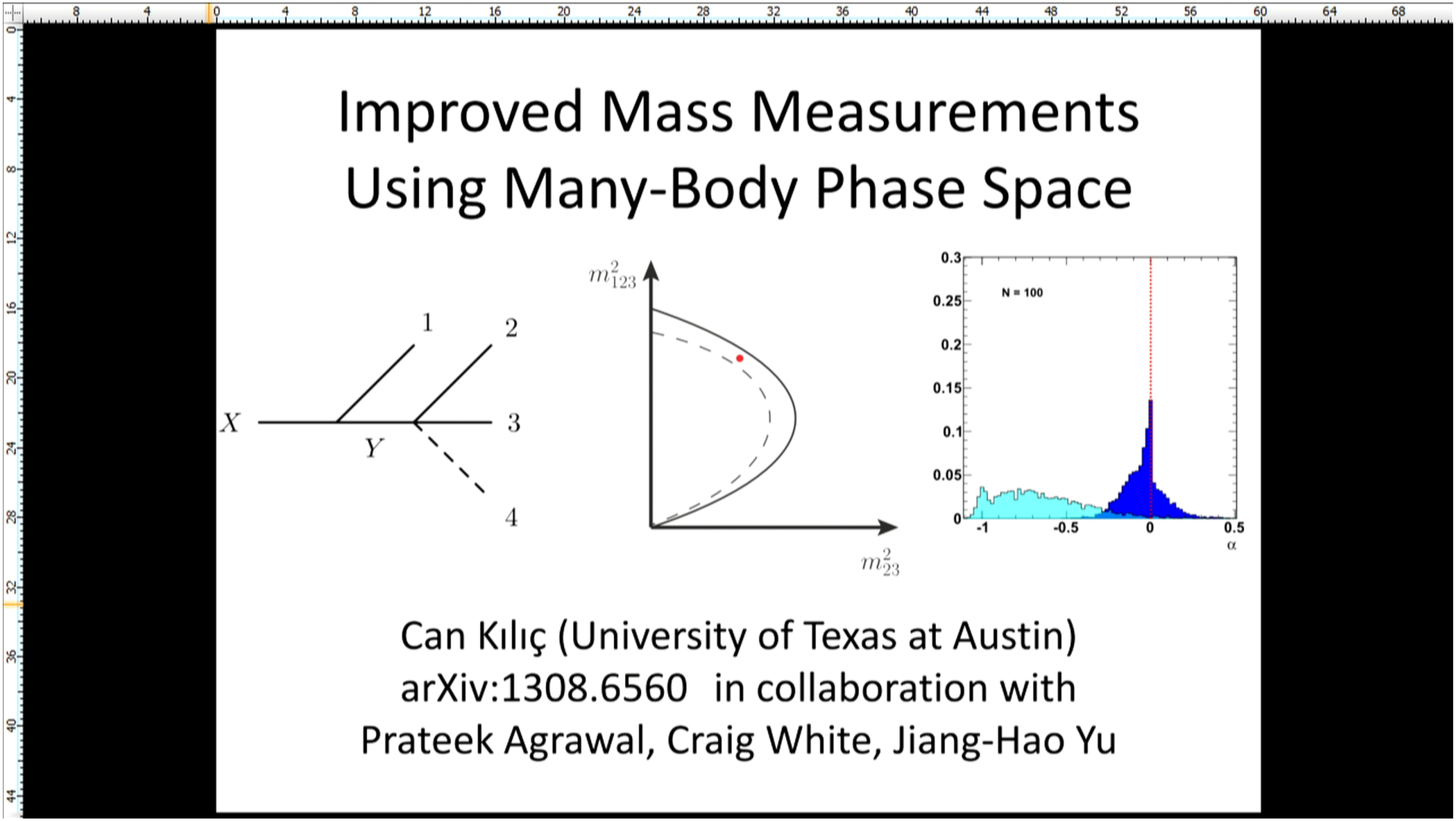
Title: Improving Mass Measurements Using Many-Body Phase Space

Date: Jul 15, 2014 01:00 PM

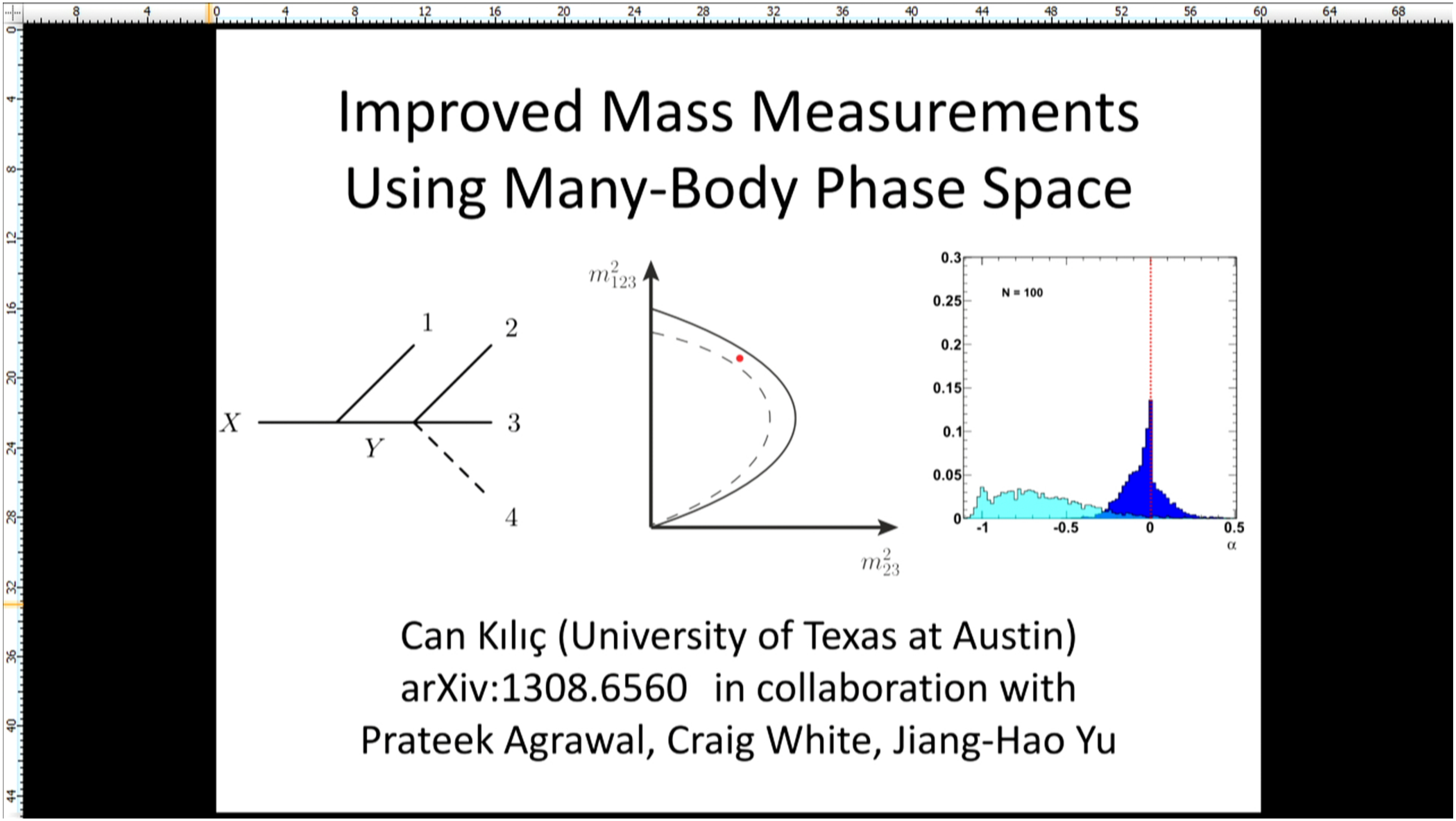
URL: http://pirsa.org/14070023

Abstract: After the 7 and 8 TeV LHC runs, we have no conclusive evidence of physics beyond the Standard Model, leading us to suspect that even if new physics is discovered during run II, the number of signal events may be limited, making it crucial to optimize measurements for the case of low statistics. I will argue that phase space correlations between subsequent on-shell decays in a cascade contain additional information compared to commonly used kinematic variables, and this can be used to significantly improve the precision and accuracy of mass measurements. The improvement is connected to the properties of the volume element of many-body phase space, and is particularly relevant to the case of low signal statistics.



Improved Mass Measurements Using Many-Body Phase Space

Can Kılıç (University of Texas at Austin)
arXiv:1308.6560 in collaboration with
Prateek Agrawal, Craig White, Jiang-Hao Yu



Improved Mass Measurements Using Many-Body Phase Space

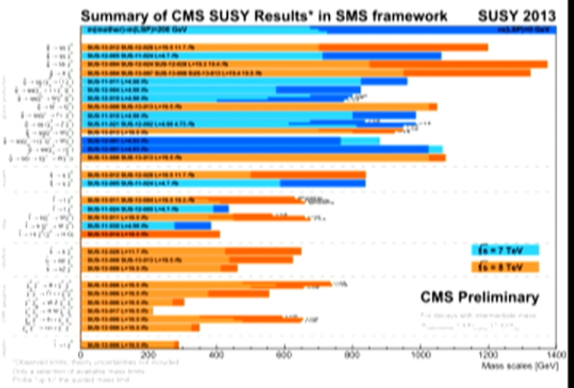
Can Kılıç (University of Texas at Austin)
arXiv:1308.6560 in collaboration with
Prateek Agrawal, Craig White, Jiang-Hao Yu

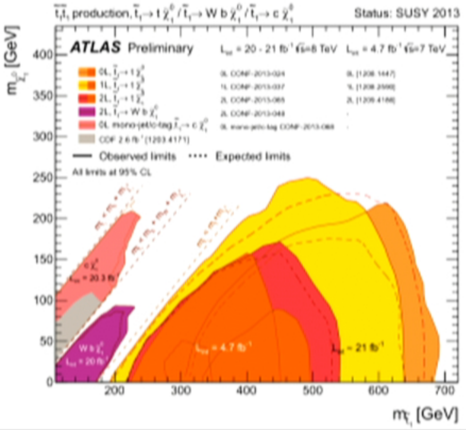
Waiting For New Physics

After 7 and 8 TeV runs, the long wait continues.

In the event of discovery during run II, we are unlikely to get many events.

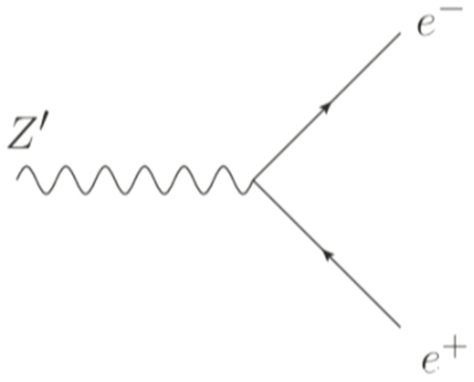
It will be crucial to make the most out of a limited amount of statistics.

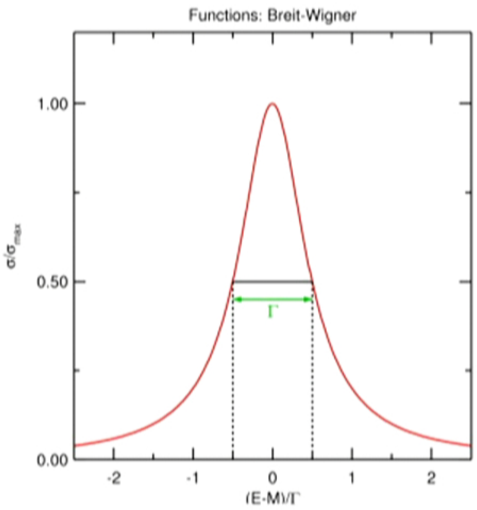




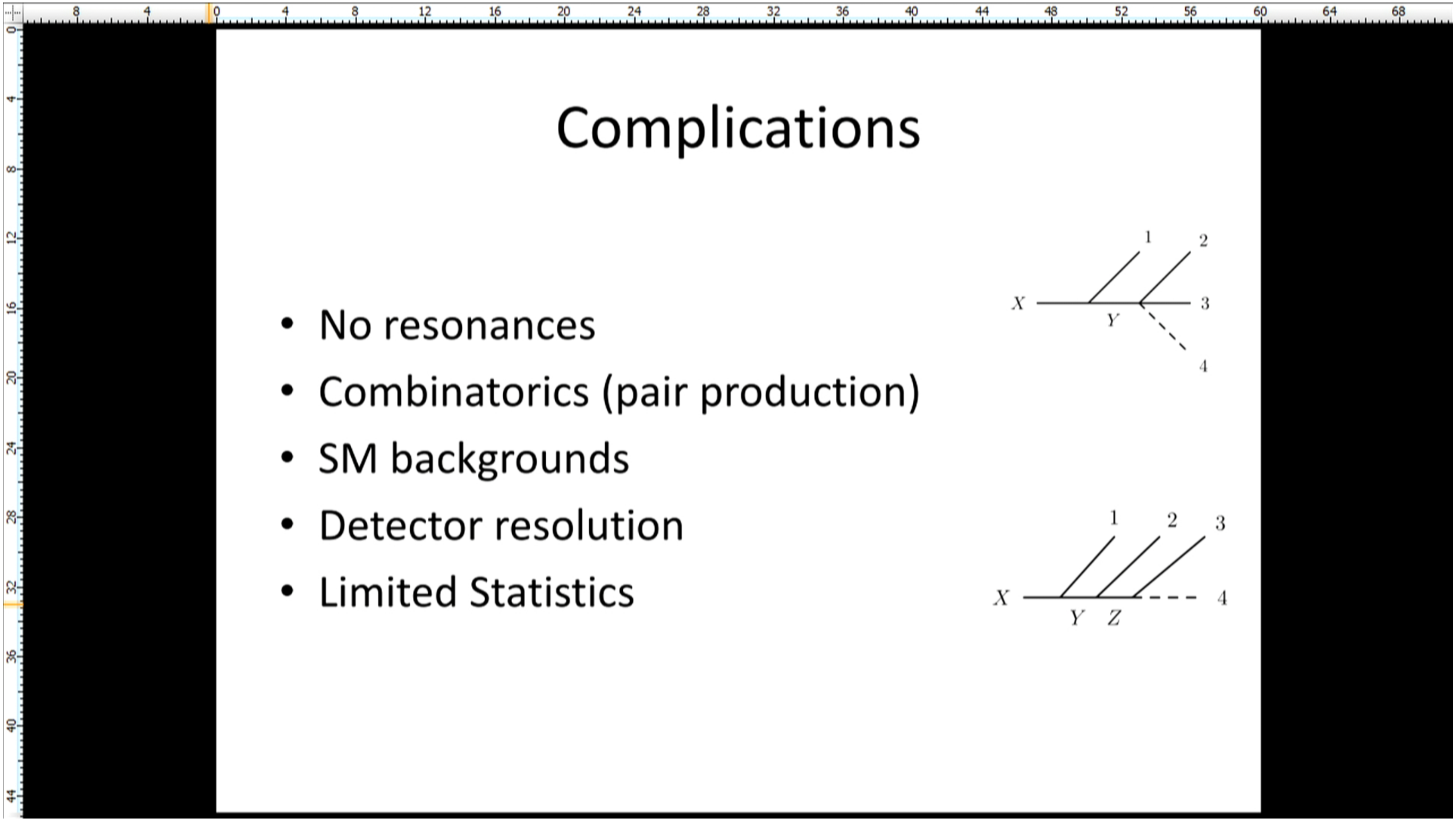
Measuring Masses

Measuring masses of new particles is one of the top priorities.



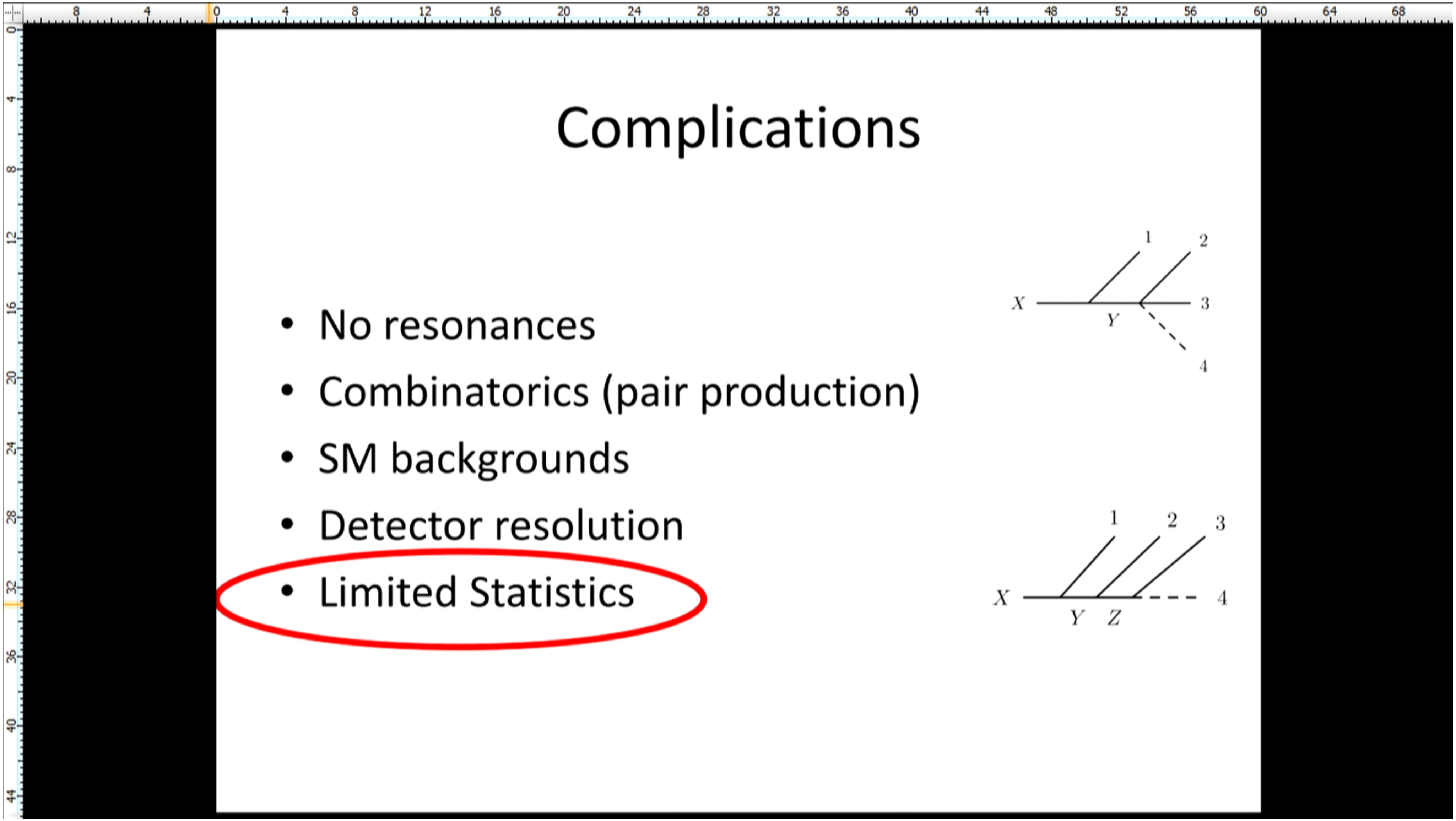


Unless we get *really* lucky, this is non-trivial.



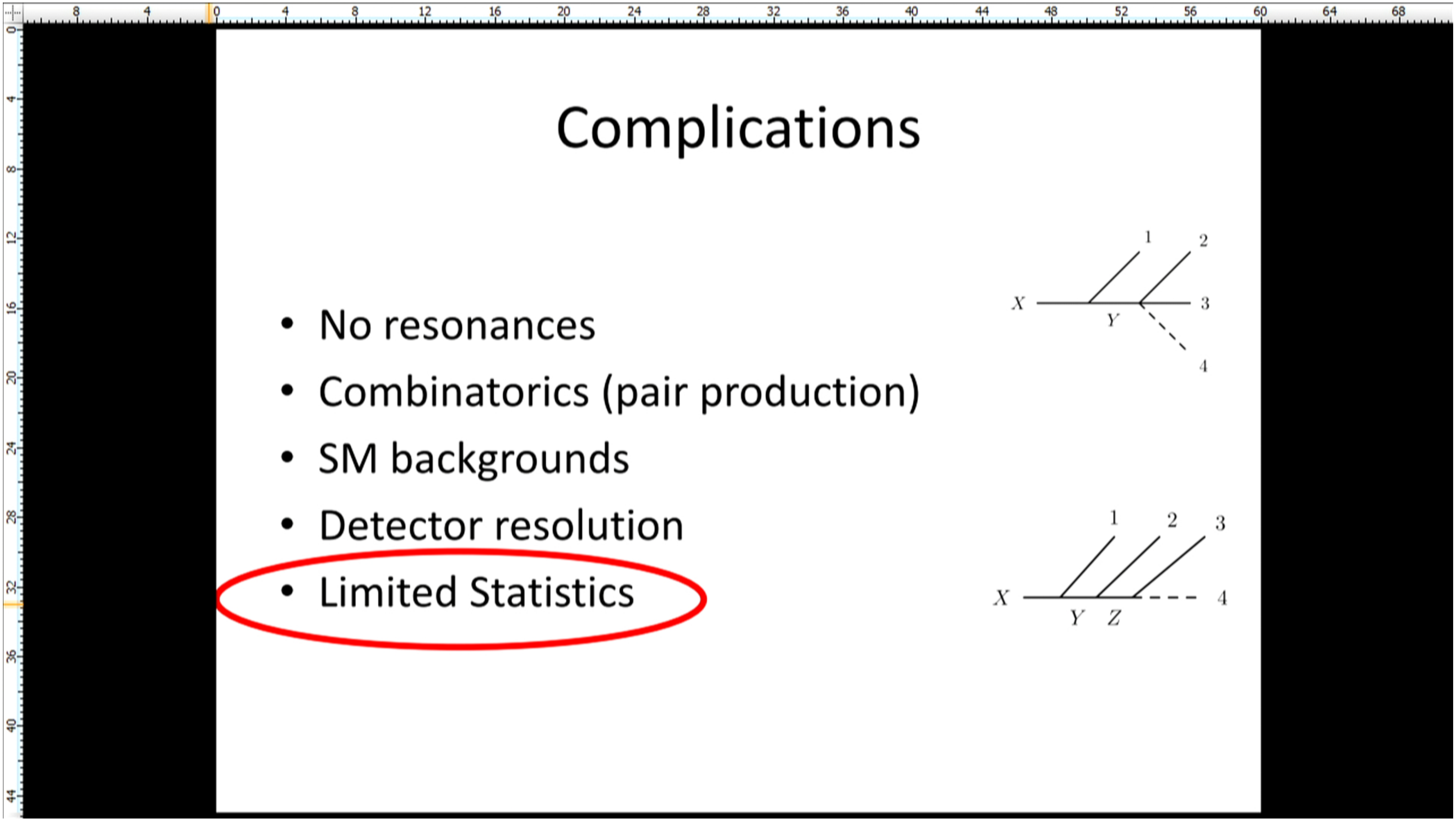
Complications

- No resonances
- Combinatorics (pair production)
- SM backgrounds
- Detector resolution
- Limited Statistics



Complications

- No resonances
- Combinatorics (pair production)
- SM backgrounds
- Detector resolution
- Limited Statistics

