Sleuth

A quasi-model-independent new physics search strategy

Motivation

Strategy

Algorithm

Results

Bruce Knuteson
Berkeley

Motivation

Most searches follow a well-defined set of steps:

- Select a model to be tested
- Find a measurable prediction of the model differing as much as possible from the prediction of the Standard Model
- Check those predictions against the data

This approach becomes problematic if the number of competing candidate theories is large . . . and it is!

Is it possible to perform some kind of "generic" search?

Sleuth

Motivation

The word "model" can connote varying degrees of generality

generality

1 - A special case of a class of models with definite parameters
  mSUGRA with $M_{1/2}=200$, $M_0=220$, $\tan\beta=2$, $\mu<0$

2 - A special case of a class of models with unspecified parameters
  mSUGRA

3 - A class of models
  SUGRA

4 - A more general class of models
  gravity-mediated supersymmetry

5 - An even more general class of models
  supersymmetry

6 - A set of even more general classes of models
  theories of electroweak symmetry breaking

Most new physics searches have generality $\approx 1\frac{1}{2}$ on this scale

We are shooting for a search strategy with a generality of ≈ 6 . . . .

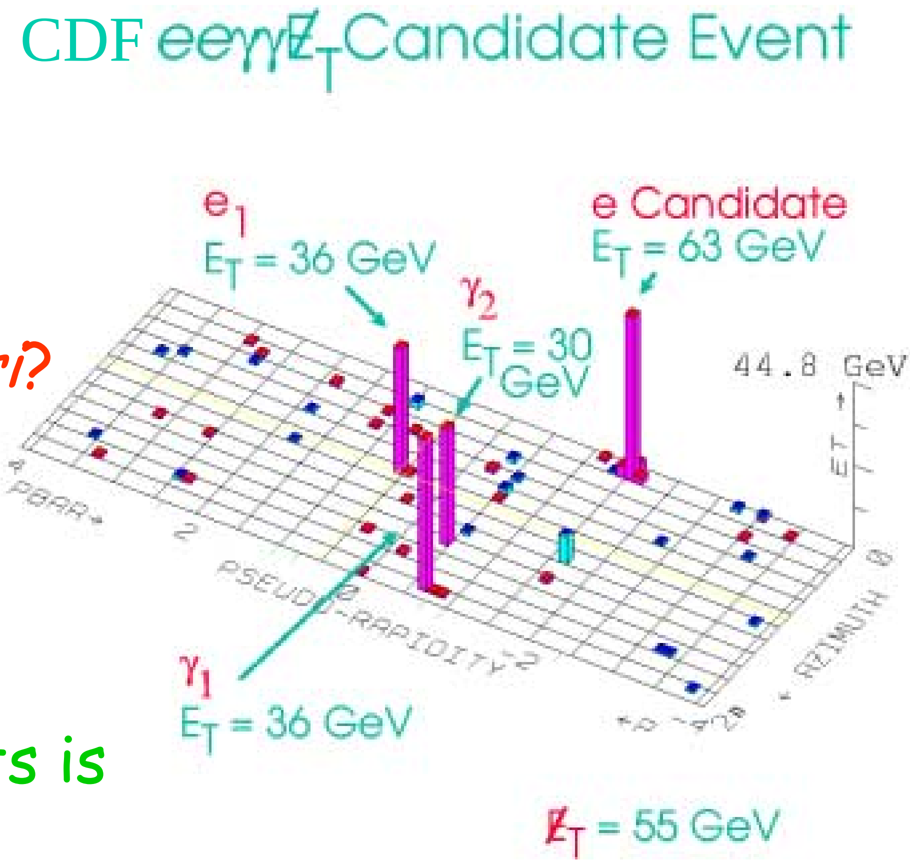
Motivation

Another, separate issue:

How do we quantify the "interestingness" of a few strange events *a posteriori*?

e.g. Barnett and Hall, PRL 77:3506 (1996)

After all, the probability of seeing <u>exactly</u> those events is zero!



How excited should we be? How can we possibly perform an unbiased analysis after seeing the data?

Sleuth

Motivation

Other advantages of **Sleuth**:

Emphasizes an understanding of the data (rather than what the data have to say about a particular model)

Provides a systematic method for analyzing the entire data set (leaving no stone unturned!)

Allows an approach that keeps attention focused on the most promising channels (rather than optimizing cuts for a signal that does not exist)

Allows for surprises . . .

Sleuth

A new generic search algorithm

- Motivation
- Strategy → Final states, Variables
- Algorithm
- Results

Initial thought:

Consider inclusive final states, such as e μ X

However:

- The presence of an extra object in an event often qualitatively changes the probable interpretation of the event
- The presence of an extra object in an event generally changes the variables that one would want to use to characterize the event
- Allowing inclusive final states leaves an ambiguity in definition

Therefore:

We consider *exclusive* final states

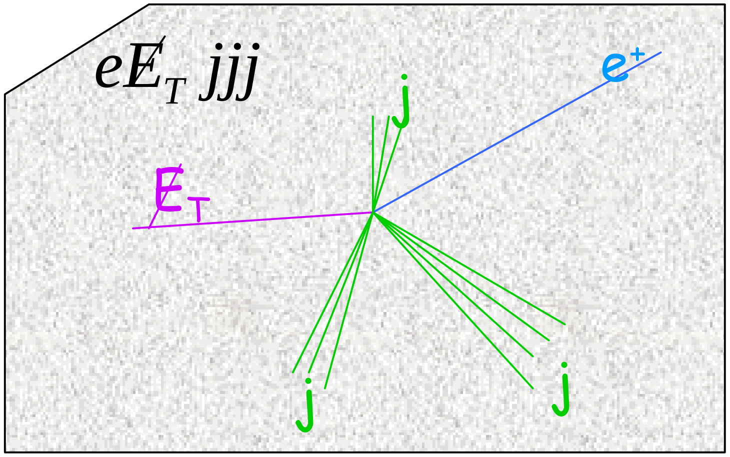
More precisely:

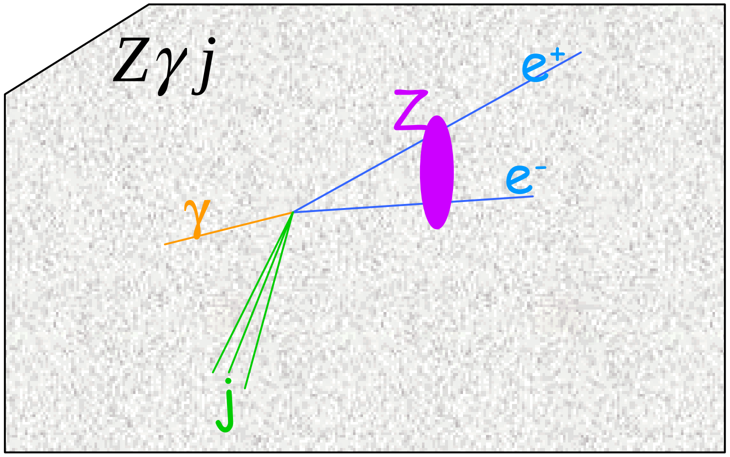
We assume the existence of standard object definitions

These define e, μ, τ, γ, j, b, c, $\not{E}_T$, W, and Z

All events that contain the same numbers of each of these objects belong to the same final state

e.g.,





Initial thought:

Construct a set of variables for each possible final state

However:

- There are a lot of final states!
  - $e\mu X$ alone comprises several final states
- Our variables need to be robust
  - Otherwise it will be too easy to change them after looking at the data!
- Our variables ought to be
  - well-motivated (sensitive to new physics)
  - simple and few

Therefore:

Instead of choosing a separate set of variables for every conceivable final state, we construct a general rule

$$\mathcal{V} : (\text{final state}) \rightarrow \{ \text{variables} \}$$

What is it we're looking for?

