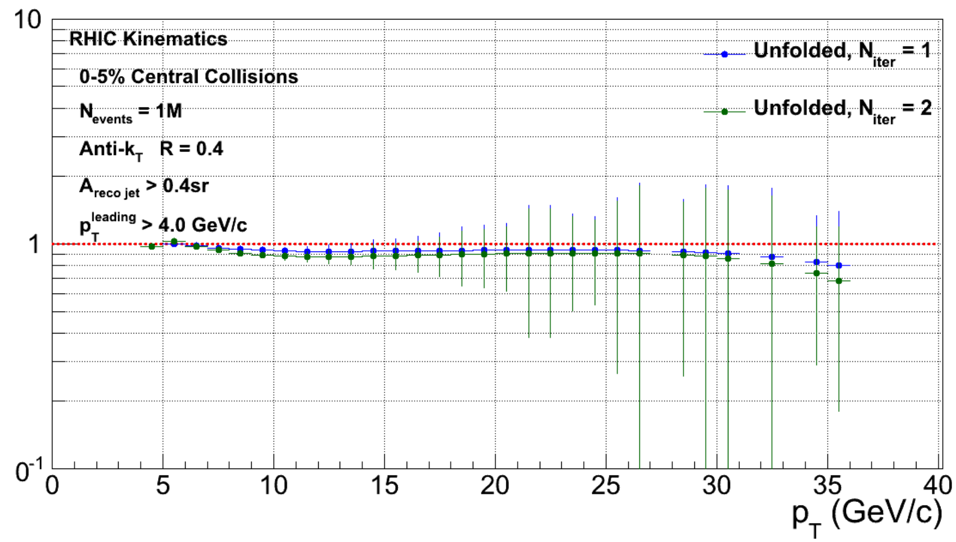




Similar result at RHIC (convergence for $p_T^{threshold}$ ~ 4 GeV)

Inclusive spectrum with $p_T^{leading}$ bias: RHIC





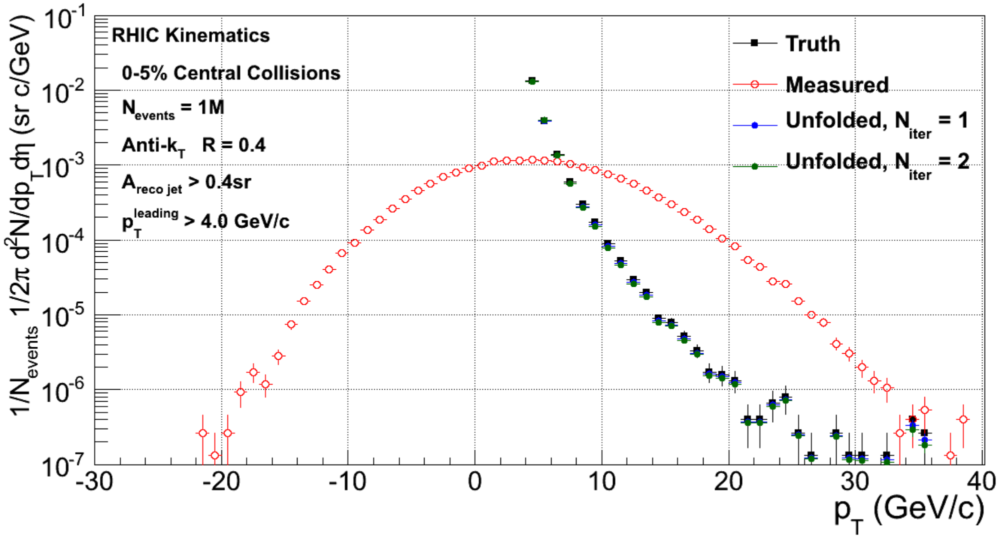
Similar result at RHIC (convergence for $p_T^{threshold}$ ~ 4 GeV)

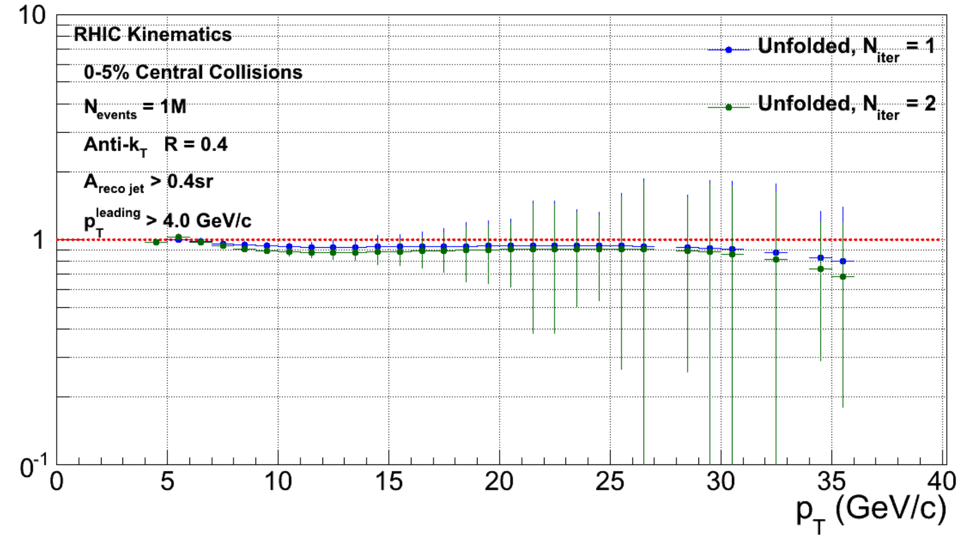
Generic observation: transition to correct unfolding solution corresponds to condition:

particle density at $p_T^{threshold}$ << 1/jet area

Rephrase: effective suppression of combinatoric jet population requires fragmentation bias with leading hadron p_T that is “rare” on the scale of jet area

Inclusive spectrum with $p_T^{leading}$ bias: RHIC





Similar result at RHIC (convergence for $p_T^{threshold}$ ~ 4 GeV)

Generic observation: transition to correct unfolding solution corresponds to condition:

particle density at $p_T^{threshold}$ << 1/jet area

Rephrase: effective suppression of combinatoric jet population requires fragmentation bias with leading hadron p_T that is “rare” on the scale of jet area

