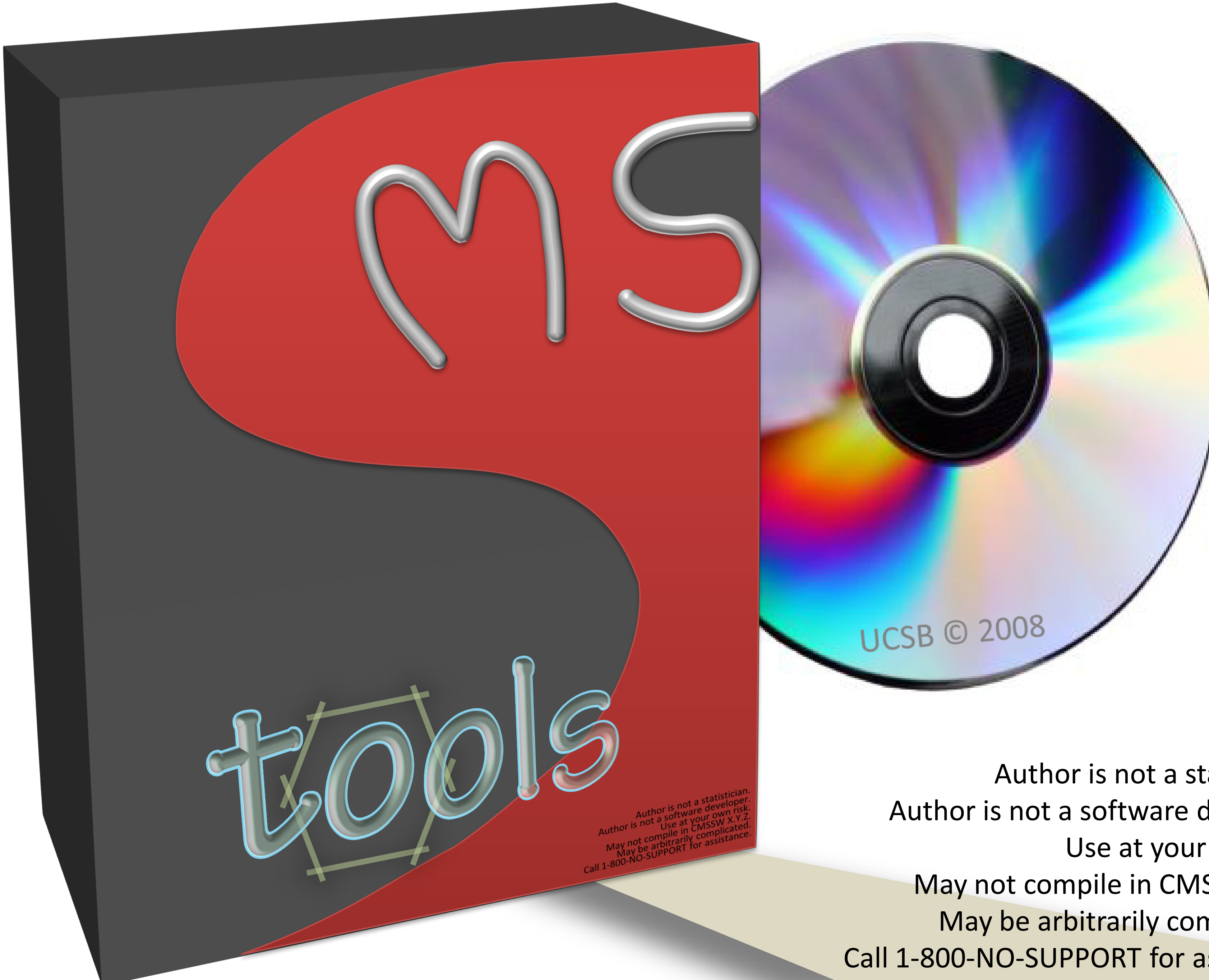
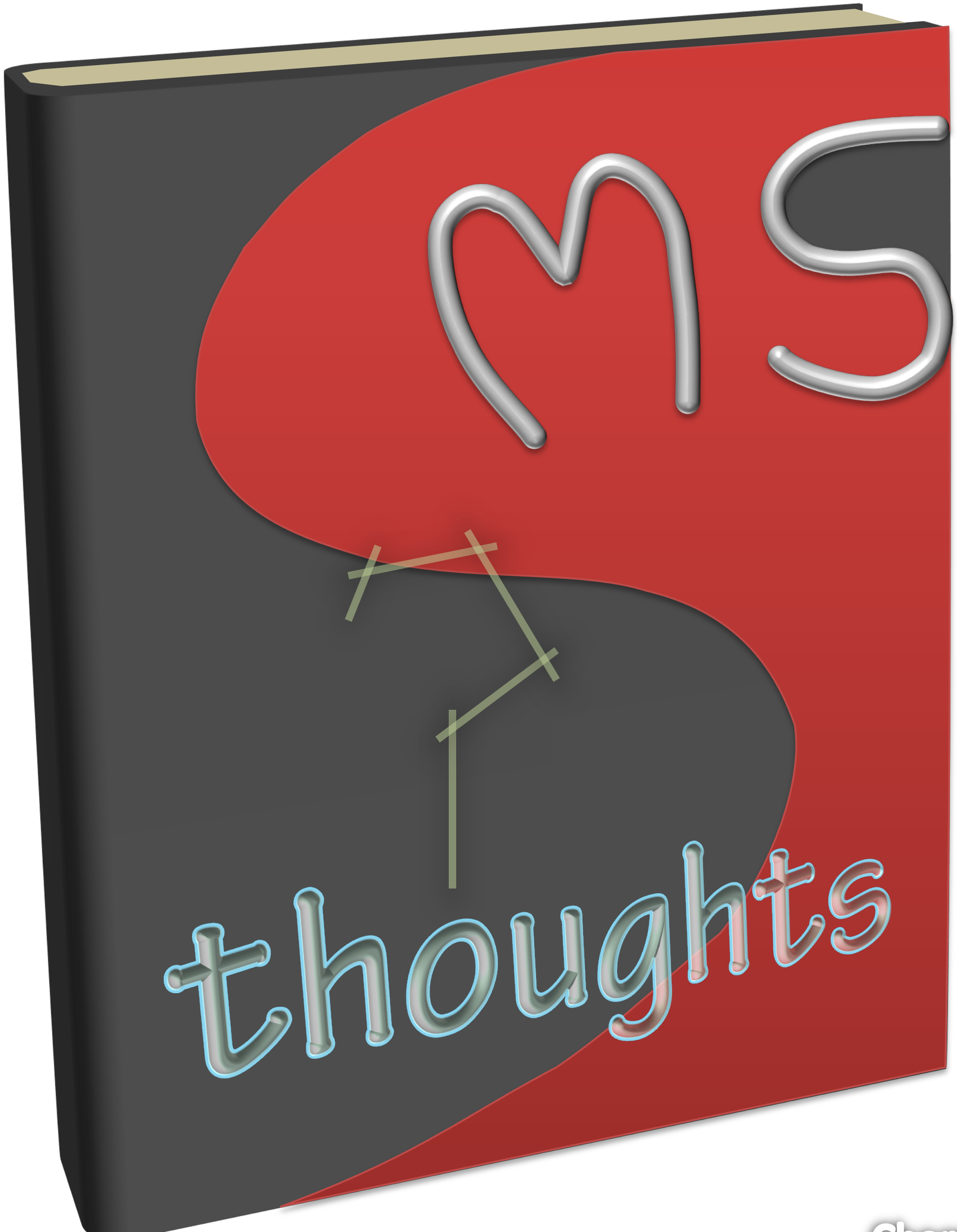


Author is not a statistician.
Author is not a software developer.
Use at your own risk.
May not compile in CMSSW X.Y.Z.
May be arbitrarily complicated.
Call 1-800-NO-SUPPORT for assistance.

```
>  Ctrl + C
>
>  reboot
_
```



J. Incandela
S. A. Koay
R. Rossin

UCSB

P. Schuster
N. Toro

PERIMETER INSTITUTE
FOR THEORETICAL PHYSICS

A Simplified Model Spectra route towards physics intuition and understanding of first observations beyond SM.

– an experimentalist

5th November 2010. CERN
Characterization of new physics at the LHC, II

Table of Contents

1. Life at a hadron collider
2. Signal characterization, bottom-up
3. A hands-on procedure
4. Stages of understanding:
   a) Upper bound on a process
   b) Composing multi-process models
   c) Deducing new particle masses
5. Models: good, bad, better

Table of Contents

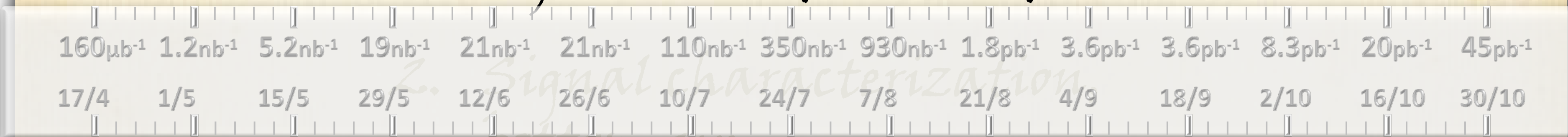
1. Life at a hadron collider

2. Signal characterization bottom-up

3. A hands-on procedure

4. Stages of understanding:
   a) Upper bound on a process
   b) Composing multi-process models
   c) Deducing new particle masses

5. Models: good, bad, better



Volume ∞, Number 0 Physical Review Dreams week ending 31 December 2008

Observation of an Excess above Standard Model Expectations in the Missing Transverse Energy and Jets Channel at $\sqrt{s}$ = ?TeV

- Generic, "catch-all" searches:
  - Cuts and variables optimized for high signal significance; not always high efficiency.
  - Rarely a hunt for mass peaks.
- Interpretation likely to be non-obvious…
- … but we want to do it anyway:
  - To gain confidence that the excess is not some strangely behaving background.
  - To have some idea where to look next.

– a SUSY analyst, 2010

Volume ∞, Number 0 | Physical Review Dreams | week ending 31 December 2008

Observation of an Excess above Standard Model Expectations in the Missing Transverse Energy and Jets Channel at $\sqrt{s}$ = ?TeV

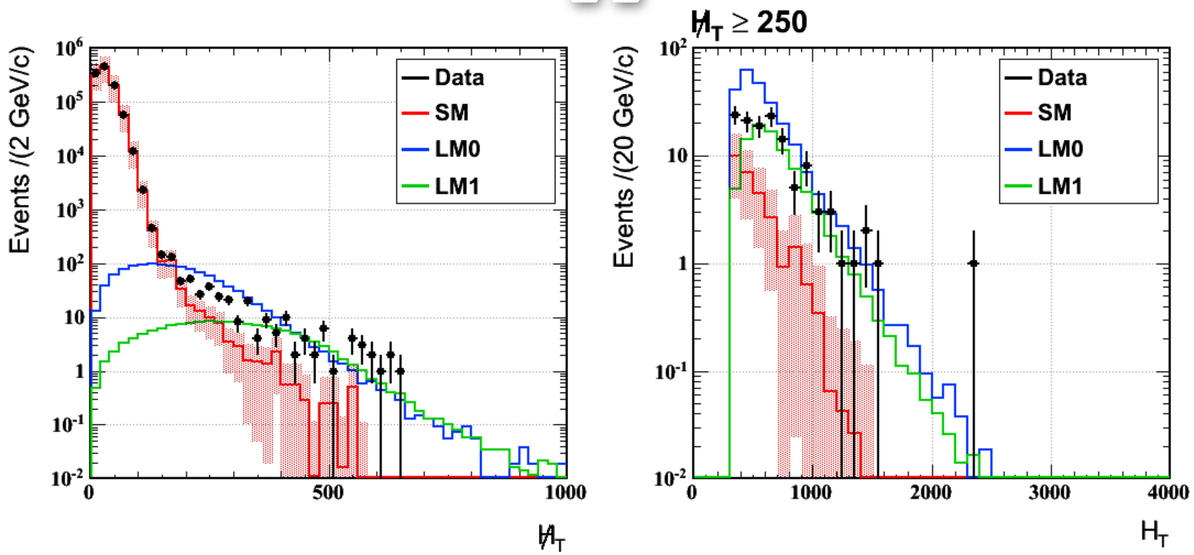


FIG 1. Distribution of $\not{H}_T$ and H_T variables in data, overlaid on the expected Standard Model (SM) background. Two SUSY benchmark points LM0 and LM1 are also shown for comparison.

Example search that publishes two distributions.

Volume ∞, Number 0 Physical Review Dreams week ending 31 December 2008

Observation of an Excess above Standard Model Expectations in the Missing Transverse Energy and Jets Channel at $\sqrt{s}$ = ?TeV

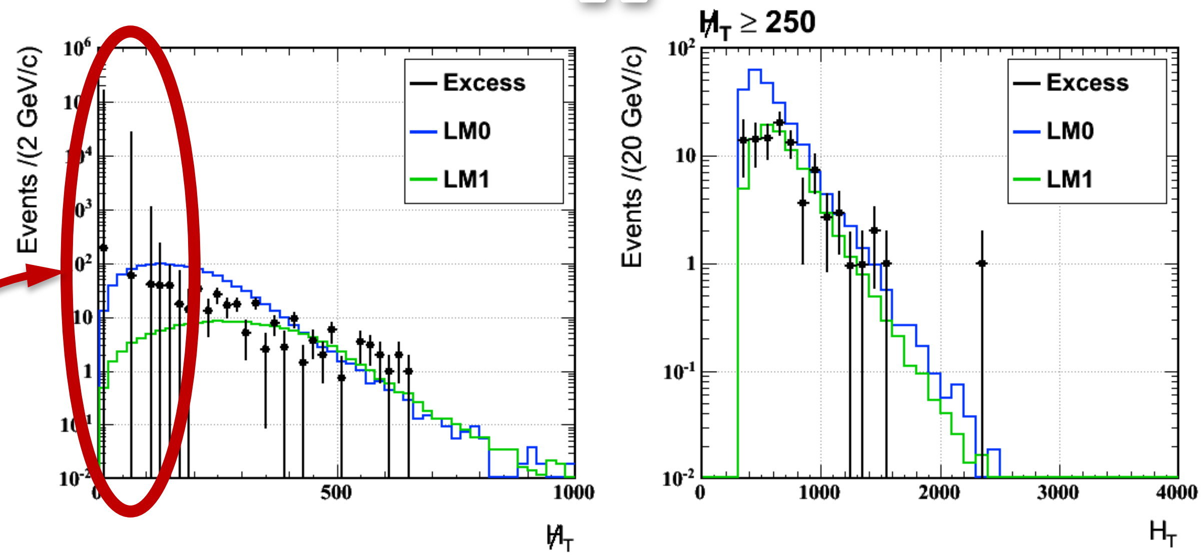


FIG 2. Excess in the $\not{H}_T$ and H_T distributions in data, after subtraction of the expected SM background. Two SUSY benchmark points LM0 and LM1 are also shown for comparison.