The physics responsible for EWSB

What do we know about it?

Its natural scale is a few hundred GeV

What characteristics will such events have?

Final state objects with large transverse momentum

What variables do we want to look at?

p_T's

Variables

If the final state contains	Then consider the variable
1 or more lepton	$\sum p_T^{\ell}$
1 or more γ/W/Z	$\sum p_T^{\gamma/W/Z}$
1 or more jet	$\sum' p_T^{j}$
missing E_T	$\not{E}_T$

(adjust slightly for idiosyncrasies of each experiment)

Sleuth

A new generic search algorithm

Motivation

Strategy

Algorithm

Results

Algorithm — Overview

For each final state . . .

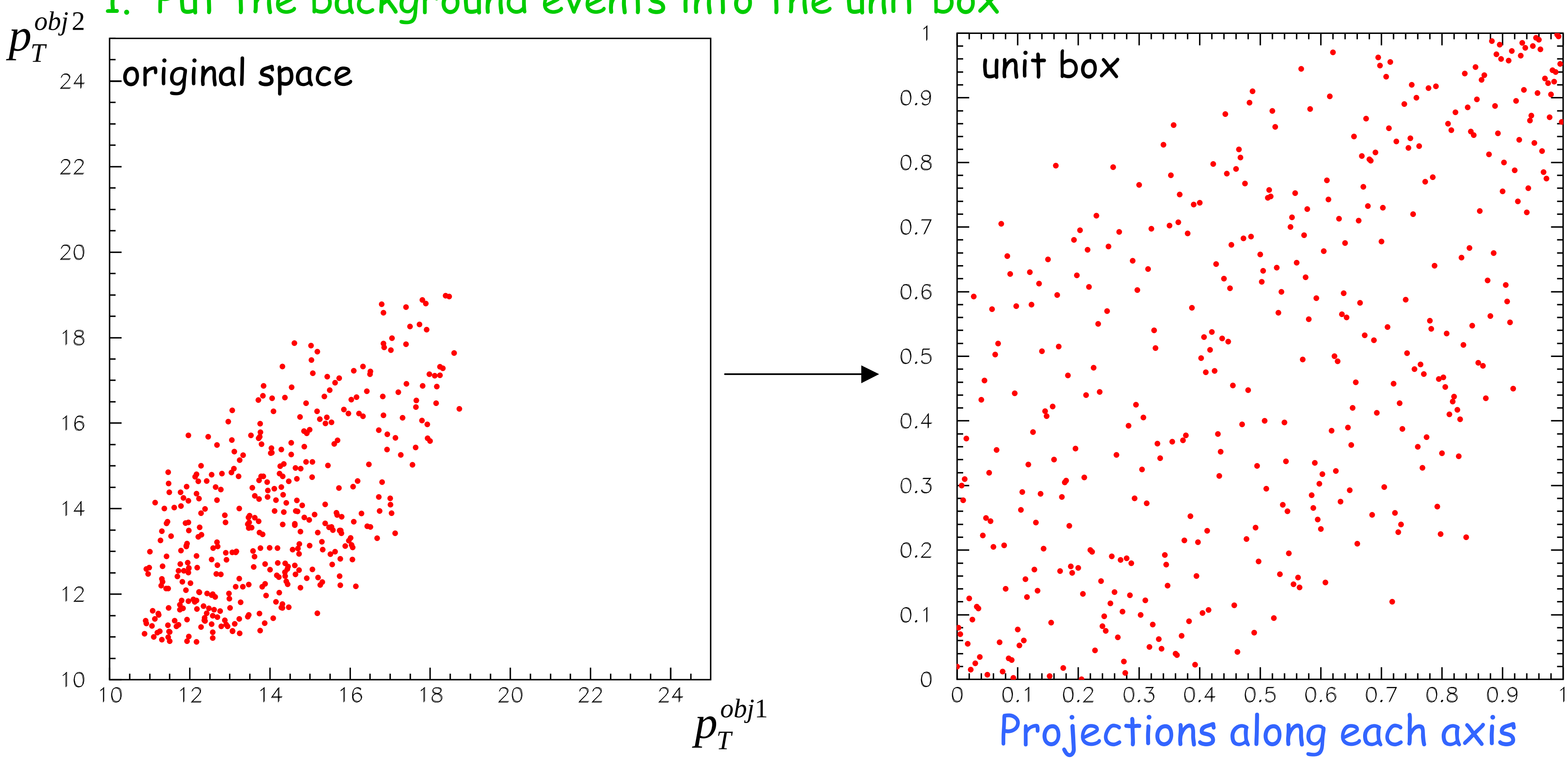
Input: 1 data file, estimated backgrounds

- transform variables into the unit box
- define regions about sets of data points
  - Voronoi diagrams
- define the "interestingness" of an arbitrary region
  - the probability that the background within that region fluctuates up to or beyond the observed number of events
- search the data to find the most interesting region, $\mathcal{R}$
- determine $\mathcal{P}$, the fraction of *hypothetical similar experiments* (hse's) in which you would see something more interesting than $\mathcal{R}$
  - Take account of the fact that we have looked in many different places

Output: $\mathcal{R}$, $\mathcal{P}$

We begin by applying a variable transformation that makes the background distribution uniform in the “unit box” — $[0,1]^d$