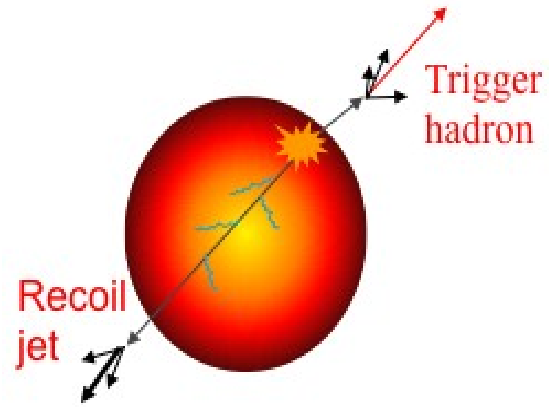
Caution: current calculation utilizes Single Particle fragmentation to isolate unfolding effects

- Spectrum of “particles” is much harder than phyiscal hadron spectrum
- Generically: expect transition in data at lower p_T (PYTHIA-based calculation in progress; quenching models? Ultimately, data will tell us..)

How to beat the fragmentation bias? hadron+jet coincidences

Coincidence of jet recoiling from hadron trigger

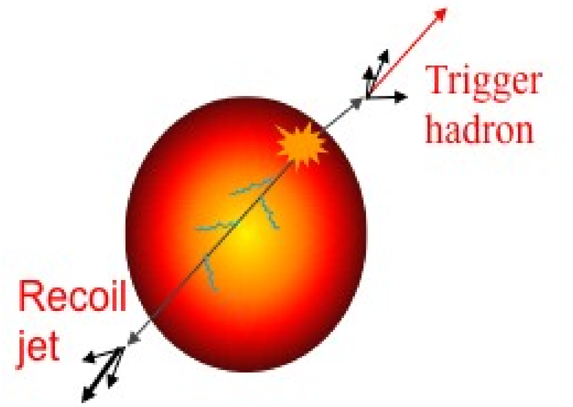
- $p_T^{trigger}$ large enough that trigger is likely leading particle of jet
- Trigger imposes surface bias → recoil jet traverses maximum path length in medium
- Observable: semi-inclusive recoil jet yield normalized per trigger

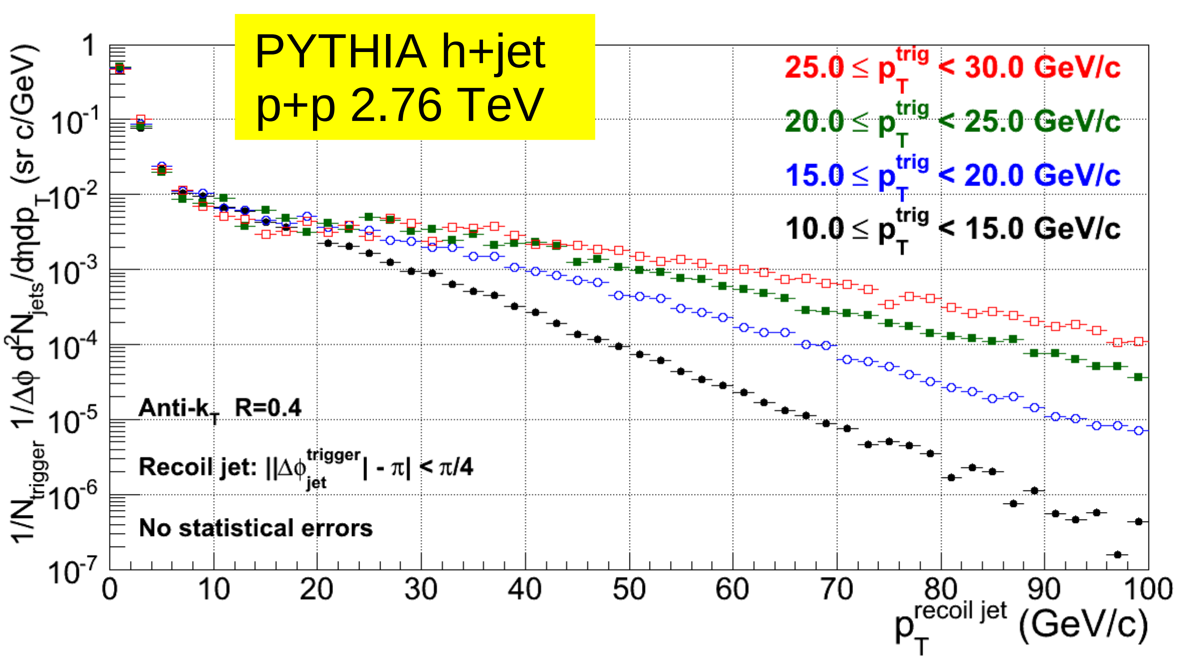


How to beat the fragmentation bias? hadron+jet coincidences

Coincidence of jet recoiling from hadron trigger

- $p_T^{trigger}$ large enough that trigger is likely leading particle of jet
- Trigger imposes surface bias → recoil jet traverses maximum path length in medium
- Observable: semi-inclusive recoil jet yield normalized per trigger



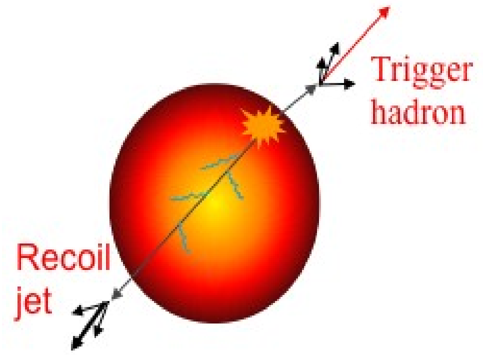


Heavy ions: hadron trigger isolates a single hard process, rest of event is background

Distribution same as in p+p except for:

- Jet quenching
- Smearing due to background fluctuations (combinatoric jets)

hadron+jet in heavy ion collisions: new observable



By definition: combinatoric jet distribution is uncorrelated with $p_T^{trigger}$

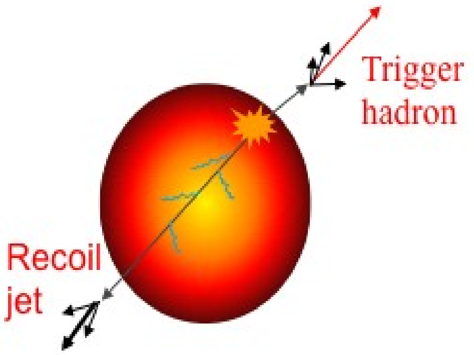
Opportunity: compare recoil jet distributions for two different (exclusive) intervals of $p_T$trigger

