Automation of the matrix element reweighting method

UIUC, 29 November 2010

Pierre Artoisenet
The Ohio State University

work in collaboration with
V. Lemaitre, F. Maltoni, O. Mattelaer [arxiv:1007.3300, accepted in JHEP]

OUTLINE

- **Motivation**
- **The matrix element method**
- **Evaluation of the matrix element weights**
- **Applications**
- **Conclusion & Perspectives**

Motivation

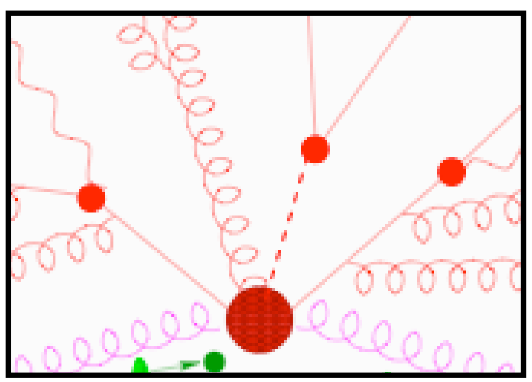
From data to theory: the inverse problem

For a given signature, try to identify an excess of events over the expected number of background events (counting analysis)

Lagrangian
L(p_1,p_2,...)

generation of MC events

Data



data analysis

Try to identify an excess of events and identify the underlying theory

From data to theory: the inverse problem

an “easy” example

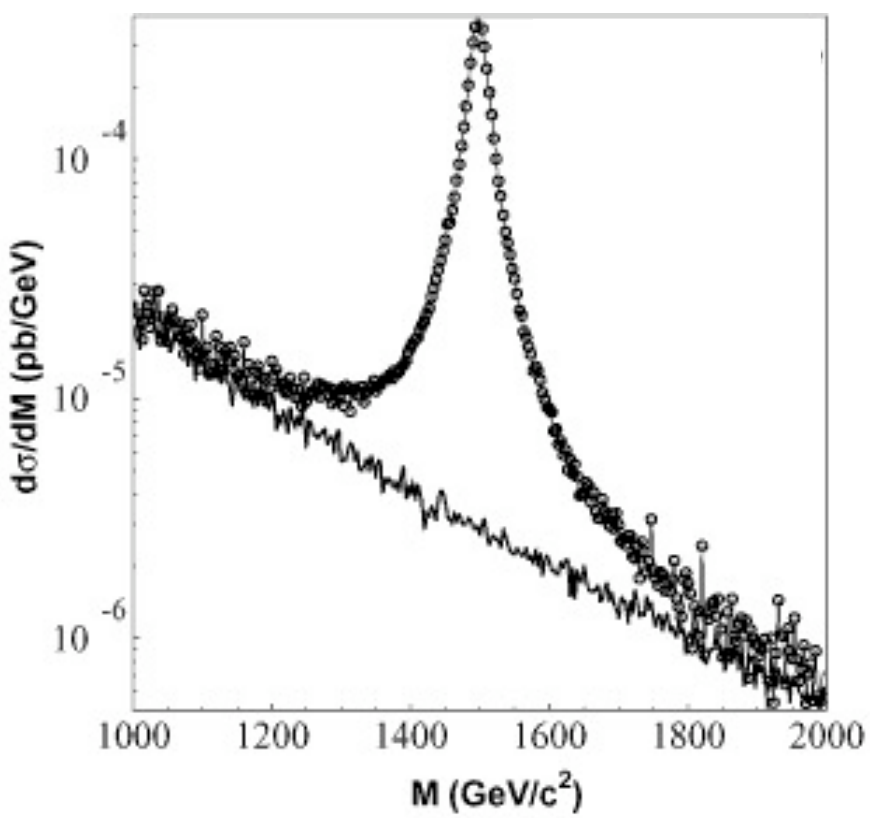
$pp \to Z' \to e^+e^-$

properties of the Z’ can be studied by analyzing one observable at the time (mass$\longleftrightarrow m_{inv(e+,e-)}$, spin$\longleftrightarrow \Omega_e$)

a “tough” example

$pp \to \tilde{g}\tilde{g}, \tilde{g}\tilde{q}, \tilde{q}\tilde{q} \to$ jets + MET

measurement of the properties of the new fields has to proceed with more complex observables



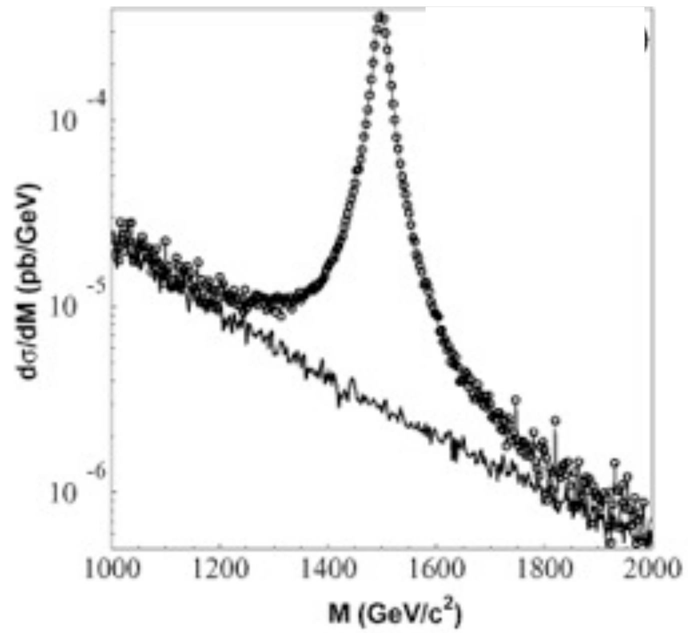
Several model-independent techniques have been introduced to handle the measurement of properties of decay chains with missing E_T at hadron colliders (end-point region, MT2, ...)