1. Put the background events into the unit box

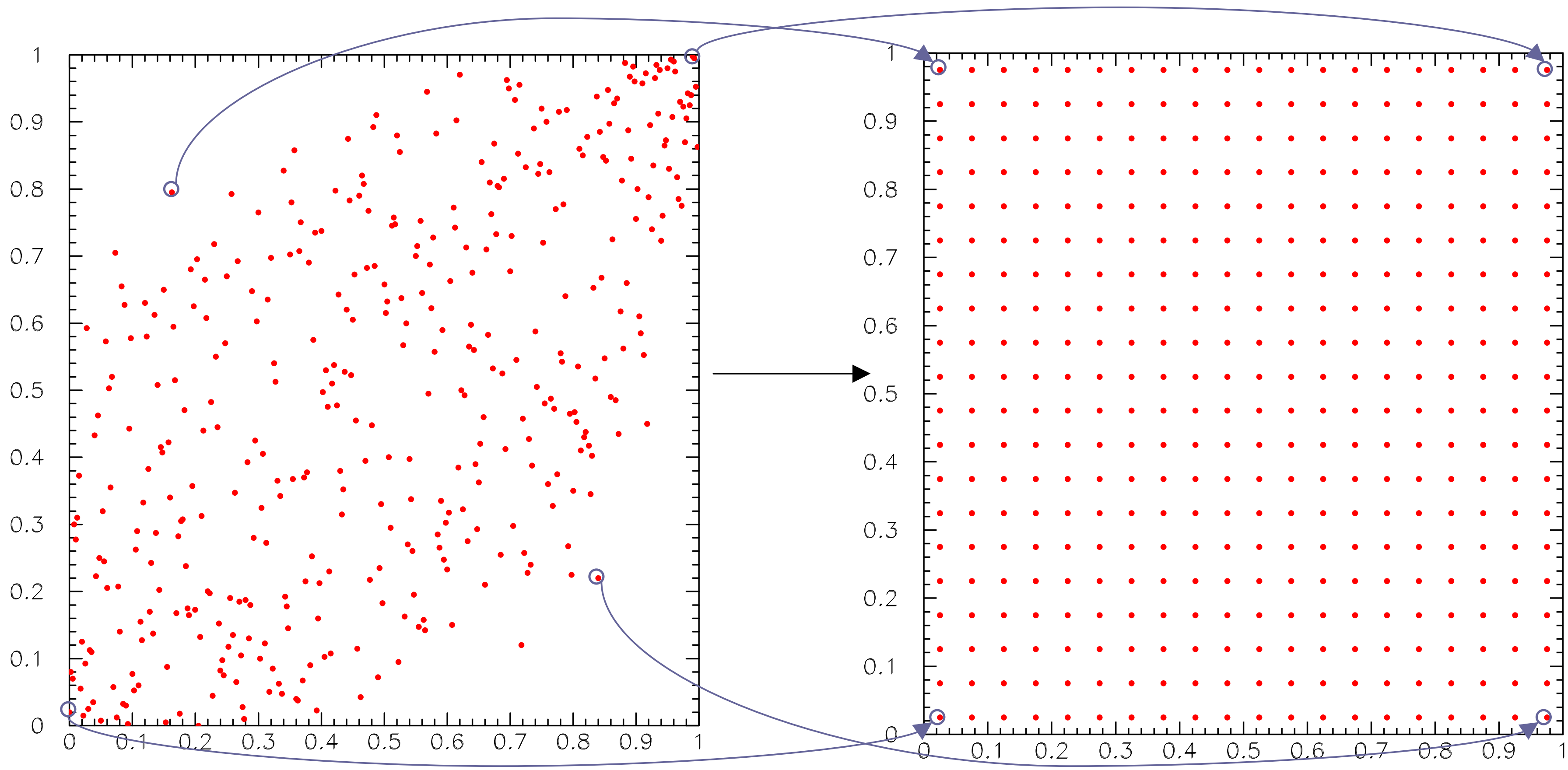


Projections along each axis are now uniform

Variable transformation

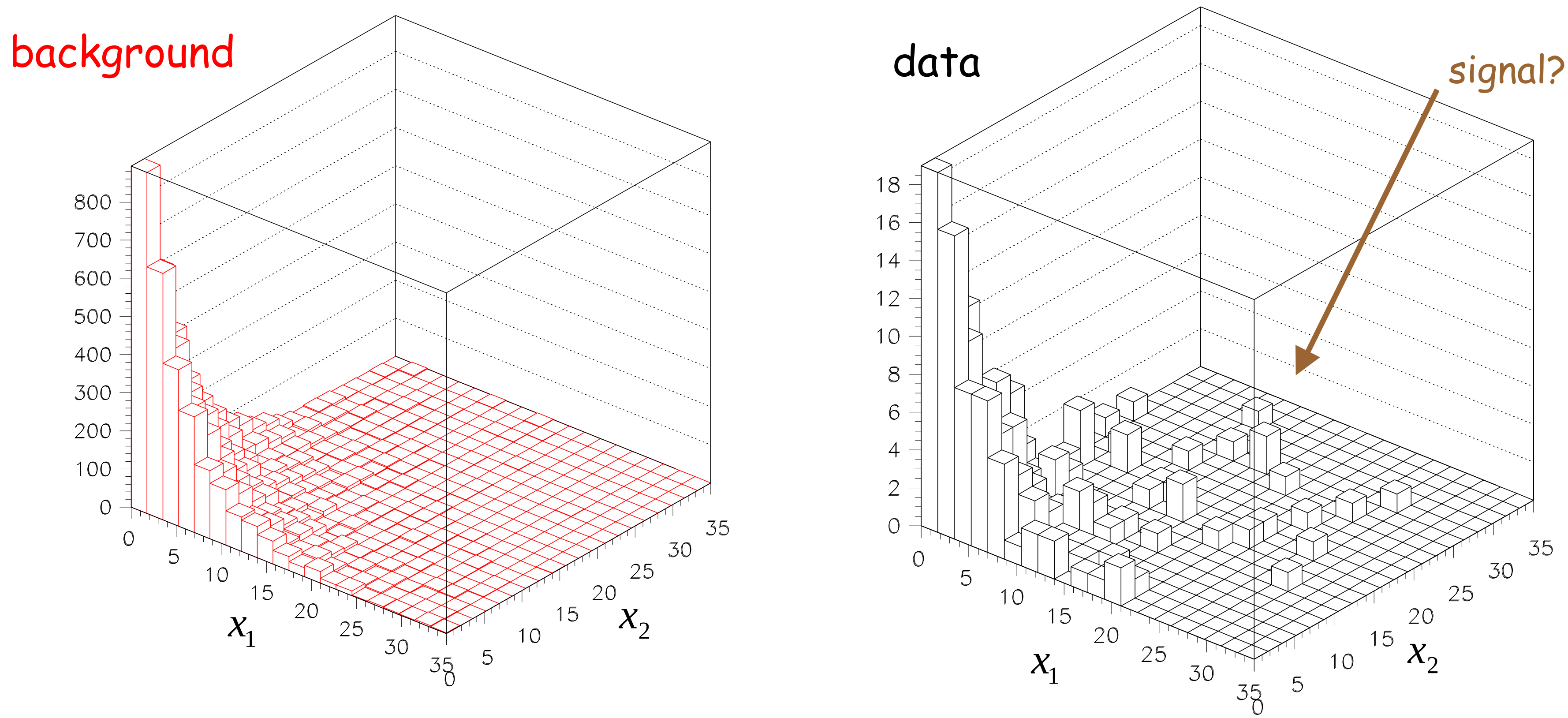
2. Map the background events onto a uniform grid

[Iteratively switch pairings to minimize the maximum distance moved]