- Combinatoric jet part should be identical

- Hard jet part should depend on $p_T^{trigger}$

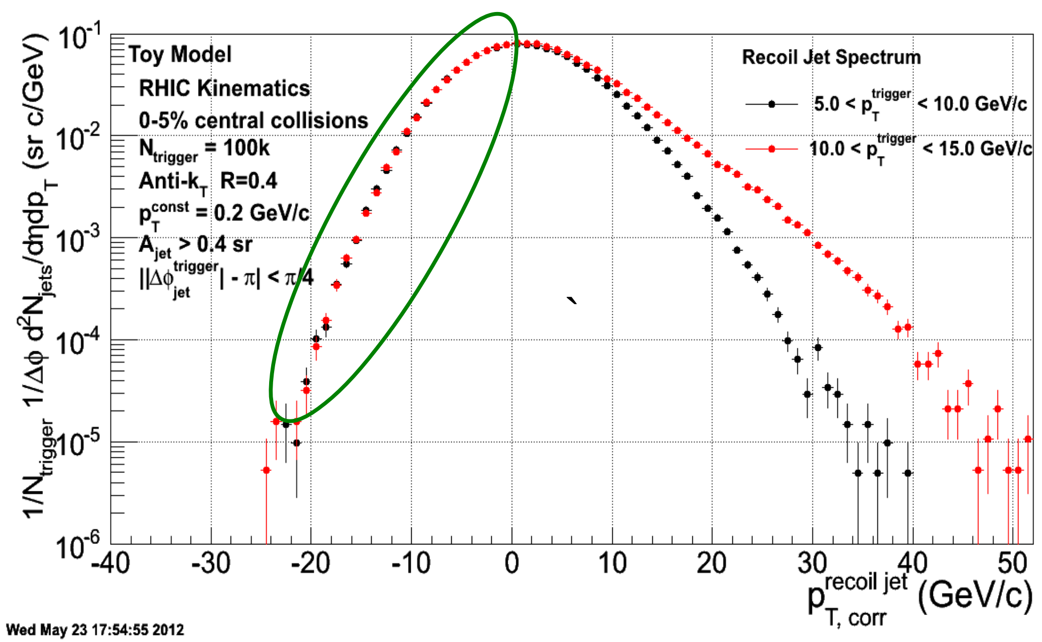
hadron+jet in heavy ion collisions: new observable

By definition: combinatoric jet distribution is uncorrelated with $p_T^{trigger}$



Opportunity: compare recoil jet distributions for two different (exclusive) intervals of $p_T$trigger

- Combinatoric jet part should be identical
- Hard jet part should depend on $p_T^{trigger}$



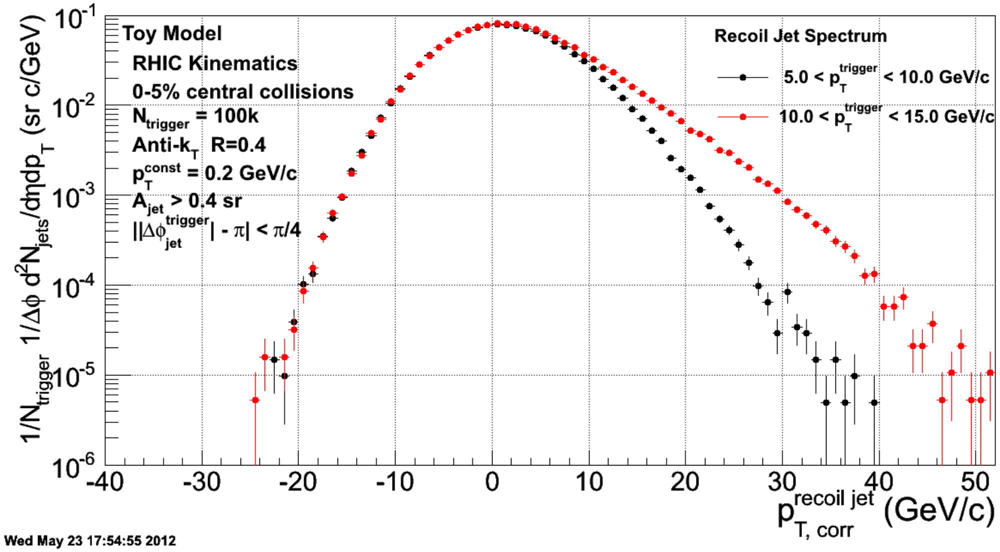
New observable: difference of the two distributions

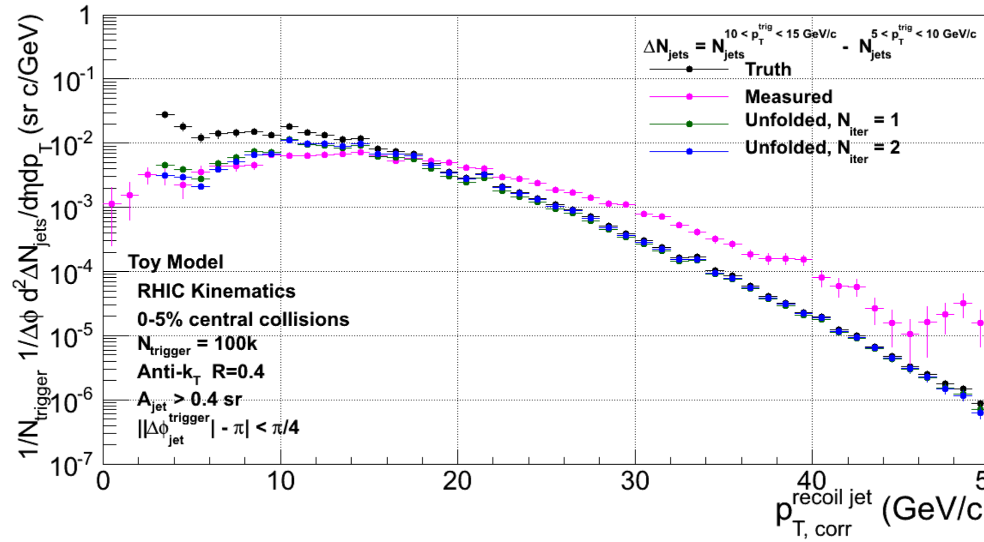
- Combinatorial jet population subtracted in fully data-driven way
- What's left: hard jet distribution smeared by background fluctuations → unfold (!)