Region unusable due to large background subtraction uncertainty

In the following discussion, assume working with background-subtracted excesses.

Volume ∞, Number 0
Physical Review Dreams
week ending
31 December 2008

Observation of an Excess above Standard Model Expectations in the Missing Transverse Energy and Jets Channel at $\sqrt{s}$ = ?TeV

What new particles?

Masses? Charges? Spins?

Production & decay modes?

Identity of WIMP?

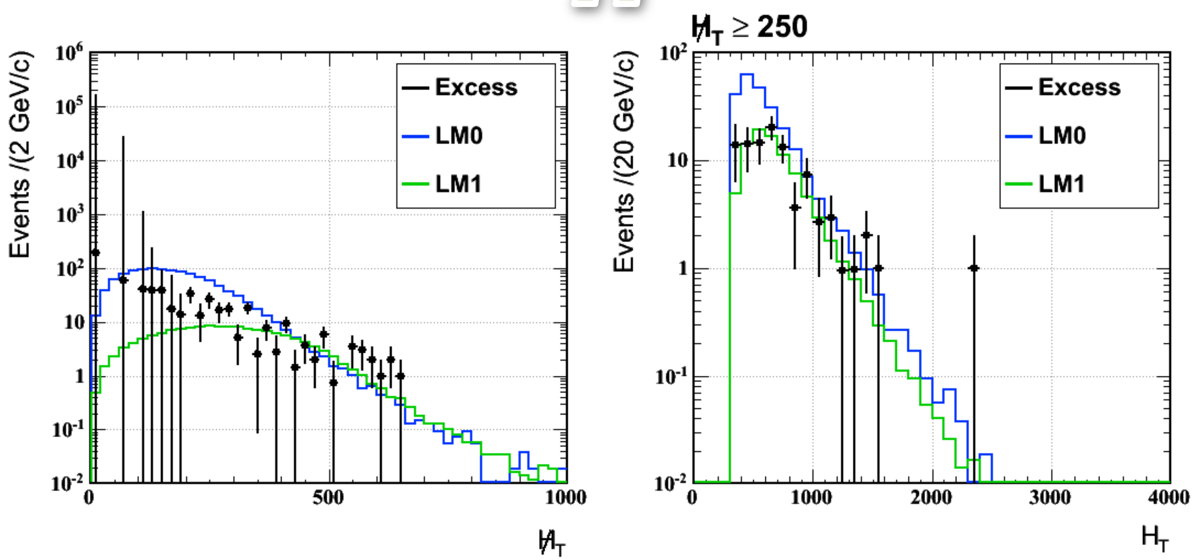


FIG 2. Excess in the $\not{H}_T$ and H_T distributions in data, after subtraction of the expected SM background. Two SUSY benchmark points LM0 and LM1 are also shown for comparison.

BSM group structure?

SUSY? UED? LHT? (your favorite TOE)?

Volume ∞, Number 0 Physical Review Dreams week ending 31 December 2008

Observation of an Excess above Standard Model Expectations in the Missing Transverse Energy and Jets Channel at $\sqrt{s}$ = ?TeV

What new particles?

Masses? Charges? Spins?

Production & decay modes?

Identity of WIMP?

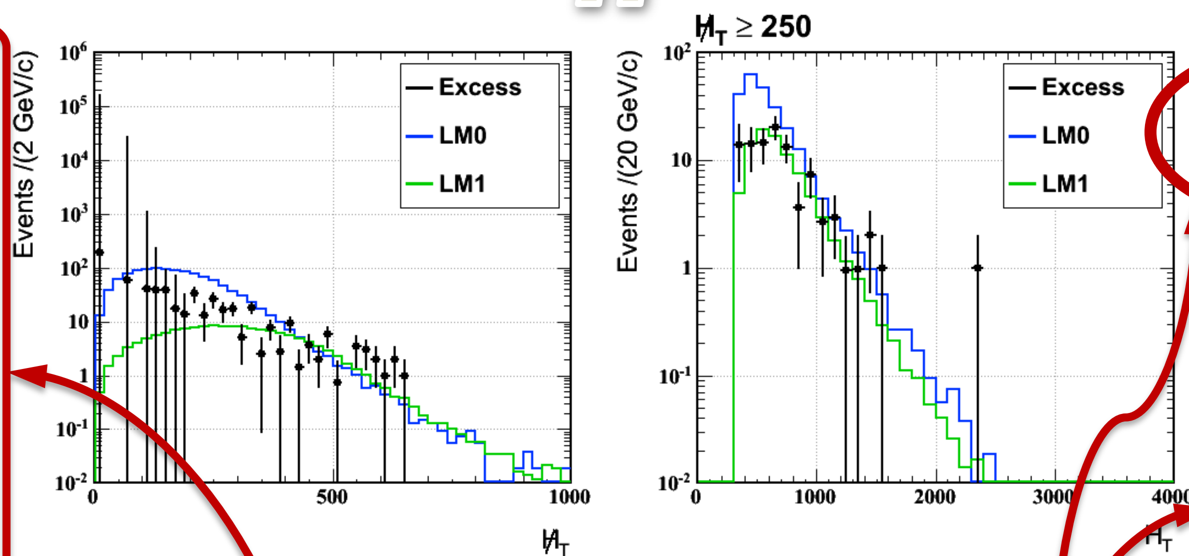


FIG 2. Excess in the $\not{H}_T$ and H_T distributions in data, after subtraction of the expected SM background. Two SUSY benchmark points LM0 and LM1 are also shown for comparison.

BSM group structure?

SUSY? UED? LHT? (your favorite TOE)?

$\mu^+\mu^-$ machine: Direct measurement of many quantities.

pp machine: Mostly *implicit* measurement via sensitive variables…

… but details will be hard to resolve early on.

Holy grail (a.k.a. far, far away)

Volume ∞, Number 0 Physical Review Dreams week ending 31 December 2008

Observation of an Excess above Standard Model Expectations in the Missing Transverse Energy and Jets Channel at $\sqrt{s}$ = ?TeV

"

The dominant backgrounds to this search–QCD multi-jet, top pair, and W boson production–are estimated using data-driven techniques.

- *Variable-specific in many cases (e.g. matrix method), for simplicity, expediency, …*
- *May not be able to predict shapes, only counts above particular set(s) of cuts.*

Volume ∞, Number 0 Physical Review Dreams week ending 31 December 2008

Observation of an Excess above Standard Model Expectations in the Missing Transverse Energy and Jets Channel at $\sqrt{s}$ = ?TeV

"

The dominant backgrounds to this search–QCD multi-jet, top pair, and W boson production–are estimated using data-driven techniques.

- *Variable-specific in many cases (e.g. matrix method), for simplicity, expediency, …*
- *May not be able to predict shapes, only counts above particular set(s) of cuts.*

- *Hard to imagine being able to characterize signal using only a count in a single bin.*
- *The following discussion assumes that a well-established excess of reasonably many events (>>1) is available in the form of one or more distributions.*

- *In general, expect a limited set of variables and regions in which signal can be tested.*

Volume ∞, Number 0 | Physical Review Dreams | week ending 31 December 2008

Observation of an Excess above Standard Model Expectations in the Missing Transverse Energy and Jets Channel at $\sqrt{s}$ = ?TeV

The dominant backgrounds to this search–QCD multi-jet, top pair, and W boson production–are estimated using data-driven techniques.

- *Variable-specific in many cases (e.g. matrix method), for simplicity, expediency, …*
- *May not be able to predict shapes, only counts above particular set(s) of cuts.*

Expect a __limited set of variables and regions__ in which signal can be tested.

Comparisons of the data-driven predictions to Monte Carlo simulated expectations are shown in Figs. 3 to 5.

“Let’s just use MC backgrounds, and take a very large uncertainty.”

- *The proper thing to do still requires validation of MC vs. data in unexplored $\sqrt{s}$ territory.*
- *Otherwise, what is a “very large” theoretical uncertainty? Scale variations? NLO? NNLO?*

And this is assuming we’ve pinned down all aspects of detector modeling!

Table of Contents

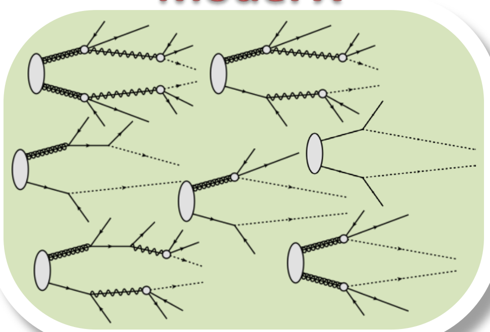
1. Life at a hadron collider
2. Signal characterization, bottom-up

3. A hands-on procedure
4. Stages of understanding:
   a) Upper bound on a process
   b) Composing multi-process models
   c) Deducing new particle masses
5. Models: good, bad, better

Signal characterization, top down

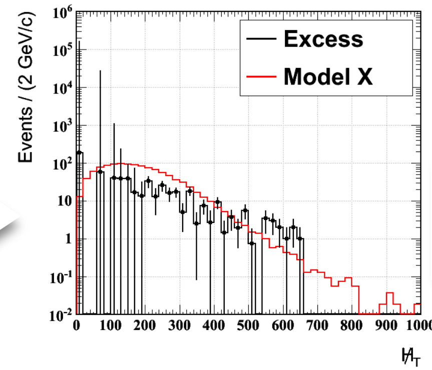
“Signal” ≡ background-subtracted data

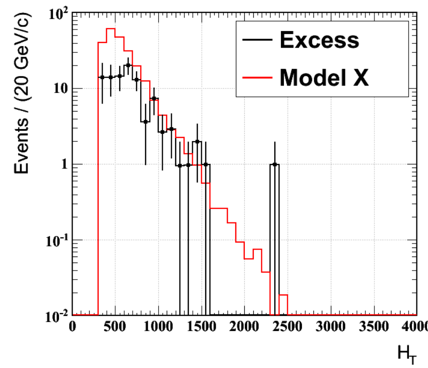
Model X



Defined by:

- Multiple processes implied by particle spectrum and decay channels.
- Cross section from matrix element calculation.





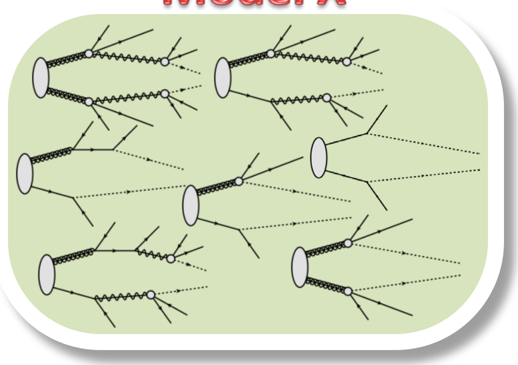
Compare distributions with a histogram compatibility test (χ^2, log-likelihood, …); or more sophisticated statistical techniques.

“Distance” ≡ degree of compatibility (χ^2 probability, …)

Signal characterization, top down

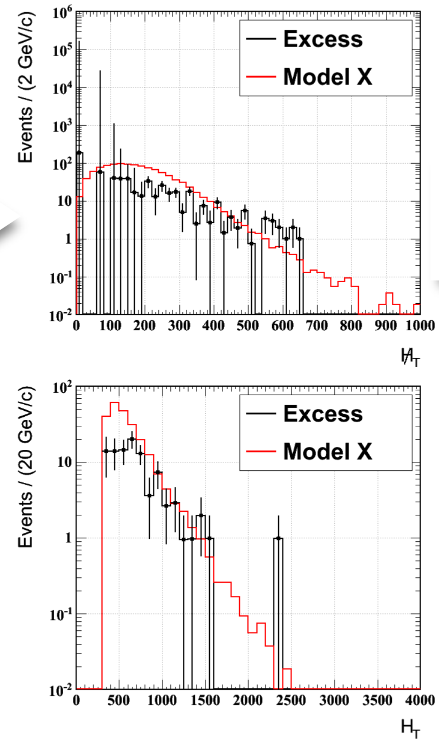
"Signal" ≡ background-subtracted data

Model X



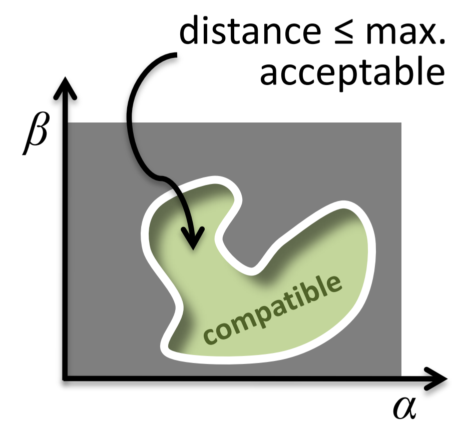
Defined by:

- Multiple processes implied by particle spectrum and decay channels.
- Cross section from matrix element calculation.



Compare distributions with a histogram compatibility test (χ^2, log-likelihood, …); or more sophisticated statistical techniques.