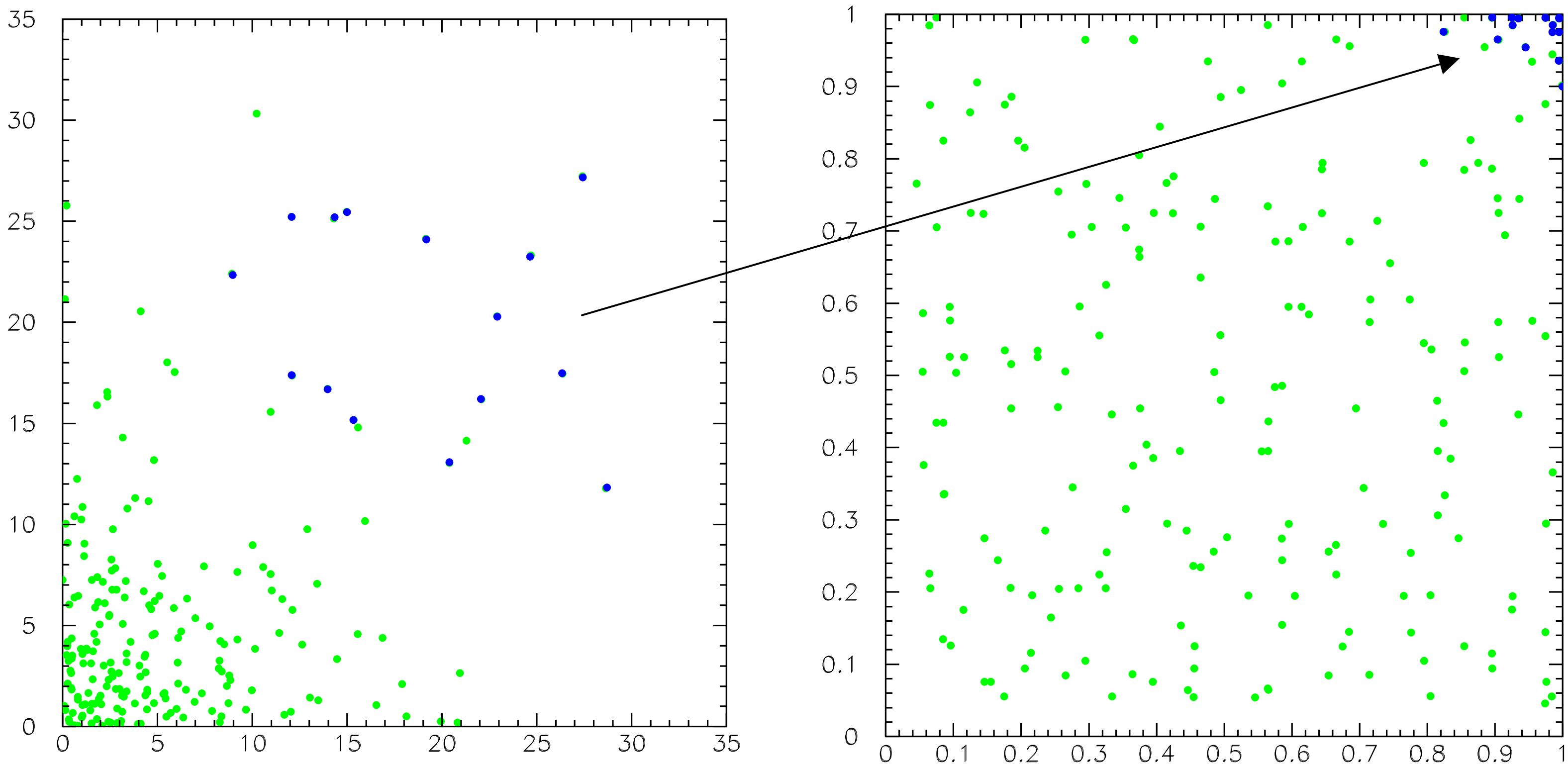


Algorithm — Variable transformation

A quick example of how this might look for data:

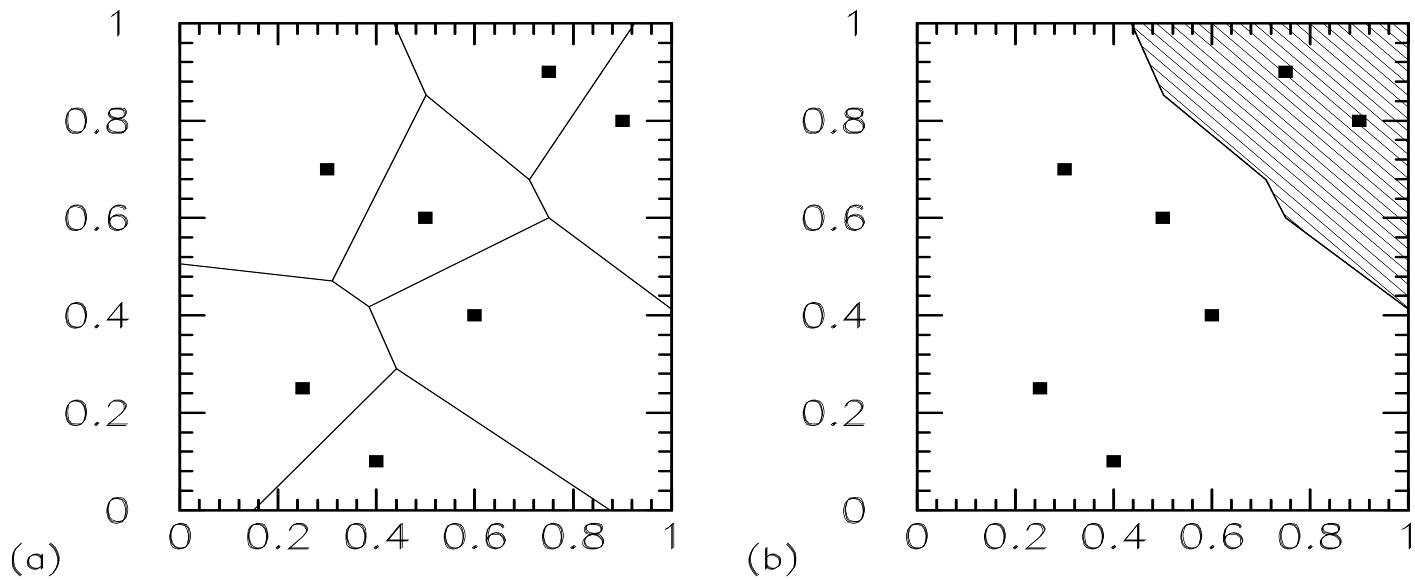


The transformation maps the signal region into the upper right-hand corner of the unit box



The background data events are uniformly distributed, as desired, and the signal cluster is “obvious”

An *N-region* (about a cluster of N data points) is the set of all values of x closer to a data point in that cluster than to any other data point in the sample.