The matrix element method corresponds to a different approach, as it makes use of strong theoretical assumptions

The matrix element method

Basic idea

- kinematic method: discriminator built on one or several reconstructed observables, e.g. the invariant mass of two leptons



1. compute the distribution of events with respect to $d=m(e^+,e^-)$ under B-only and S+B hypotheses,

2. compare with the distribution of exp. events with respect to d

- matrix element method: likelihood method built on the matrix elements $|M_\alpha|^2$ associated with the theoretical assumptions $\alpha=h_1,h_2$ ... using the complete kinematics of the events

Basic idea

- so the matrix element method is a standard likelihood procedure, with the probability density distribution provided by the hard scattering amplitude

Define a per-event probability using matrix elements → Evaluate the probability for each event under the hypotheses $\alpha = h_1, h_2, \ldots$ → Combine the weights into one likelihood

$$P(x|\alpha) \quad \Rightarrow \quad w_i(\alpha) = P(x_i|\alpha) \quad \Rightarrow \quad L(\alpha)$$

x : kinematics of the reconstructed event

α : theoretical assumption

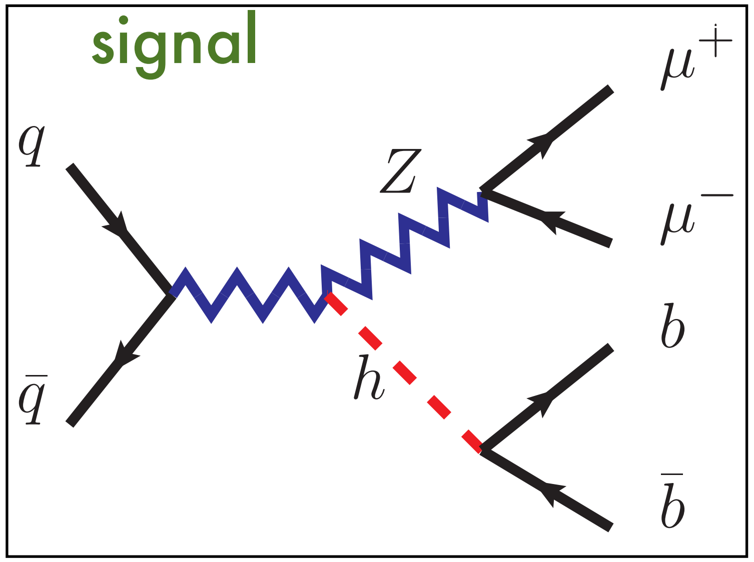
ME method for an ideal detector

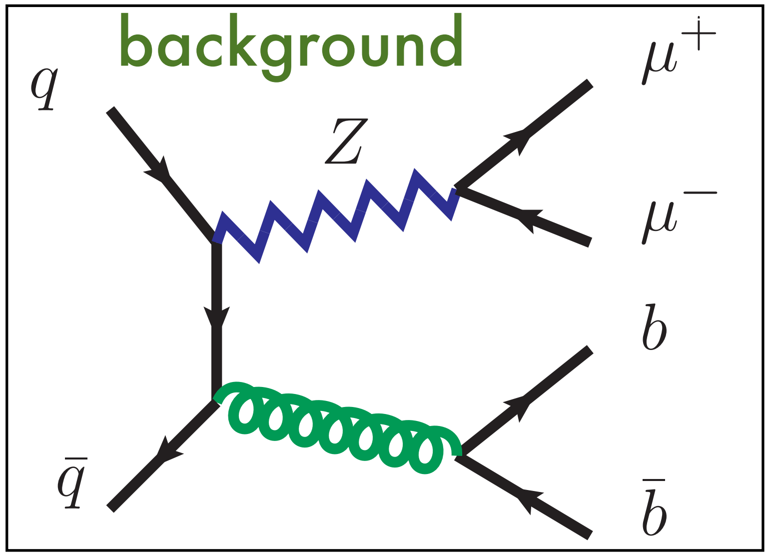
- imagine we have a ideal detector that reconstruct

  - all the final state objects
  - at the scale Q= scale of the hard interaction
  - with an infinite resolution

ME method for an ideal detector

- under these conditions, consider the following Higgs search at the Tevatron:





in this analysis, an event x corresponds to $p_{\mu^+}, p_{\mu^-}, p_b, p_{\bar{b}}$

Define a per-event probability using matrix elements

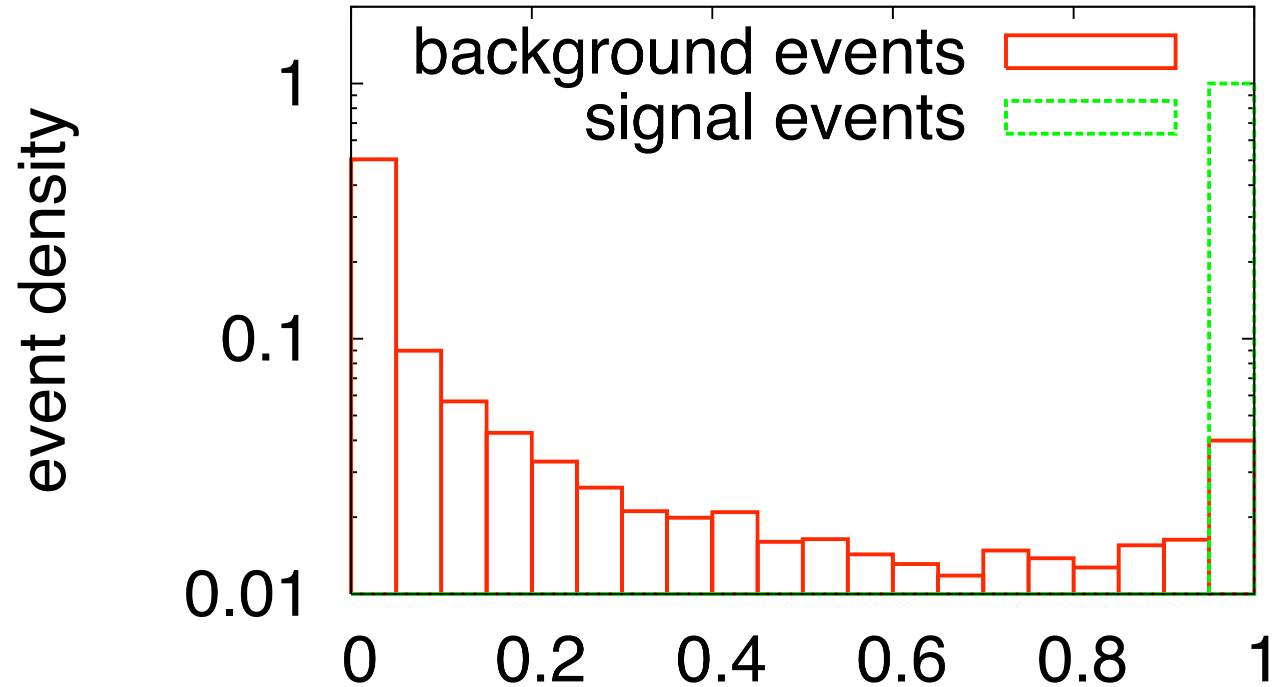
$$P(x|S) = \frac{\phi(x)}{\sigma_S}|M_S(x)|^2 \qquad P(x|B) = \frac{\phi(x)}{\sigma_B}|M_B(x)|^2$$