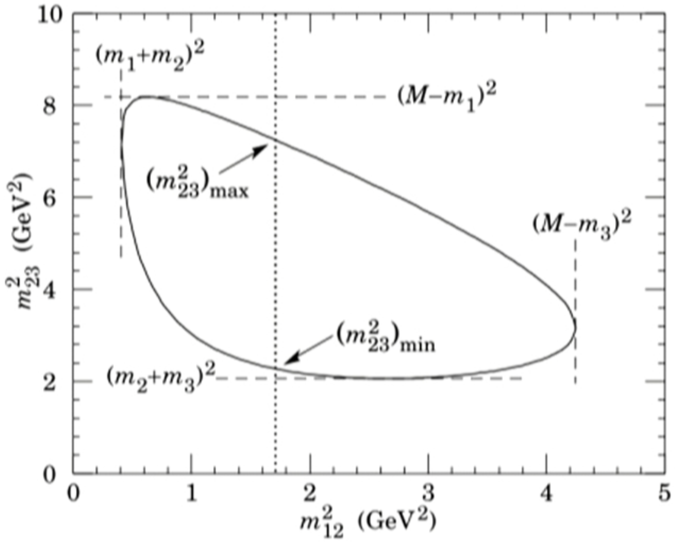


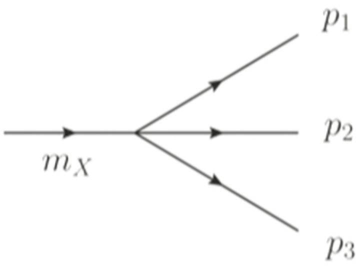
Complications

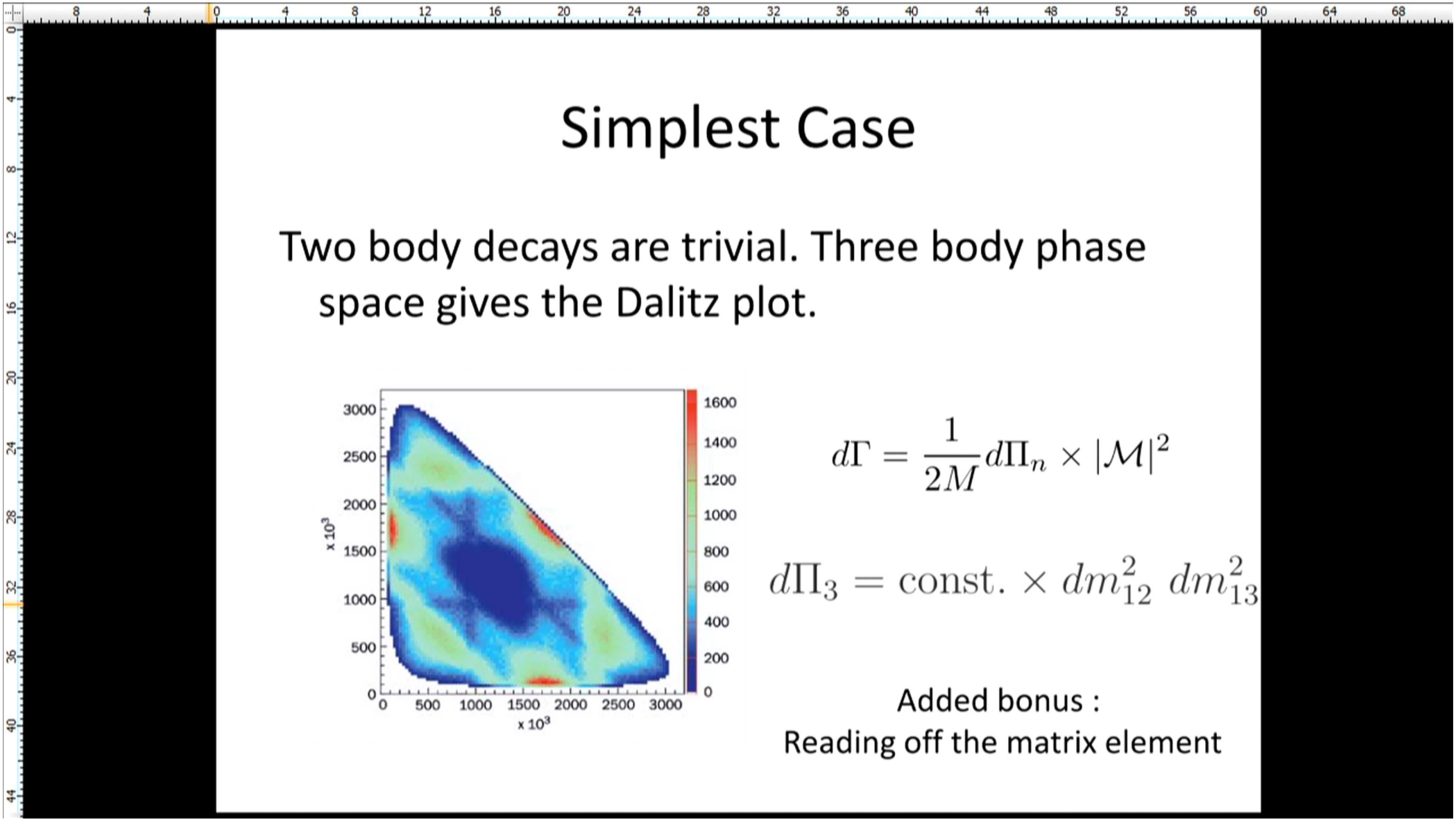
- No resonances
- Combinatorics (pair production)
- SM backgrounds
- Detector resolution
- Limited Statistics

Simplest Case

Two body decays are trivial. Three body phase space gives the Dalitz plot.







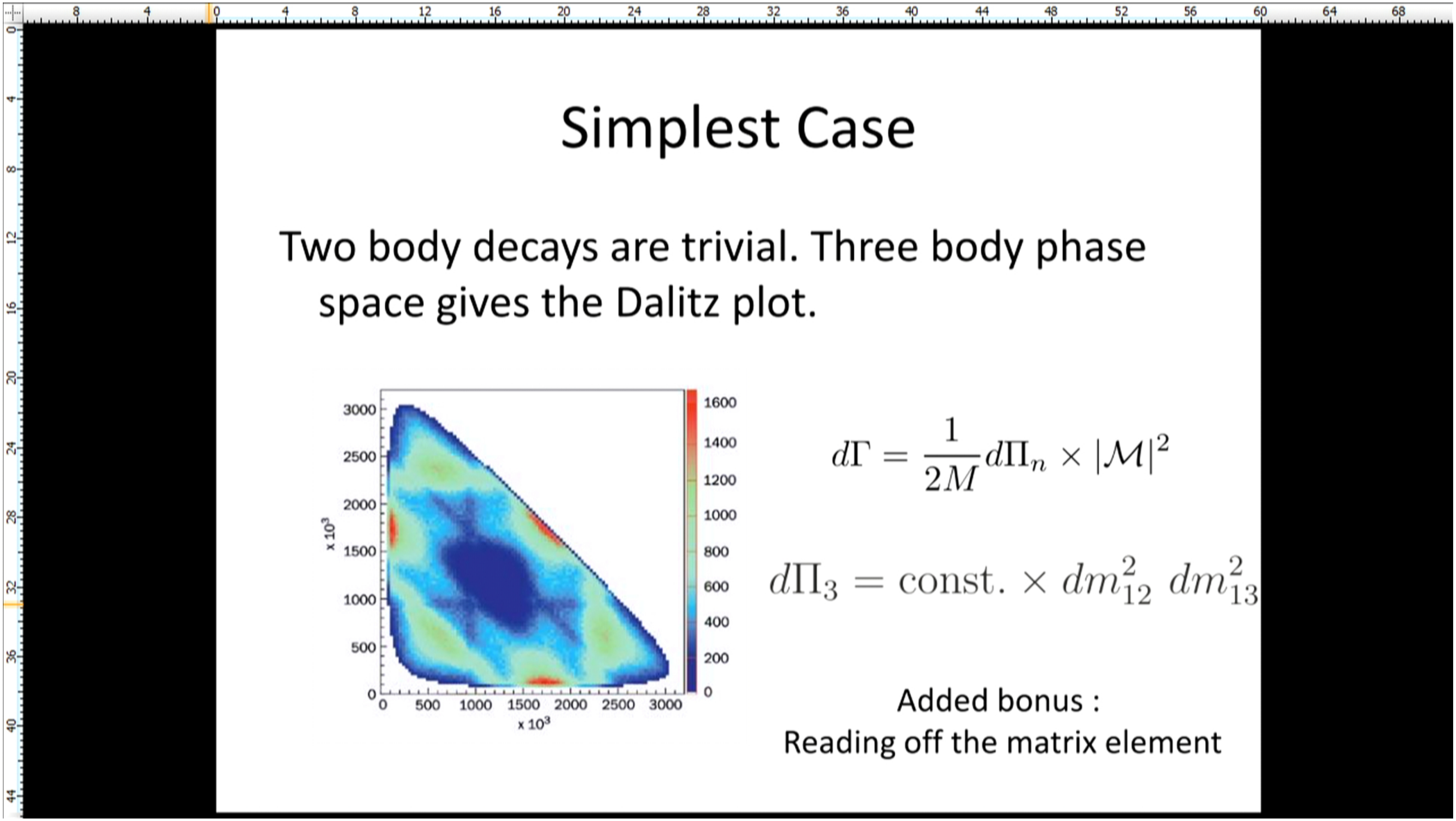
Simplest Case

Two body decays are trivial. Three body phase space gives the Dalitz plot.

$$d\Gamma = \frac{1}{2M} d\Pi_n \times |\mathcal{M}|^2$$

$$d\Pi_3 = \text{const.} \times dm_{12}^2 \; dm_{13}^2$$

Added bonus :
Reading off the matrix element



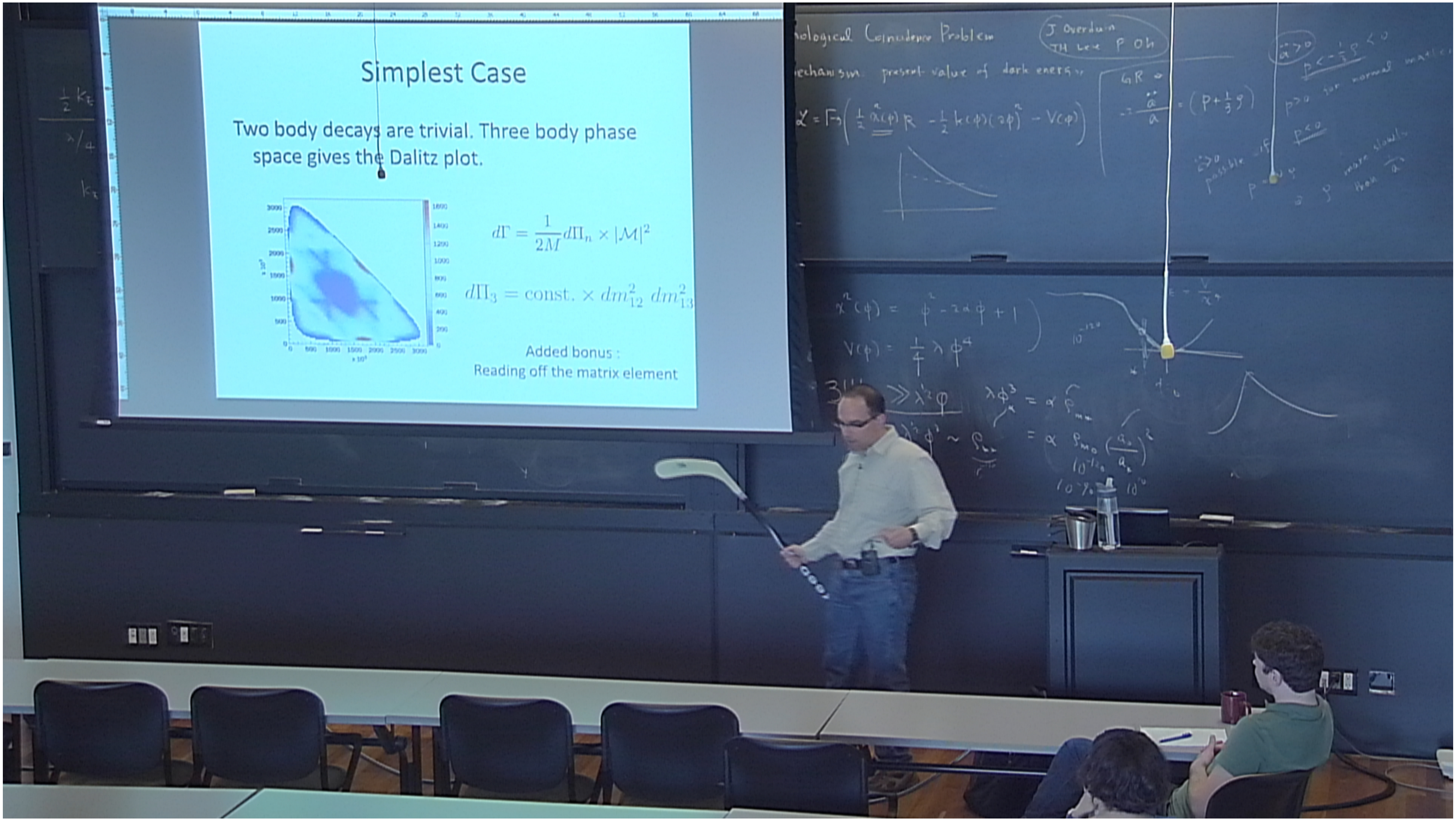
Simplest Case

Two body decays are trivial. Three body phase space gives the Dalitz plot.

$$d\Gamma = \frac{1}{2M} d\Pi_n \times |\mathcal{M}|^2$$

$$d\Pi_3 = \text{const.} \times dm_{12}^2 \; dm_{13}^2$$

Added bonus :
Reading off the matrix element



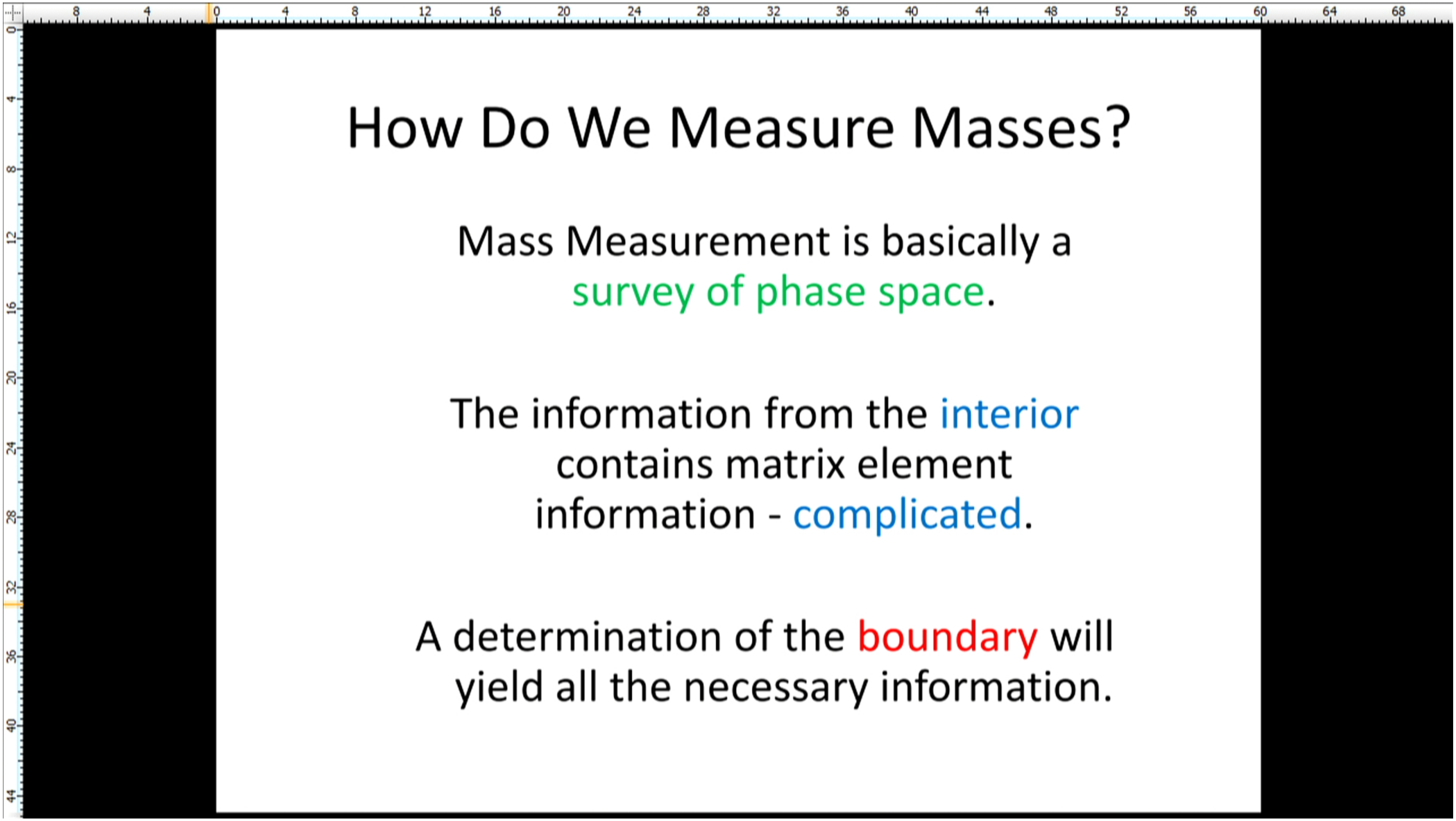
Simplest Case

Two body decays are trivial. Three body phase space gives the Dalitz plot.

$$d\Gamma = \frac{1}{2M} d\Pi_n \times |\mathcal{M}|^2$$

$$d\Pi_3 = \text{const.} \times dm_{12}^2 \; dm_{13}^2$$

Added bonus :
Reading off the matrix element



How Do We Measure Masses?

Mass Measurement is basically a survey of phase space.

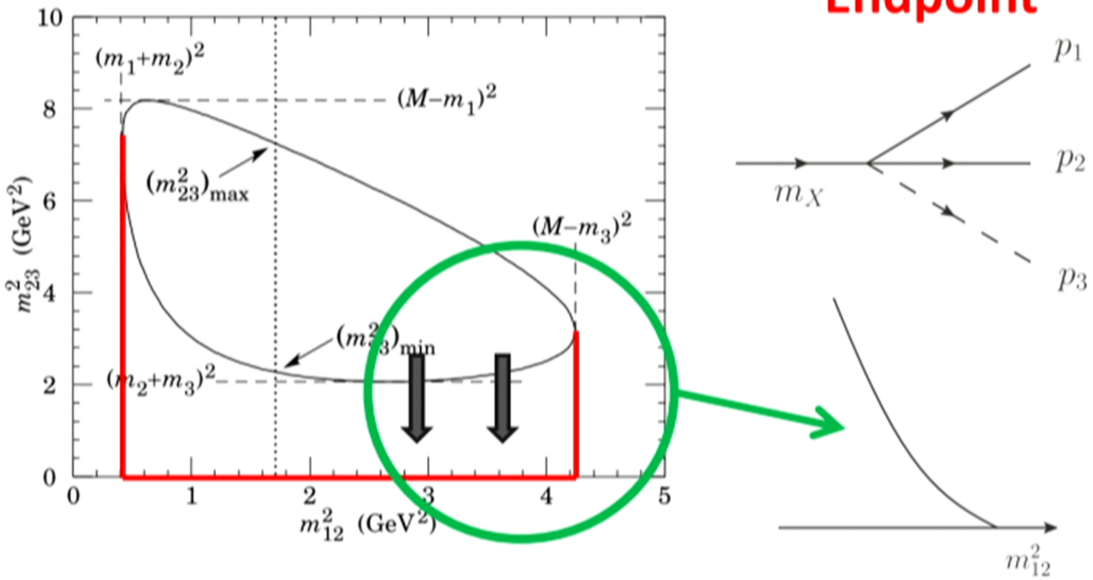
The information from the interior contains matrix element information - complicated.

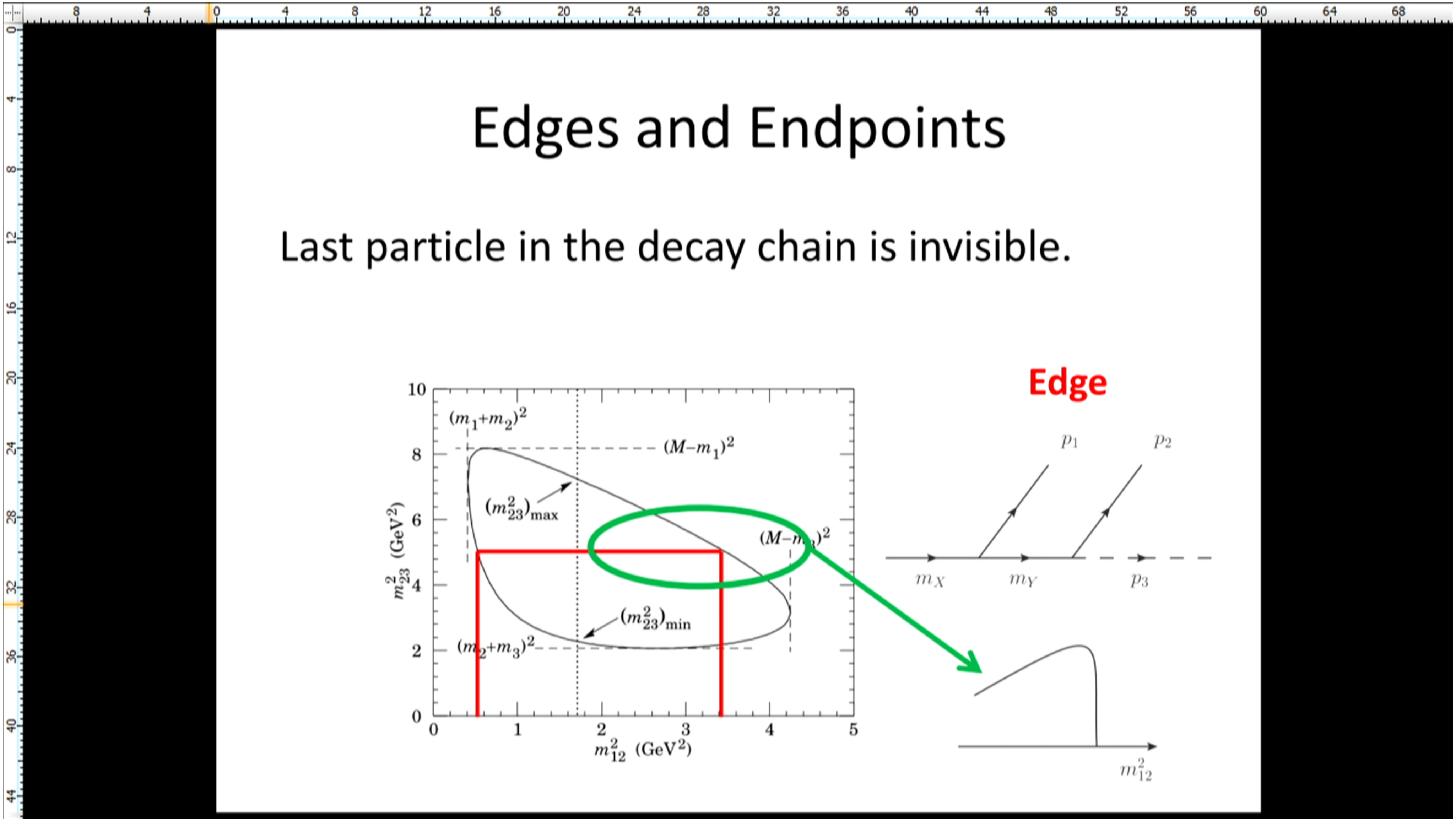
A determination of the boundary will yield all the necessary information.

Edges and Endpoints

Last particle in the decay chain is invisible.

Endpoint





Edges and Endpoints

Last particle in the decay chain is invisible.