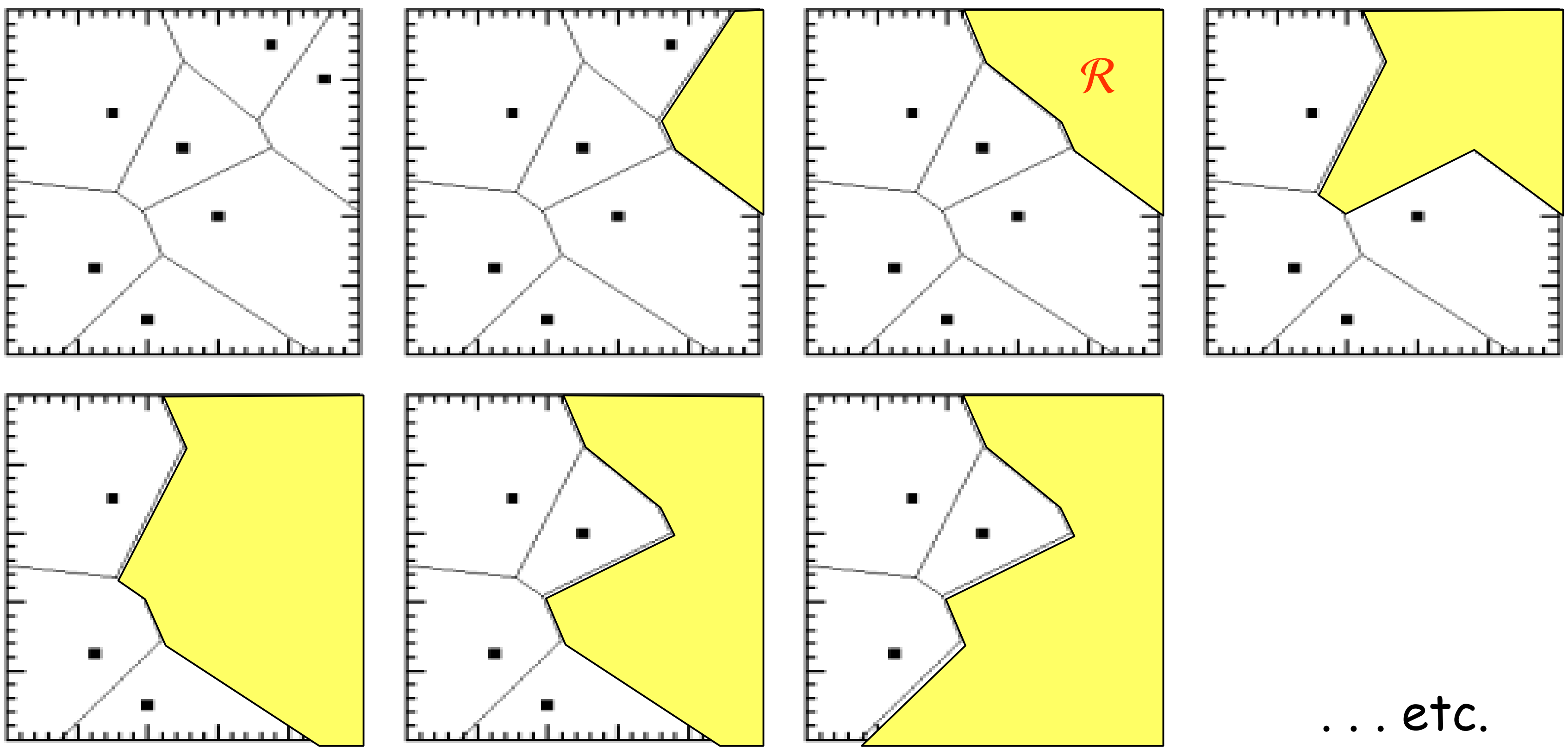


seven 1-regions

one 2-region

Voronoi diagrams

Search the space to find the region of greatest excess, $\mathcal{R}$



Perform many hypothetical similar experiments

- generate "data samples" from the background distributions
  - Allow numbers of events from each background source to vary according to statistical and systematic errors
- find the most interesting region in each pseudo sample
  - Use same searching algorithm as for the actual data
- compare the most interesting region in each pseudo sample with $\mathcal{R}$
- Determine $\mathcal{P}$, the fraction of *hypothetical similar experiments* in which you see something more interesting than $\mathcal{R}$

Sleuth

A new generic search algorithm

Motivation

Strategy

Algorithm

Results

Let the backgrounds include

1)
- fakes
- $Z \to \tau\tau$
- WW
- tt

DØ data

Data Set	$\mathcal{P}$
$e\mu\not{E}_T$	2.4σ
$e\mu\not{E}_T j$	0.4σ
$e\mu\not{E}_T jj$	2.3σ
$e\mu\not{E}_T jjj$	0.3σ
Combined	1.9σ

Excesses corresponding (presumably) to WW and $t\bar{t}$

2)
- fakes
- $Z \to \tau\tau$
- WW
- tt

DØ data

Data Set	$\mathcal{P}$
$e\mu\not{E}_T$	1.1σ
$e\mu\not{E}_T j$	0.1σ
$e\mu\not{E}_T jj$	1.9σ
$e\mu\not{E}_T jjj$	0.2σ
Combined	1.2σ

Excess corresponding (presumably) to $t\bar{t}$

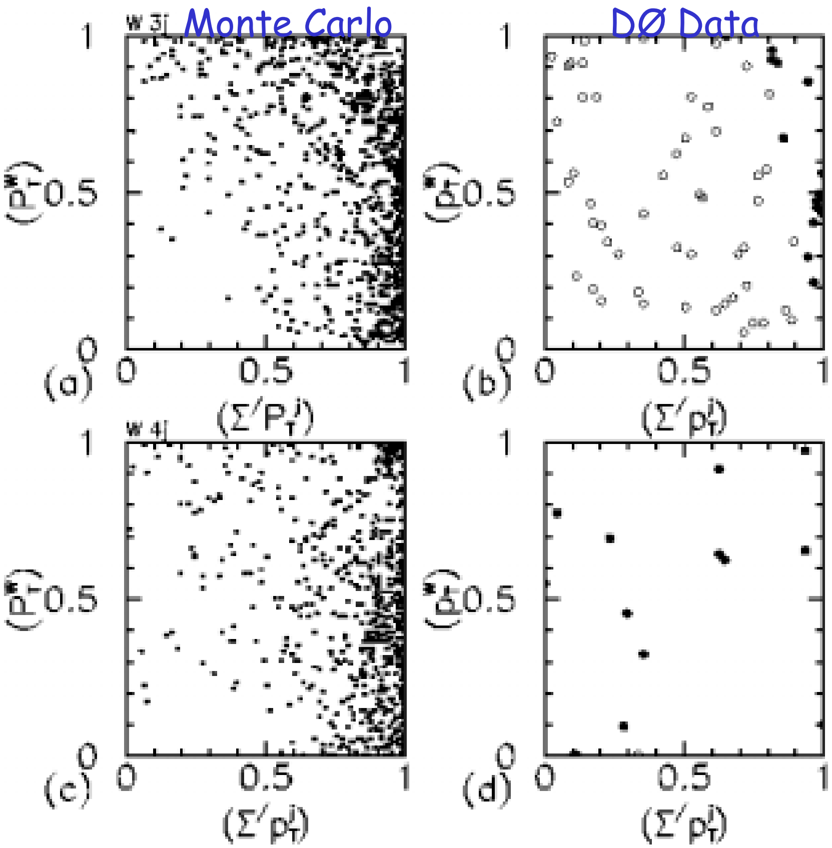
3)
- fakes
- $Z \to \tau\tau$
- WW
- $t\bar{t}$