"Distance"
≡
degree of compatibility
(χ^2 probability, …)



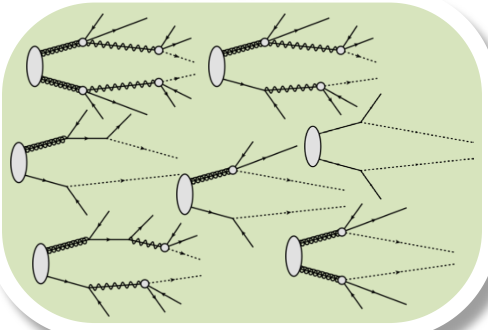
(theory with 2 parameters)

Typically a point in parameter space of some larger theory

Signal characterization, bottom up

“Signal” ≡ background-subtracted data

Model X



Defined by:

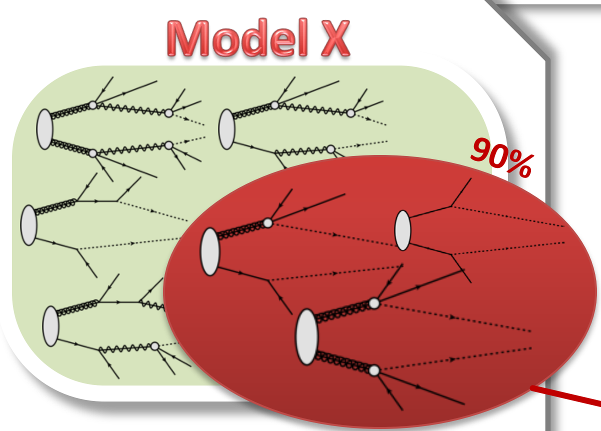
- Multiple processes implied by particle spectrum and decay channels.
- Cross section from matrix element calculation.

- The bottom up approach differs in how one selects/evolves the models to test.
- Statistics treatment of whether a model is feasible (or not) remains the same.

Rather than checking all models in an O(100)-dimensional TOE parameter space, the idea is to *progressively use information from data* to narrow down possibilities:

Signal characterization, bottom up

"Signal" ≡ background-subtracted data



Defined by:
- Multiple processes implied by particle spectrum and decay channels.
- Cross section from matrix element calculation.

- The bottom up approach differs in how one selects/evolves the models to test.
- Statistics treatment of whether a model is feasible (or not) remains the same.

Rather than checking all models in an O(100)-dimensional TOE parameter space, the idea is to *progressively use information from data* to narrow down possibilities:

- *Condense indistinguishable processes into representative classes.*
- *Regions can differ only by change in fractions of each topology.*

⇒ *Fit for fractions in data.*

- *Typically the most constraining quantity.*
- *Depends strongly on quantities with little observable impact–spins of produced particles, species and masses of internal propagators.*

⇒ *Extract directly from data.*

Table of Contents

1. Life at a hadron collider

2. Signal characterization, bottom-up

3. A hands-on procedure

$160\mu b^{-1}$	$1.2nb^{-1}$	$5.2nb^{-1}$	$19nb^{-1}$	$21nb^{-1}$	$21nb^{-1}$	$110nb^{-1}$	$350nb^{-1}$	$930nb^{-1}$	$1.8pb^{-1}$	$3.6pb^{-1}$	$3.6pb^{-1}$	$8.3pb^{-1}$	$20pb^{-1}$	$45pb^{-1}$
17/4	1/5	15/5	29/5	12/6	26/6	10/7	24/7	7/8	21/8	4/9	18/9	2/10	16/10	30/10

4. Stages of understanding:
   a) Upper bound on a process
   b) Composing multi-process models
   c) Deducing new particle masses

5. Models: good, bad, better

But hasn't this all already been said and done?

Bard: Interpreting New Frontier Energy Collider Physics
(arXiv:hep-ph/0602101v1)

MUSiC – An Automated Scan for Deviations between Data and Monte Carlo Simulation
(http://cms-physics.web.cern.ch/cms-physics/public/EXO-08-005-pas.pdf)

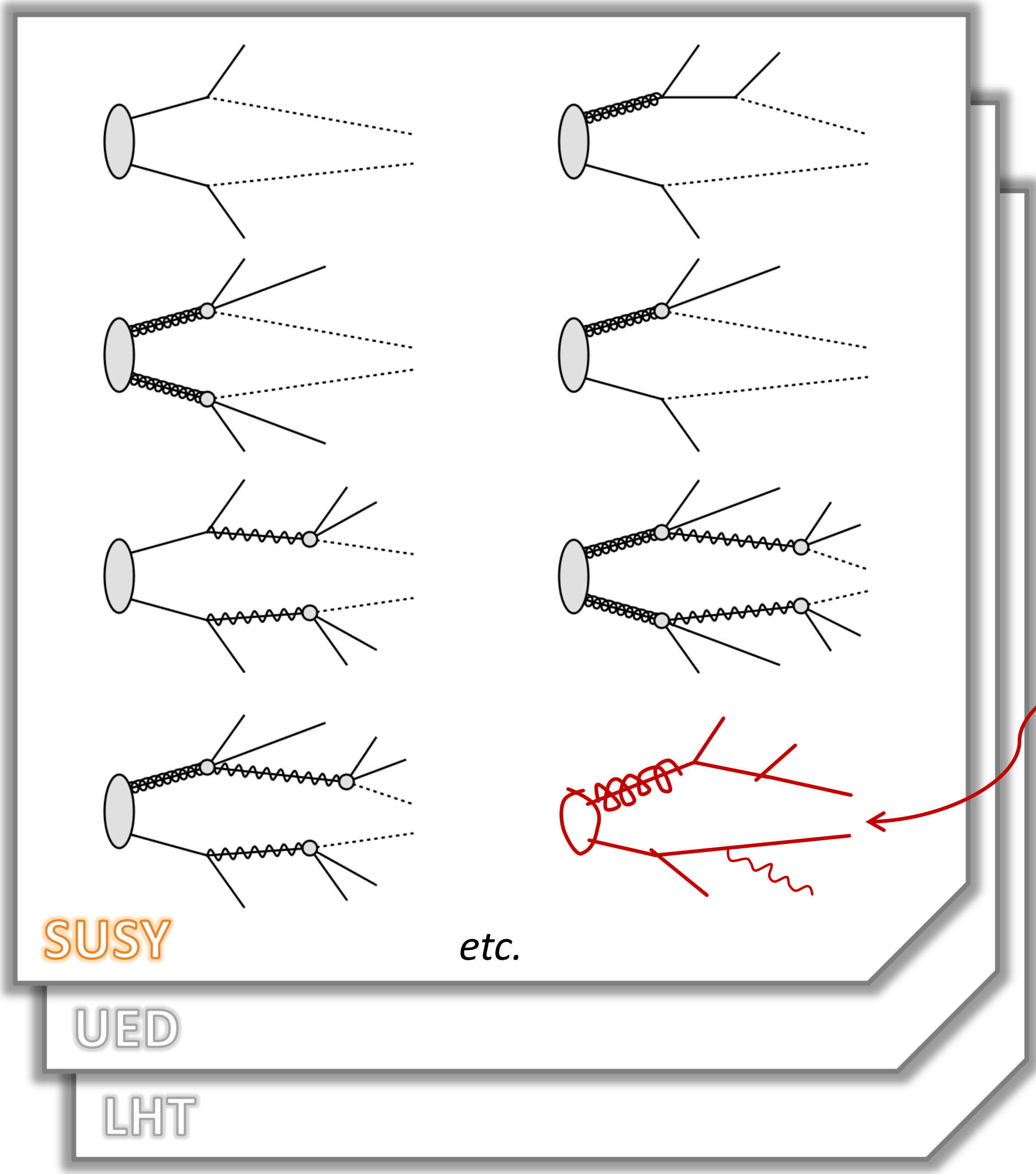
MARMOSET: The Path from LHC Data to the New Standard Model via On-Shell Effective Theories
(arXiv:hep-ph/0703088v1)

The writing and people who initiated this work.

The strategy shown here follows in this spirit; think of the choice as a matter of taste (or ignorance).

An “informed search” through BSM space

SKETCH



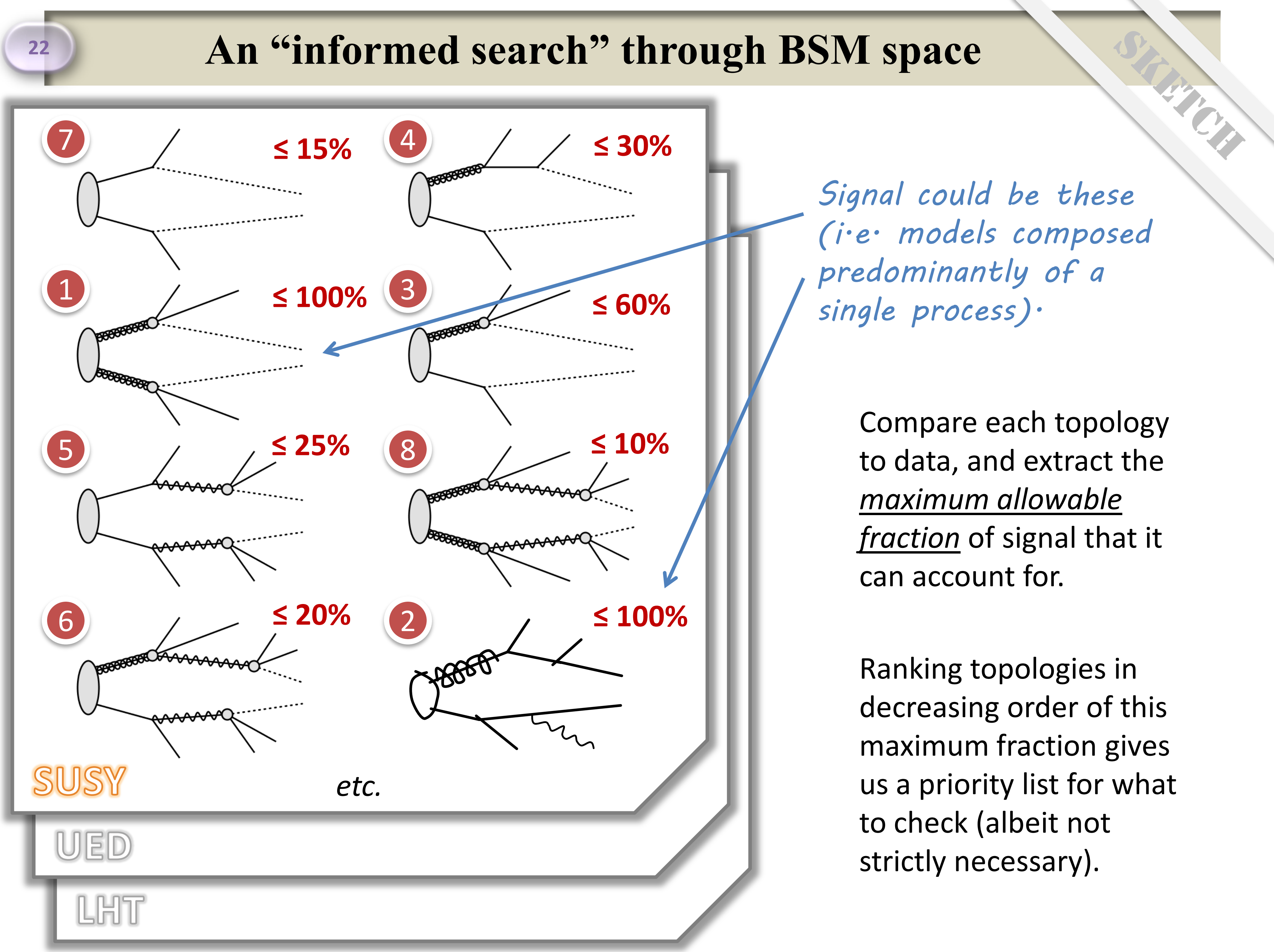
Start from a “comprehensive” list of SMS topologies.

Most of these will come from the sets proposed a priori by theorists.

Some may be inspired during the course of investigation.

We focus on a SUSY-inspired example; the treatment of other theories is analogous. Some topologies can anyway be interpreted in the language of several theories.