What is it?

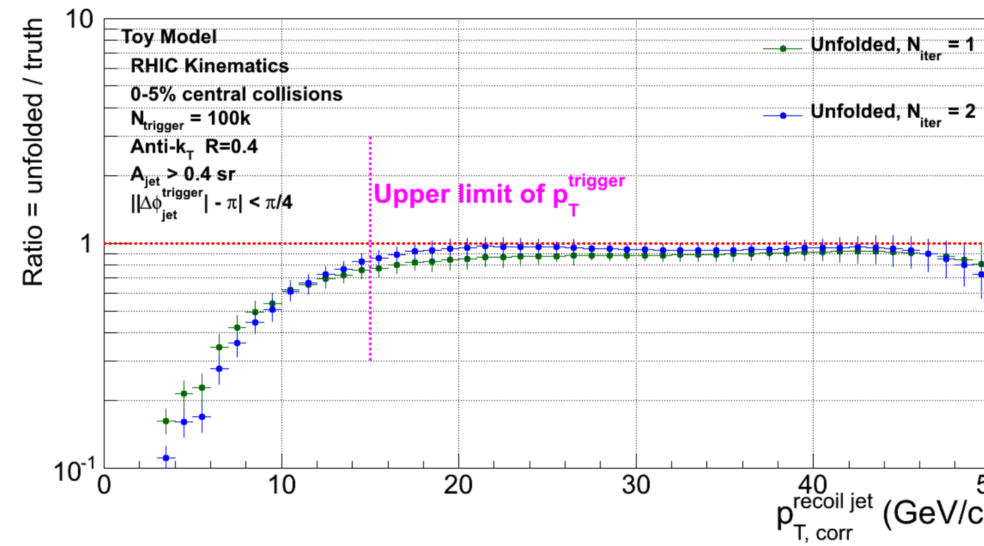
- The evolution of hard jet distribution as $p_T^{trigger}$ changes
- Unusual, but perfectly legitimate and perturbatively calculable

Differential h+jet coincidence at RHIC: unfolding





Unfolding converges stably to Truth above $p_T^{trigger}$ threshold



Correct result: subtraction is designed to suppress jet yield below threshold

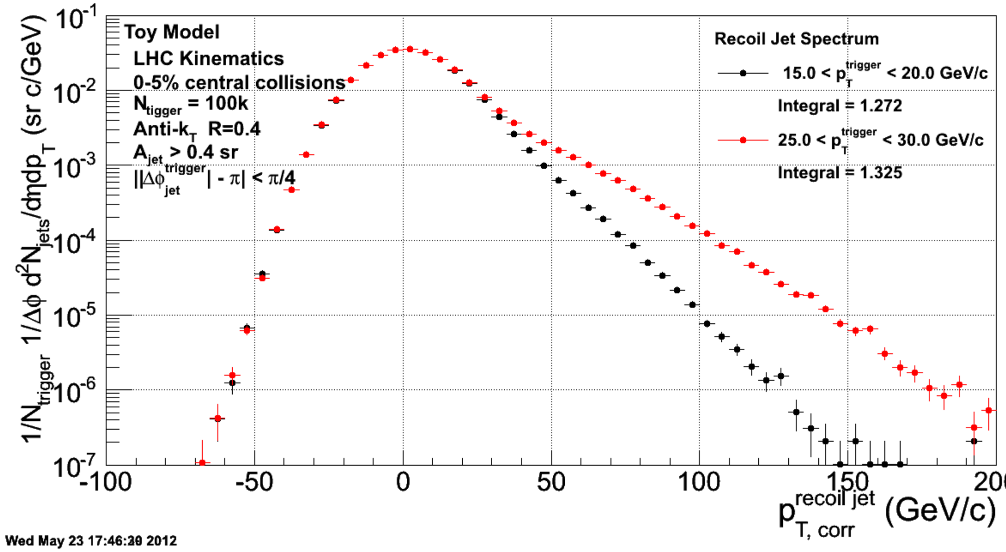
- physics bias (which is fine, and interesting)