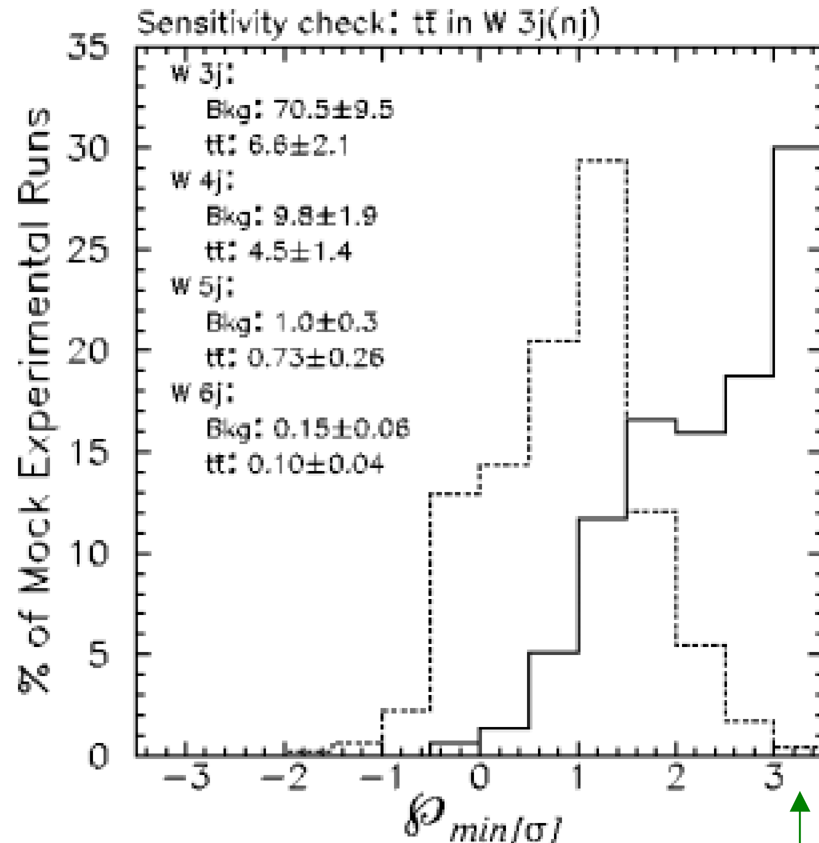
DØ data

Data Set	$\mathcal{P}$
$e\mu\not{E}_T$	1.1σ
$e\mu\not{E}_T j$	0.1σ
$e\mu\not{E}_T jj$	0.5σ
$e\mu\not{E}_T jjj$	-0.5σ
Combined	-0.6σ

No evidence for new physics

Could Sleuth have found $t\bar{t}$ in the lepton+jets channel?





← All over-flows in last bin

Sleuth finds $\mathcal{P}_{min}$ > 3σ in 30% of an ensemble of mock experimental runs

Sensitivity check: Leptoquarks in eejj

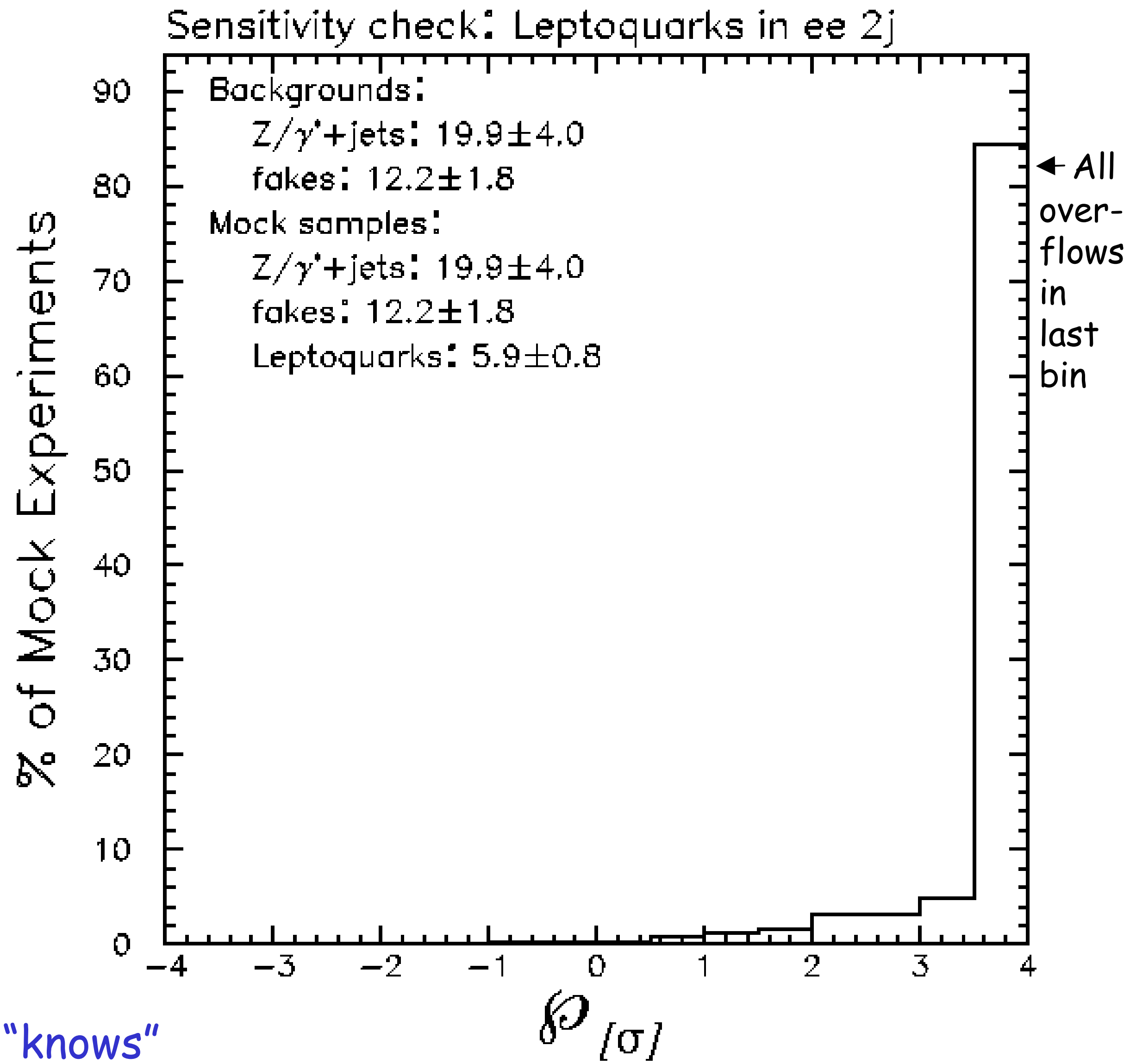
We can run mock experiments with hypothetical signals, too

What if our data contained leptoquarks?