Edge

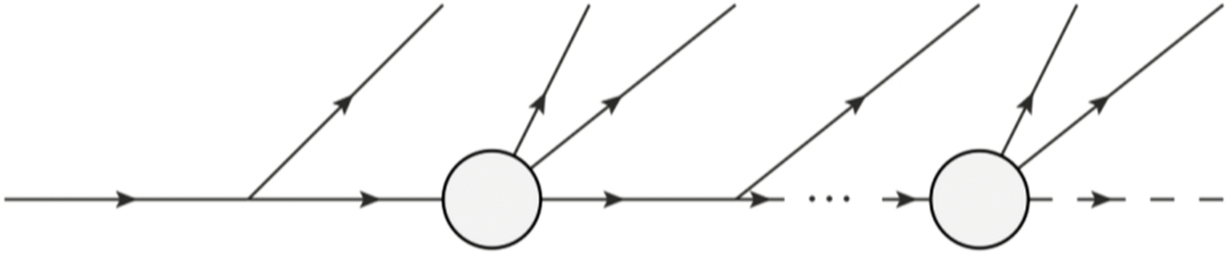
Beyond 3-Body Phase Space

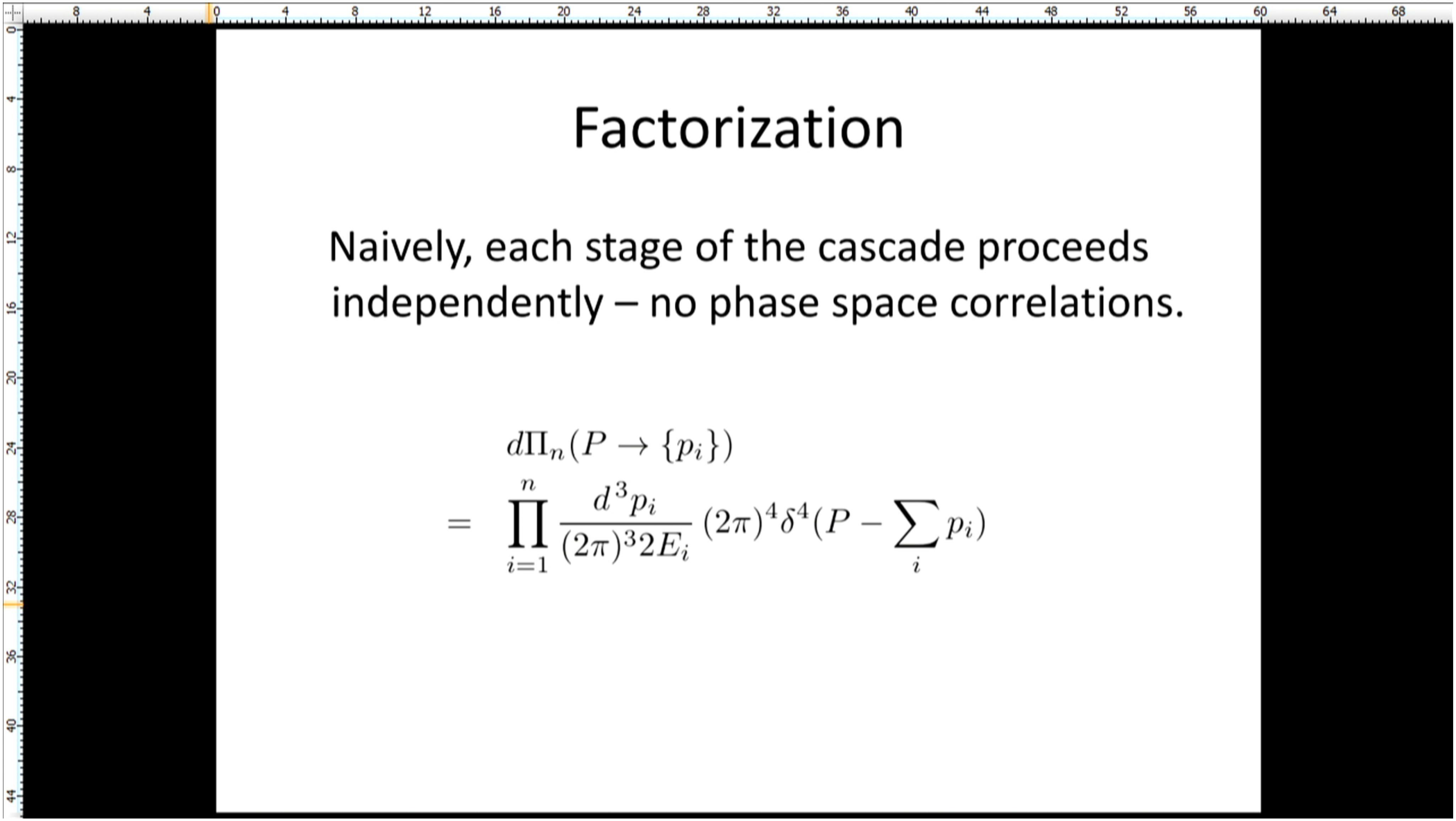
Basically all known (and hypothesized) particles will decay either to two or three final state particles.

Any longer decay chain can be decomposed into a cascade.

Isn't it good enough to analyze the cascade step by step, looking for edges and endpoints?

Flat direction when all masses increase.

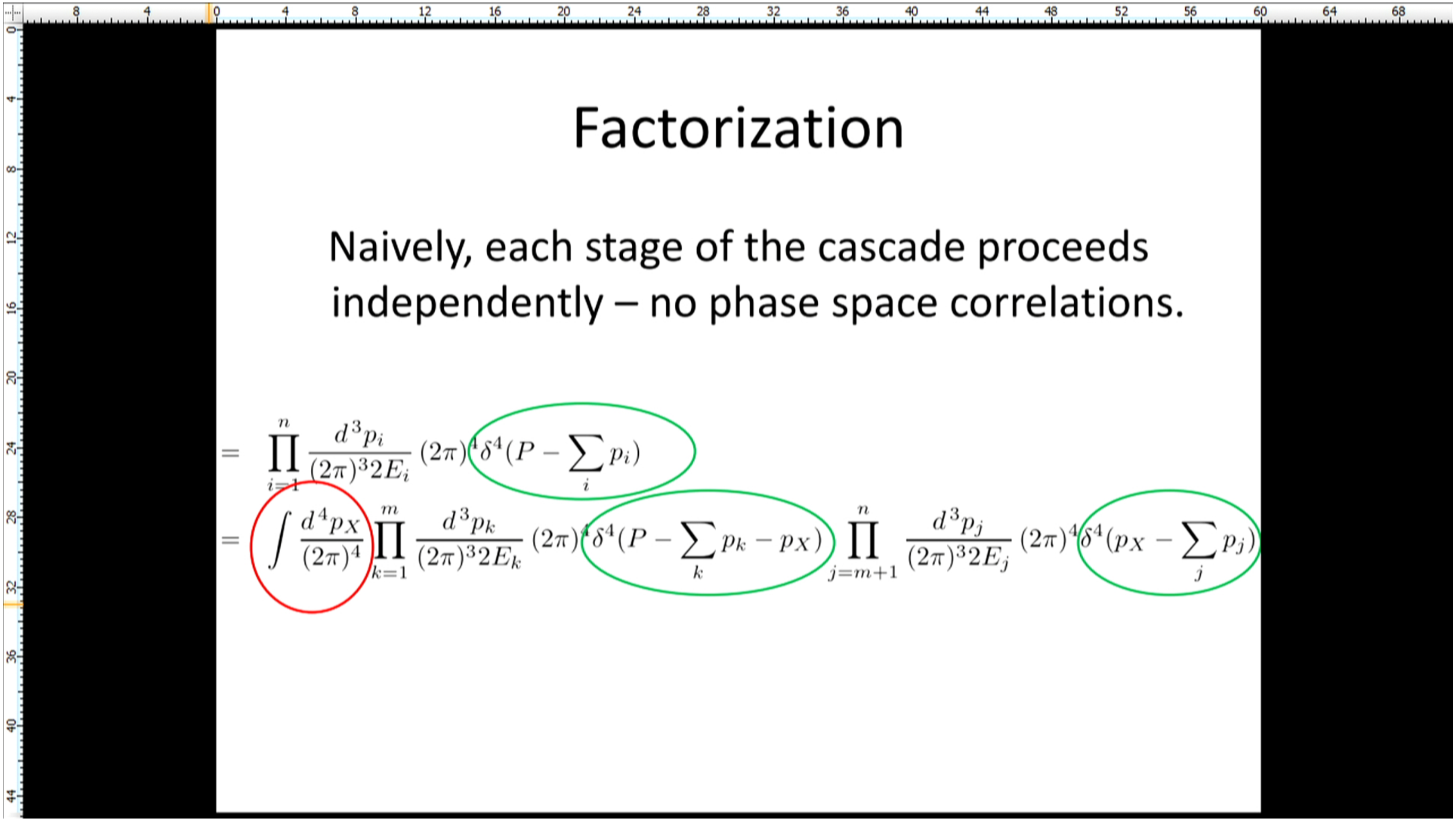




Factorization

Naively, each stage of the cascade proceeds independently – no phase space correlations.

$$
\begin{aligned}
& d\Pi_n(P \to \{p_i\}) \\
= \quad & \prod_{i=1}^{n} \frac{d^3 p_i}{(2\pi)^3 2E_i}\,(2\pi)^4 \delta^4(P - \sum_i p_i)
\end{aligned}
$$

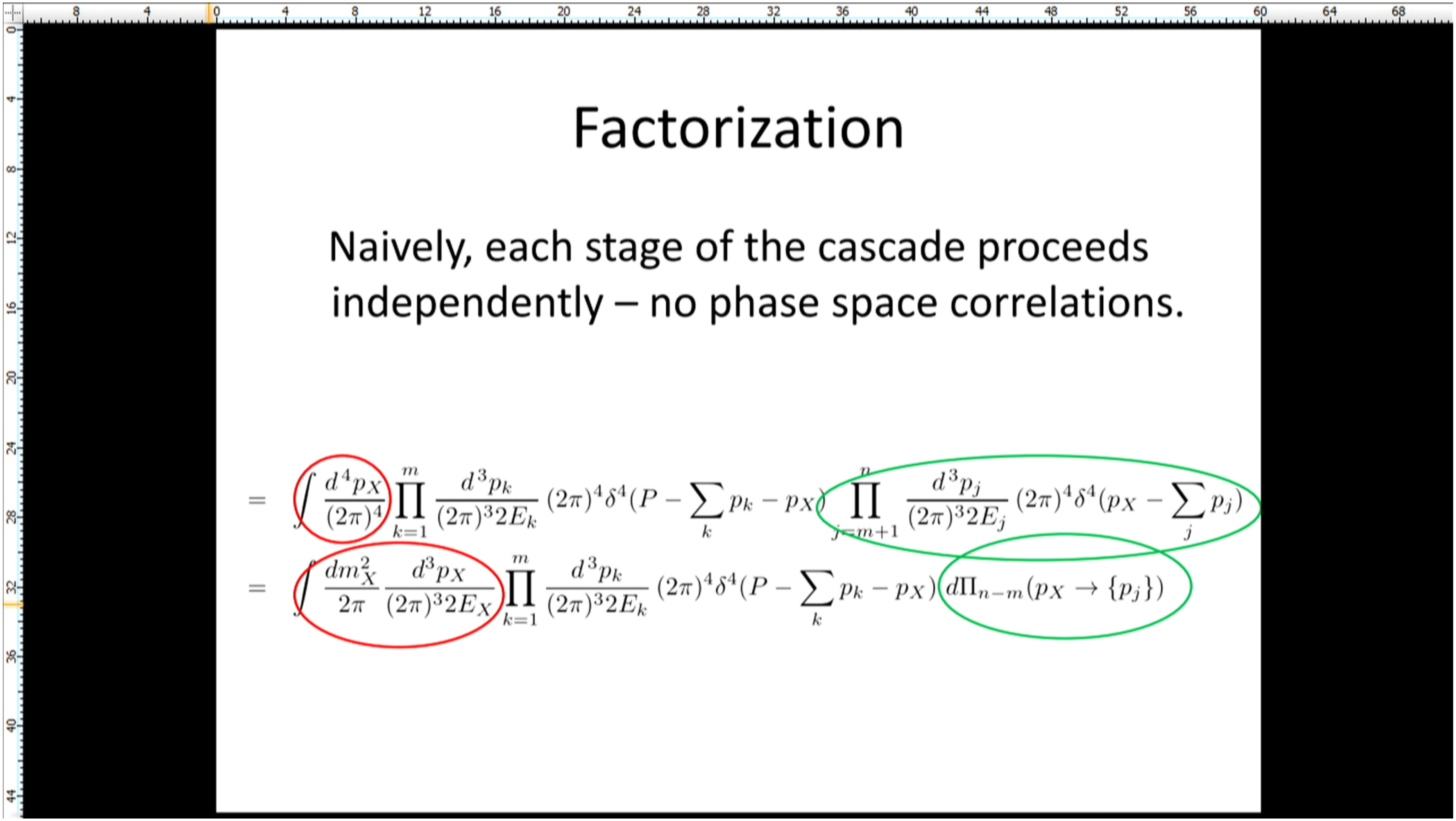


Factorization

Naively, each stage of the cascade proceeds independently – no phase space correlations.

$$= \prod_{i=1}^{n} \frac{d^3p_i}{(2\pi)^3 2E_i} (2\pi)^4 \delta^4(P - \sum_i p_i)$$

$$= \int \frac{d^4p_X}{(2\pi)^4} \prod_{k=1}^{m} \frac{d^3p_k}{(2\pi)^3 2E_k} (2\pi)^4 \delta^4(P - \sum_k p_k - p_X) \prod_{j=m+1}^{n} \frac{d^3p_j}{(2\pi)^3 2E_j} (2\pi)^4 \delta^4(p_X - \sum_j p_j)$$

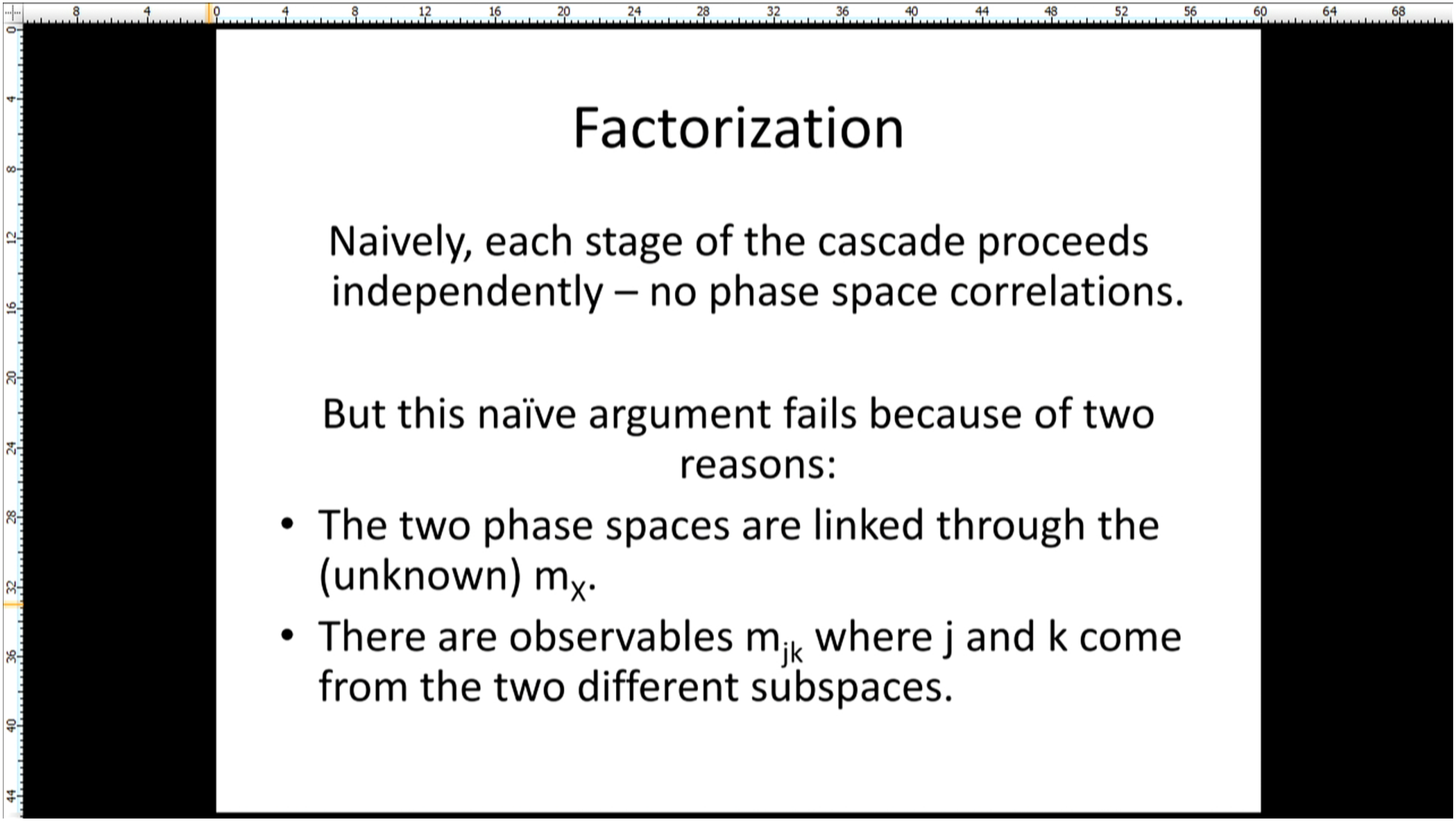


Factorization

Naively, each stage of the cascade proceeds independently – no phase space correlations.

$$= \int \frac{d^4p_X}{(2\pi)^4} \prod_{k=1}^{m} \frac{d^3p_k}{(2\pi)^3 2E_k} (2\pi)^4 \delta^4(P - \sum_k p_k - p_X) \prod_{j=m+1}^{n} \frac{d^3p_j}{(2\pi)^3 2E_j} (2\pi)^4 \delta^4(p_X - \sum_j p_j)$$

$$= \int \frac{dm_X^2}{2\pi} \frac{d^3p_X}{(2\pi)^3 2E_X} \prod_{k=1}^{m} \frac{d^3p_k}{(2\pi)^3 2E_k} (2\pi)^4 \delta^4(P - \sum_k p_k - p_X) \, d\Pi_{n-m}(p_X \to \{p_j\})$$

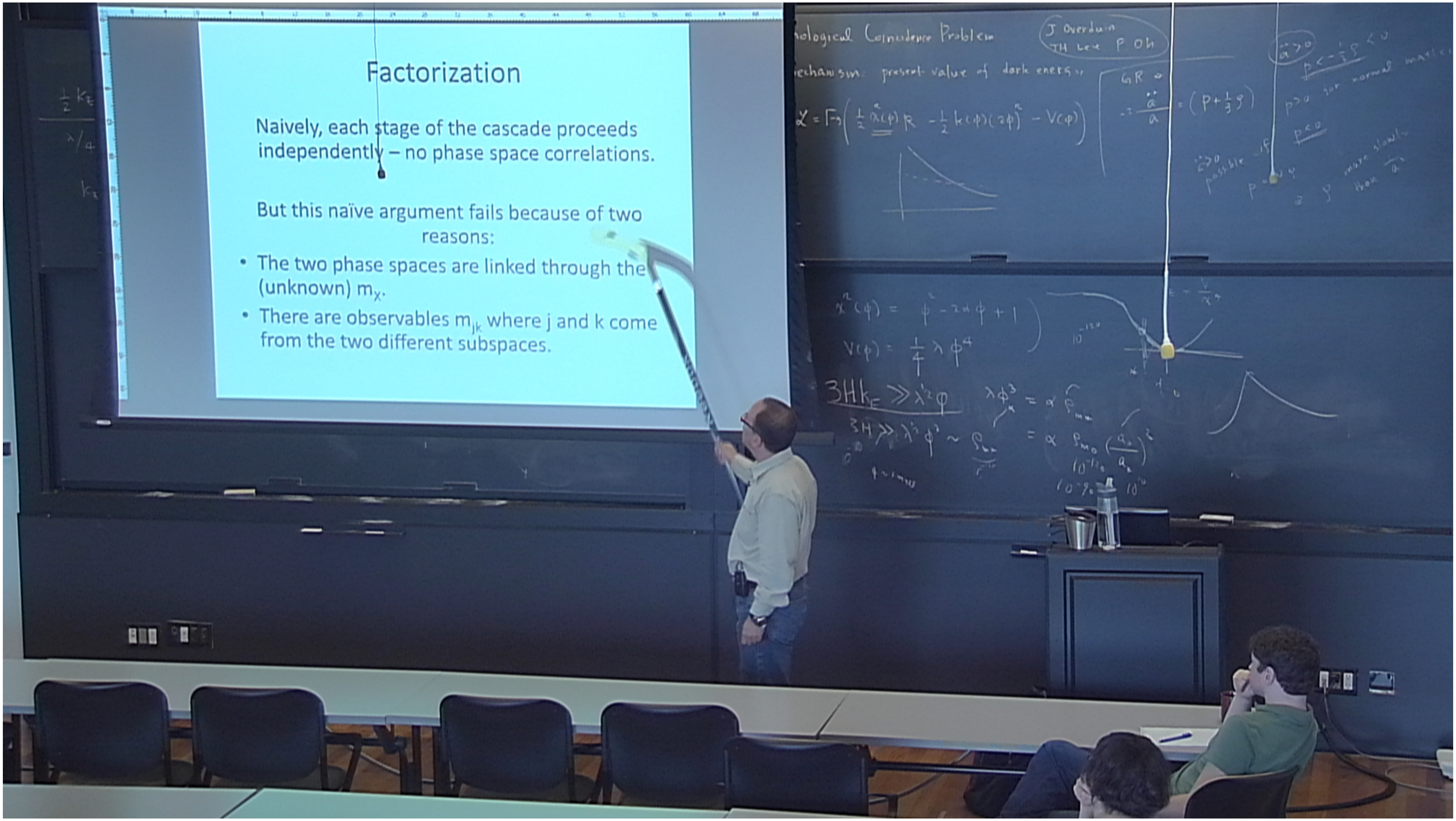


Factorization

Naively, each stage of the cascade proceeds independently – no phase space correlations.

But this naïve argument fails because of two reasons:

- The two phase spaces are linked through the (unknown) m_X.
- There are observables m_{jk} where j and k come from the two different subspaces.

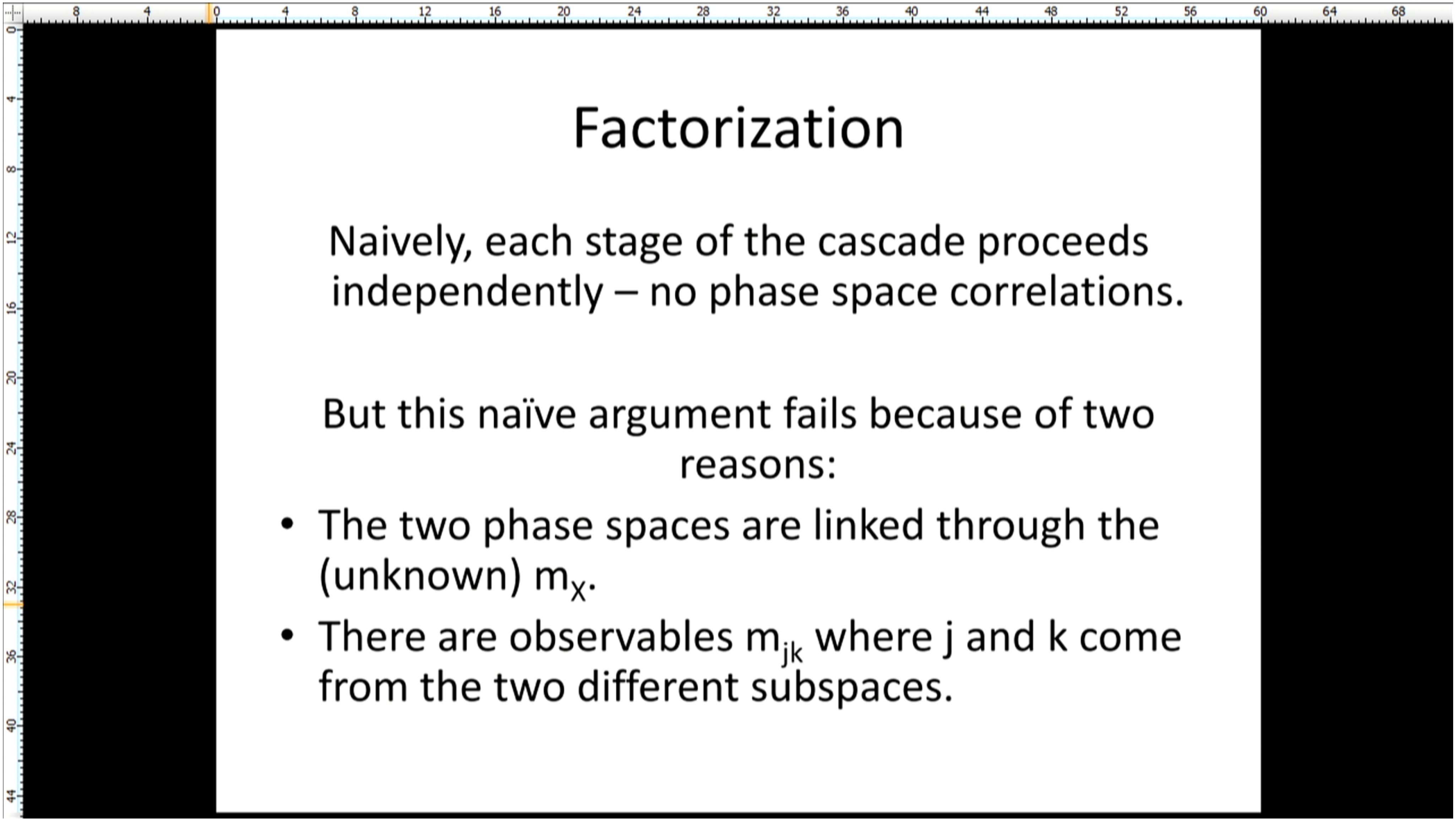


Factorization

Naively, each stage of the cascade proceeds independently – no phase space correlations.