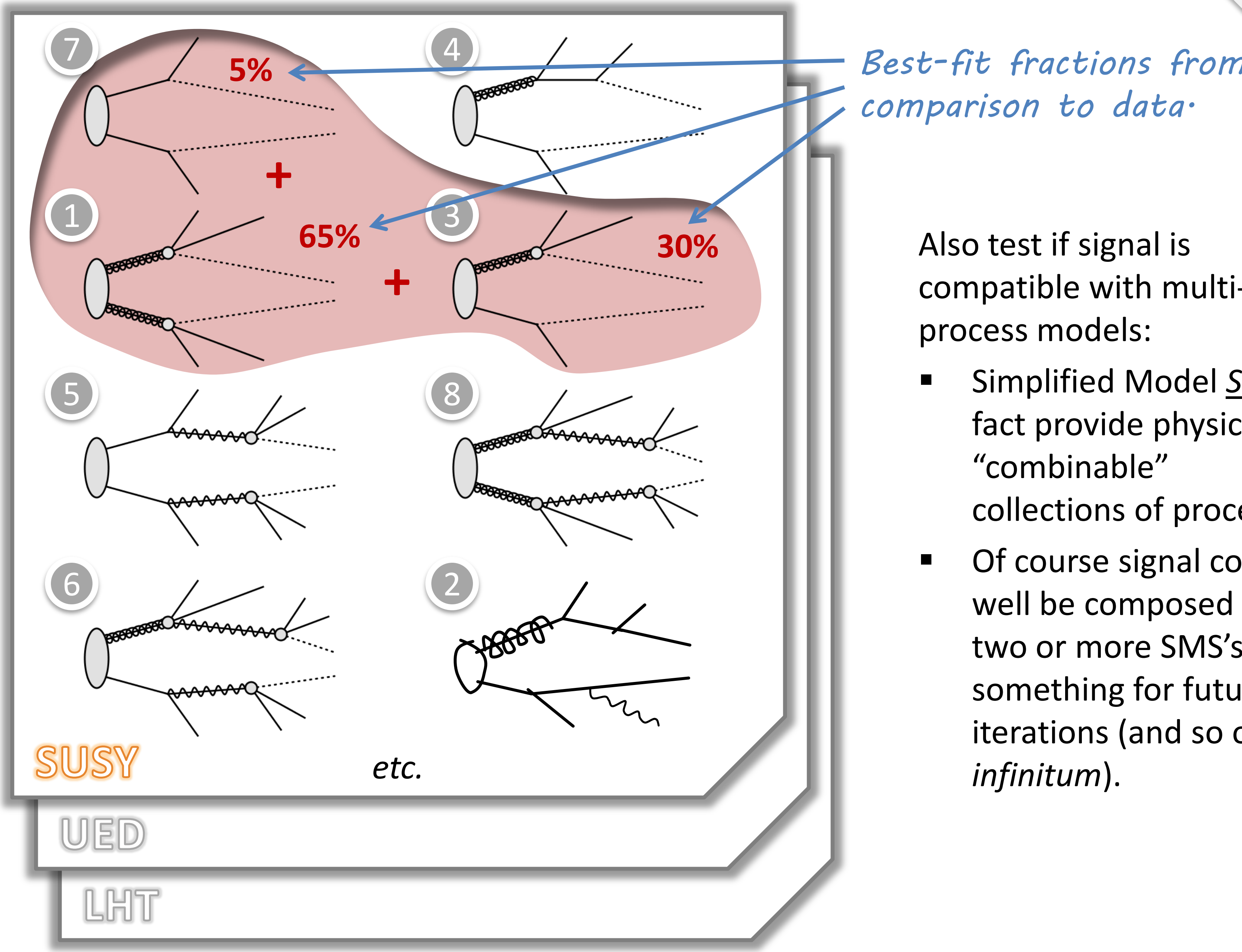
An "informed search" through BSM space



Compare each topology to data, and extract the *maximum allowable fraction* of signal that it can account for.

Ranking topologies in decreasing order of this maximum fraction gives us a priority list for what to check (albeit not strictly necessary).

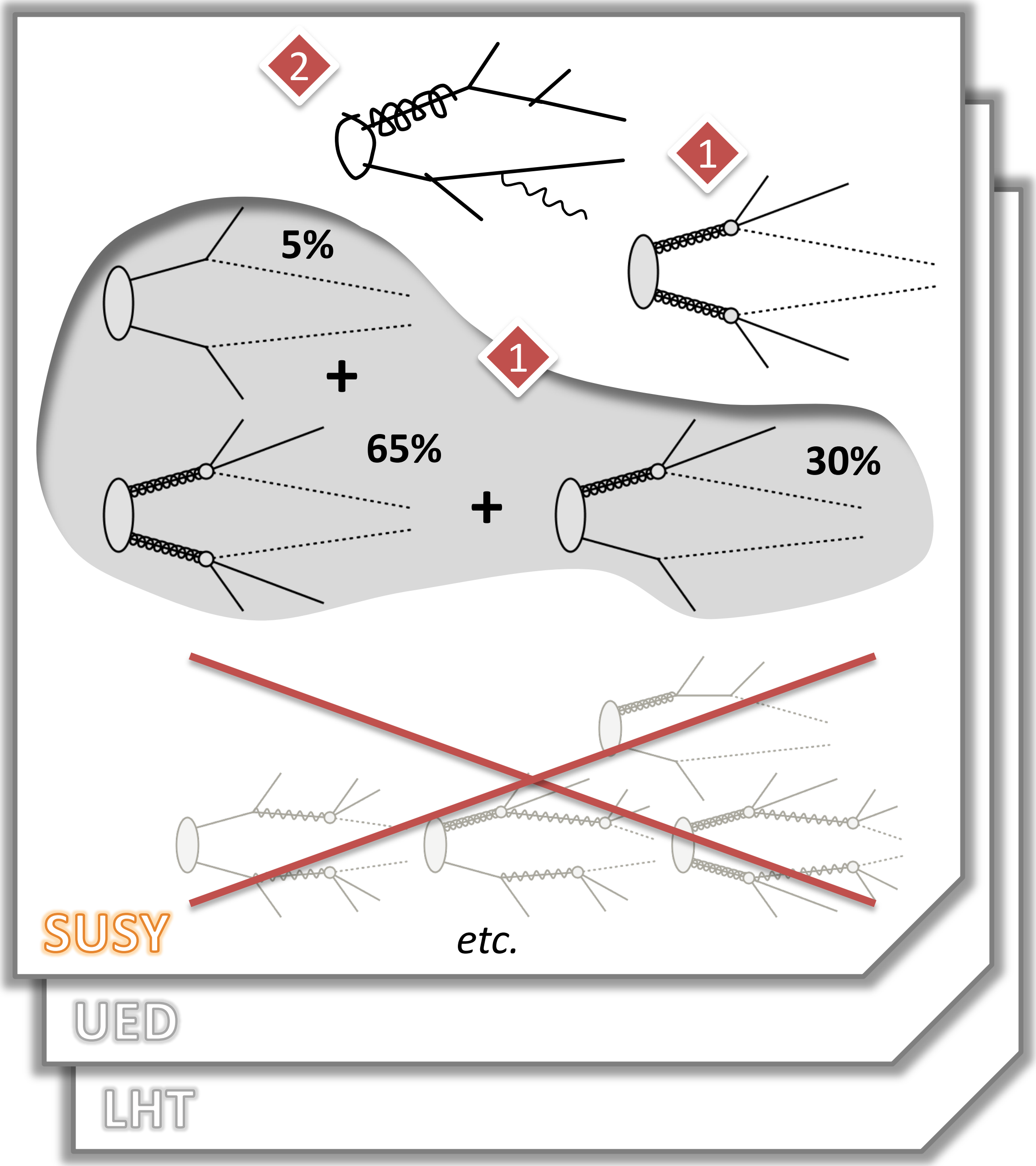
An "informed search" through BSM space



Also test if signal is compatible with multi-process models:

- Simplified Model *Sets* in fact provide physically "combinable" collections of processes.
- Of course signal could as well be composed of two or more SMS's – something for future iterations (and so on *ad infinitum*).

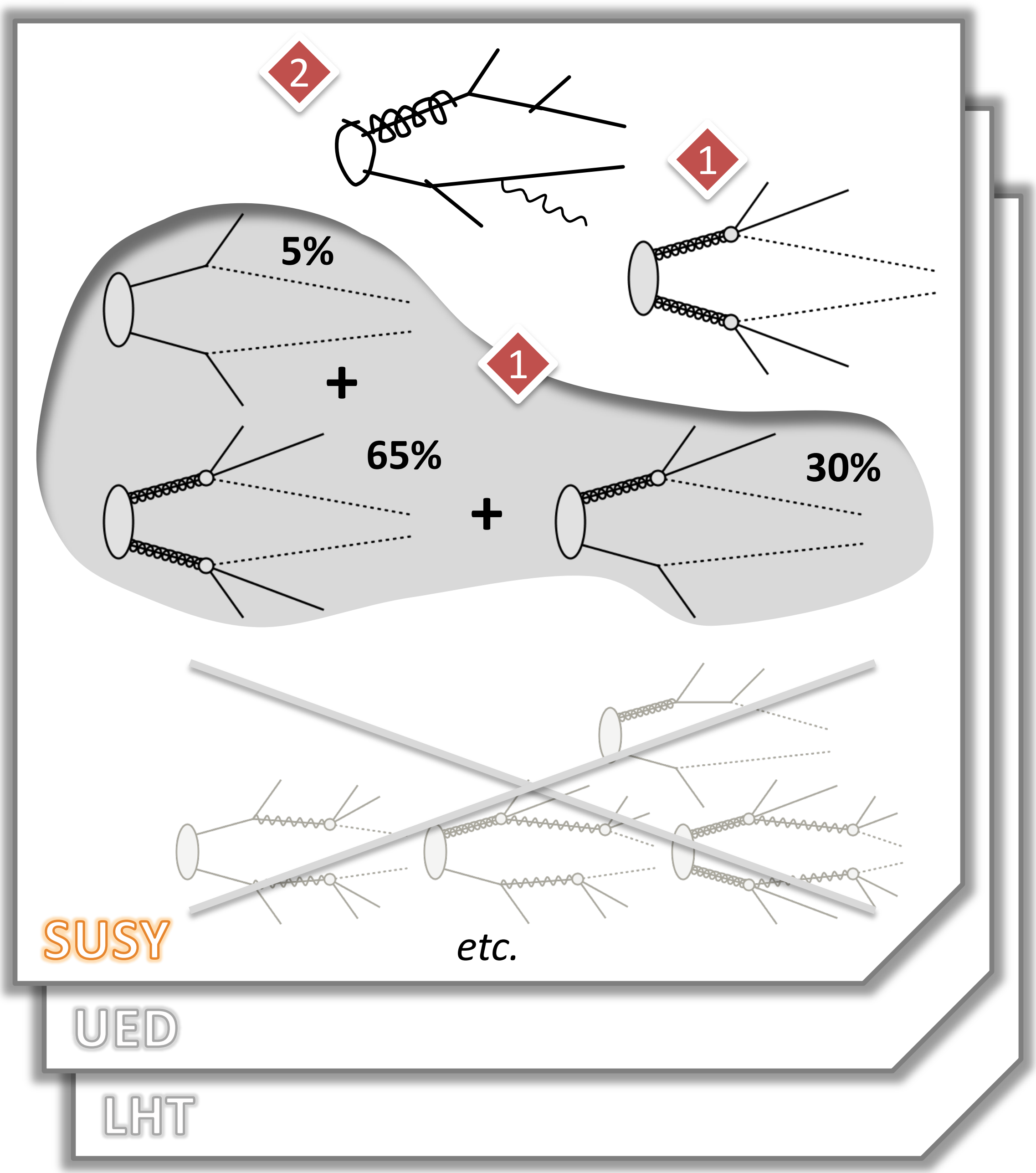
An "informed search" through BSM space



Perform a "final" ranking of all the tested models:

- Models that are not compatible with signal are eliminated.
- Remaining viable models are ordered by decreasing goodness-of-fit to signal.

An "informed search" through BSM space



Why bother to rank? Because then an experimentalist could pick a top few off this list to start looking.

Perform a "final" ranking of all the tested models:

- Models that are not compatible with signal are eliminated.
- Remaining viable models are ordered by decreasing goodness-of-fit to signal.

Table of Contents

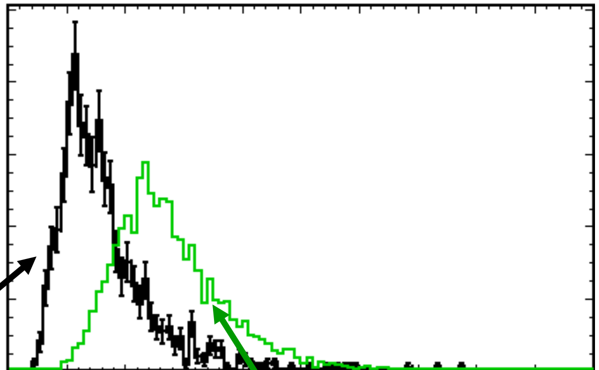
1. Life at a hadron collider
2. Signal characterization, bottom-up
3. A hands-on procedure
4. Stages of understanding:
   a) Upper bound on a process
   b) Comparing results to process models
   c) Deducing new particle masses
5. Models: good, bad, better

160μb^{-1}	1.2nb^{-1}	5.2nb^{-1}	19nb^{-1}	21nb^{-1}	21nb^{-1}	110nb^{-1}	350nb^{-1}	930nb^{-1}	1.8pb^{-1}	3.6pb^{-1}	3.6pb^{-1}	8.3pb^{-1}	20pb^{-1}	45pb^{-1}
17/4	1/5	15/5	29/5	12/6	26/6	10/7	24/7	7/8	21/8	4/9	18/9	2/10	16/10	30/10

Some terminology

Template = 1.0 × Signal

Template scaled to number of signal events

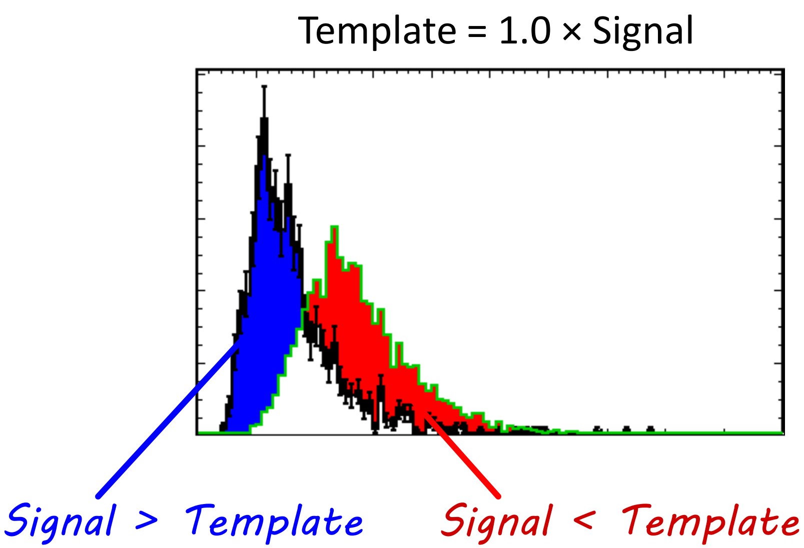


Signal (≡ background subtracted data)

Template (from a particular SMS model hypothesis)

- Our staple tool is the simple histogram comparison.
  - Assume χ^2 test, to be concrete.