(Assume scalar, $\beta = 1$, $m_{LQ} = 170$ GeV)

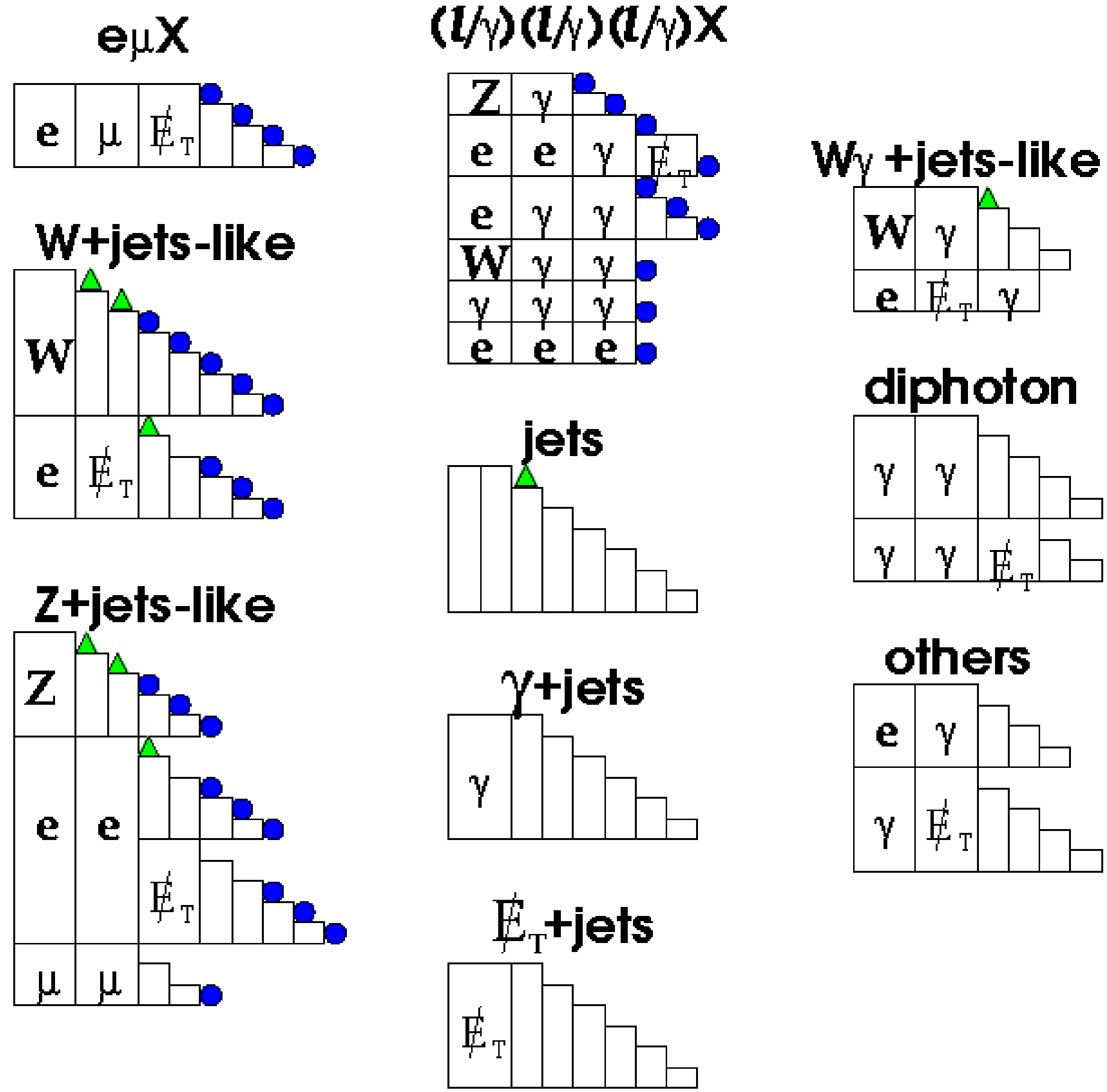
Sherlock finds $\mathcal{P} > 3.5\sigma$ in > 80% of the mock experiments

(Remember that Sherlock "knows" nothing about leptoquarks!)



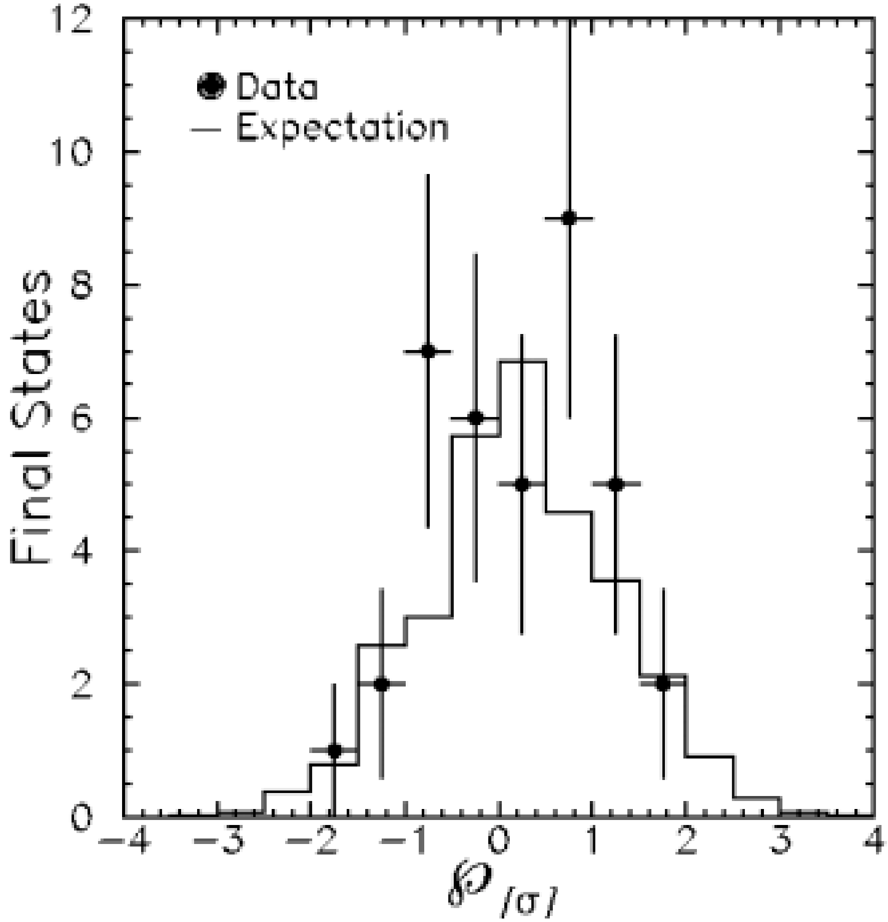
There were ≈ 80 populated final states at DØ in Run I.

We have applied Sleuth to roughly half of these final states.