But this naïve argument fails because of two reasons:

- The two phase spaces are linked through the (unknown) m_X.
- There are observables m_{jk} where j and k come from the two different subspaces.

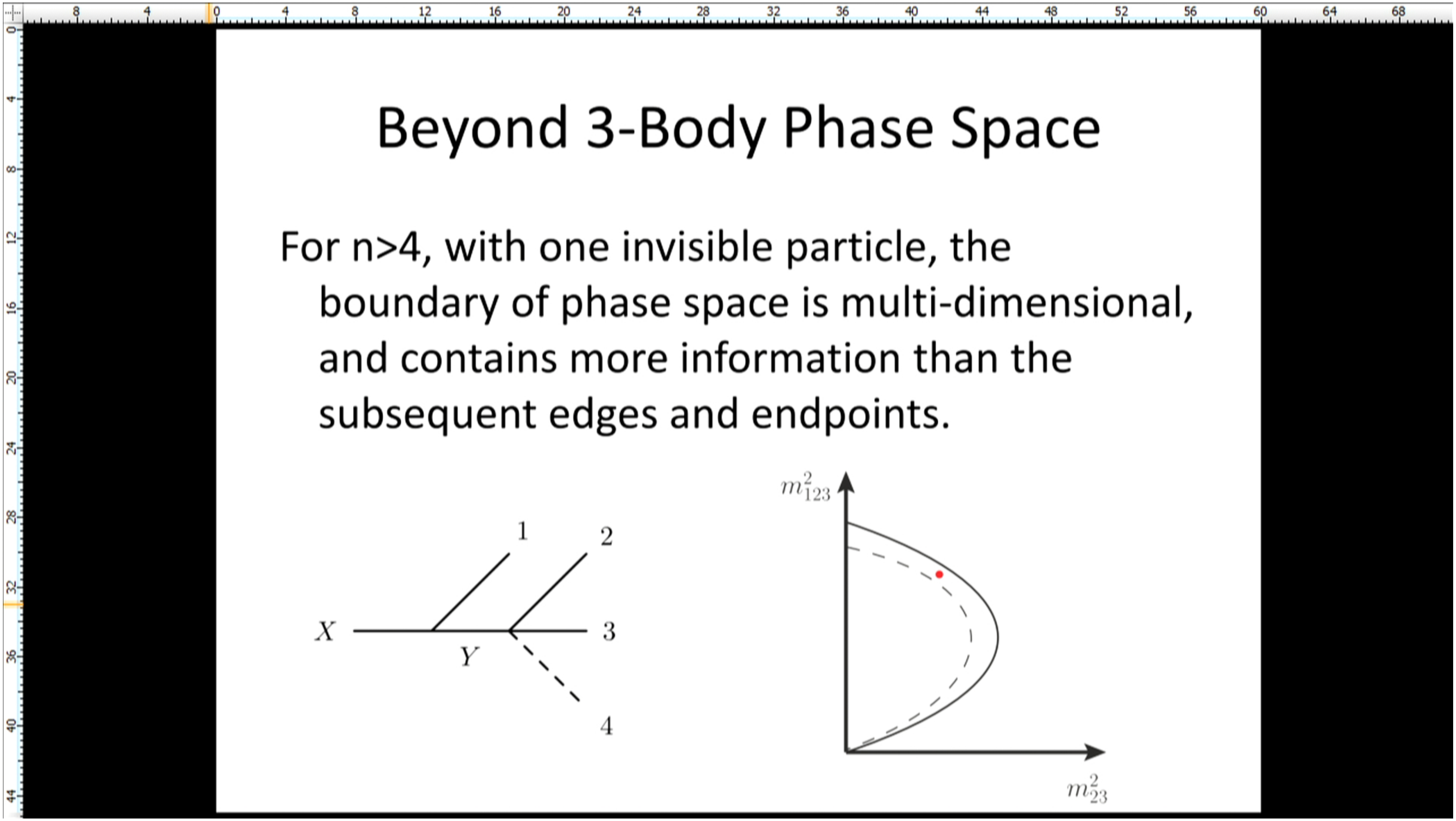


Factorization

Naively, each stage of the cascade proceeds independently – no phase space correlations.

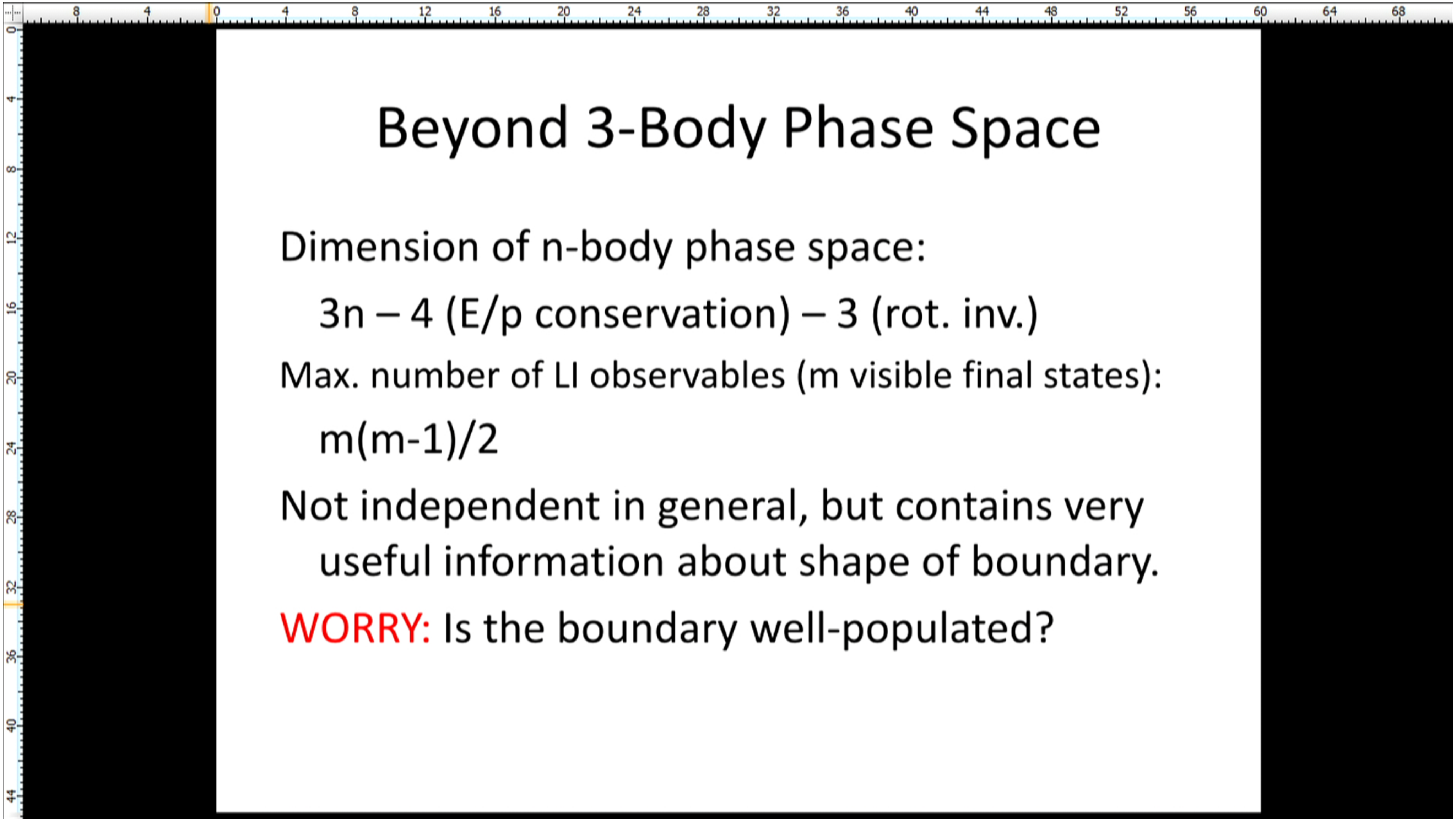
But this naïve argument fails because of two reasons:

- The two phase spaces are linked through the (unknown) m_X.
- There are observables m_{jk} where j and k come from the two different subspaces.



Beyond 3-Body Phase Space

For n>4, with one invisible particle, the boundary of phase space is multi-dimensional, and contains more information than the subsequent edges and endpoints.



Beyond 3-Body Phase Space

Dimension of n-body phase space:

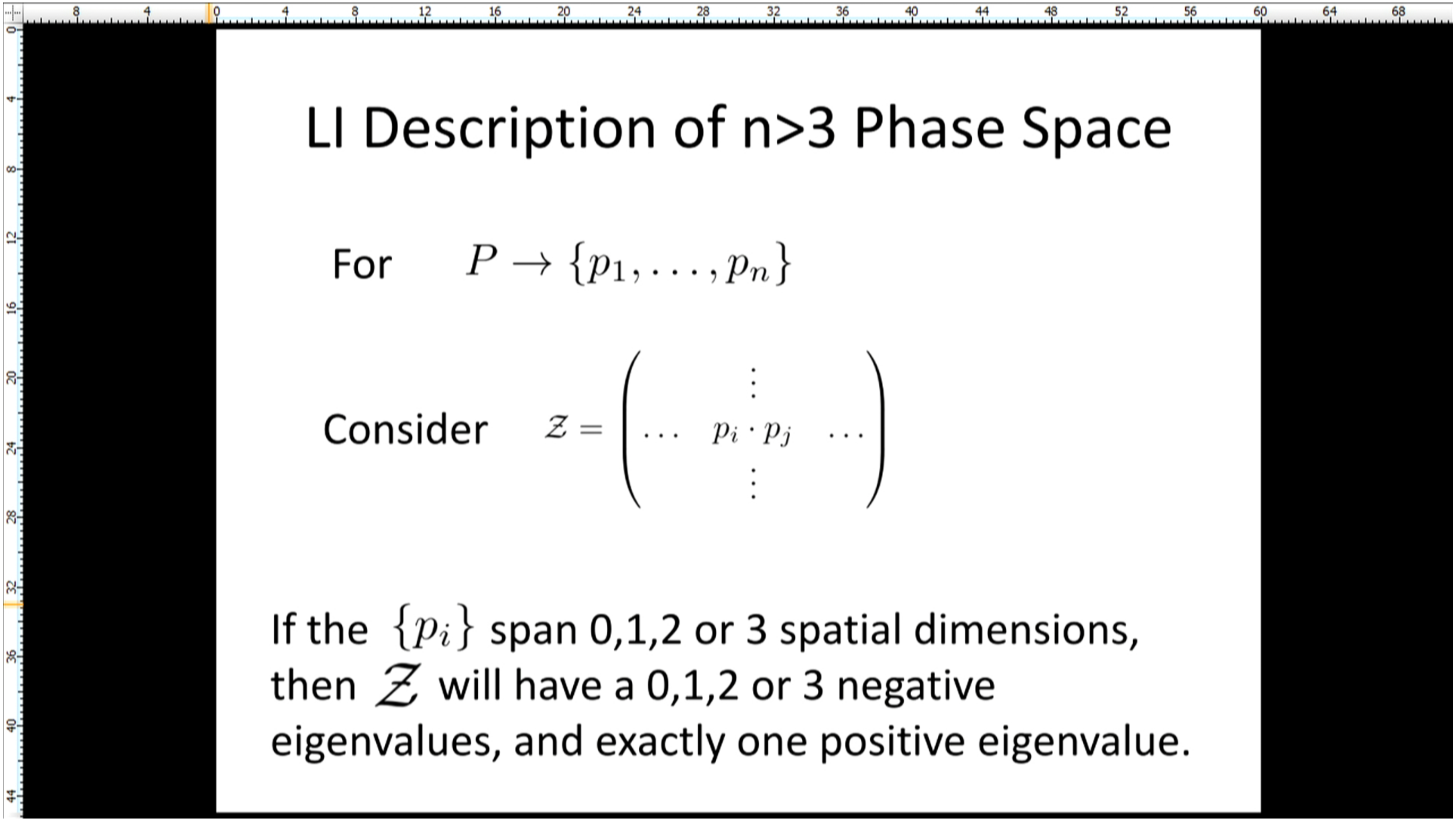
3n – 4 (E/p conservation) – 3 (rot. inv.)

Max. number of LI observables (m visible final states):

m(m-1)/2

Not independent in general, but contains very useful information about shape of boundary.

WORRY: Is the boundary well-populated?

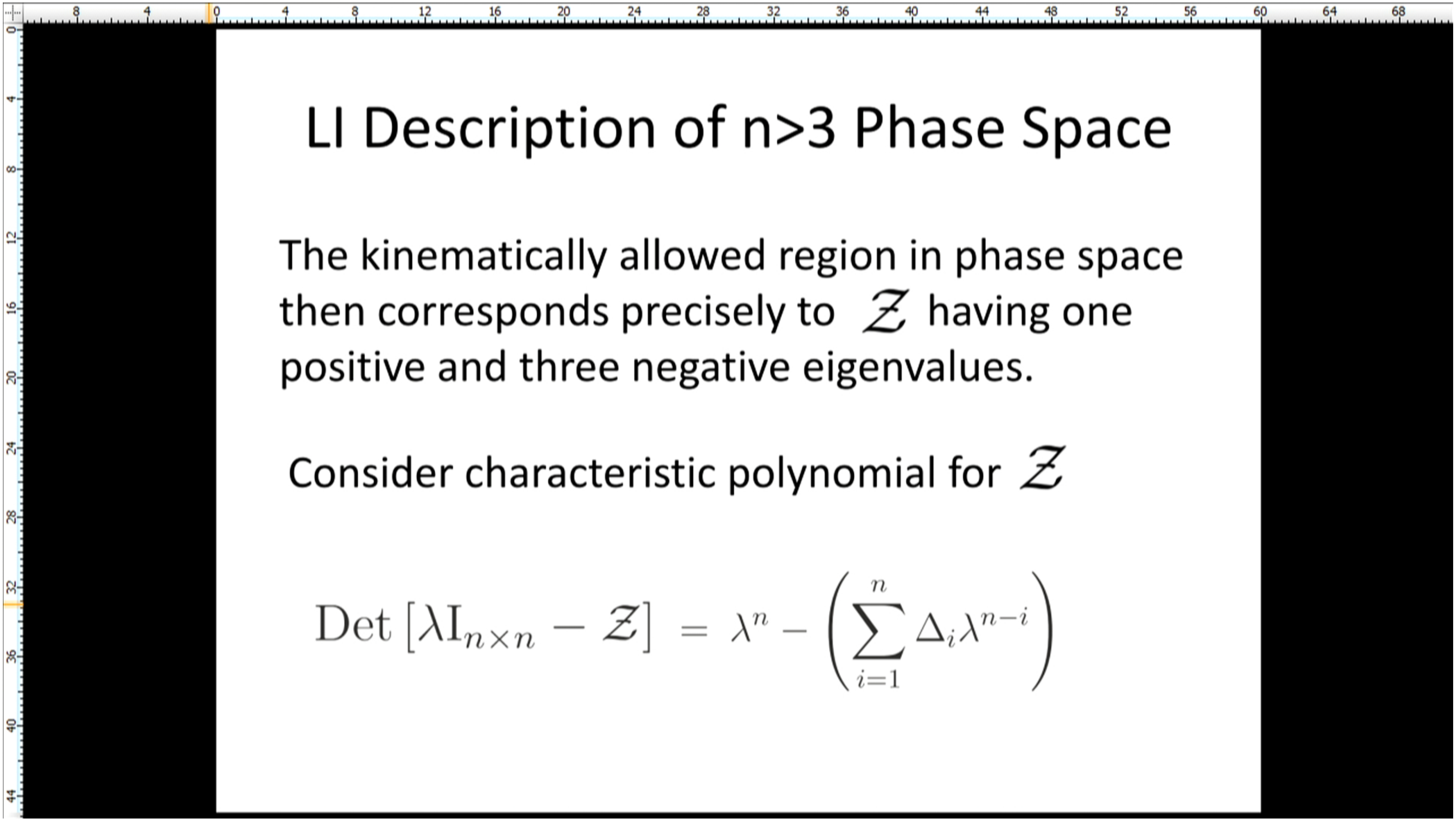


LI Description of n>3 Phase Space

For $P \rightarrow \{p_1, \ldots, p_n\}$

Consider $$\mathcal{Z} = \begin{pmatrix} & \vdots & \\ \cdots & p_i \cdot p_j & \cdots \\ & \vdots & \end{pmatrix}$$

If the $\{p_i\}$ span 0,1,2 or 3 spatial dimensions, then $\mathcal{Z}$ will have a 0,1,2 or 3 negative eigenvalues, and exactly one positive eigenvalue.

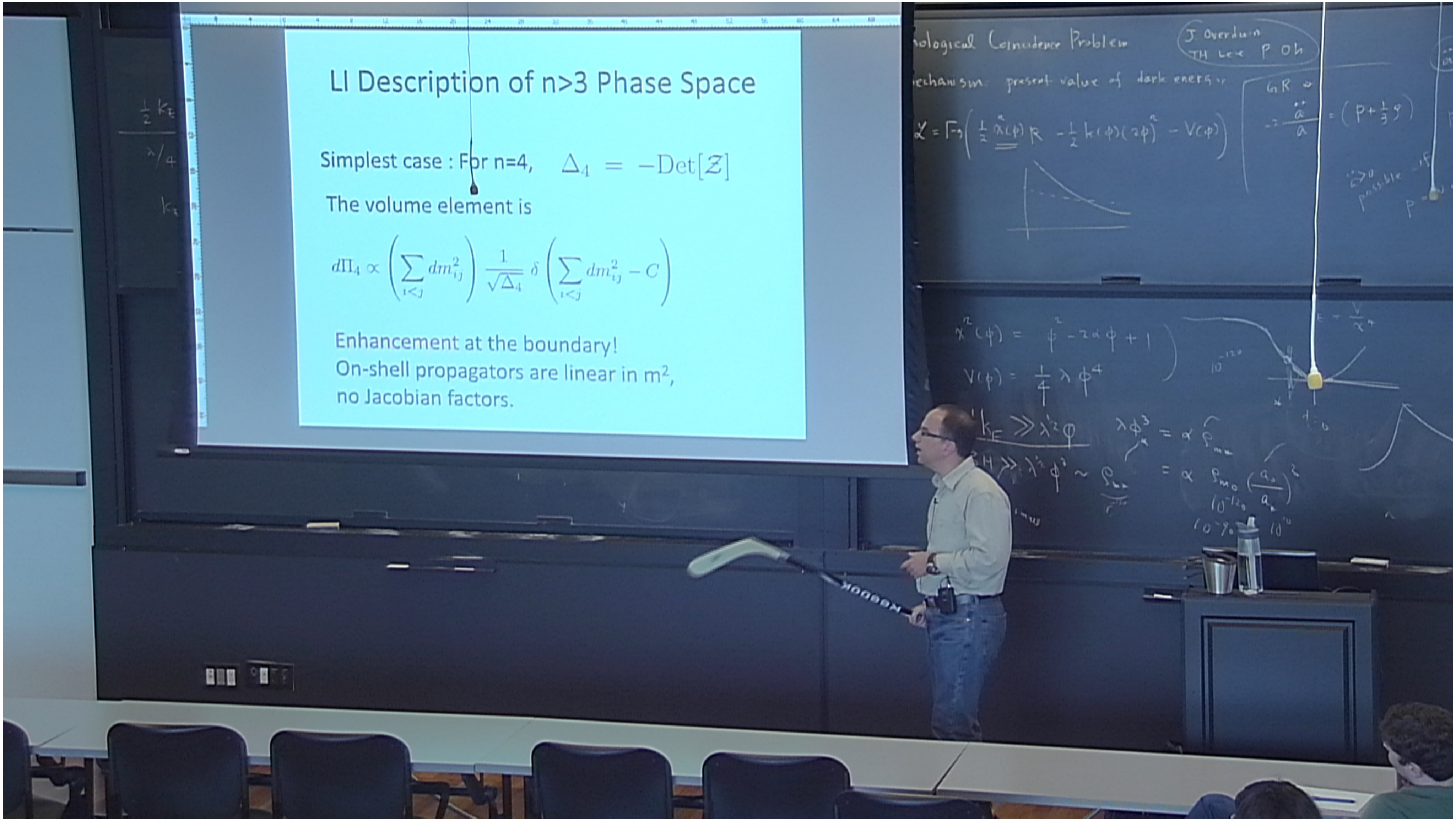


LI Description of n>3 Phase Space

The kinematically allowed region in phase space then corresponds precisely to $\mathcal{Z}$ having one positive and three negative eigenvalues.

Consider characteristic polynomial for $\mathcal{Z}$

$$\mathrm{Det}\left[\lambda \mathrm{I}_{n\times n} - \mathcal{Z}\right] = \lambda^n - \left(\sum_{i=1}^{n} \Delta_i \lambda^{n-i}\right)$$



LI Description of n>3 Phase Space

Simplest case : For n=4, $\Delta_4 = -\mathrm{Det}[\mathcal{Z}]$

The volume element is

$$d\Pi_4 \propto \left(\sum_{i<j} dm_{ij}^2\right) \frac{1}{\sqrt{\Delta_4}}\, \delta\left(\sum_{i<j} dm_{ij}^2 - C\right)$$

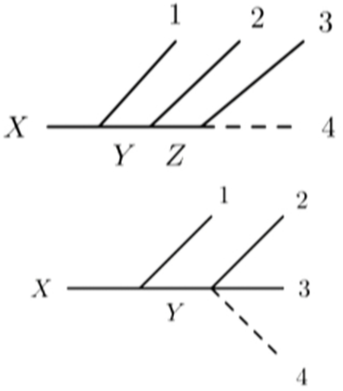
Enhancement at the boundary!
On-shell propagators are linear in m^2,
no Jacobian factors.