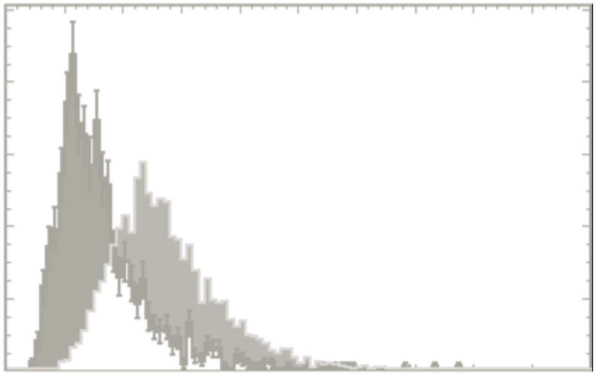
Upper bound on a process X



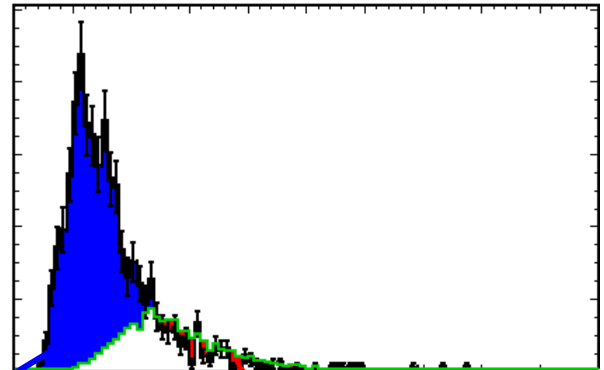
- For this SMS process, a shape comparison of this distribution indicates that it is not compatible with the observed signal.
- However the above does ***not*** mean that process X cannot be present:
- It only means that it cannot be the ***only*** source of the signal.

Upper bound on process X

Template = 1.0 × Signal



Template = 0.3 × Signal



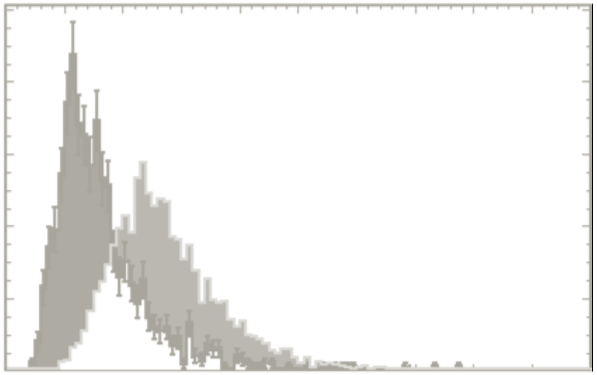
This part could be due to other process(es)

The maximum fractional contribution of process X is what can manage to fit under the entire distribution…

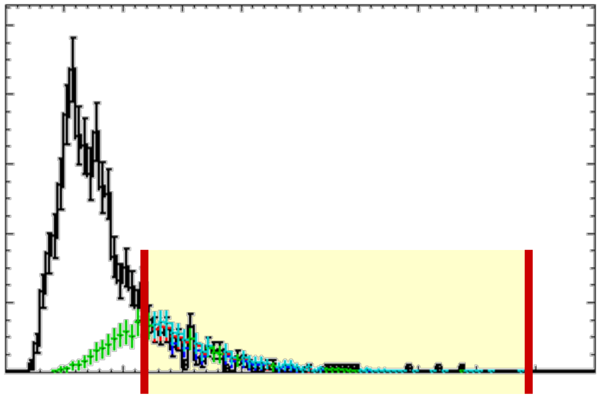
… with a ~2σ excess allowed in case there was a downwards Poisson fluctuation.

Upper bound on process X

Template = 1.0 × Signal



Template = 0.3 × Signal

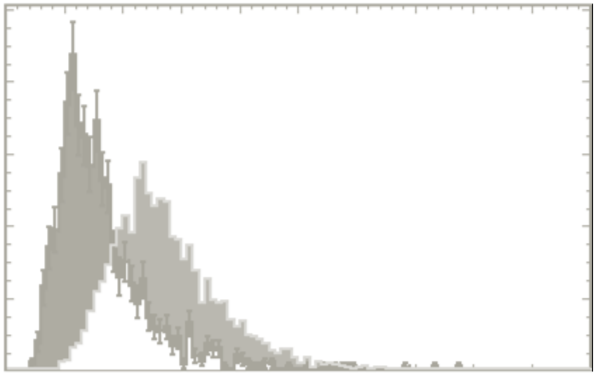


Procedure

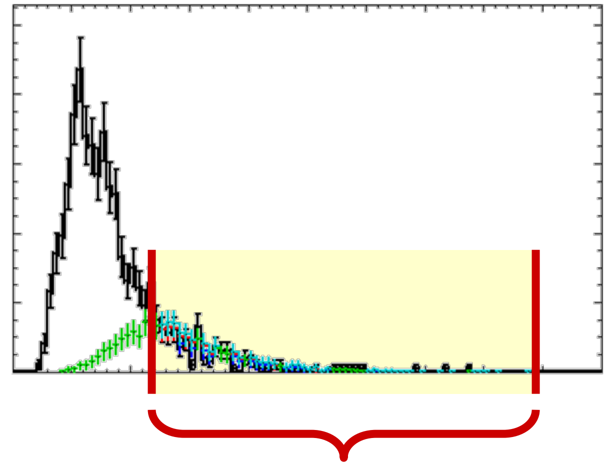
Compute a "constrained distance" using only those bins where template > signal

Upper bound on process X

Template = 1.0 × Signal



Template = 0.3 × Signal



Procedure

Compute a “constrained distance” using only those bins where template > signal

- This upper bound must be checked as a function of the new particle masses.
- Provides a crude first pass to see which processes can be the dominant source of signal, versus those that cannot be present except in conjunction with some other process(es).

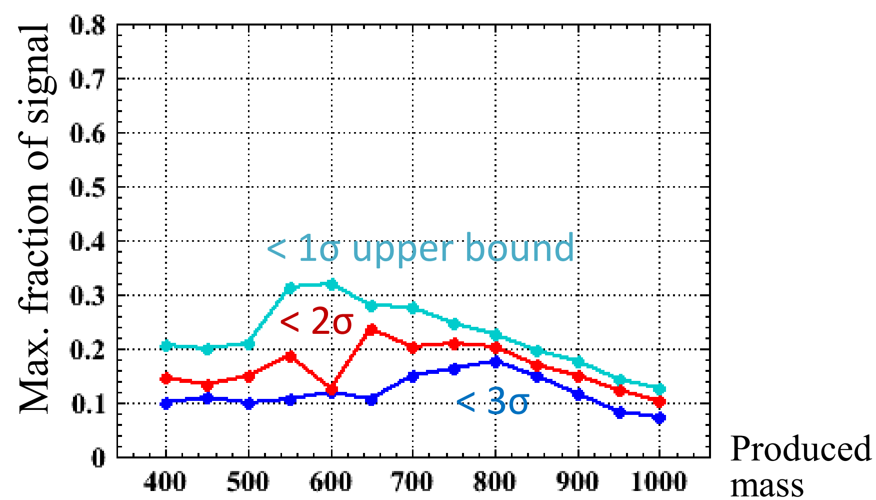
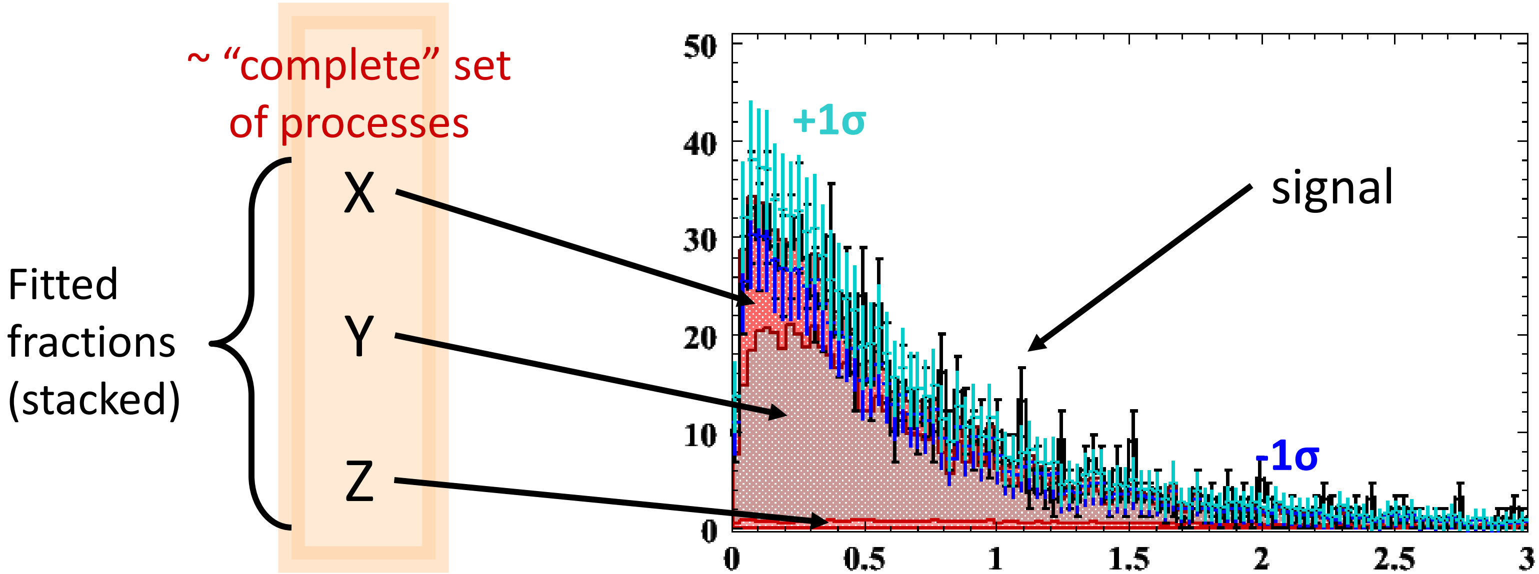


Table of Contents

1. Life at a hadron collider
2. Signal characterization, bottom-up
3. A hands-on procedure
4. Stages of understanding:
   a) Upper bound on a process
   b) Composing multi-process models
   c) Deducing new particle masses
5. Models: good, bad, better

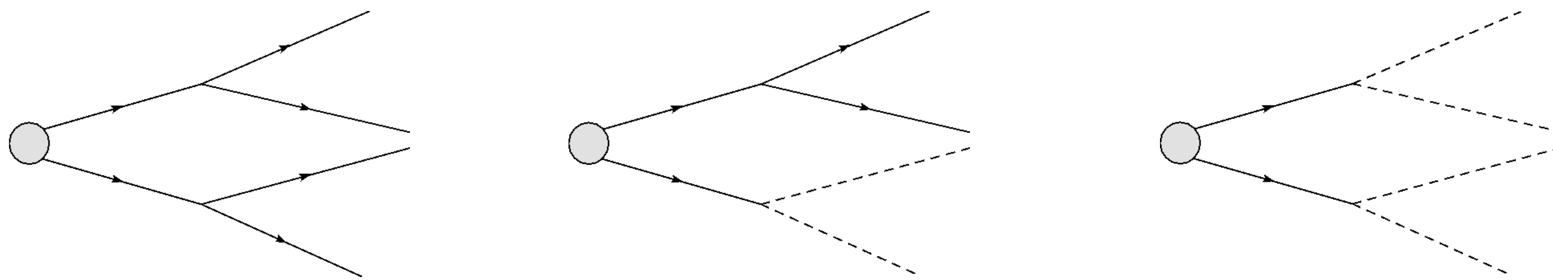
160μb^{-1}	1.2nb^{-1}	5.2nb^{-1}	19nb^{-1}	21nb^{-1}	21nb^{-1}	110nb^{-1}	350nb^{-1}	930nb^{-1}	1.8pb^{-1}	3.6pb^{-1}	3.6pb^{-1}	8.3pb^{-1}	20pb^{-1}	45pb^{-1}
17/4	1/5	15/5	29/5	12/6	26/6	10/7	24/7	7/8	21/8	4/9	18/9	2/10	16/10	30/10

Fitting for fractions of processes in a SMS



- To compose a viable BSM model, the processes must sum up to be compatible with the entire signal in all distributions.
- Standard template-fitting techniques (e.g. **`TFractionFitter`** @ ROOT) can be used to extract the best composition of processes to explain signal.
  - These methods typically implicitly or explicitly constrain the total yield (≡ crosssection × search cut efficiency) of the model to be equal to the number of observed signal events.
  - The cross-section deduced in this way can provide clues towards interpreting the SMS in terms of full theories (SUSY, UED, LTH, …).

Fitting for fractions of processes in a SMS



~BR( ) ~BR( ) BR( ) ~[BR( )]^2

- The main subtlety here is if the fractions of the processes are constrained by some model relationship, e.g. a single BSM particle with exactly two decay modes and .
- Unitarity requires

  BR( ) = 1 − BR( )

  which is a constraint that must be used to eliminate one degree of freedom in the otherwise 3-template fraction fit problem illustrated here.