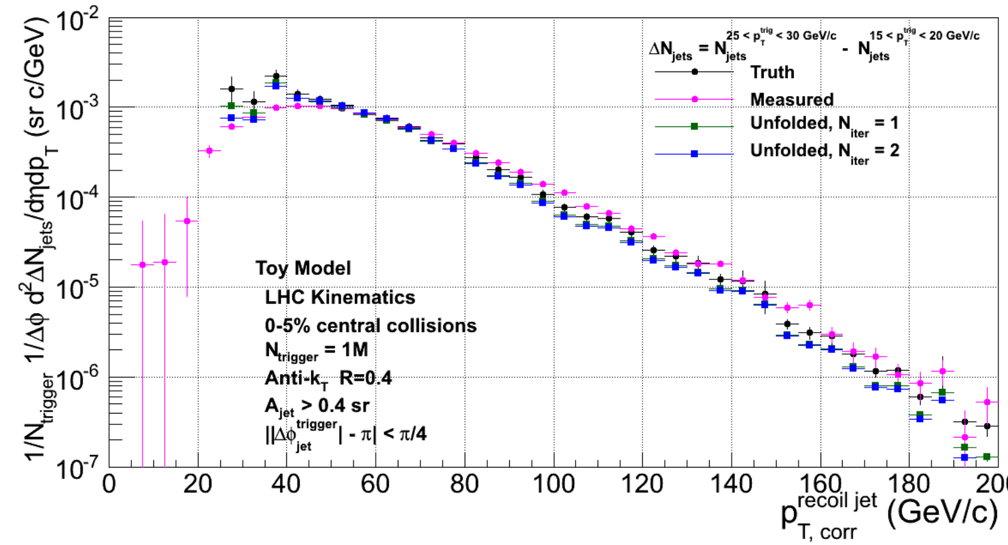
Unfolding bias: minimal

- Unfolded/Truth~1 to better than 10%

Fragmentation bias: none, by design

Differential h+jet coincidence at LHC

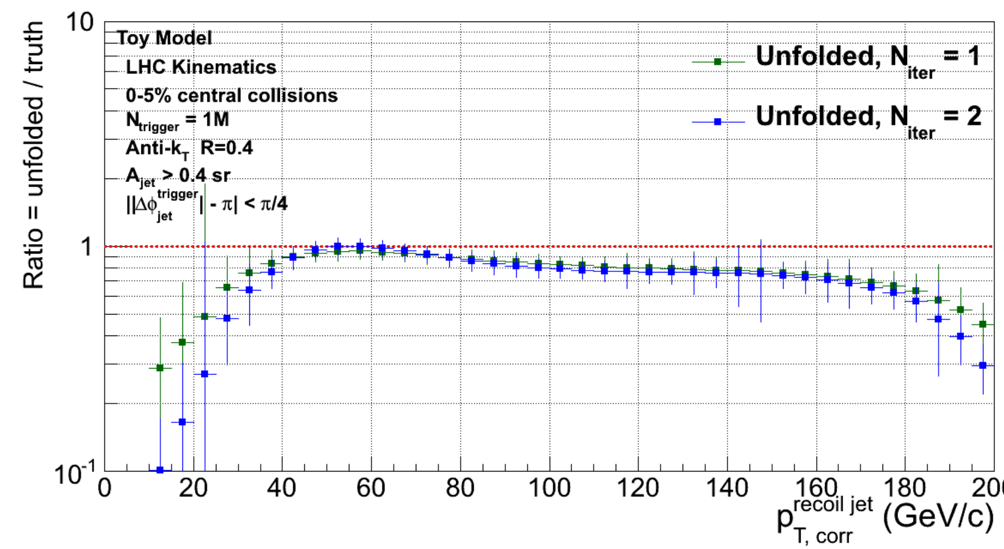




Similar picture to RHIC: unfolding converges stably to Truth above $p_T^{trigger}$ threshold

Residual unfolding bias ~10% (work in progress)

Fragmentation bias: none, by design



Summary and Outlook

We have reassessed the problem of jet reconstruction in heavy ion collisions

- Key issue: precision with which background fluctuations can be measured and corrected
- Problem recast as minimization of both unfolding and fragmentation biases

New analysis methods proposed to minimize these biases in a transparent way

- Both quasi-inclusive and coincidence observables
- Optimally implemented via STAR/ALICE approach to jet measurements
- Methods were tested on Model studies representative of data
- Methods work well over full jet kinematic range at RHIC and LHC

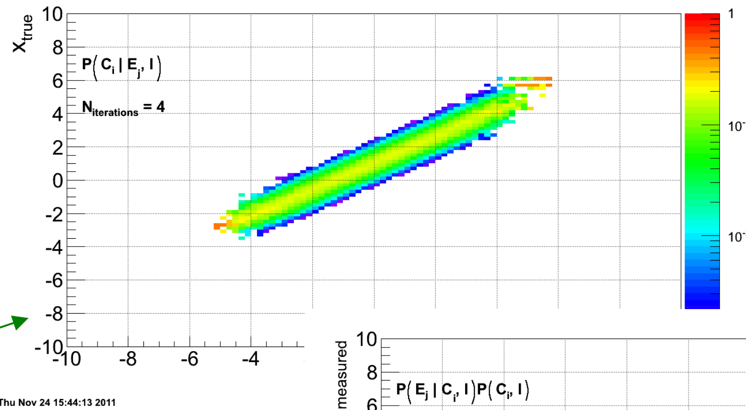
Next step: apply to data

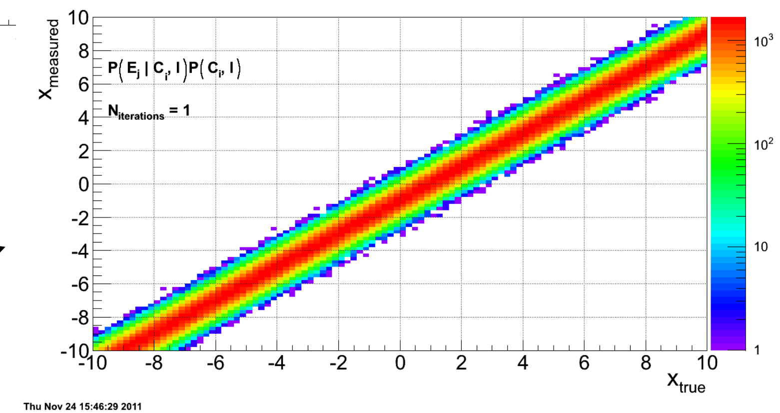
Backup Slides

Bayesian Unfolding

P(E_j|C_i,I): probability of **effect j (E_j)** has been caused due **cause i (C_i)**

P(C_i|E_j,I): probability of **cause i (C_i)** comes from **effect j (E_j)**



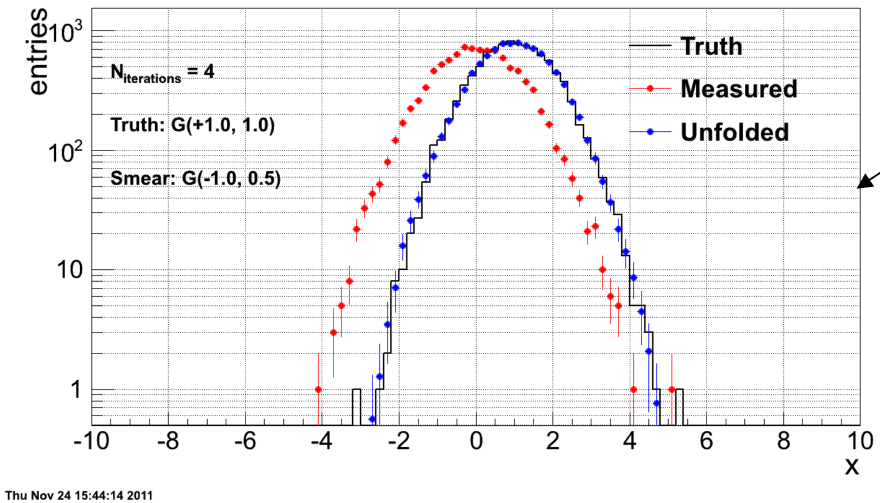


By the knowledge of **P(E_j|C_i)** and a choice of prior distribution, one obtains **P(C_i|E_j):**

$$P(C_i|E_j,I) = \frac{P(E_j|C_i,I) \cdot P_0(C_i,I)}{\sum_{l=0}^{n_C} P(E_j|C_l,I) \cdot P_0(C_l,I)}$$