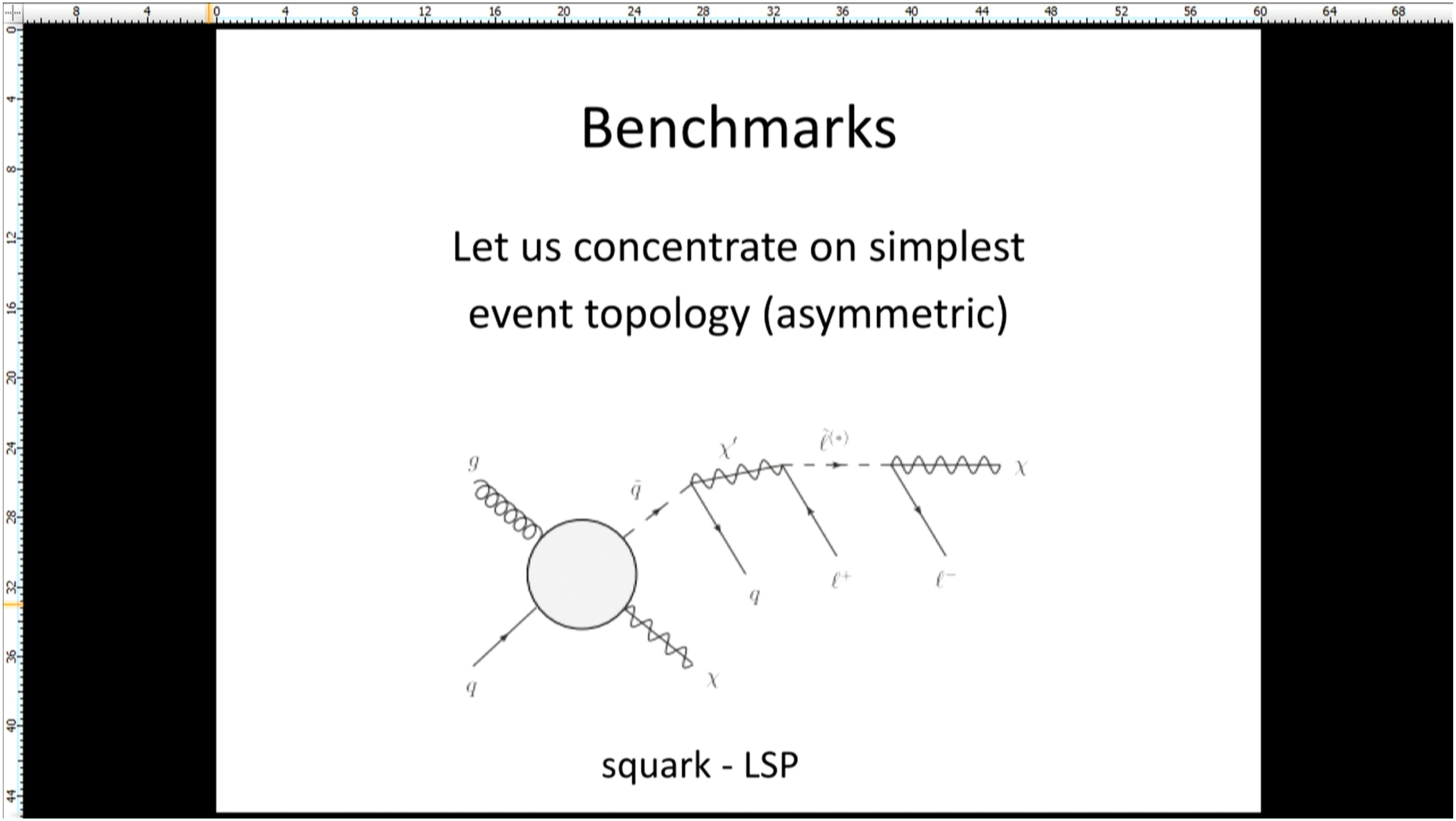
Classification of n=4 Cascades

- X→1+2+3+4 (no examples)
- X → (Y →1+2)+(Z →3+4)
  - 2,4 invisible (different topology)
  - only 4 invisible (secretly n=3)
- X →1+(Y →2+(Z →3+4))

  (solvable, but good warm-up)
- X →1+(Y →(2+3+4))

  (interesting! polynomial methods don't work.)
- X →1+2+(Y →3+4) (similar to above case)

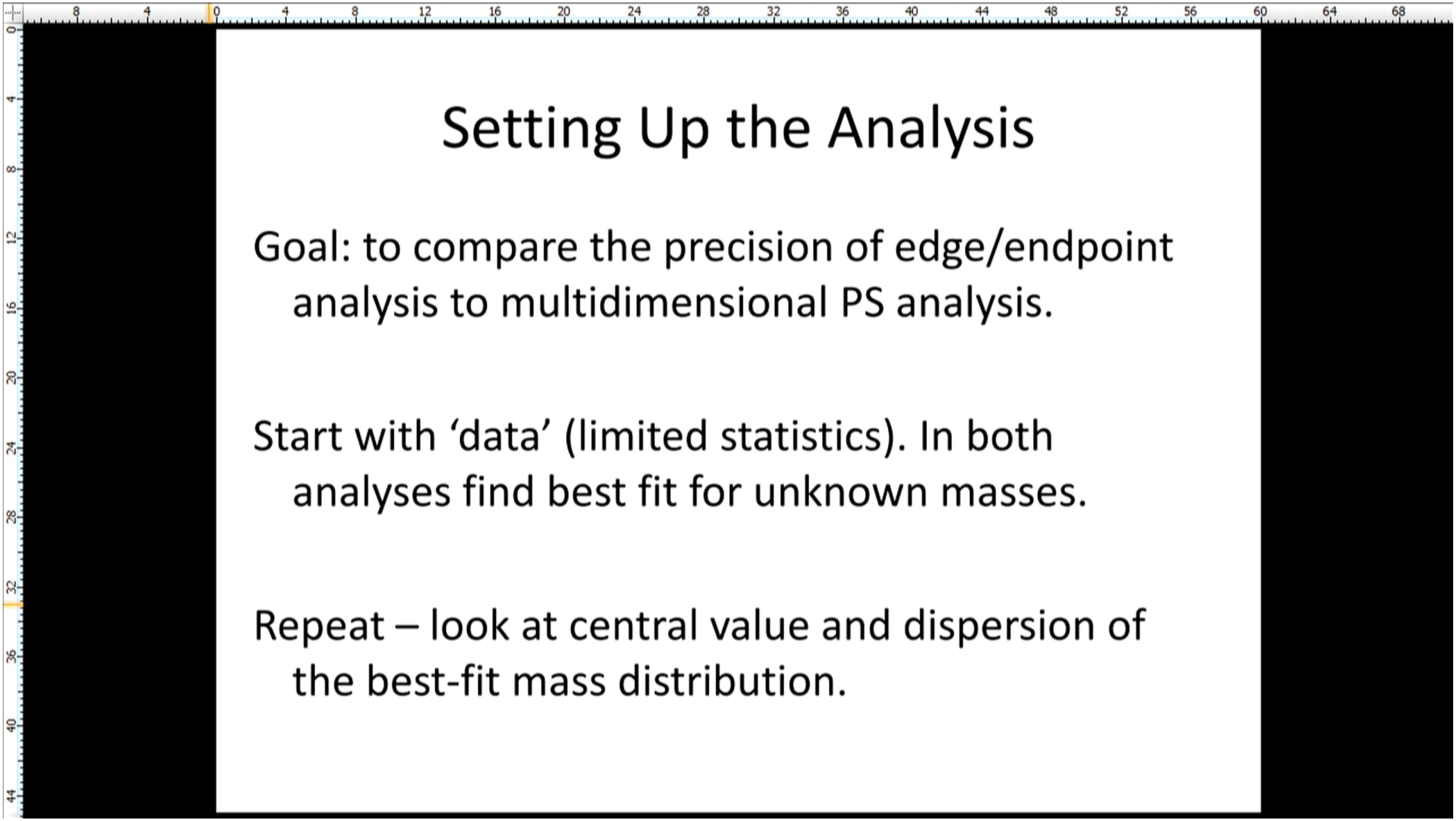




Benchmarks

Let us concentrate on simplest event topology (asymmetric)

squark - LSP

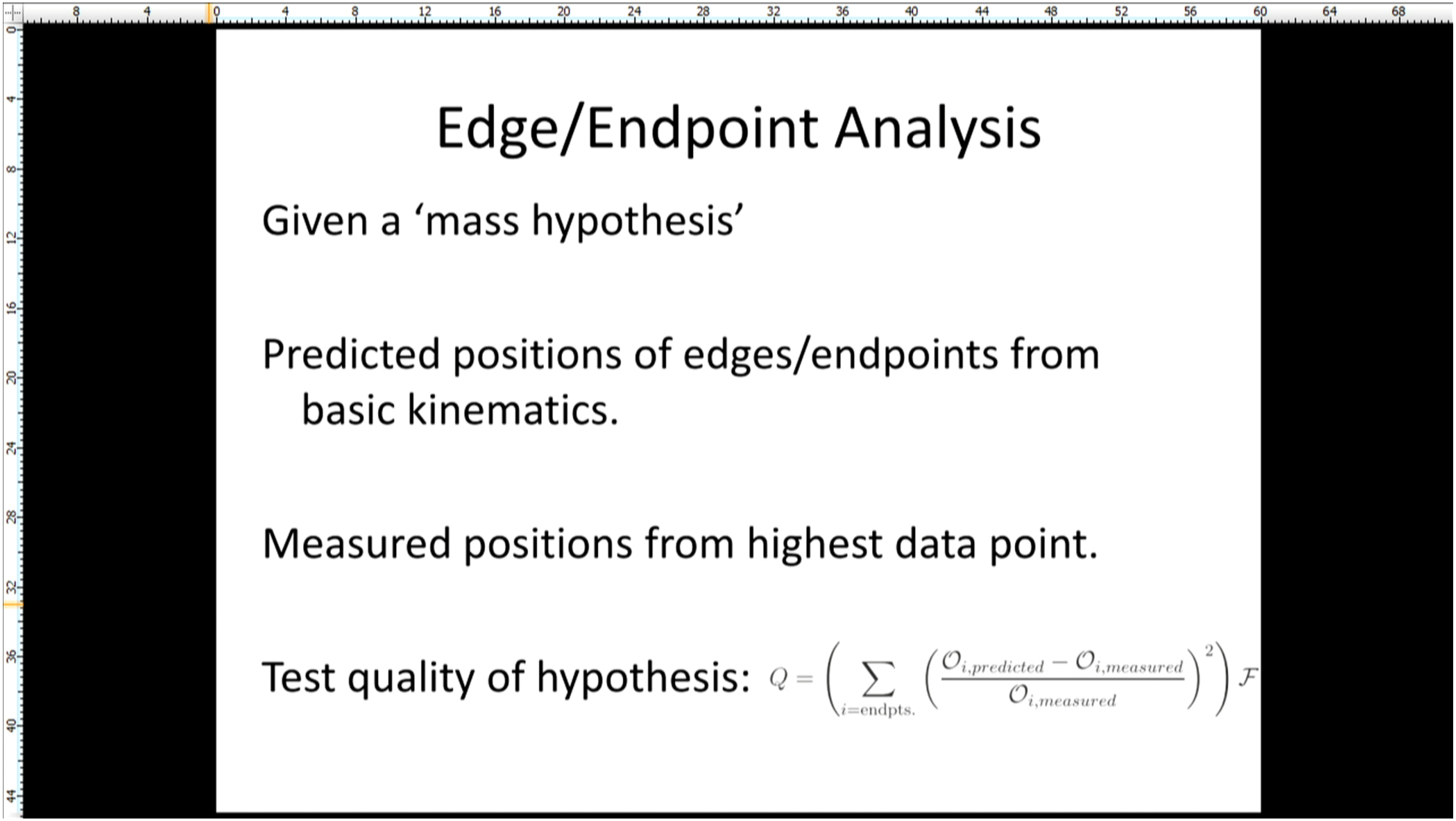


Setting Up the Analysis

Goal: to compare the precision of edge/endpoint analysis to multidimensional PS analysis.

Start with ‘data’ (limited statistics). In both analyses find best fit for unknown masses.

Repeat – look at central value and dispersion of the best-fit mass distribution.



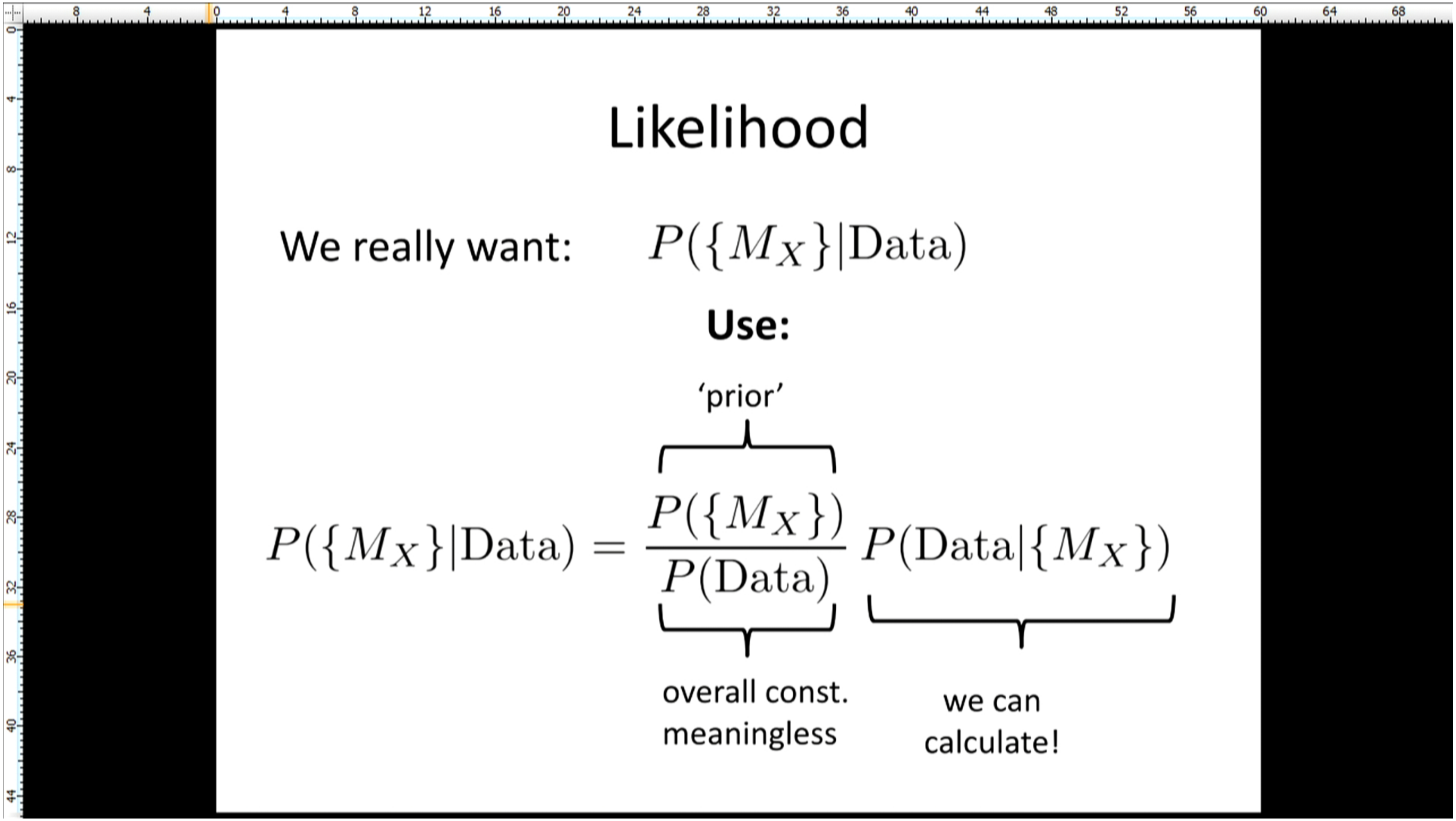
Edge/Endpoint Analysis

Given a ‘mass hypothesis’

Predicted positions of edges/endpoints from basic kinematics.

Measured positions from highest data point.

Test quality of hypothesis: $Q = \left(\sum_{i=\mathrm{endpts.}} \left(\frac{\mathcal{O}_{i,predicted} - \mathcal{O}_{i,measured}}{\mathcal{O}_{i,measured}} \right)^2 \right) \mathcal{F}$

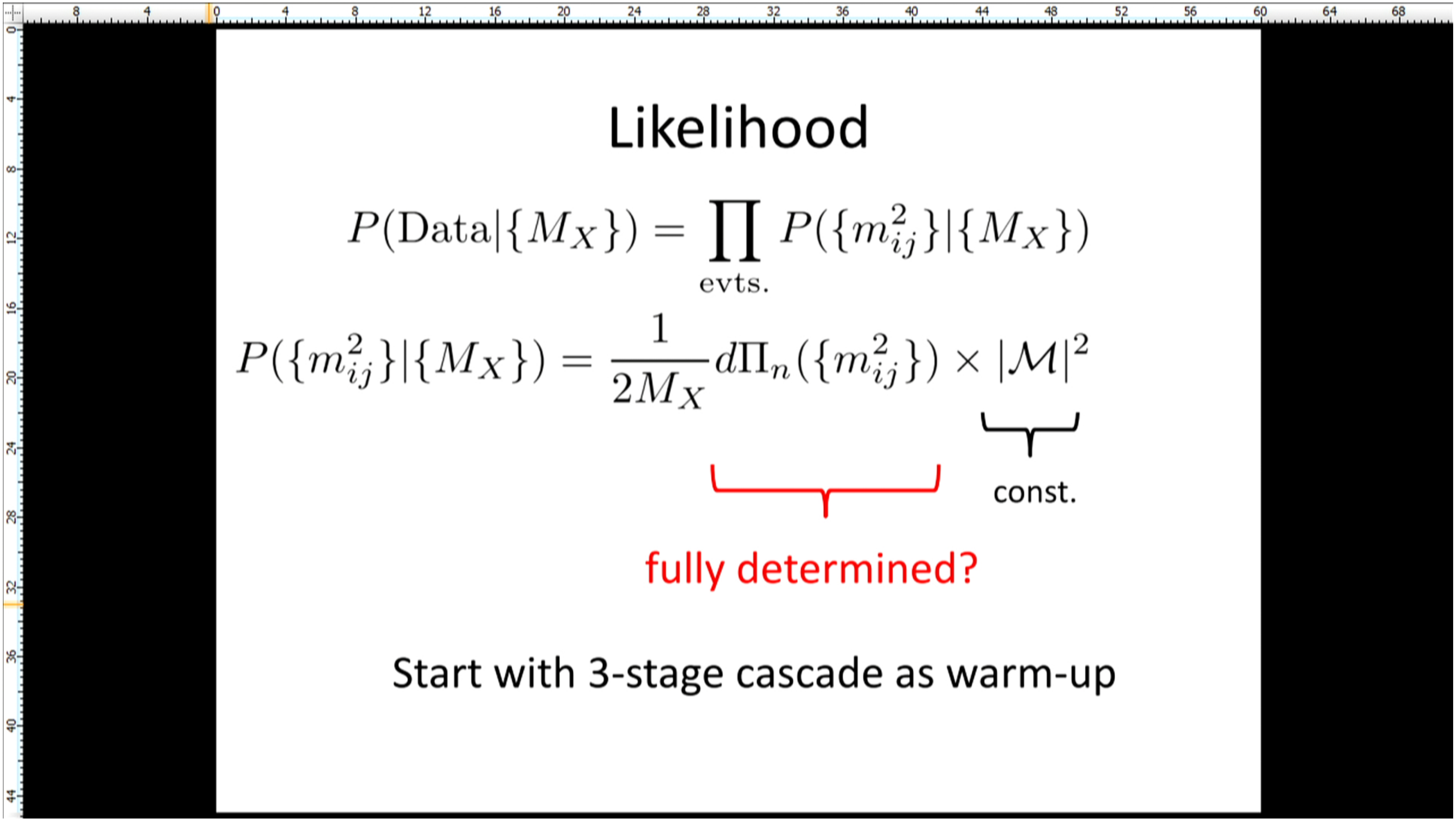


Likelihood

We really want: $P(\{M_X\}|\mathrm{Data})$

Use:

$$P(\{M_X\}|\mathrm{Data}) = \frac{\overbrace{P(\{M_X\})}^{\text{‘prior’}}}{\underbrace{P(\mathrm{Data})}_{\substack{\text{overall const.}\\ \text{meaningless}}}} \underbrace{P(\mathrm{Data}|\{M_X\})}_{\substack{\text{we can}\\ \text{calculate!}}}$$

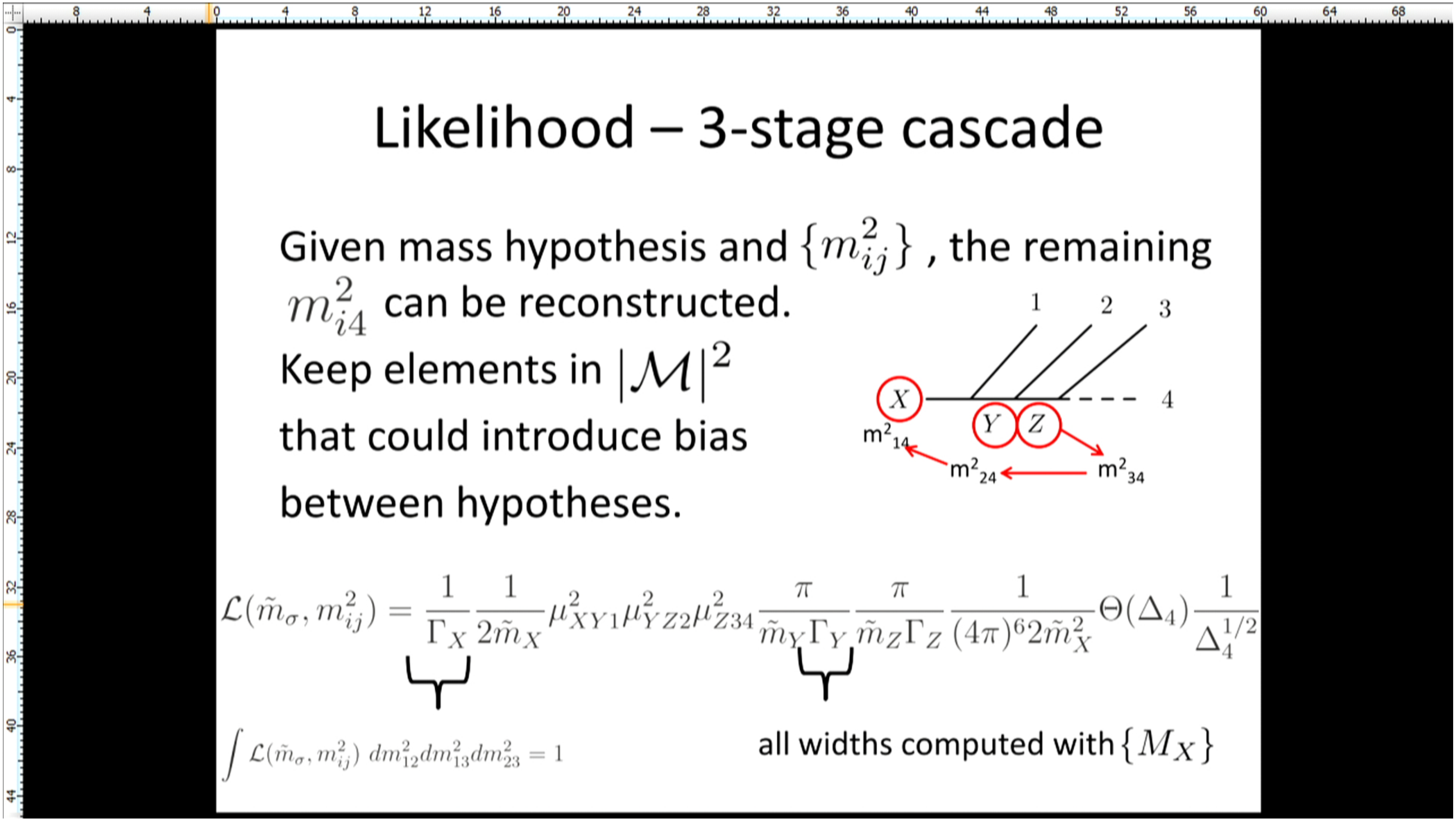


Likelihood

$$P(\mathrm{Data}|\{M_X\}) = \prod_{\mathrm{evts.}} P(\{m_{ij}^2\}|\{M_X\})$$

$$P(\{m_{ij}^2\}|\{M_X\}) = \frac{1}{2M_X} \underbrace{d\Pi_n(\{m_{ij}^2\})}_{\text{fully determined?}} \times \underbrace{|\mathcal{M}|^2}_{\text{const.}}$$