Table of Contents

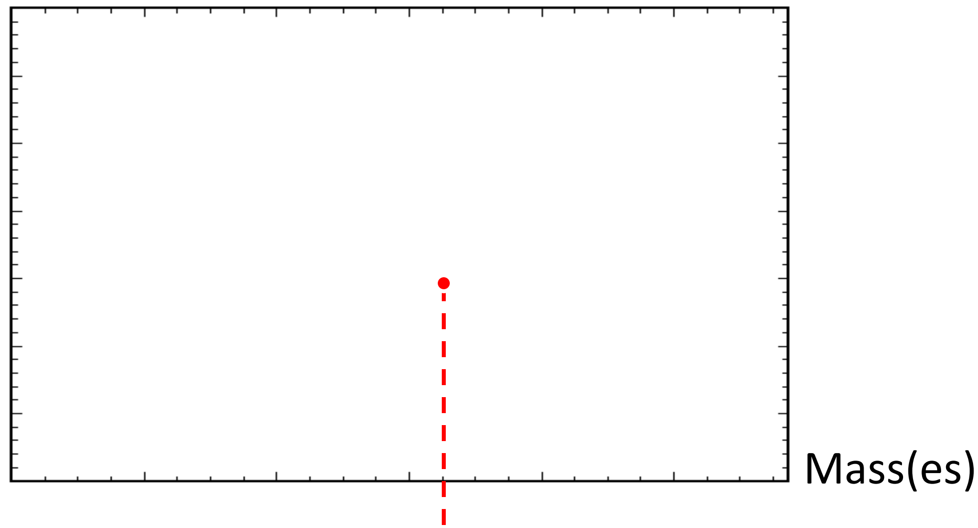
1. Life at a hadron collider
2. Signal characterization, bottom-up
3. A hands-on procedure
4. Stages of understanding:
   a) Upper bound on a process
   b) Composing multi-process models
   c) Deducing new particle masses

$160\mu b^{-1}$	$1.2 nb^{-1}$	$5.2 nb^{-1}$	$19 nb^{-1}$	$21 nb^{-1}$	$21 nb^{-1}$	$110 nb^{-1}$	$350 nb^{-1}$	$930 nb^{-1}$	$1.8 pb^{-1}$	$3.6 pb^{-1}$	$3.6 pb^{-1}$	$8.3 pb^{-1}$	$20 pb^{-1}$	$45 pb^{-1}$
17/4	1/5	15/5	29/5	12/6	26/6	10/7	24/7	7/8	21/8	4/9	18/9	2/10	16/10	30/10

5. Models: good, bad, better

Best-fit masses for each SMS model

So far:



mass(1) = 80 GeV
mass(2) = 100 GeV
…

At each set of masses that we generate MC for:

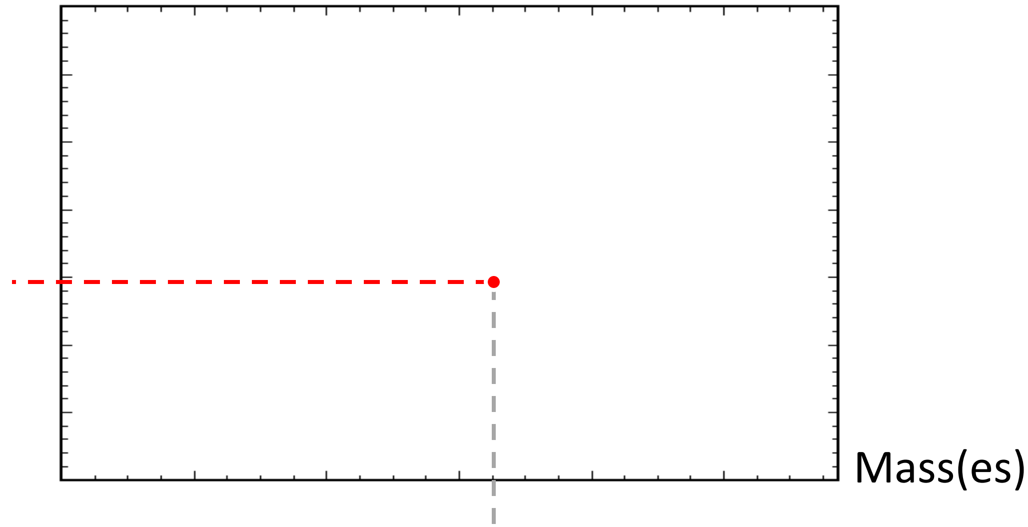
Perform a fit for other parameters (σ, BR)

(a cartoon)

Best-fit masses for each SMS model

So far:

“Height” of mass point
≡ Goodness of best fit
(χ^2 probability, …)



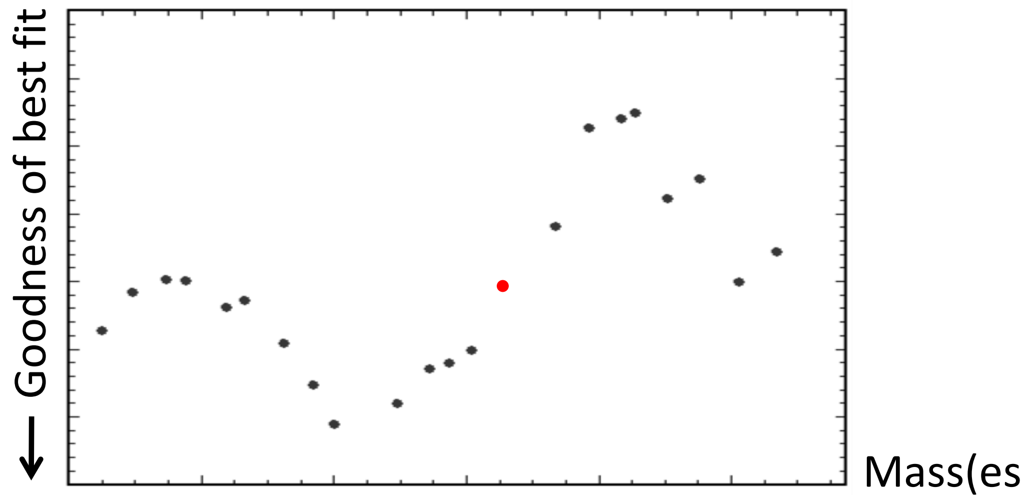
At each set of masses that we generate MC for:

mass(1) = 80 GeV
mass(2) = 100 GeV
…

Perform a fit for other parameters (σ, BR)

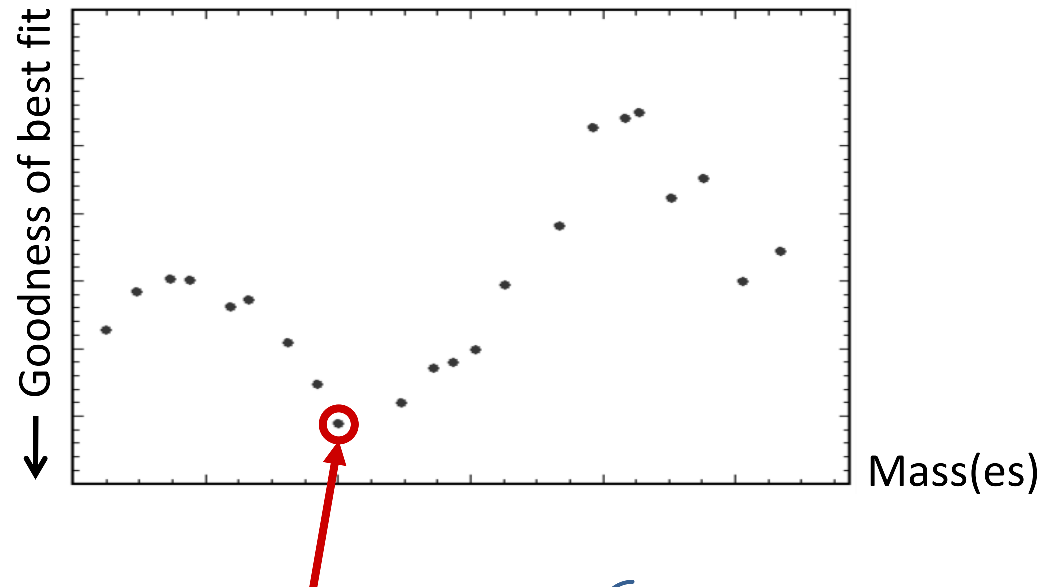
(a cartoon)

Best-fit masses for each SMS model



- Mass fitting is technically different (and more challenging) due to the computational burden of generating MC events for each mass hypothesis.
- One is likely restricted to a finite, discrete grid of mass points, and cannot directly use a minimization package to locate the "true" minimum.

Best-fit masses for each SMS model



Procedure 1

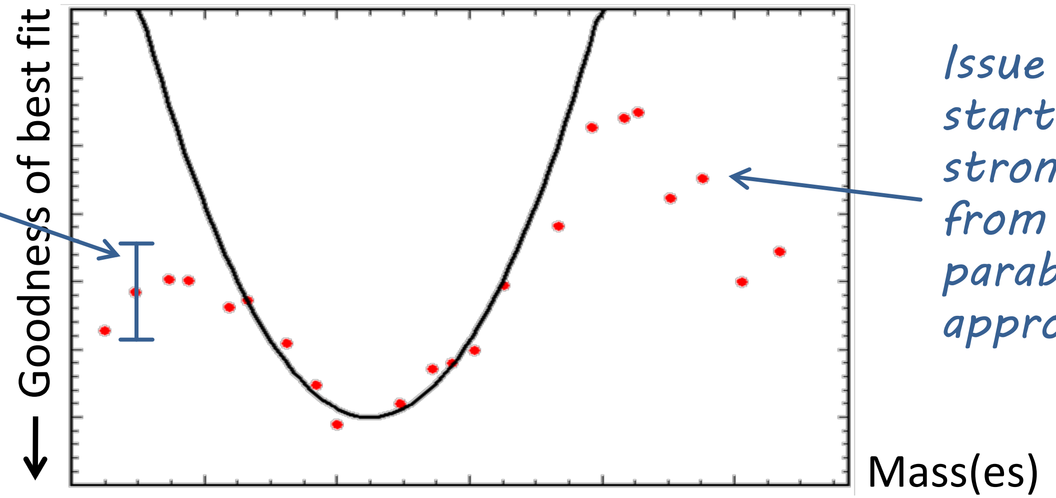
Pick the best generated point.

Not bad. But can do better by using information from neighboring points.

- Mass fitting is technically different (and more challenging) due to the computational burden of generating MC events for each mass hypothesis.
- One is likely restricted to a finite, discrete grid of mass points, and cannot directly use a minimization package to locate the "true" minimum.

Best-fit masses for each SMS model

Issue #1: What "error bars" can we assign to a goodness of fit?



Issue #2: Points start to diverge strongly away from the parabolic approximation.

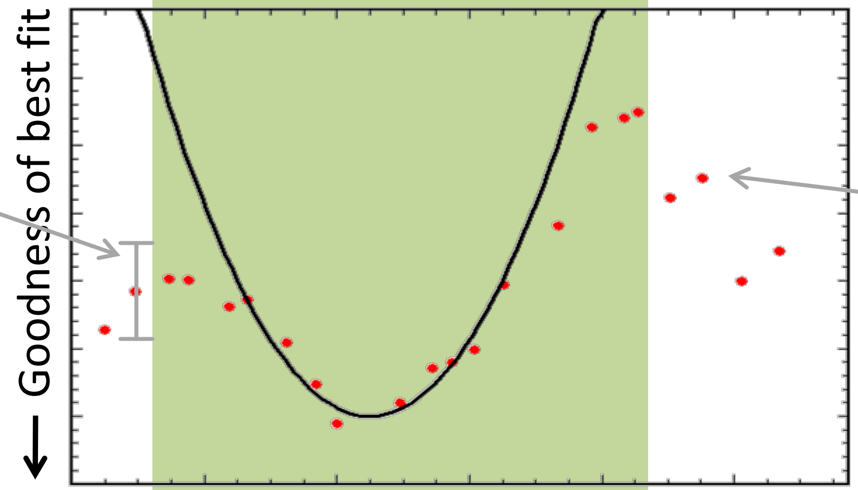
Procedure 2

Fit for parabolic expansion around minimum.

Issue #3: Can have multiple minima.

Best-fit masses for each SMS model

Issue #1: What "error bars" can we assign to a goodness of fit?



Issue #2: Points start to diverge strongly away from the parabolic approximation.

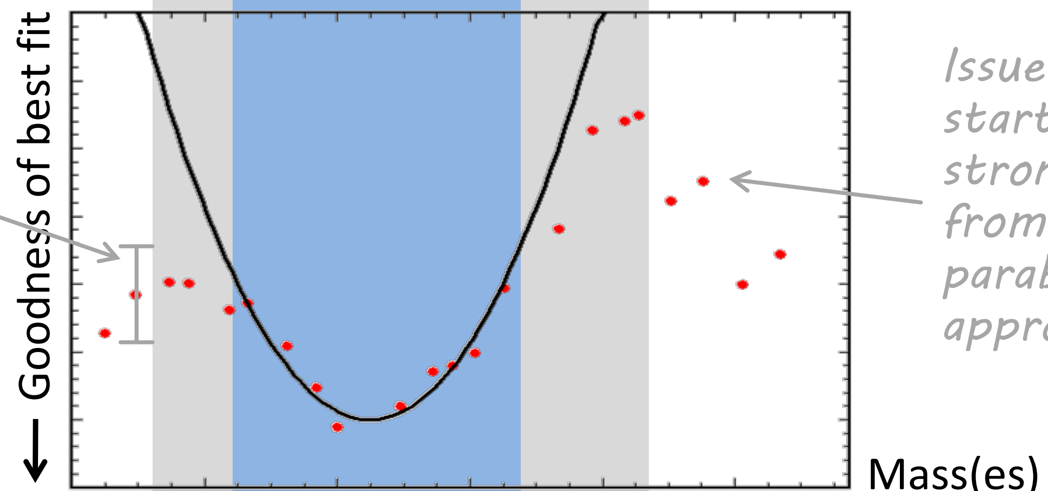
Procedure 2 Fit for parabolic expansion around minimum.

Issue #3: Can have multiple minima.

An "obvious" solution (at least for a 1D problem):

- Group points into regions containing exactly one minimum.
-

Best-fit masses for each SMS model



Issue #1: What "error bars" can we assign to a goodness of fit?

Issue #2: Points start to diverge strongly away from the parabolic approximation.

Procedure 2

Fit for parabolic expansion around minimum.

Issue #3: Can have multiple minima.