M_S : matrix element under the signal hypothesis

M_B : matrix element under the background hypothesis

ME method for an ideal detector

Evaluate the probability for each event under the hypotheses α=S or B



$$d(x) = \left[1 + \frac{P(x|B)}{P(x|S)}\right]^{-1}$$

d is an optimal discriminator based on the phase-space distribution of the events

ME method for an ideal detector

Combine the weights into one likelihood

Given N experimental events, you can test the S+B hypothesis versus the B-only hypothesis

If s,b =expected numbers of signal and background events is known, you can also use this information to improve the discriminating power

Likelihood for the B-only hypothesis: $$\mathrm{Pois}(N|b)\prod_{i=1}^{N} P(x_i|B)$$

Likelihood for S+B hypothesis:
$$\mathrm{Pois}(N|s+b)\prod_{i=1}^{N}[sP(x_i|S)+bP(x_i|B)]/(s+b)$$

The likelihood ratio is the most discriminating variable for this test

see K. Cranmer, T. Plehn, Eur. Phys. J. C 51, 415-420

ME method for a real experiment

in a real experiment, a reconstructed event cannot be weighted by a unique matrix element:

1. missing energy

some particles escape from the detector
without any interaction (neutrino, wimp, ...)

ME method for a real experiment

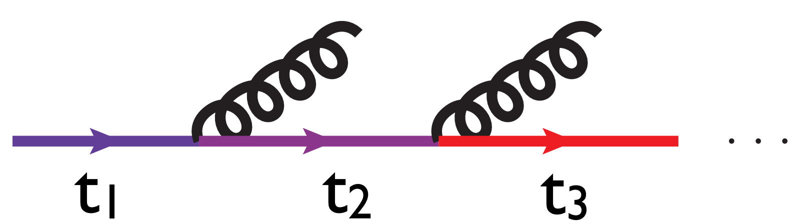
in a real experiment, a reconstructed event cannot be weighted by a unique matrix element:

2. showering/hadronization effects

a high energy collision is a multi-scale process, but a fixed-order matrix element provides a relevant description only for the hard scale Q

Q ⟶ ~1 GeV

physics	hard scattering	showering	hadronization
description tool	matrix element at fixed order in α_s	Sudakov form factors	simulation model tuned to the data



$$\Delta(t_1, t_2) = \exp\left\{ - \int_{t_2}^{t_1} \frac{dt'}{t'} \int dz \frac{\alpha_s(t')}{2\pi} P(z) \right\}$$

non-branching probability between scales t_1 and t_2

ME method for a real experiment

in a real experiment, a reconstructed event cannot be weighted by a unique matrix element:

3. experimental resolution/reconstruction algorithm

the final state objects (hadrons, leptons) are reconstructed with a finite resolution

ME method for a real experiment

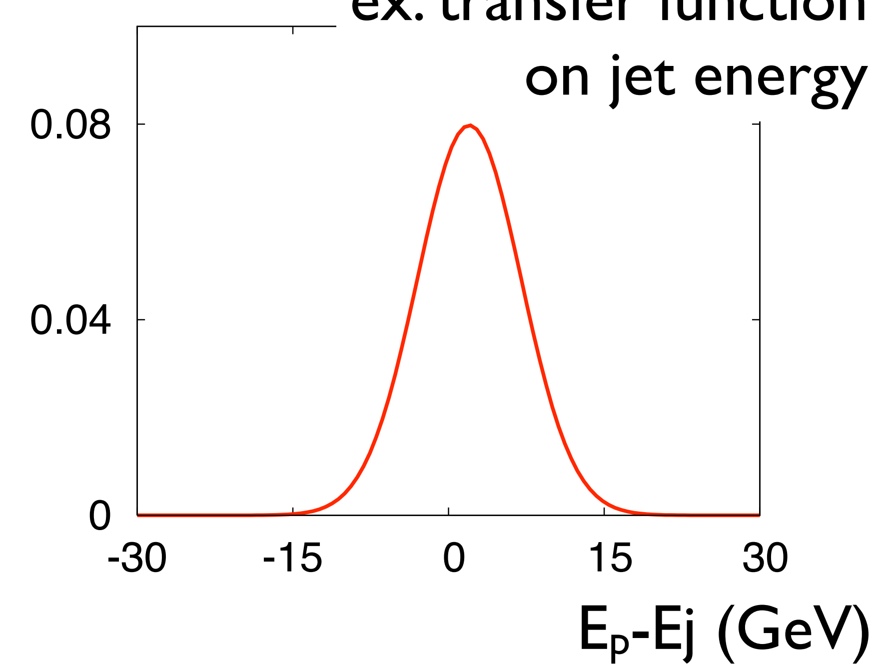
in a real experiment, a reconstructed event cannot be weighted by a unique matrix element:

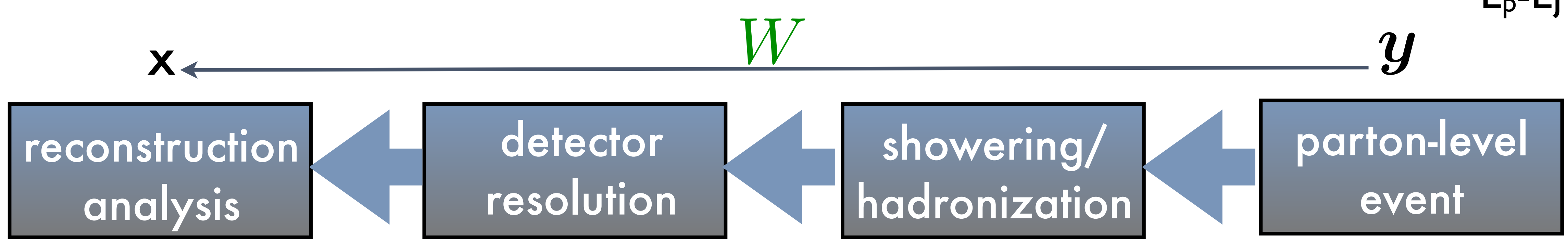
1. missing energy ⟷ $P(x,\alpha)$ must be summed over the unobserved degrees of freedom