Start with 3-stage cascade as warm-up



Likelihood – 3-stage cascade

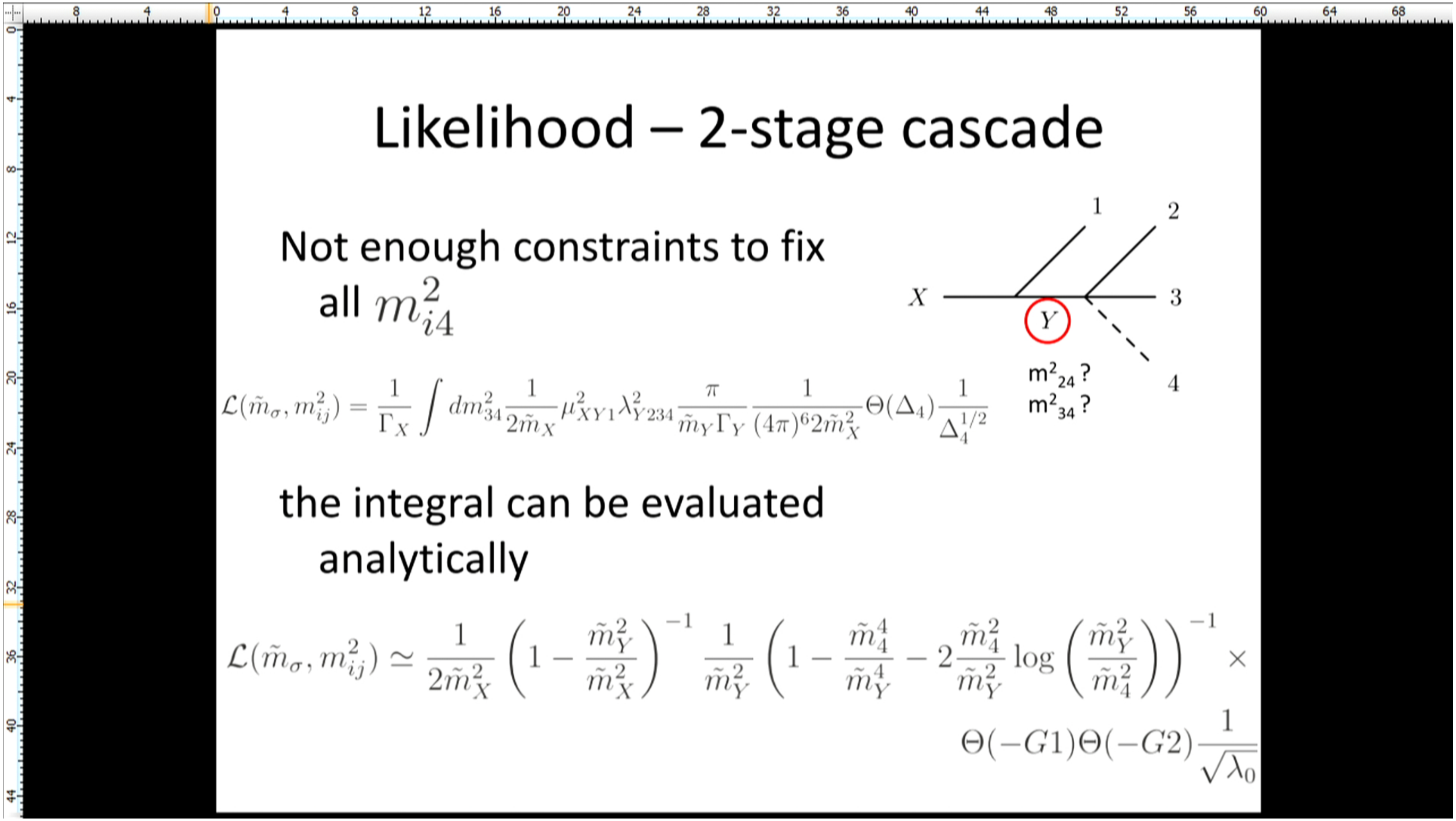
Given mass hypothesis and $\{m_{ij}^2\}$, the remaining m_{i4}^2 can be reconstructed.

Keep elements in $|\mathcal{M}|^2$ that could introduce bias between hypotheses.

$$\mathcal{L}(\tilde{m}_\sigma, m_{ij}^2) = \underbrace{\frac{1}{\Gamma_X}}_{} \frac{1}{2\tilde{m}_X}\mu^2_{XY1}\mu^2_{YZ2}\mu^2_{Z34}\frac{\pi}{\underbrace{\tilde{m}_Y\Gamma_Y}_{}}\frac{\pi}{\tilde{m}_Z\Gamma_Z}\frac{1}{(4\pi)^6 2\tilde{m}_X^2}\Theta(\Delta_4)\frac{1}{\Delta_4^{1/2}}$$

$$\int \mathcal{L}(\tilde{m}_\sigma, m_{ij}^2)\; dm_{12}^2 dm_{13}^2 dm_{23}^2 = 1$$

all widths computed with $\{M_X\}$



Likelihood – 2-stage cascade

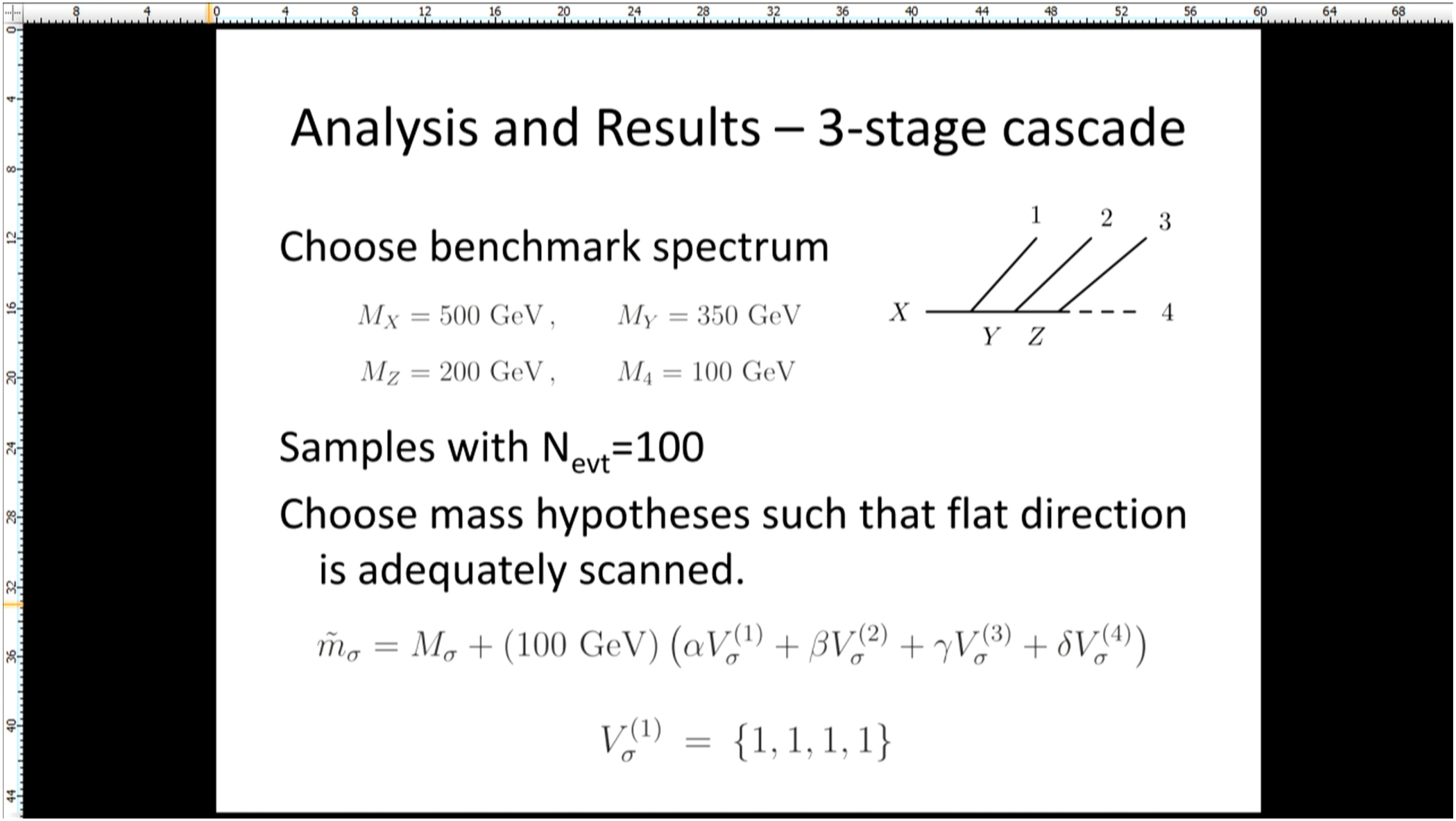
Not enough constraints to fix all m_{i4}^2

$$\mathcal{L}(\tilde{m}_\sigma, m_{ij}^2) = \frac{1}{\Gamma_X}\int dm_{34}^2 \frac{1}{2\tilde{m}_X}\mu_{XY1}^2\lambda_{Y234}^2\frac{\pi}{\tilde{m}_Y\Gamma_Y}\frac{1}{(4\pi)^6 2\tilde{m}_X^2}\Theta(\Delta_4)\frac{1}{\Delta_4^{1/2}}$$

the integral can be evaluated analytically

$$\mathcal{L}(\tilde{m}_\sigma, m_{ij}^2) \simeq \frac{1}{2\tilde{m}_X^2}\left(1-\frac{\tilde{m}_Y^2}{\tilde{m}_X^2}\right)^{-1}\frac{1}{\tilde{m}_Y^2}\left(1-\frac{\tilde{m}_4^4}{\tilde{m}_Y^4}-2\frac{\tilde{m}_4^2}{\tilde{m}_Y^2}\log\left(\frac{\tilde{m}_Y^2}{\tilde{m}_4^2}\right)\right)^{-1}\times$$

$$\Theta(-G1)\Theta(-G2)\frac{1}{\sqrt{\lambda_0}}$$



Analysis and Results – 3-stage cascade

Choose benchmark spectrum

$$M_X = 500 \text{ GeV}\,, \qquad M_Y = 350 \text{ GeV}$$

$$M_Z = 200 \text{ GeV}\,, \qquad M_4 = 100 \text{ GeV}$$

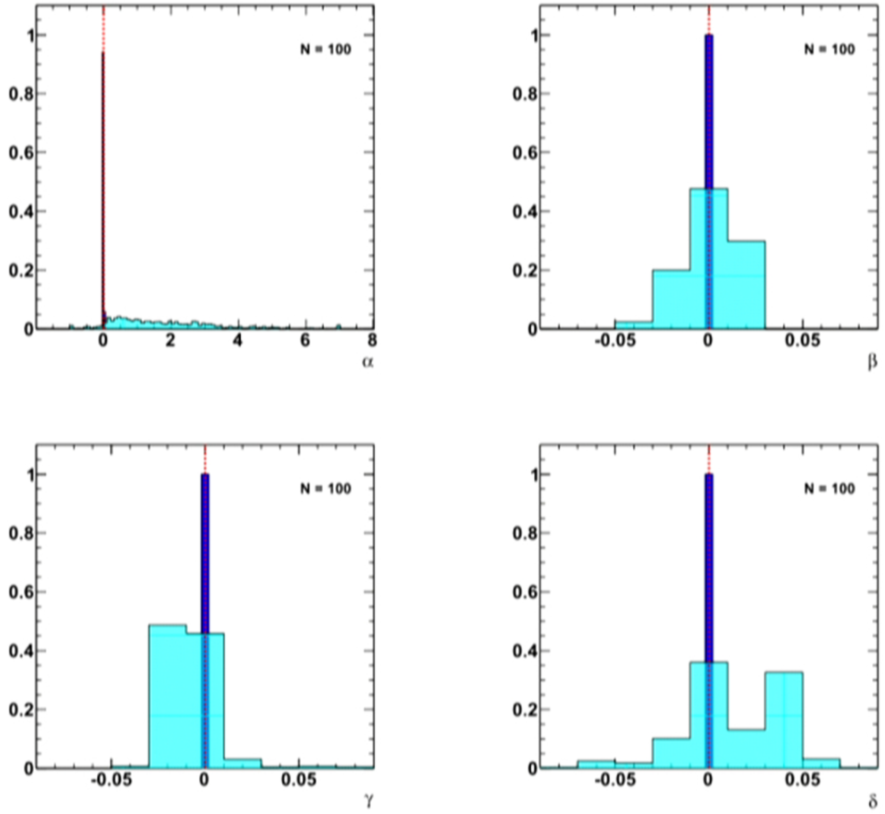
Samples with N_{evt}=100

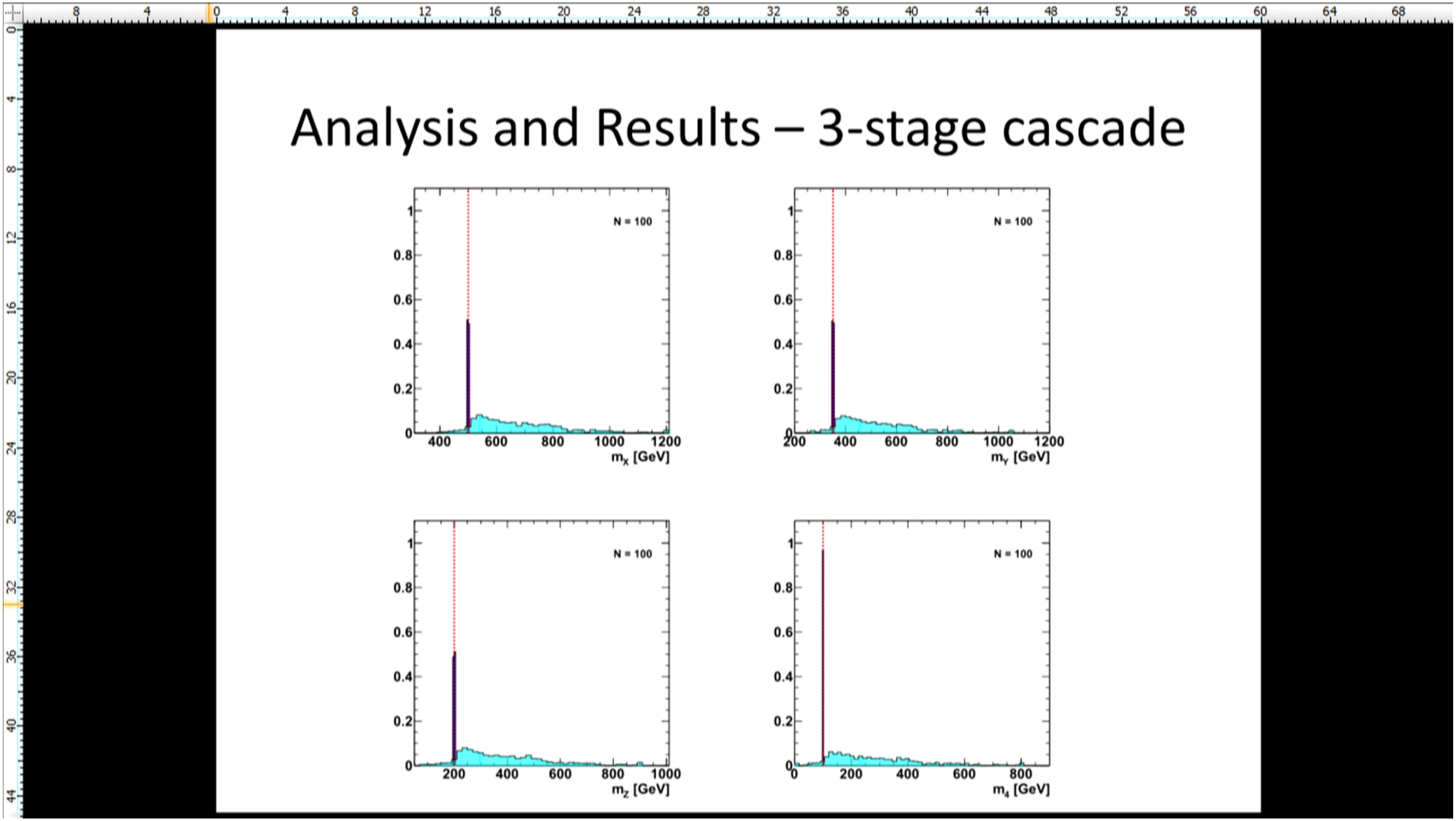
Choose mass hypotheses such that flat direction is adequately scanned.

$$\tilde{m}_\sigma = M_\sigma + (100 \text{ GeV}) \left(\alpha V_\sigma^{(1)} + \beta V_\sigma^{(2)} + \gamma V_\sigma^{(3)} + \delta V_\sigma^{(4)} \right)$$

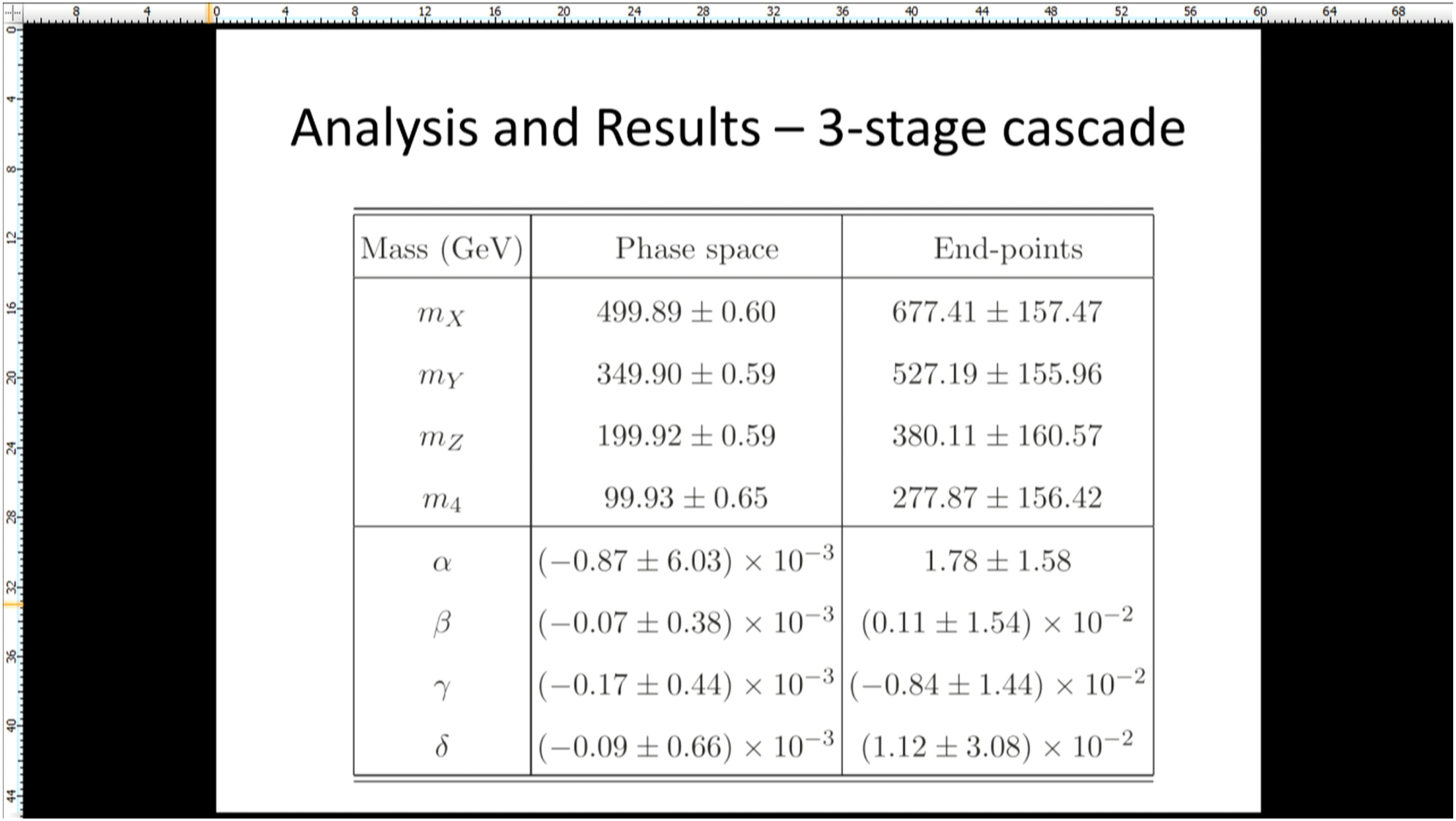
$$V_\sigma^{(1)} \; = \; \{1, 1, 1, 1\}$$

Analysis and Results – 3-stage cascade



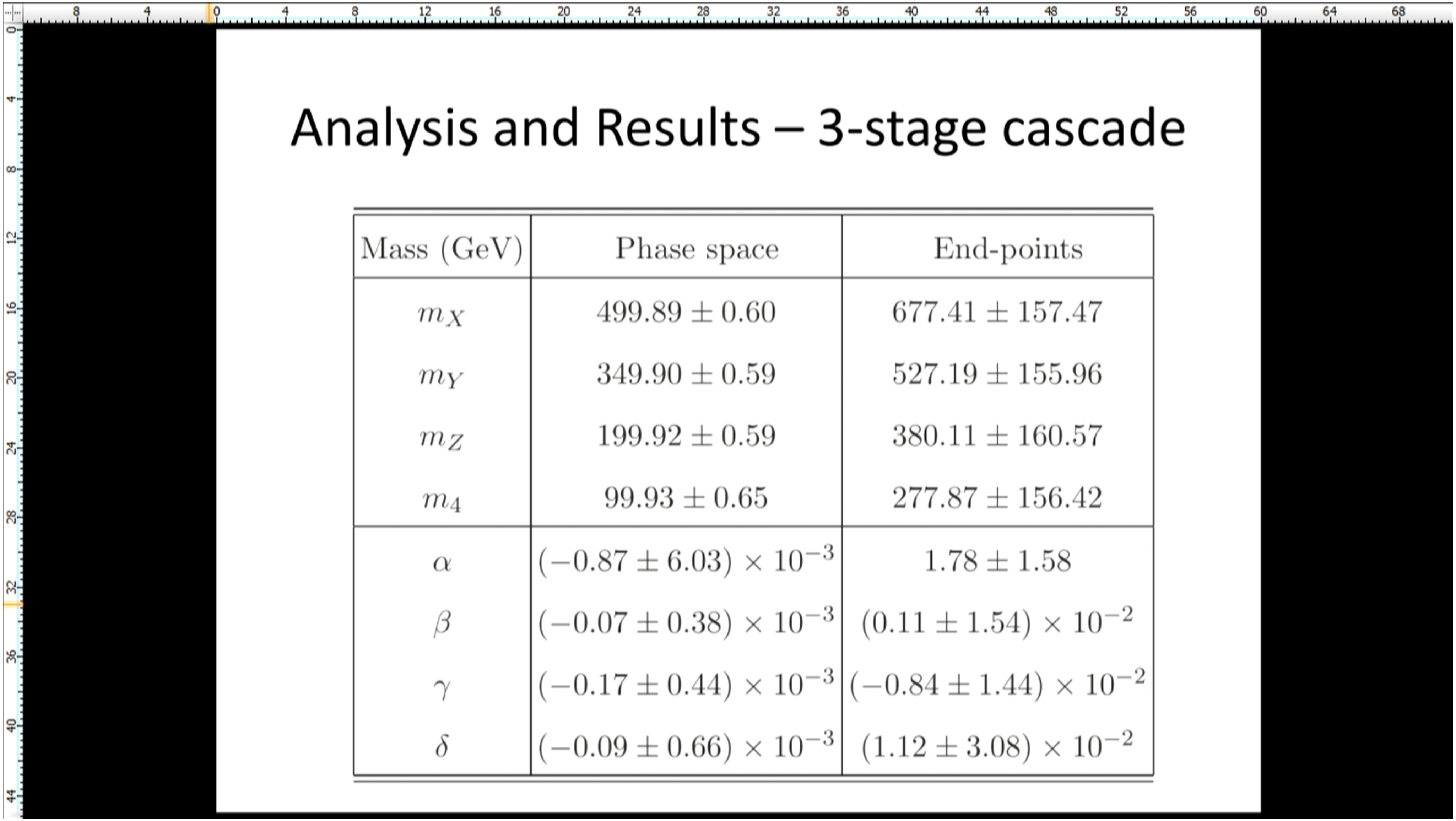


Analysis and Results – 3-stage cascade



Analysis and Results – 3-stage cascade

Mass (GeV)	Phase space	End-points
m_X	499.89 ± 0.60	677.41 ± 157.47
m_Y	349.90 ± 0.59	527.19 ± 155.96
m_Z	199.92 ± 0.59	380.11 ± 160.57
m_4	99.93 ± 0.65	277.87 ± 156.42
α	$(-0.87 \pm 6.03) \times 10^{-3}$	1.78 ± 1.58
β	$(-0.07 \pm 0.38) \times 10^{-3}$	$(0.11 \pm 1.54) \times 10^{-2}$
γ	$(-0.17 \pm 0.44) \times 10^{-3}$	$(-0.84 \pm 1.44) \times 10^{-2}$
δ	$(-0.09 \pm 0.66) \times 10^{-3}$	$(1.12 \pm 3.08) \times 10^{-2}$