An "obvious" solution (at least for a 1D problem):

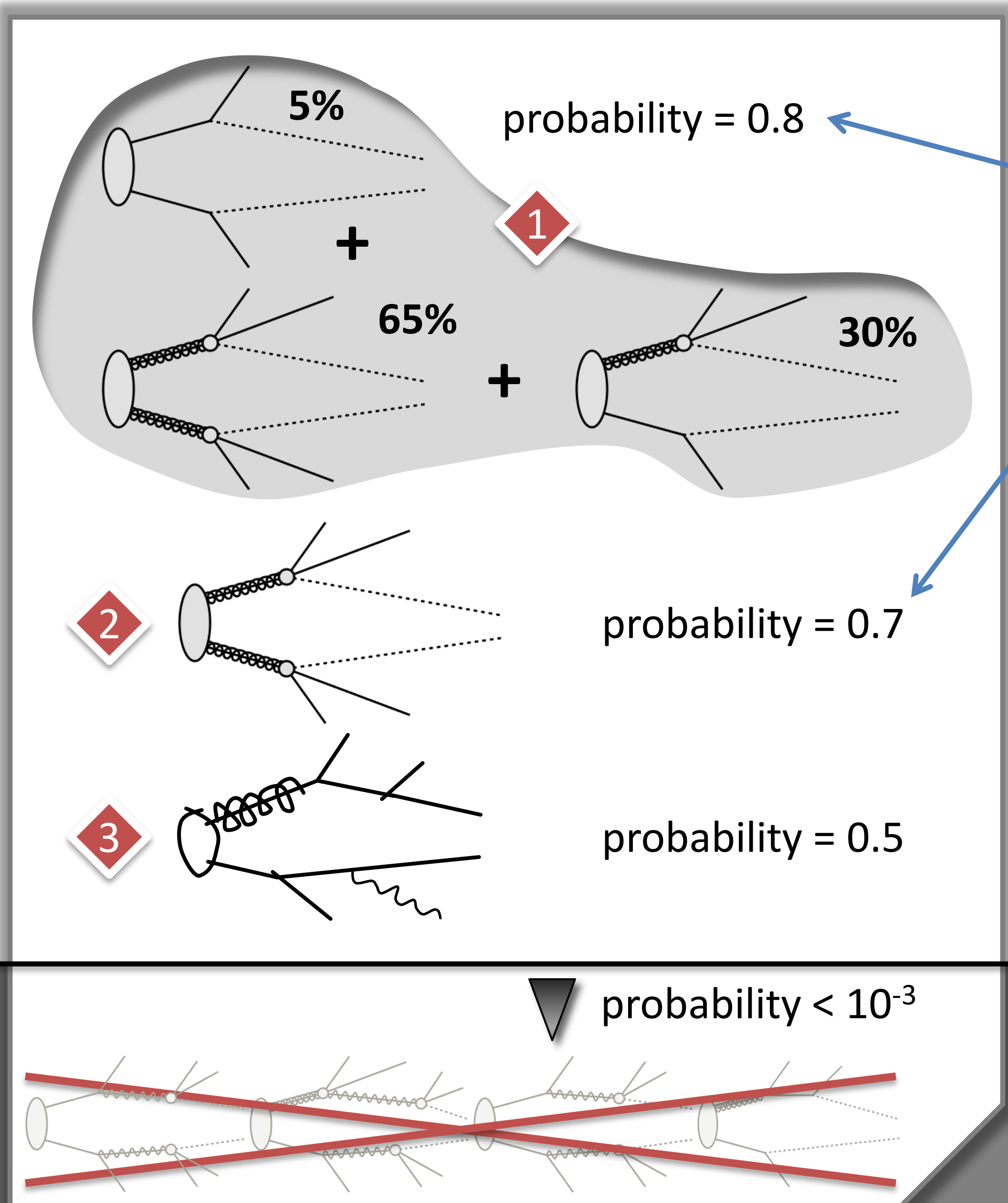
- Group points into regions containing exactly one minimum.
- Use only those points that are reasonably within the desired approximation in the parabola fit.

Table of Contents

1. Life at a hadron collider
2. Signal characterization, bottom-up
3. A hands-on procedure
4. Stages of understanding:
   a) Upper bound on a process
   b) Composing multi-process models
   c) Deducing new particle masses
5. **Models: good, bad, better**

160μb^{-1}	1.2nb^{-1}	5.2nb^{-1}	19nb^{-1}	21nb^{-1}	21nb^{-1}	110nb^{-1}	350nb^{-1}	930nb^{-1}	1.8pb^{-1}	3.6pb^{-1}	3.6pb^{-1}	8.3pb^{-1}	20pb^{-1}	45pb^{-1}
17/4	1/5	15/5	29/5	12/6	26/6	10/7	24/7	7/8	21/8	4/9	18/9	2/10	16/10	30/10

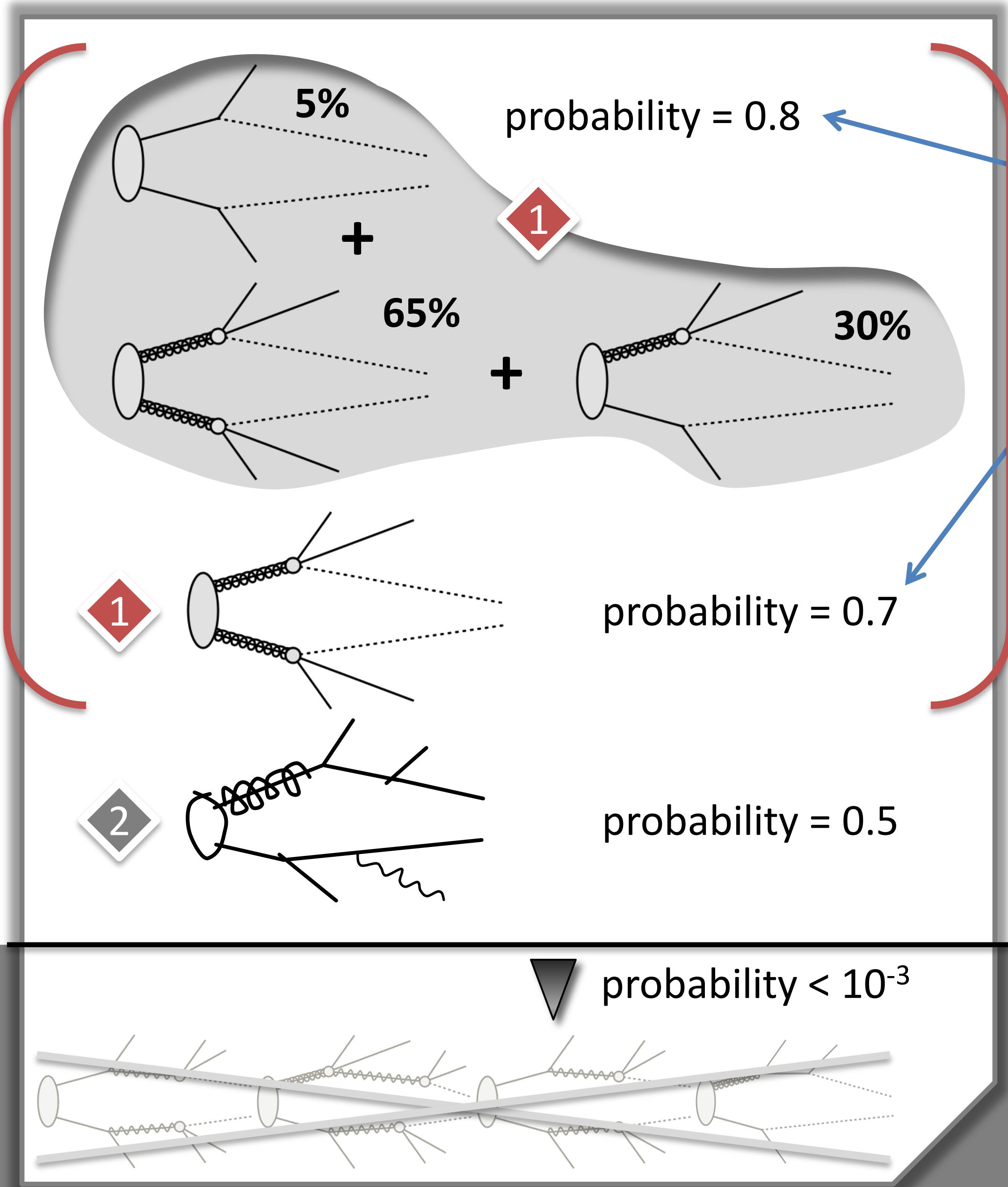
Ranking models by compatibility with signal



Increasing the number of free parameters (e.g. processes) always increases goodness of fit, but this may be a spurious effect.

All models with "low" probability are discarded as incompatible.

Ranking models by compatibility with signal



Increasing the number of free parameters (e.g. processes) always increases goodness of fit, but this may be a spurious effect.

Instead need to check if the two viable models are actually mutually compatible (within data uncertainties); if so they should be "tied" in rank.

All models with "low" probability are discarded as incompatible.

Table of Contents

jets + jets + ... + jets
= takes many pb^{-1} to resolve details
= first pass will be macroscopic

1. Life at a hadron collider
2. Signal characterization, bottom-up
3. A hands-on procedure
4. Stages of understanding:
   a) Upper bound on a process
   b) Composing multi-process models
   c) Deducing new particle masses
5. Models: good, bad, better

More flexible than top-down tests for specific models; quickly map out signal possibilities

Unsophisticated techniques using histogram comparison

Complementary to direct measurement (mass peaks)

Rank models by degree of compatibility with observation

Table of Contents

Need variables that can discriminate topologies…
… But can our searches predict all of these **distributions?**

1. Life at a hadron collider
2. Signal characterization, bottom-up
3. A hands-on procedure
4. Stages of understanding:
   a) Upper bound on a process
   b) Composing multi-process models
   c) Deducing new particle masses
5. Models: good, bad, better

jets + jets + … + jets
= takes many pb^{-1} to resolve details
= first pass will be macroscopic

More flexible than top-down tests for specific models; quickly map out signal possibilities

Unsophisticated techniques using histogram comparison

Complementary to direct measurement (mass peaks)

Rank models by degree of compatibility with observation

Need variables that can discriminate topologies...
... But can our searches predict all of these distributions?

$$\frac{300 \text{ recorded/s}}{4\times10^6 \text{ events/s}}$$
→ 99·99% events discarded:

Signal in some new, powerful corner of phase space need not be sitting around waiting to be analyzed at our convenience·

⇒ **Plan on planning ahead!**

Table of Contents

1. Life at a hadron collider
2. Signal characterization, bottom-up
3. A hands-on procedure
4. Stages of understanding:
   a) Upper bound on a process
   b) Composing multi-process models
   c) Deducing new particle masses
5. Models: good, bad, better

jets + jets + ... + jets
= takes many pb^{-1} to resolve details
= first pass will be macroscopic

More flexible than top-down tests for specific models; quickly map out signal possibilities

Unsophisticated techniques using histogram comparison

Complementary to direct measurement (mass peaks)

Rank models by degree of compatibility with observation

Chapter 1

It was a dark and stormy night. A team of physicists (armed with tall flasks of coffee) were watching with utmost care the monitors at ____, when all of a sudden...

a.k.a. The story is just beginning. ☺

"Salame" fit : slice N-dimensional mass parameter space up into all possible 1-D cells, within which we perform simple polynomial fits. Then collect the 1-D minima into an N-D result.

A problem encountered even for very simple pseudo-experiments.

Fit with higher-order polynomials (or a more efficient basis expansion).

BUT…

- How do we know what region falls within the parabolic approximation?
- Are we assured of there being exactly one minimum?

Goodness-of-fit:

- Compare mass hypotheses to locate most likely spectrum

"Salame" fit : slice N-dimensional mass parameter space up into all possible 1-D cells, within which we perform simple polynomial fits. Then collect the 1-D minima into an N-D result.

"Good" points can later be used to

Fit with higher-order polynomials (or a more efficient basis expansion).

A problem encountered even for very simple pseudo-experiments.

BUT…

- How do we know what region falls within the parabolic approximation?
- Are we assured of there being exactly one minimum?

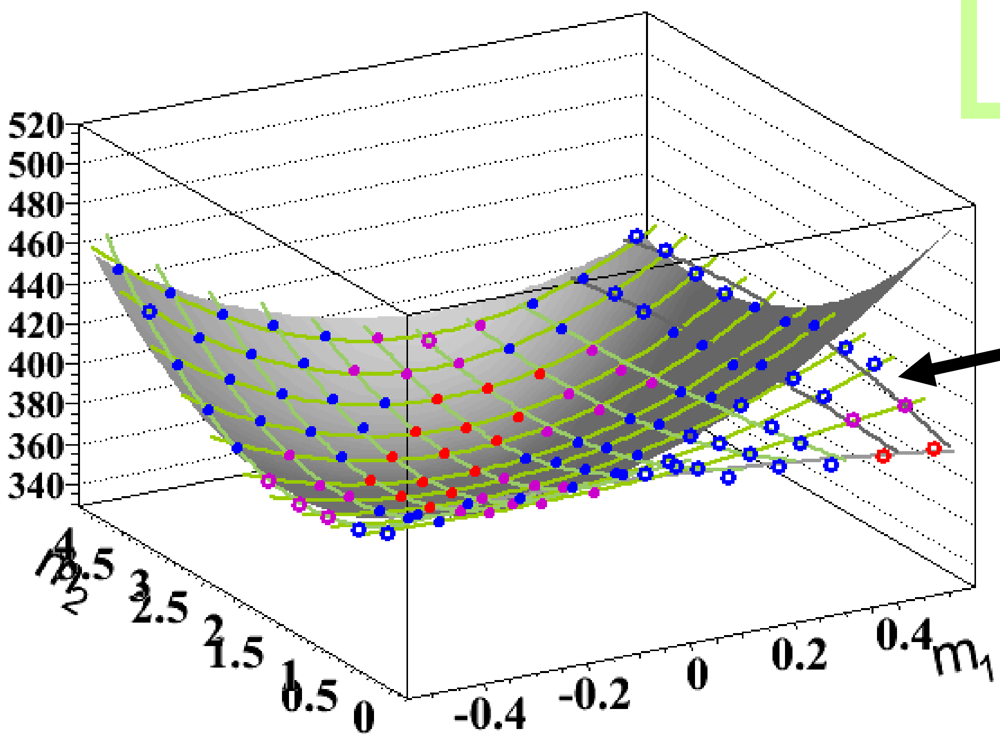
Goodness-of-fit:

- Compare mass hypotheses to locate most likely spectrum

e.g. Particular cell in a 3-mass parameter space

"Salame" fit

$m_1 = x$, $m_2 = y$, $m_3 = 5$: 1D fits



The (mostly) green mesh are the 1-D slice fits to the shape function: $a + b\,x + c\,x^2$

If a slice is convex ($c < 0$), it likely does not belong to an N-D *minimum* — an expansion around a minimum must have positive 2nd order derivatives.