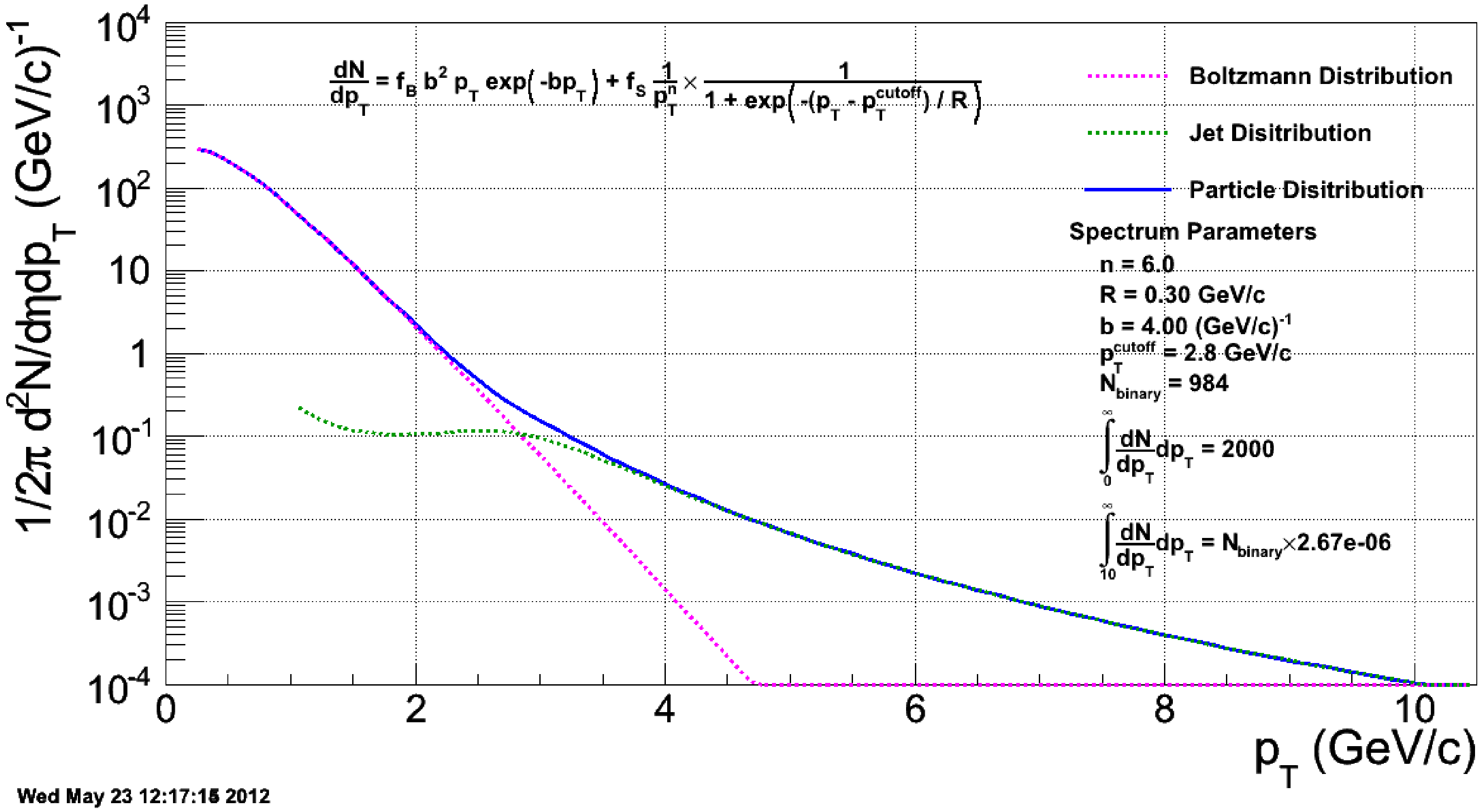
The **unfolded spectrum n(C_i)** is obtained from the **measured spectrum n(E_j)** by:

$$n(C_i) = \sum_{j=0}^{n_E} P(C_i|E_j,I) \cdot n(E_j)$$

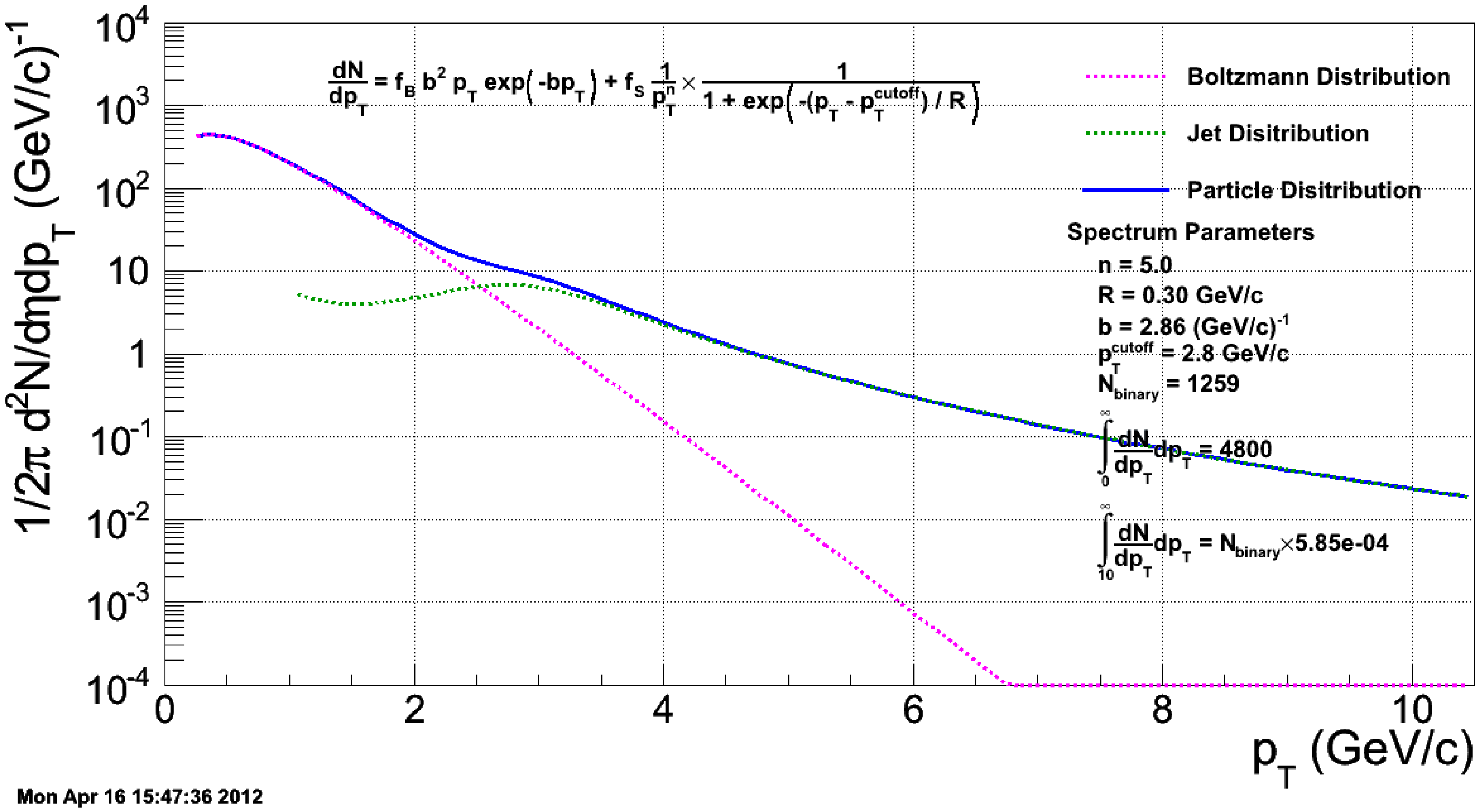


n(C_i) can be used as a **new input for the prior** and unfolding can be done again (iterations) → the number of iterations is the regularization parameter