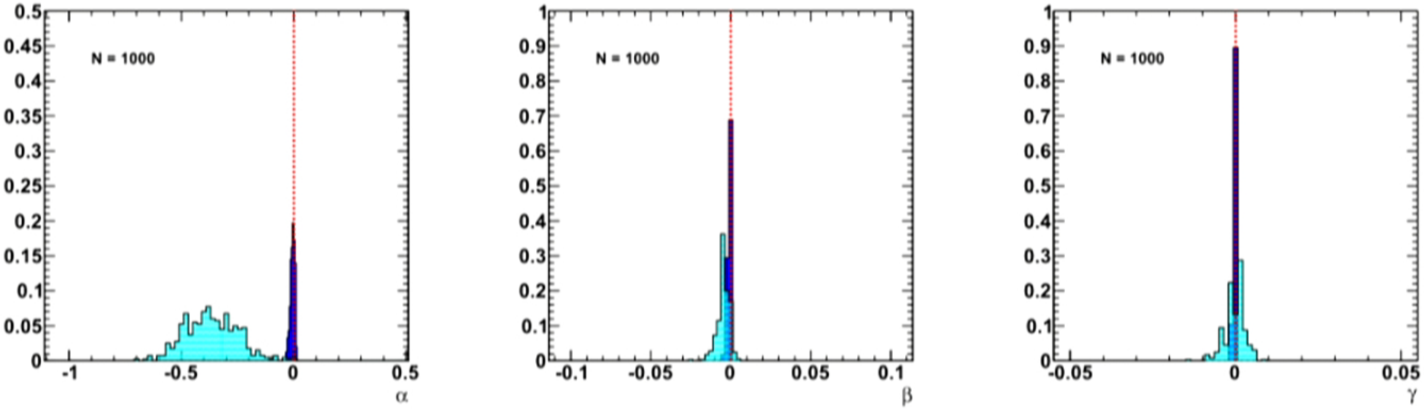


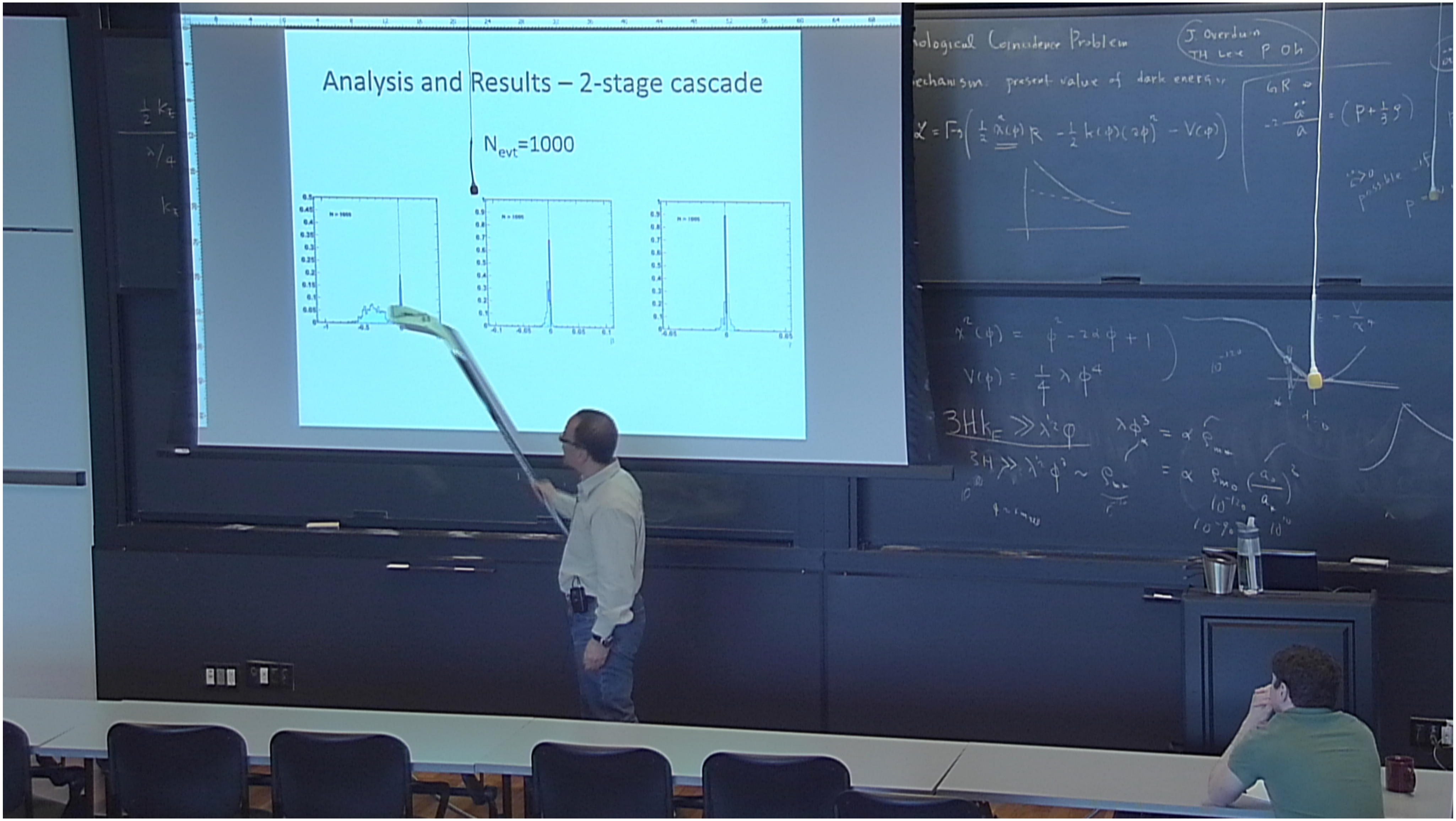
Analysis and Results – 3-stage cascade

Mass (GeV)	Phase space	End-points
m_X	499.89 ± 0.60	677.41 ± 157.47
m_Y	349.90 ± 0.59	527.19 ± 155.96
m_Z	199.92 ± 0.59	380.11 ± 160.57
m_4	99.93 ± 0.65	277.87 ± 156.42
α	$(-0.87 \pm 6.03) \times 10^{-3}$	1.78 ± 1.58
β	$(-0.07 \pm 0.38) \times 10^{-3}$	$(0.11 \pm 1.54) \times 10^{-2}$
γ	$(-0.17 \pm 0.44) \times 10^{-3}$	$(-0.84 \pm 1.44) \times 10^{-2}$
δ	$(-0.09 \pm 0.66) \times 10^{-3}$	$(1.12 \pm 3.08) \times 10^{-2}$

Analysis and Results – 2-stage cascade

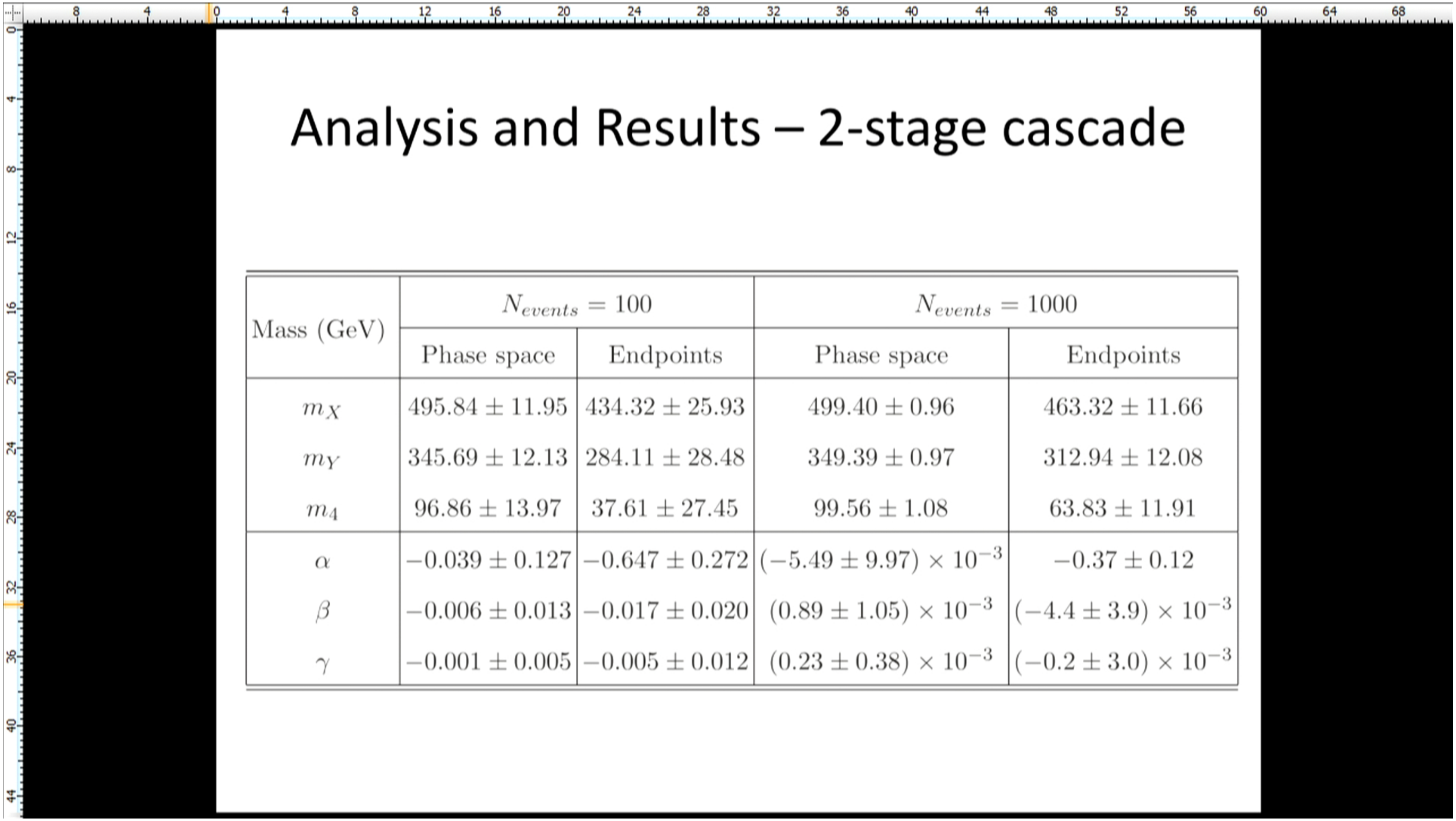
N_{evt}=1000





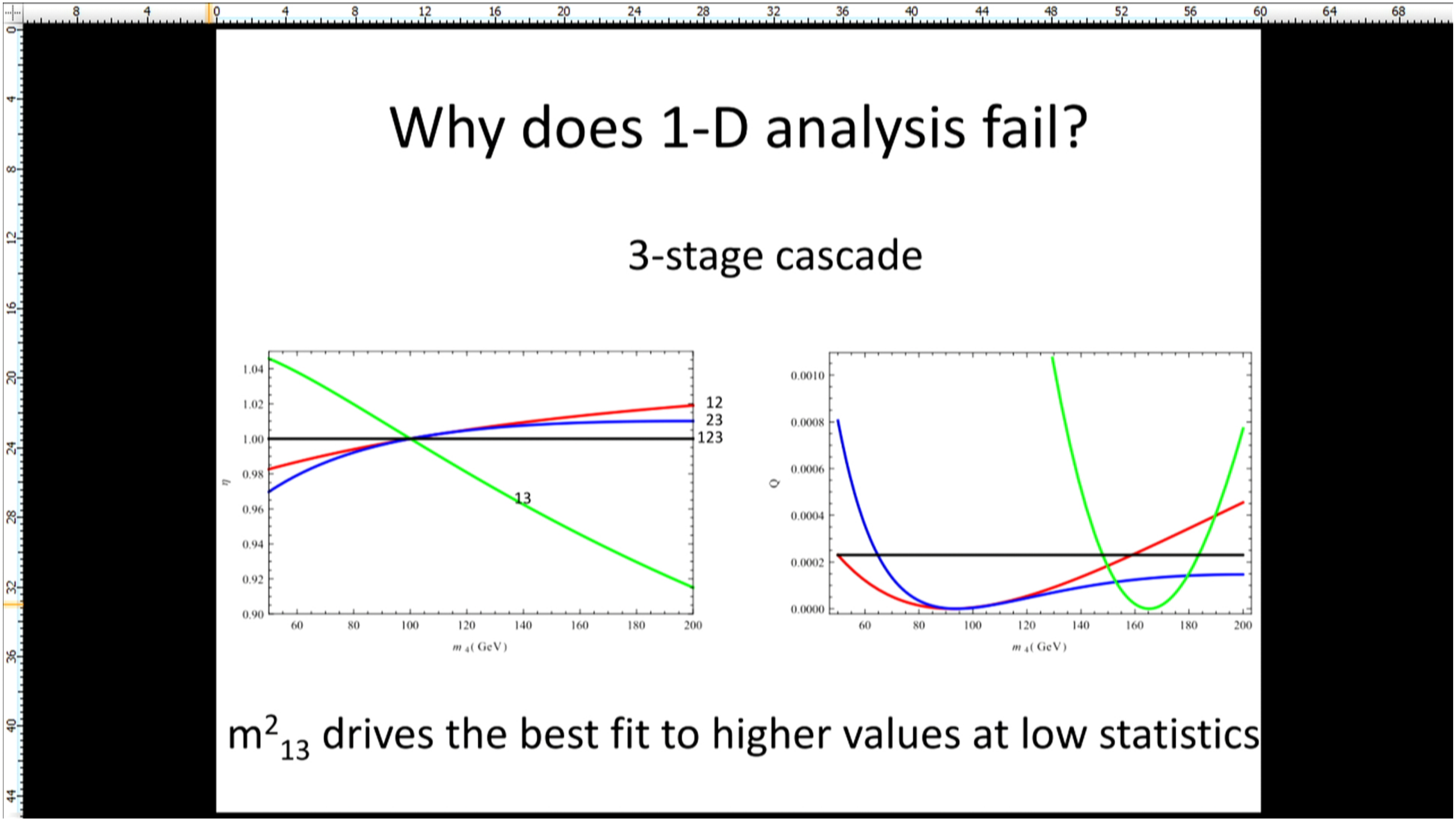
Analysis and Results – 2-stage cascade

N_{evt}=1000



Analysis and Results – 2-stage cascade

Mass (GeV)	$N_{events} = 100$		$N_{events} = 1000$	
	Phase space	Endpoints	Phase space	Endpoints
m_X	495.84 ± 11.95	434.32 ± 25.93	499.40 ± 0.96	463.32 ± 11.66
m_Y	345.69 ± 12.13	284.11 ± 28.48	349.39 ± 0.97	312.94 ± 12.08
m_4	96.86 ± 13.97	37.61 ± 27.45	99.56 ± 1.08	63.83 ± 11.91
α	-0.039 ± 0.127	-0.647 ± 0.272	$(-5.49 \pm 9.97) \times 10^{-3}$	-0.37 ± 0.12
β	-0.006 ± 0.013	-0.017 ± 0.020	$(0.89 \pm 1.05) \times 10^{-3}$	$(-4.4 \pm 3.9) \times 10^{-3}$
γ	-0.001 ± 0.005	-0.005 ± 0.012	$(0.23 \pm 0.38) \times 10^{-3}$	$(-0.2 \pm 3.0) \times 10^{-3}$



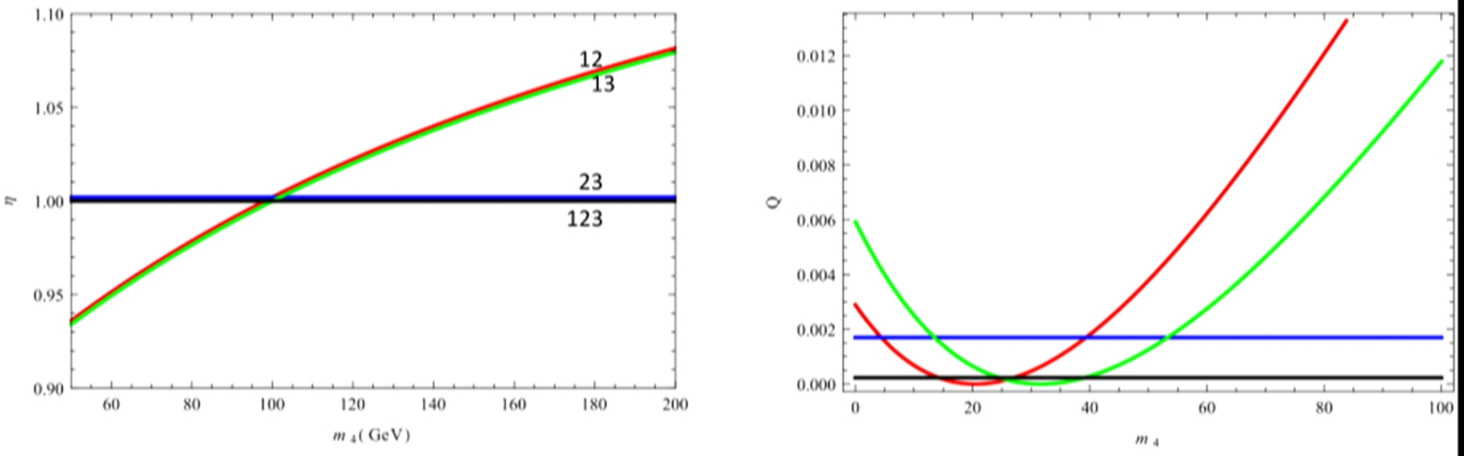
Why does 1-D analysis fail?

3-stage cascade

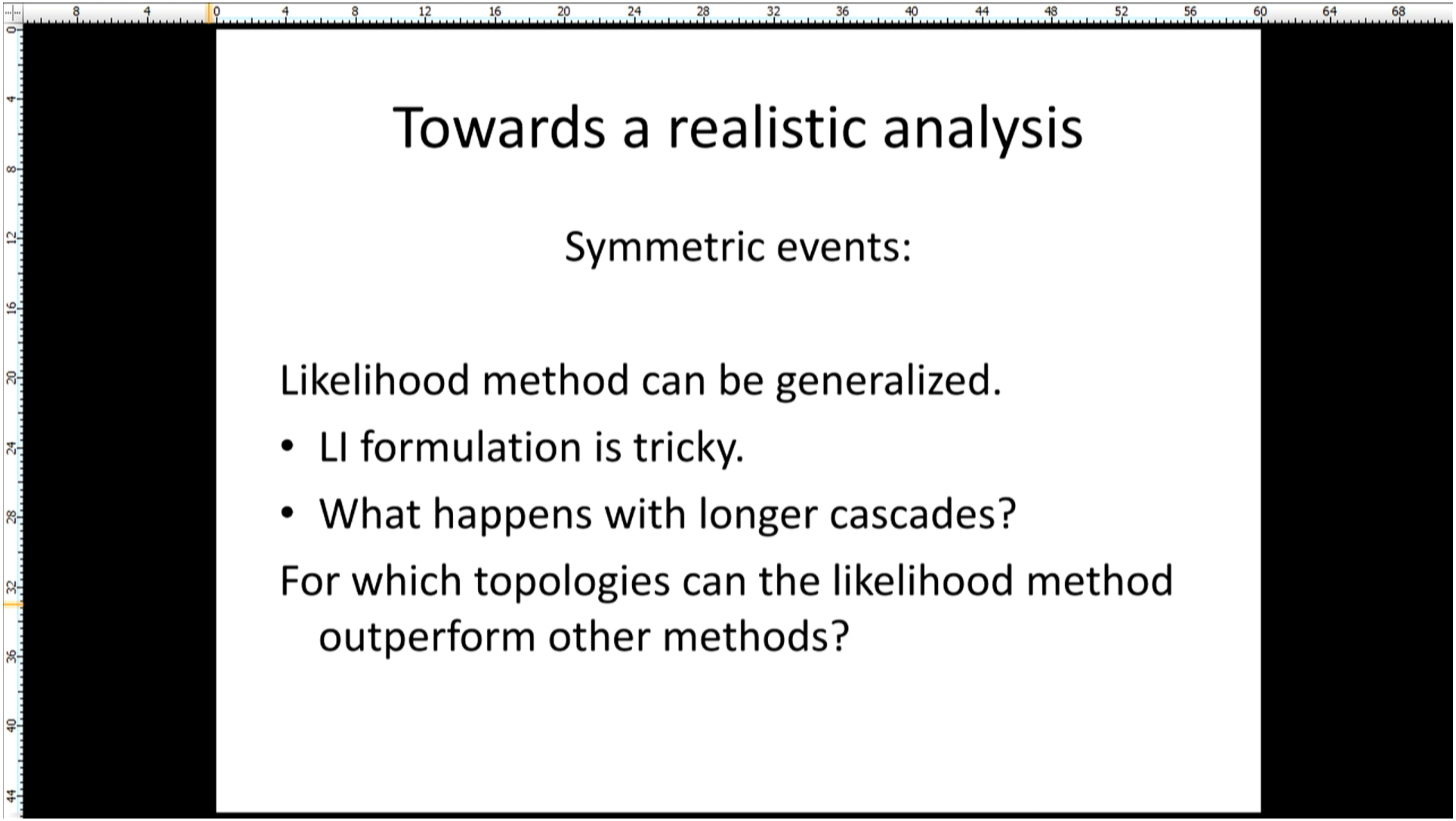
m^2_{13} drives the best fit to higher values at low statistics

Why does 1-D analysis fail?