Omit these points from N-D fit

BUT…

- How do we know what region falls within the parabolic approximation?
- Are we assured of there being exactly one minimum?

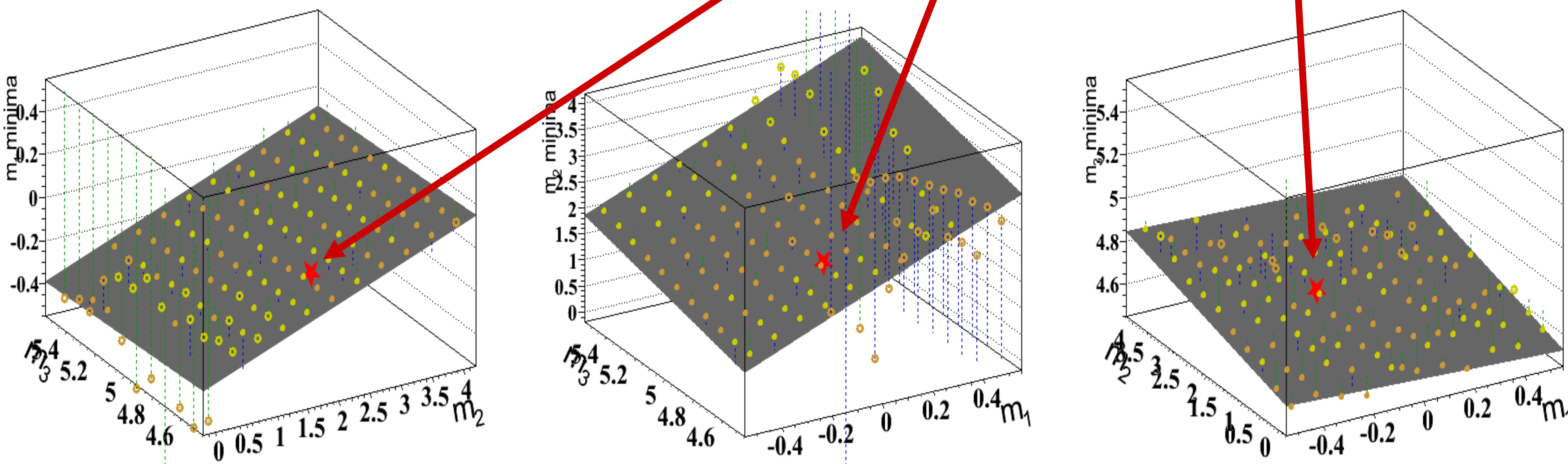
Goodness-of-fit:

- Compare mass hypotheses to locate most likely spectrum

“Salame” fit

The 1-D minima form N “planes” that are close to the principal axes of the N-D parabola around the global minimum.

The intersection of all these planes are a none-too-shabby estimator for the global minimum.



BUT…

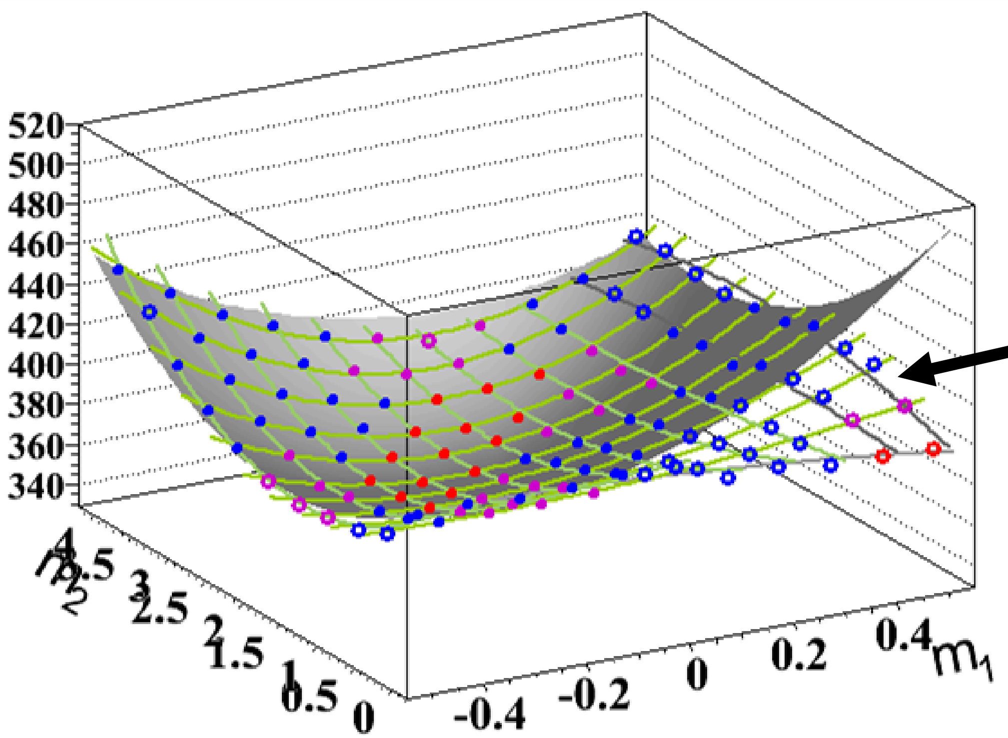
- How do we know what region falls within the parabolic approximation?
- Are we assured of there being exactly one minimum?

Goodness-of-fit:

- Compare mass hypotheses to locate most likely spectrum

"Salame" fit

For a better estimator, we can also use the 1-D fits to more correctly seed the N-D fit.



If a slice is convex ($c < 0$), it likely does not belong to an N-D *minimum*

Omit these points from N-D fit

BUT…

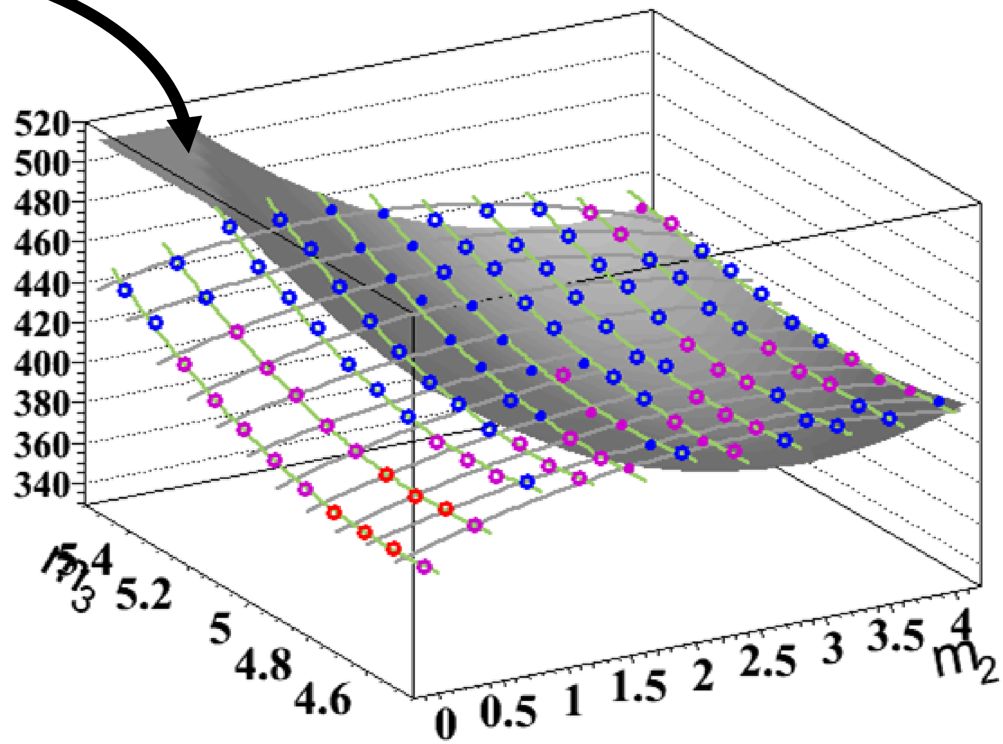
- How do we know what region falls within the parabolic approximation?
- Are we assured of there being exactly one minimum?

Goodness-of-fit:

- Compare mass hypotheses to locate most likely spectrum

“Salame” fit

The N-D fit (gray surface) successfully ignores points that can’t reasonably lie within the region where the parabolic expansion is valid (empty circles)



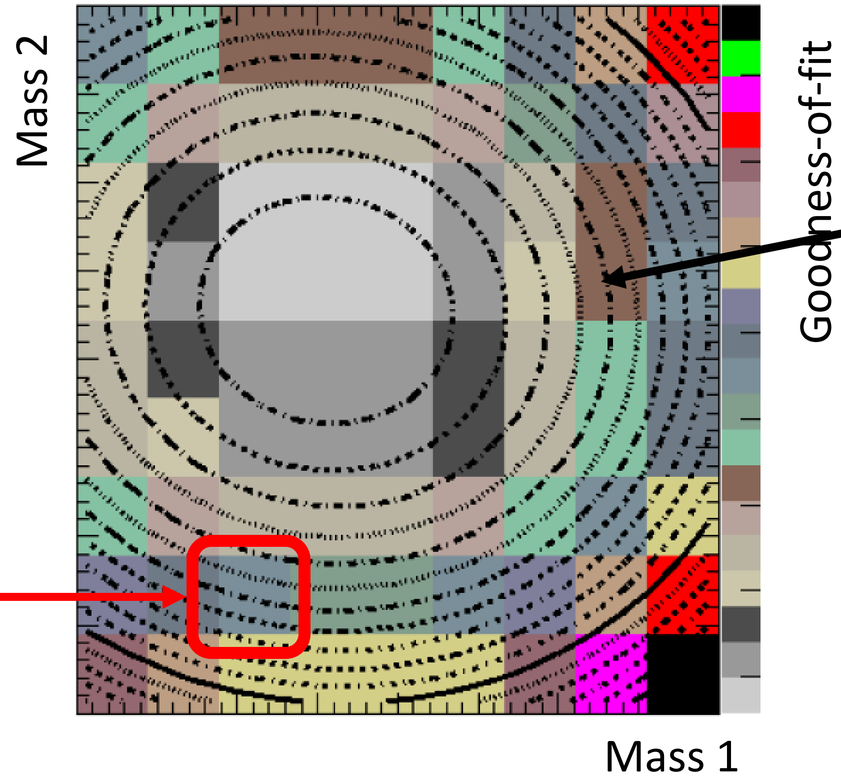
BUT…

- How do we know what region falls within the parabolic approximation?
- Are we assured of there being exactly one minimum?

Goodness-of-fit:

- Compare mass hypotheses to locate most likely spectrum

For a specific set of mass hypotheses (at this point of the mass grid): record how well we can fit the three processes to explain signal.



Contours of parabolic fit interpolating between points in (possibly coarse) grid

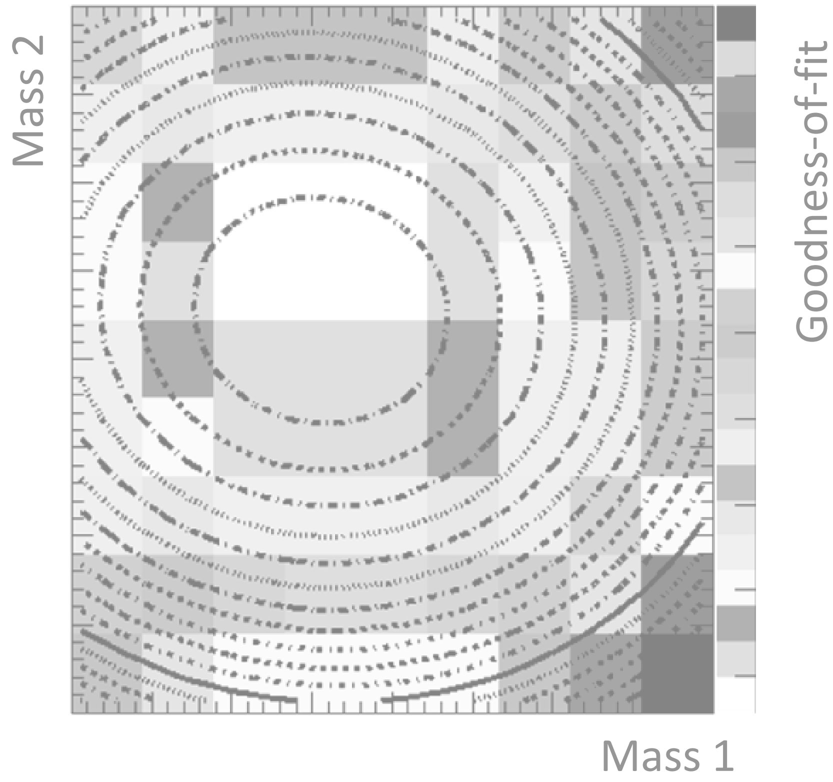
Refine grid after roughly locating minimum

For each particular model:

The goodness-of-fit (for the various processes) as we vary the mass parameters can be used to locate the most probable mass spectrum.

Goodness-of-fit:

- Compare mass hypotheses to locate most likely spectrum



BUT…

- How do we know what region falls within the parabolic approximation?
- Are we assured of there being exactly one minimum?

Goodness-of-fit:

- Compare mass hypotheses to locate most likely spectrum