Results

DØ data



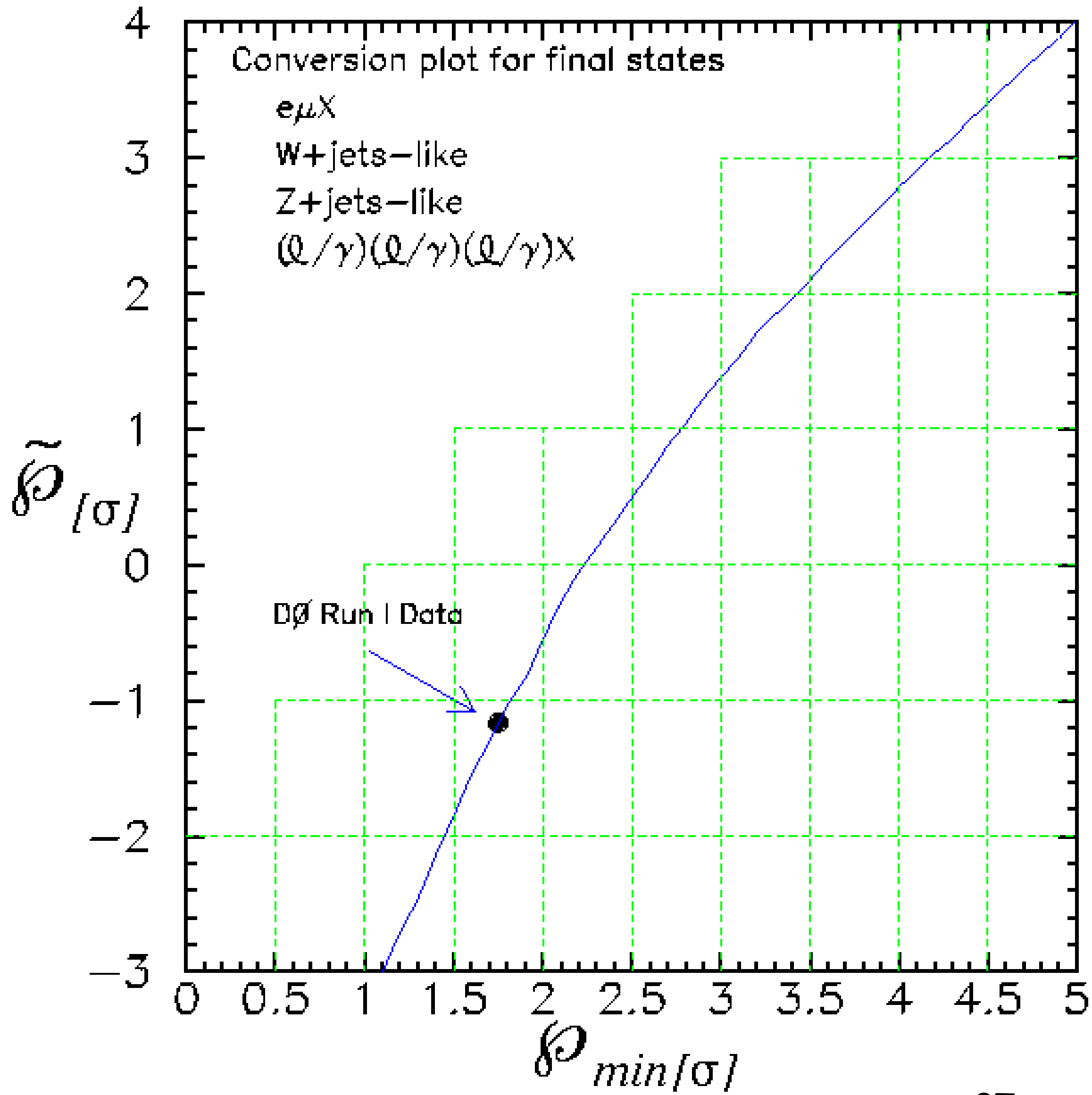
Data set	$\mathcal{P}$
$e\mu X$	
$e\mu\not\!\!E_T$	0.14 ($+1.08\sigma$)
$e\mu\not\!\!E_T j$	0.45 ($+0.13\sigma$)
$e\mu\not\!\!E_T\ 2j$	0.31 ($+0.50\sigma$)
$e\mu\not\!\!E_T\ 3j$	0.71 (-0.55σ)
W+jets-like	
$W\ 2j$	0.29 ($+0.55\sigma$)
$W\ 3j$	0.23 ($+0.74\sigma$)
$W\ 4j$	0.53 (-0.08σ)
$W\ 5j$	0.81 (-0.88σ)
$W\ 6j$	0.22 ($+0.77\sigma$)
$e\not\!\!E_T\ 2j$	0.76 (-0.71σ)
$e\not\!\!E_T\ 3j$	0.17 ($+0.95\sigma$)
$e\not\!\!E_T\ 4j$	0.13 ($+1.13\sigma$)
Z+jets-like	
$Z\ 2j$	0.52 (-0.05σ)
$Z\ 3j$	0.71 (-0.55σ)
$Z\ 4j$	0.83 (-0.95σ)
$ee\ 2j$	0.72 (-0.58σ)
$ee\ 3j$	0.61 (-0.28σ)
$ee\ 4j$	0.04 ($+1.75\sigma$)
$ee\not\!\!E_T\ 2j$	0.68 (-0.47σ)
$ee\not\!\!E_T\ 3j$	0.36 ($+0.36\sigma$)
$ee\not\!\!E_T\ 4j$	0.06 ($+1.55\sigma$)
$\mu\mu\ 2j$	0.08 ($+1.41\sigma$)
$(\ell/\gamma)(\ell/\gamma)(\ell/\gamma)X$	
eee	0.89 (-1.23σ)
$Z\gamma$	0.84 (-0.99σ)
$Z\gamma j$	0.63 (-0.33σ)
$ee\gamma$	0.88 (-1.17σ)
$ee\gamma\not\!\!E_T$	0.23 ($+0.74\sigma$)
$e\gamma\gamma$	0.66 (-0.41σ)
$e\gamma\gamma j$	0.21 ($+0.81\sigma$)
$e\gamma\gamma\ 2j$	0.30 ($+0.52\sigma$)
$W\gamma\gamma$	0.18 ($+0.92\sigma$)
$\gamma\gamma\gamma$	0.41 ($+0.23\sigma$)
$\tilde{\mathcal{P}}$	0.89 (-1.23σ)

Results agree well with expectation

No evidence of new physics is observed

We can account for the fact that we have looked at many different final states by computing $\tilde{\mathcal{P}}$

The correspondence between $\tilde{\mathcal{P}}$ and the minimum $\mathcal{P}$ found for the final states that we have considered is shown here



Conclusions

- **Sleuth** is a quasi-model-independent search strategy for new high p_T physics
  - Defines final states and variables
  - Systematically searches for and quantifies regions of excess
- Allows an *a posteriori* analysis of interesting events
- **Sleuth** appears sensitive to new physics
- But finds no evidence of new physics in DØ data
- Should be a useful data-driven search engine in Run II

hep-ex/0006011