2. showering/hadronization effets
3. experimental resolution/reconstruction algorithm

⟷ convolute with a transfer function $W(x,y)$

ex: transfer function on jet energy





ME method for a real experiment

- real detector: we need to marginalize over unconstrained information and to convolute with the resolution function W for the measured quantities

$$P(\boldsymbol{x}_i, \alpha) = \frac{1}{\sigma^{obs}} \frac{1}{N} \sum_{\text{jet perm.}} \int d\phi_{\boldsymbol{y}} |M|^2(\boldsymbol{y}) W(\boldsymbol{x}_i, \boldsymbol{y}) Acc(x)$$

integration on the parton-level phase-space

tree-level matrix element

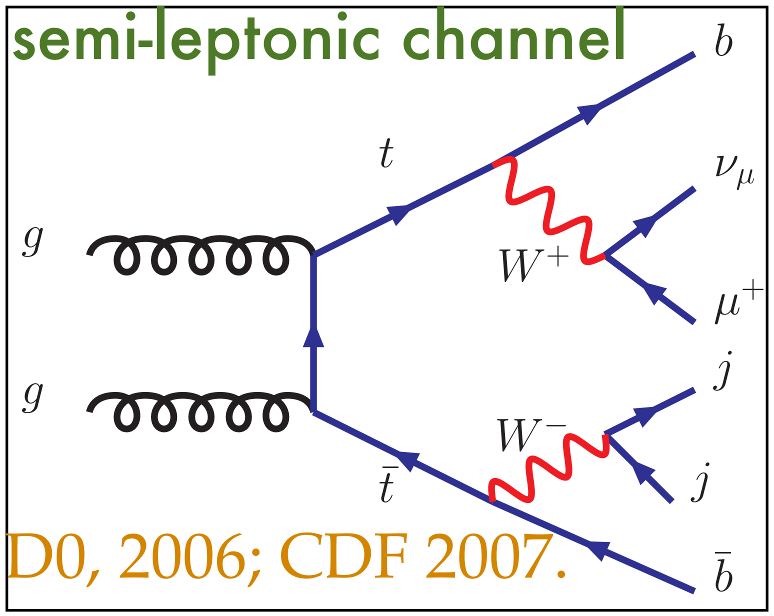
transfer function extracted from MC simulation

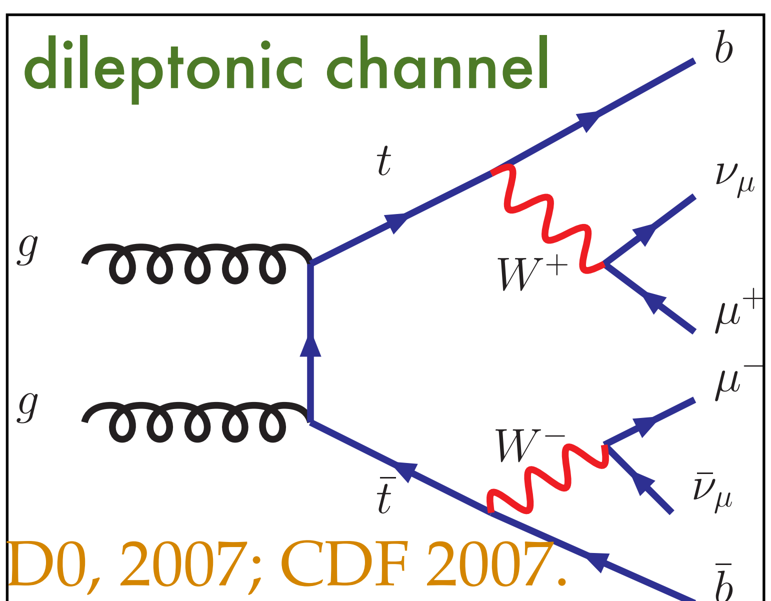
normalization: $\int dx W(x, y) = 1$

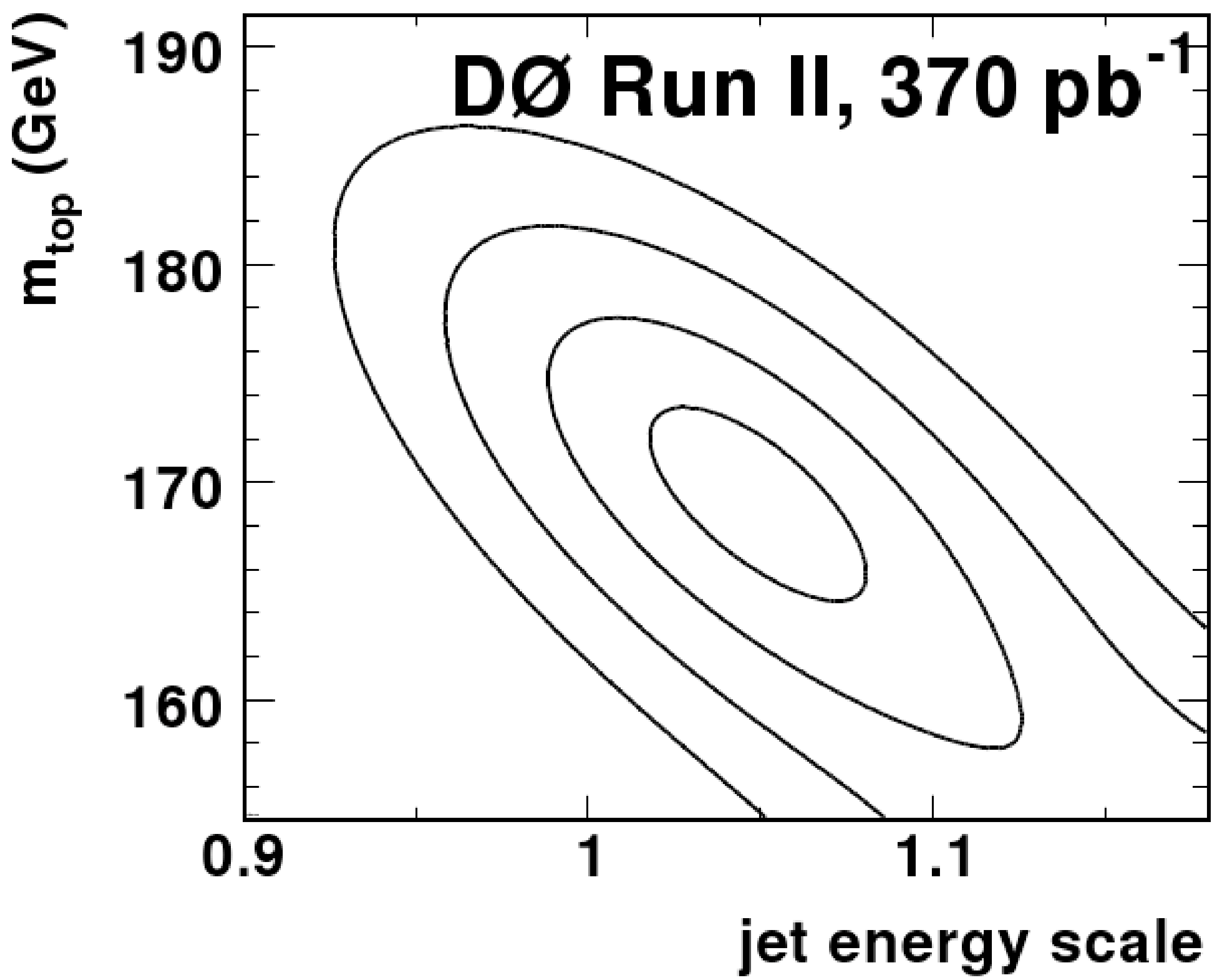
⇒ the probability density P(x| α) is normalized to 1