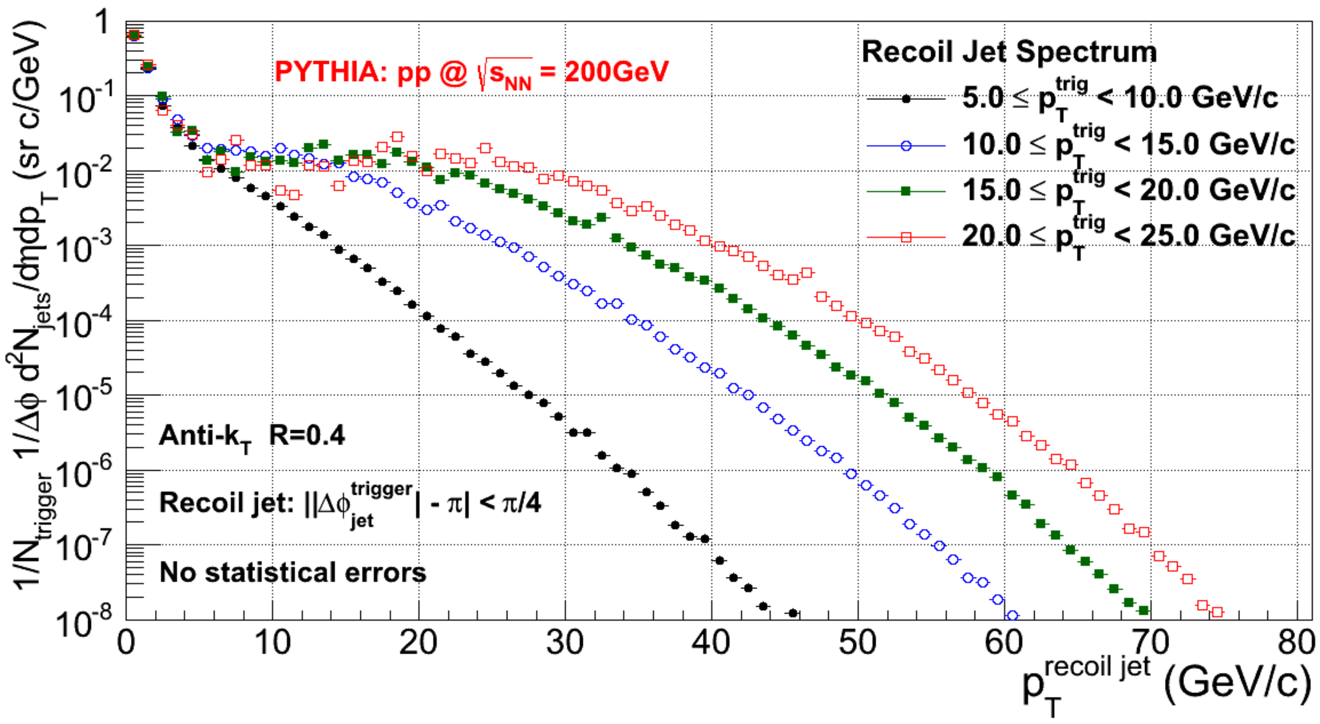
Toy Model – RHIC



Toy Model – LHC

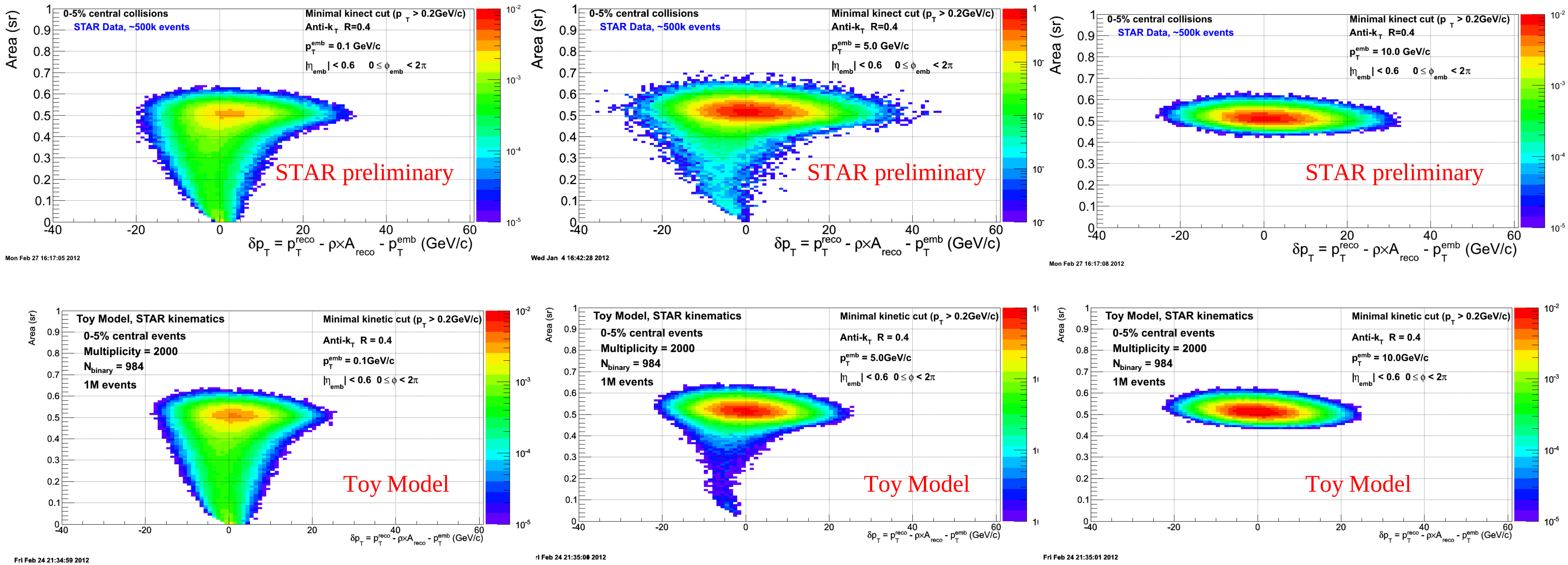


Coincidence Analysis: h+jet (I)



- Trigger on a high pT particle and report recoil jet
- Utilize exclusive trigger classes
- Allows differential analysis

Toy Model Validation

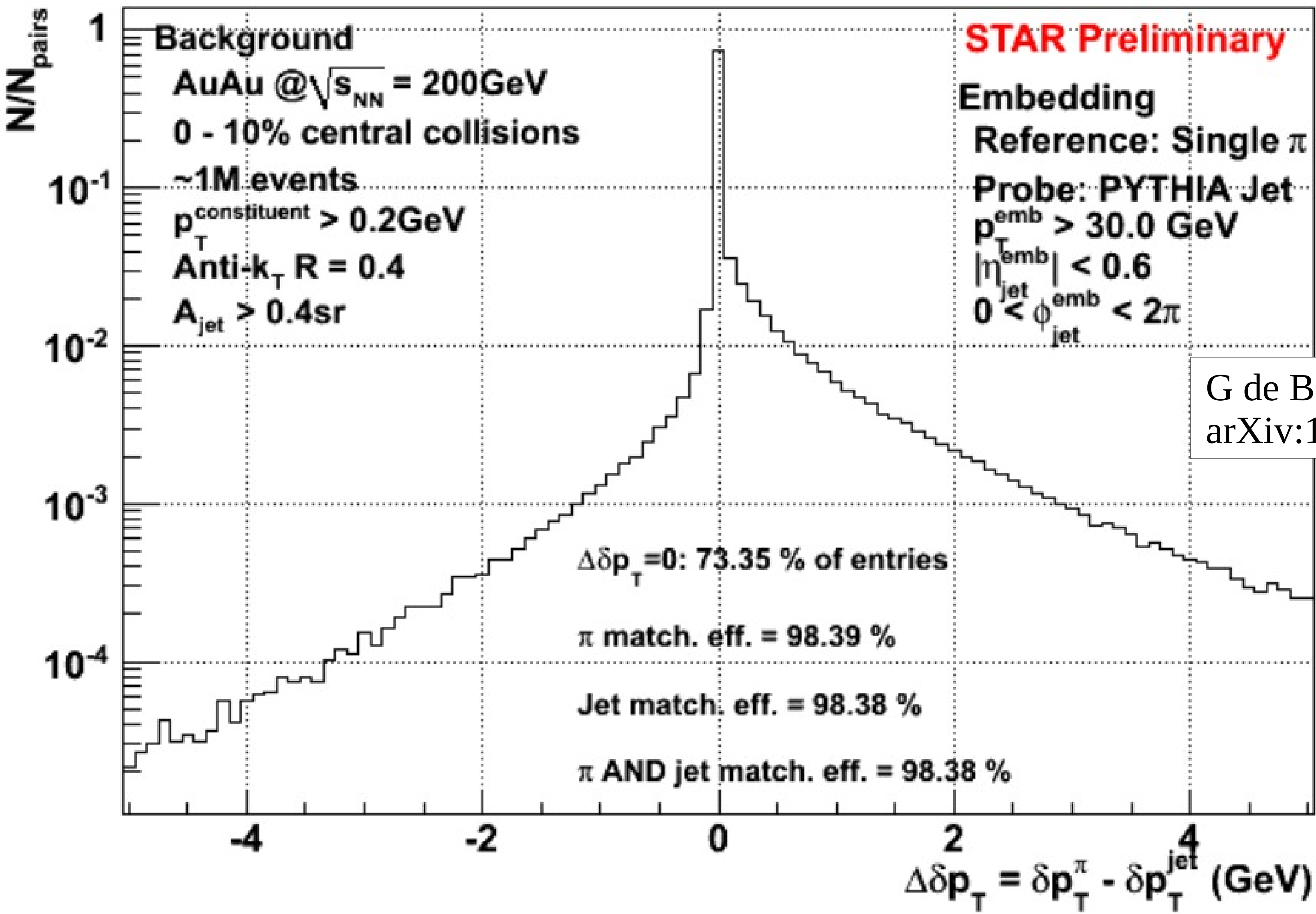


Comparison: delta pT X Jet area:

Same behavior for toy model events as real data

Note transition at pTemb = 5GeV/c.

$$\Delta\delta p_T = \delta p_T^{\pi} - \delta p_T^{\rm jet}$$



G de Barros (STAR)
arXiv:1109.4386