2-stage cascade



m^2_{13} and m^2_{12} drive the best fit to lower values



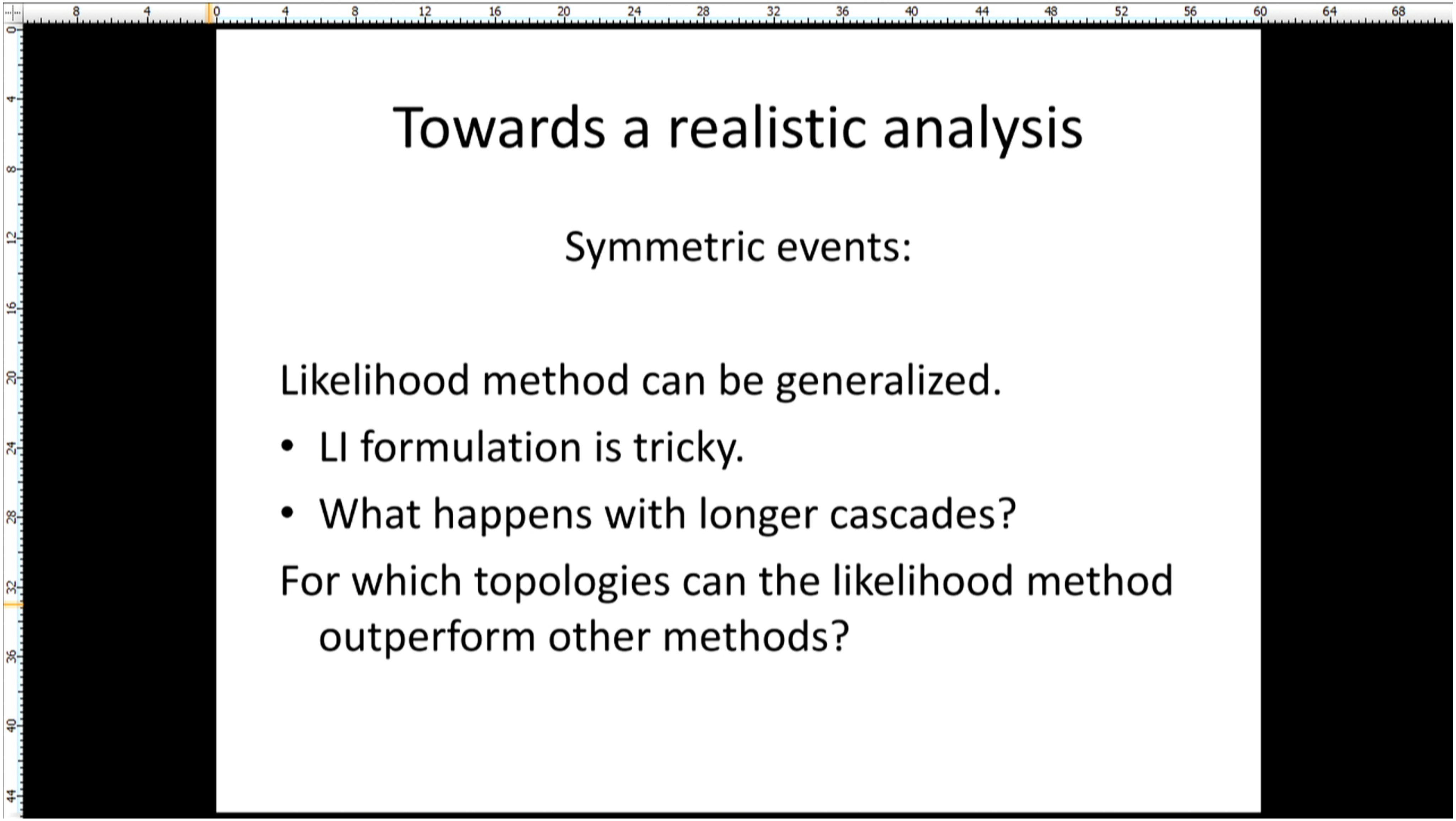
Towards a realistic analysis

Symmetric events:

Likelihood method can be generalized.

- LI formulation is tricky.
- What happens with longer cascades?

For which topologies can the likelihood method outperform other methods?



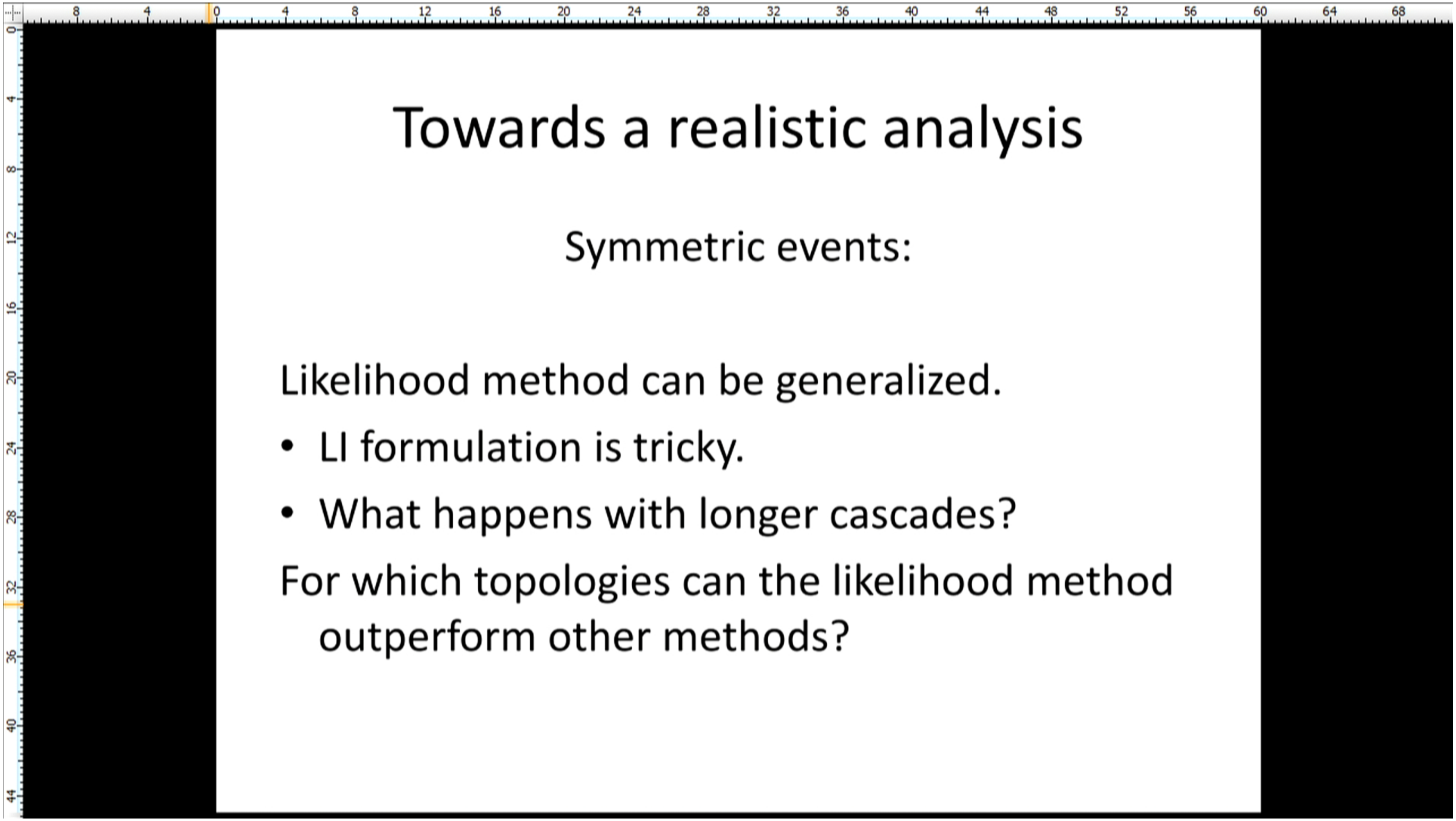
Towards a realistic analysis

Symmetric events:

Likelihood method can be generalized.

- LI formulation is tricky.
- What happens with longer cascades?

For which topologies can the likelihood method outperform other methods?



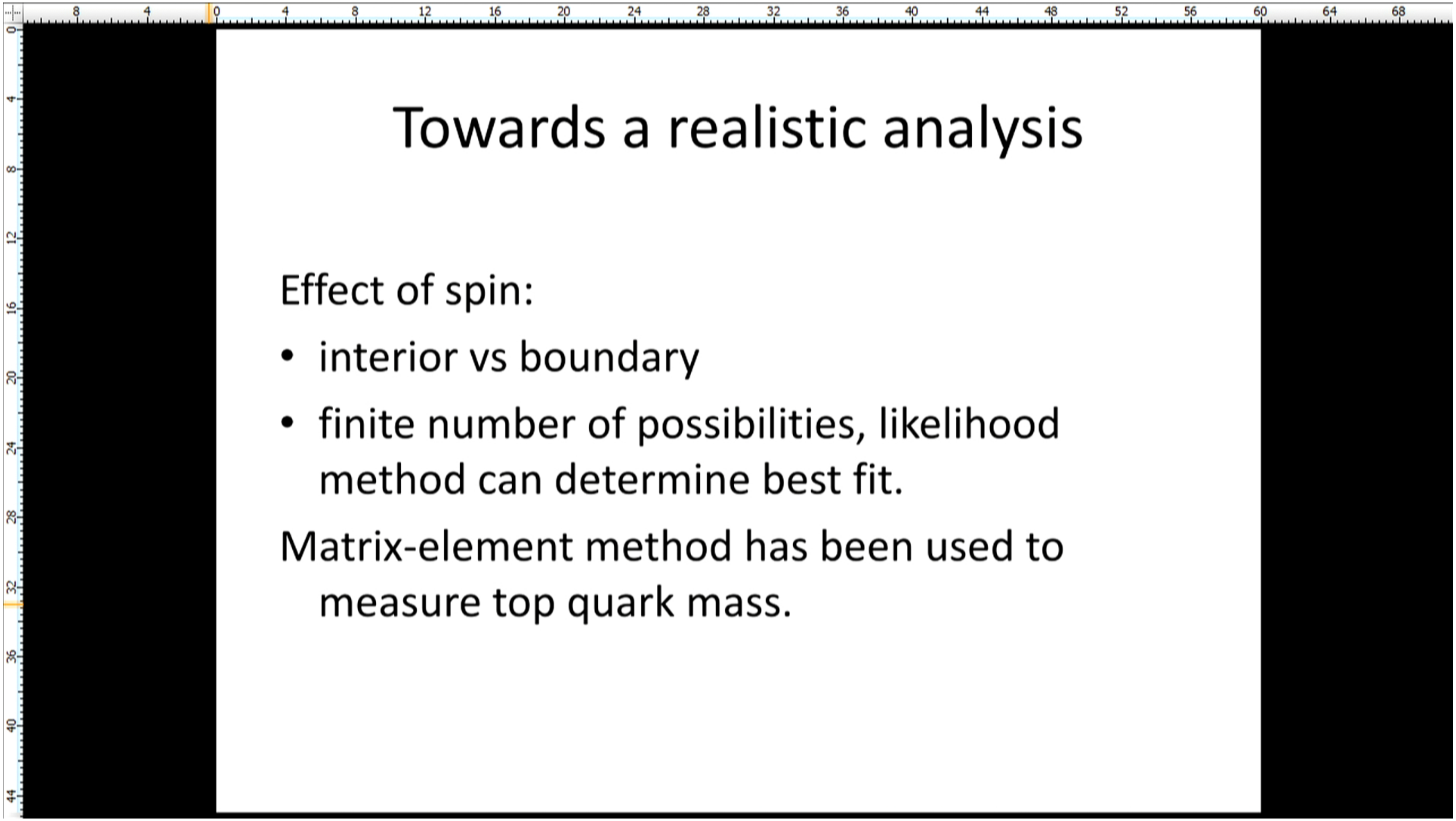
Towards a realistic analysis

Symmetric events:

Likelihood method can be generalized.

- LI formulation is tricky.
- What happens with longer cascades?

For which topologies can the likelihood method outperform other methods?

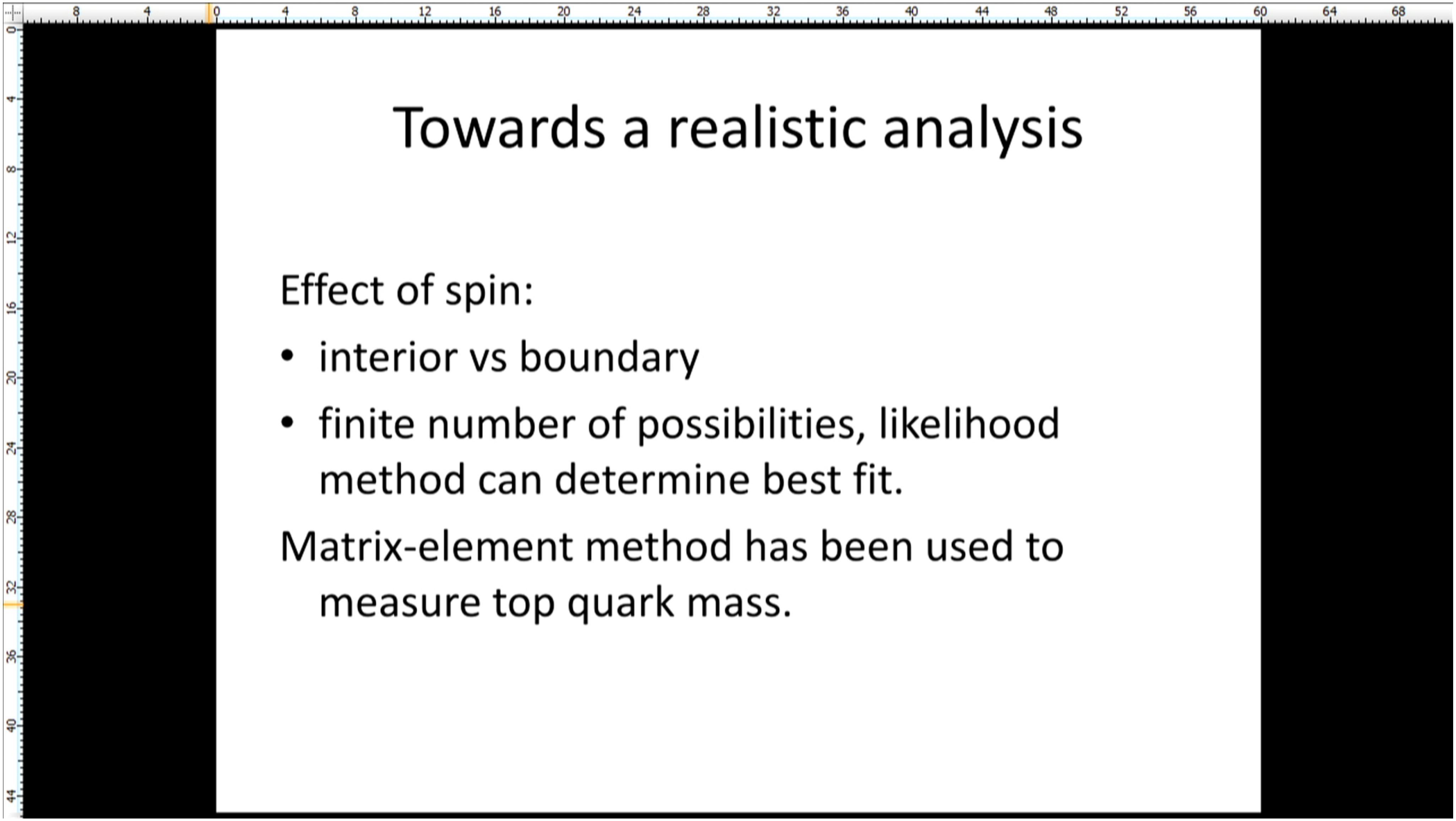


Towards a realistic analysis

Effect of spin:

- interior vs boundary
- finite number of possibilities, likelihood method can determine best fit.

Matrix-element method has been used to measure top quark mass.



Towards a realistic analysis

Effect of spin:

- interior vs boundary
- finite number of possibilities, likelihood method can determine best fit.

Matrix-element method has been used to measure top quark mass.

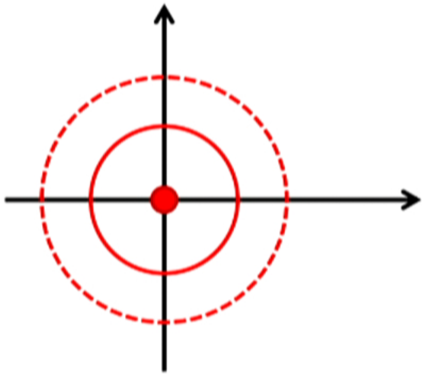
Longer Cascades

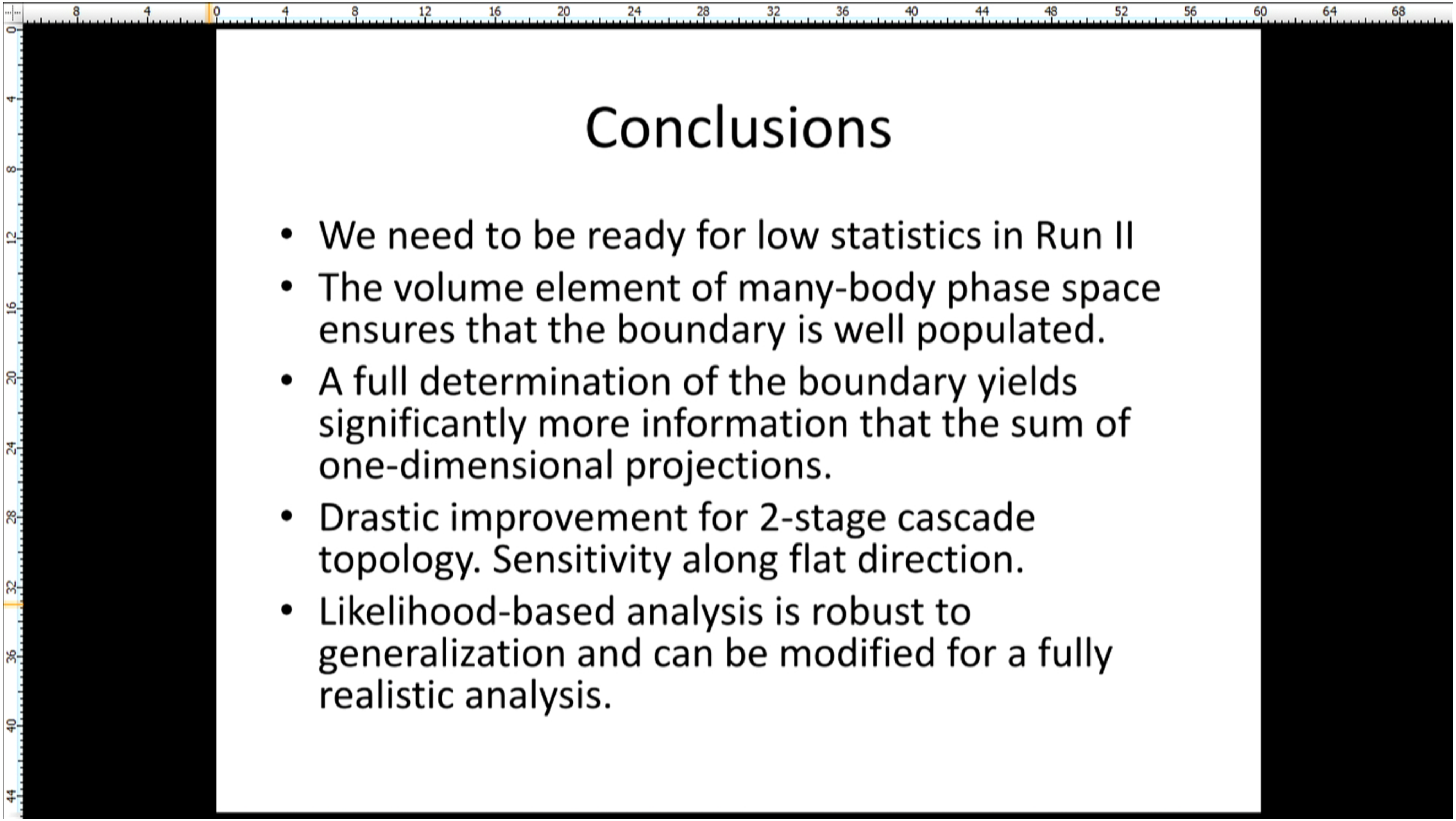
Is the enhancement for n=4 a generic feature?

$$d\Pi_n \propto \left(\sum_{i<j} dm_{ij}^2\right) (\sqrt{\Delta_4})^{n-5}\ \delta(\Delta_5)\cdots\delta(\Delta_n)\ \delta\left(\sum_{i<j} dm_{ij}^2 - C\right)$$

Unusual δ-functions

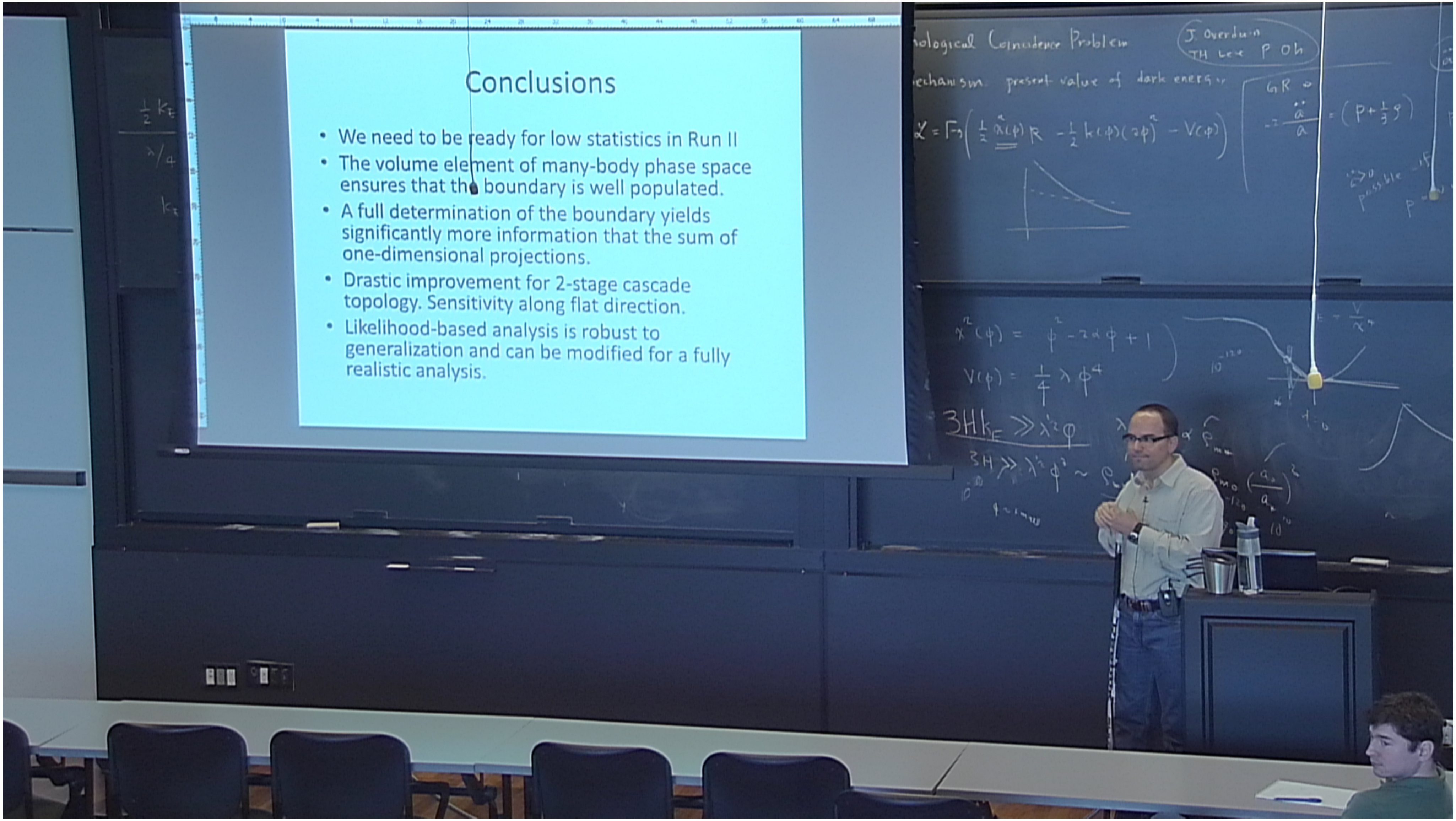
$$\lim_{R\to 0}\int dxdyf(x,y)\delta(x^2+y^2-R^2)$$





Conclusions

- We need to be ready for low statistics in Run II
- The volume element of many-body phase space ensures that the boundary is well populated.
- A full determination of the boundary yields significantly more information that the sum of one-dimensional projections.
- Drastic improvement for 2-stage cascade topology. Sensitivity along flat direction.
- Likelihood-based analysis is robust to generalization and can be modified for a fully realistic analysis.



Conclusions

- We need to be ready for low statistics in Run II
- The volume element of many-body phase space ensures that the boundary is well populated.
- A full determination of the boundary yields significantly more information that the sum of one-dimensional projections.
- Drastic improvement for 2-stage cascade topology. Sensitivity along flat direction.
- Likelihood-based analysis is robust to generalization and can be modified for a fully realistic analysis.