First ME analysis at the Tevatron

Top-quark mass measurement from $t\bar{t}$ production in hadron collisions







[DO Phys. Rev. D75 092005, 2006]

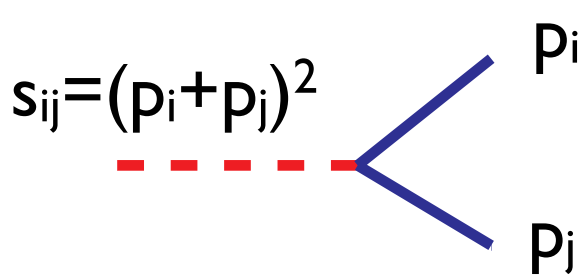
Significant improvement for the measurement of the top-quark mass

Evaluation of the weights

Practical Evaluation of the weights

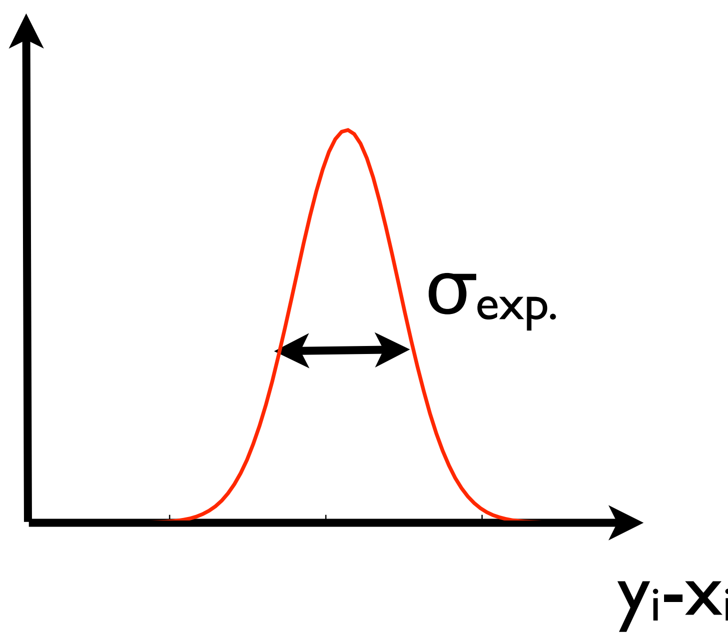
$$P(x,\alpha) \propto \int d\phi_{\boldsymbol{y}} \boxed{|M|^2(\boldsymbol{y})} \boxed{W(\boldsymbol{x},\boldsymbol{y})}$$

highly non-uniform, especially in the presence of resonances



Breit-Wigner distr. in s_{ij}

highly non-uniform, especially when the resolution associated with a reconstructed quantity x_i is high:



when the dimension of the phase-space is large, this structure in "peaks" complicates the numerical evaluation of the weights

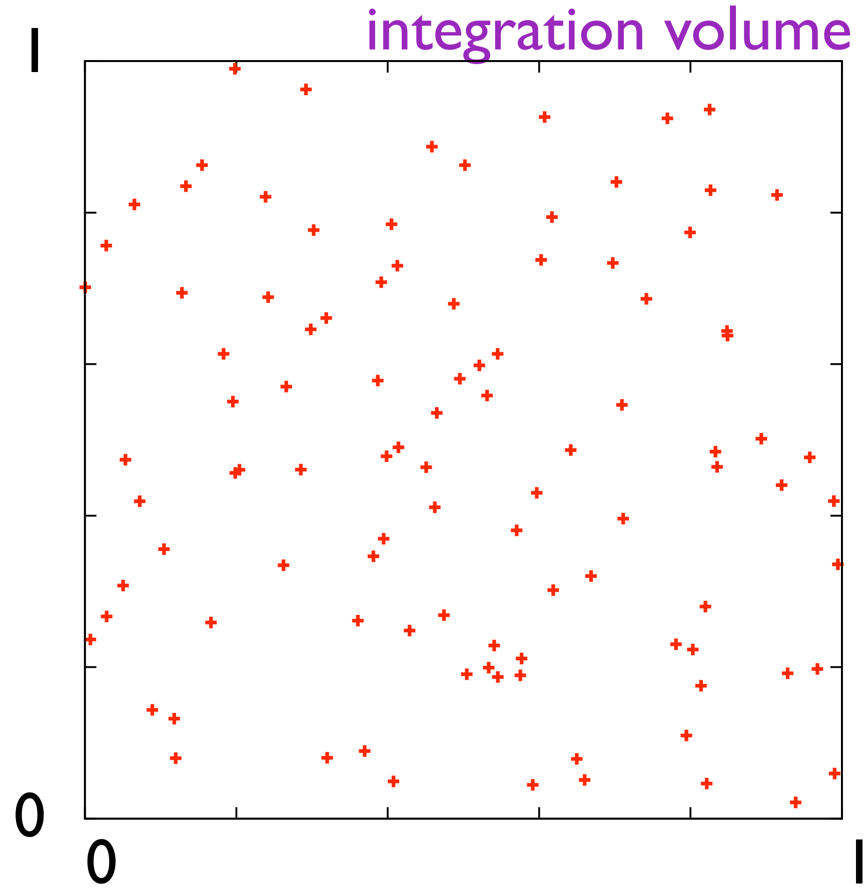
⇒ need for an algorithm that is sufficiently fast (large number of weights must be evaluated)

Monte Carlo integration

1. basic idea: $I = \int_V dz\, f(\boldsymbol{z})$ is estimated by sampling the volume V=[0,1]^d

with N uniformly distributed random points: $E = \frac{1}{N}\sum_{n=1}^{N} f(\boldsymbol{z}_n)$

Std deviation: $\sigma_I \approx \frac{S}{\sqrt{N}}$



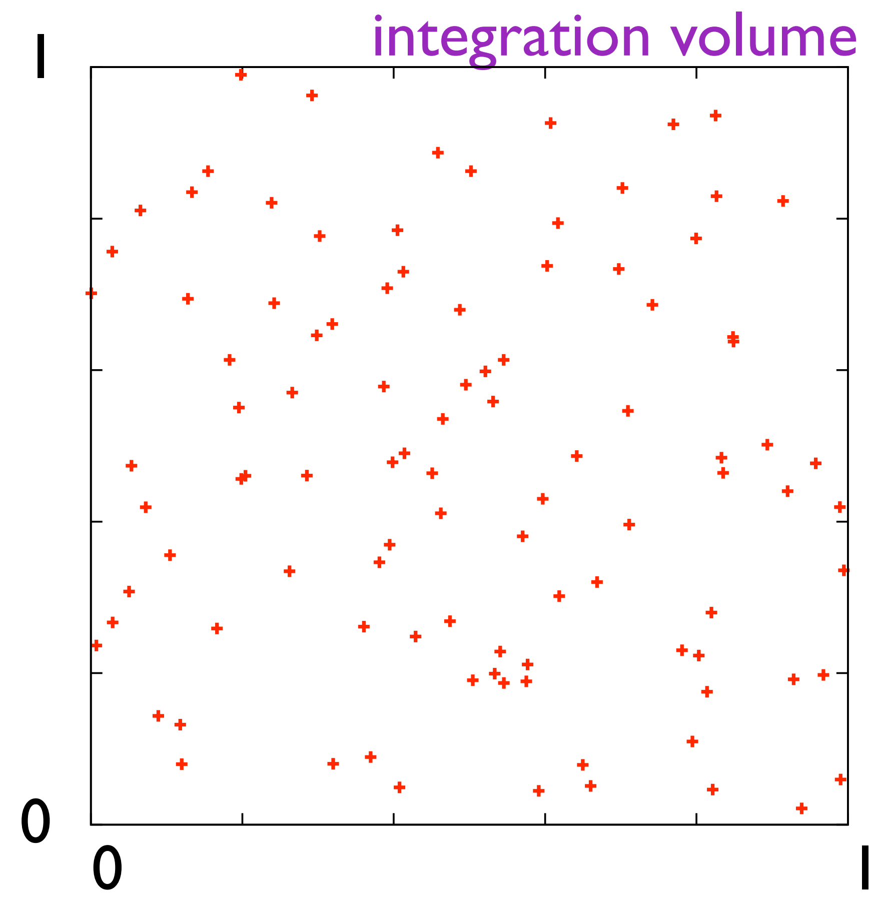
Monte Carlo integration

1. basic idea: $I = \int_V d\boldsymbol{z}\, f(\boldsymbol{z})$ is estimated by sampling the volume V=[0,1]^d

with N uniformly distributed random points: $E = \frac{1}{N}\sum_{n=1}^{N} f(\boldsymbol{z}_n)$

Std deviation: $\sigma_I \approx \frac{S}{\sqrt{N}}$

$$S^2 = \mathrm{var}(f) = \frac{1}{N-1}\sum_{n=1}^{N}\left[f(\boldsymbol{z}_n) - E\right]^2$$



Monte Carlo integration

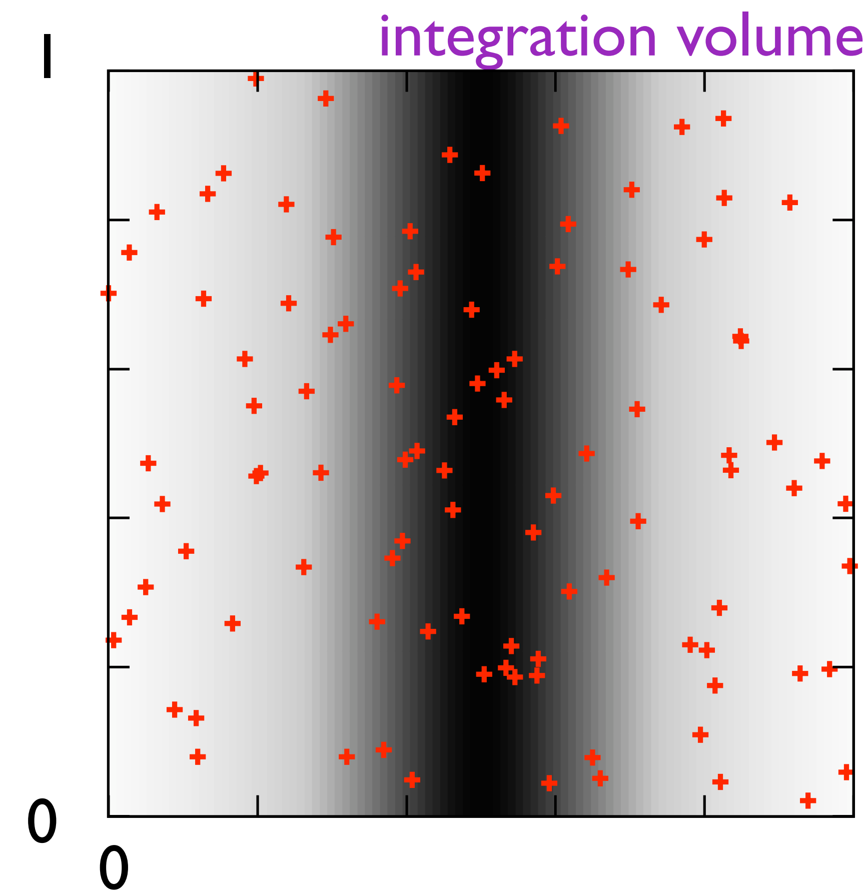
1. basic idea: $I = \int_V dz\, f(z)$ is estimated by sampling the volume V=[0,1]^d

with N uniformly distributed random points: $E = \frac{1}{N} \sum_{n=1}^{N} f(z_n)$

Std deviation: $\sigma_I \approx \frac{S}{\sqrt{N}}$

$$S^2 = \mathrm{var}(f) = \frac{1}{N-1} \sum_{n=1}^{N} [f(z_n) - E]^2$$

S large → poor convergence

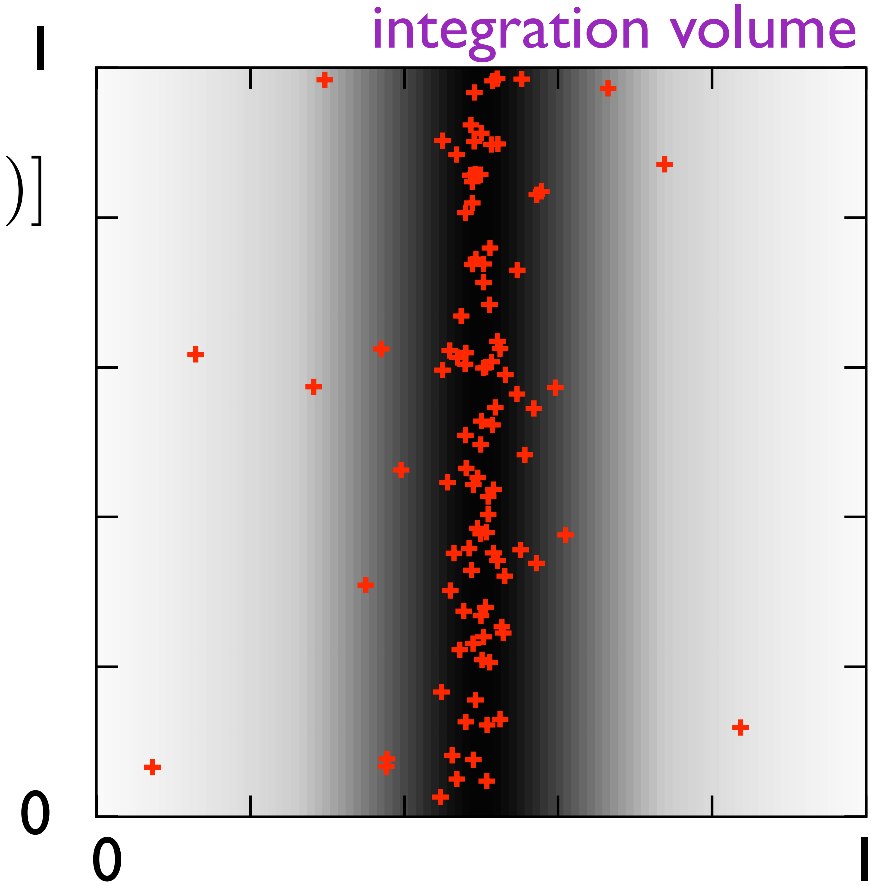


Monte Carlo integration

1. basic idea: $I = \int_V dz\, f(\boldsymbol{z})$ is estimated by sampling the volume V=[0,1]^d

with N uniformly distributed random points: $E = \frac{1}{N}\sum_{n=1}^{N} f(\boldsymbol{z}_n)$

Std deviation: $\sigma_I \approx \frac{S}{\sqrt{N}}$



2. importance sampling: $\boldsymbol{z}' = \boldsymbol{P}(\boldsymbol{z}),\ p(\boldsymbol{z}) = Jac[\boldsymbol{P}(\boldsymbol{z})]$

$$\int dz\, f(\boldsymbol{z}) = \int \frac{f[\boldsymbol{P}^{-1}(\boldsymbol{z}')]}{p[\boldsymbol{P}^{-1}(\boldsymbol{z}')]} d\boldsymbol{z}' = \int \boxed{\frac{f(\boldsymbol{z})}{p(\boldsymbol{z})}} \boxed{p(\boldsymbol{z}) d\boldsymbol{z}}$$

new integrand

new integr. measure

if $\{\boldsymbol{z}_n\}$ distributed according to $p(\boldsymbol{z})$ then

$$E \rightarrow \frac{1}{N}\sum_{n=1}^{N} \frac{f(\boldsymbol{z}_n)}{p(\boldsymbol{z}_n)}$$

$$S^2 \rightarrow \frac{1}{N-1}\sum_{n=1}^{N} \left[\frac{f(\boldsymbol{z}_n)}{p(\boldsymbol{z}_n)} - E\right]^2$$

S is decreased if $p(\boldsymbol{z}) \approx f(\boldsymbol{z})/E$

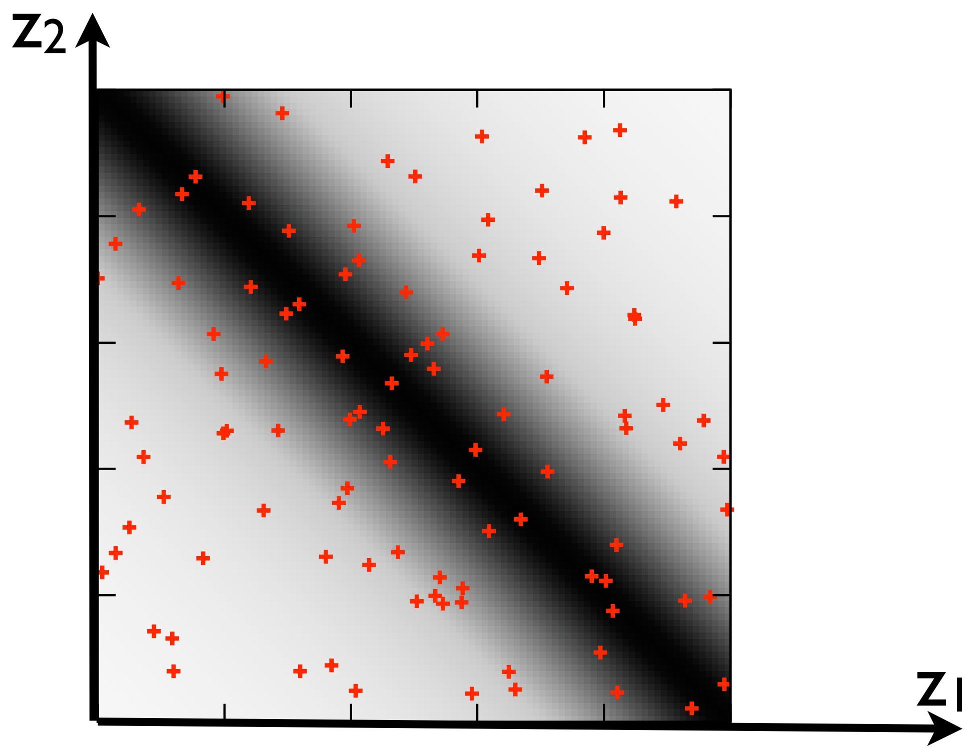
3. adaptive Monte Carlo integration:

$p(\boldsymbol{z}) = p_1(z^1)p_2(z^2)\ldots p_d(z^d)$ (grid)

optimized using an iteration procedure

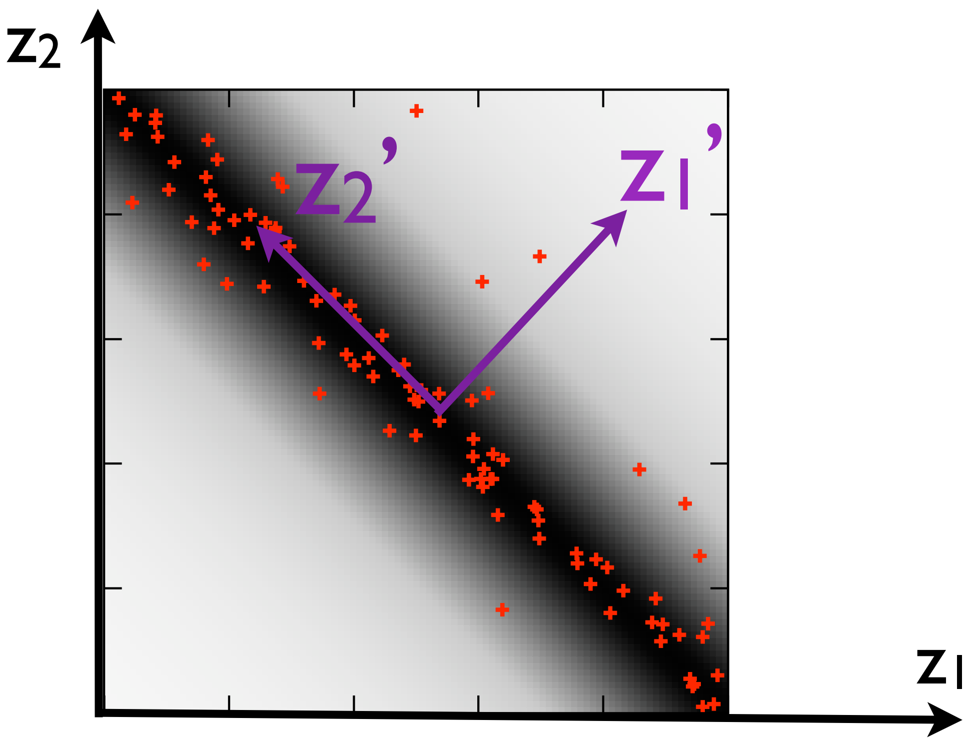
Adaptive Monte Carlo integration

the efficiency of the adaptive MC integration depends on the choice of variables of integrations



variables z_1, z_2:

the grid cannot be adjusted efficiently to the shape of the integrand because the strength of the "peak" in the integrand is not controlled by a single variable of integration

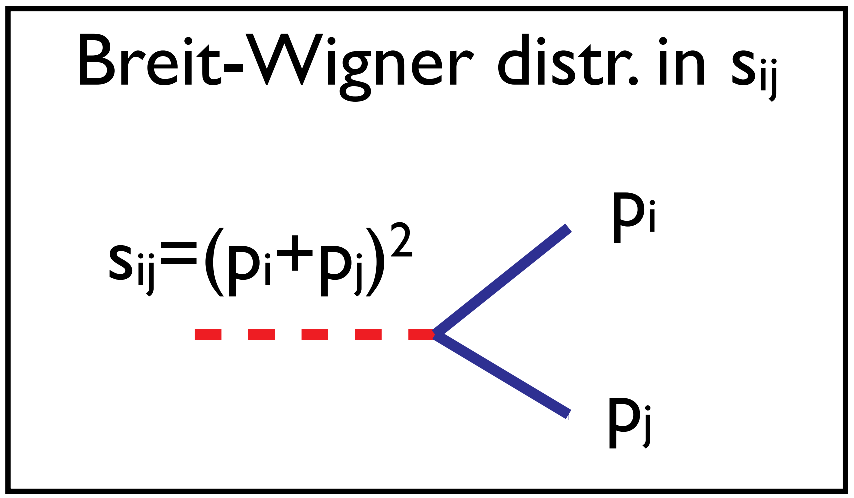


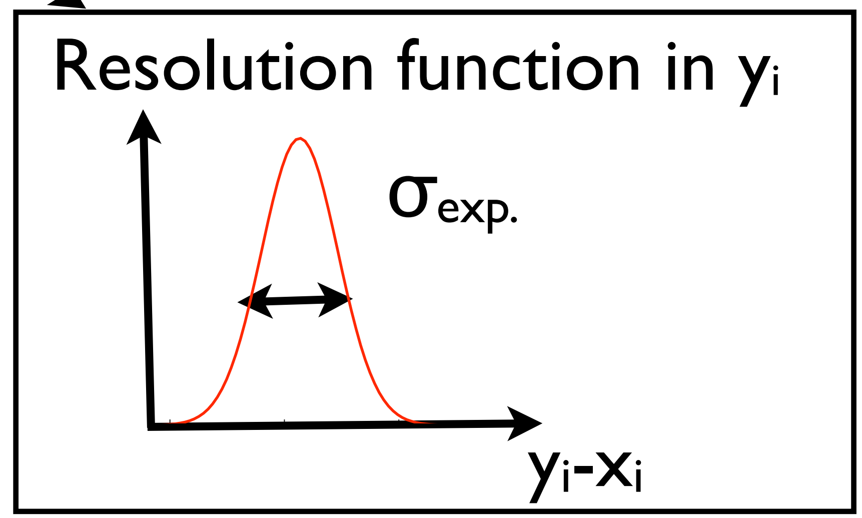
variables z_1', z_2':

the probability density along z_1' (= variable that controls the strength of the "peak") can be adapted to probe the integration region where the integrand is the largest

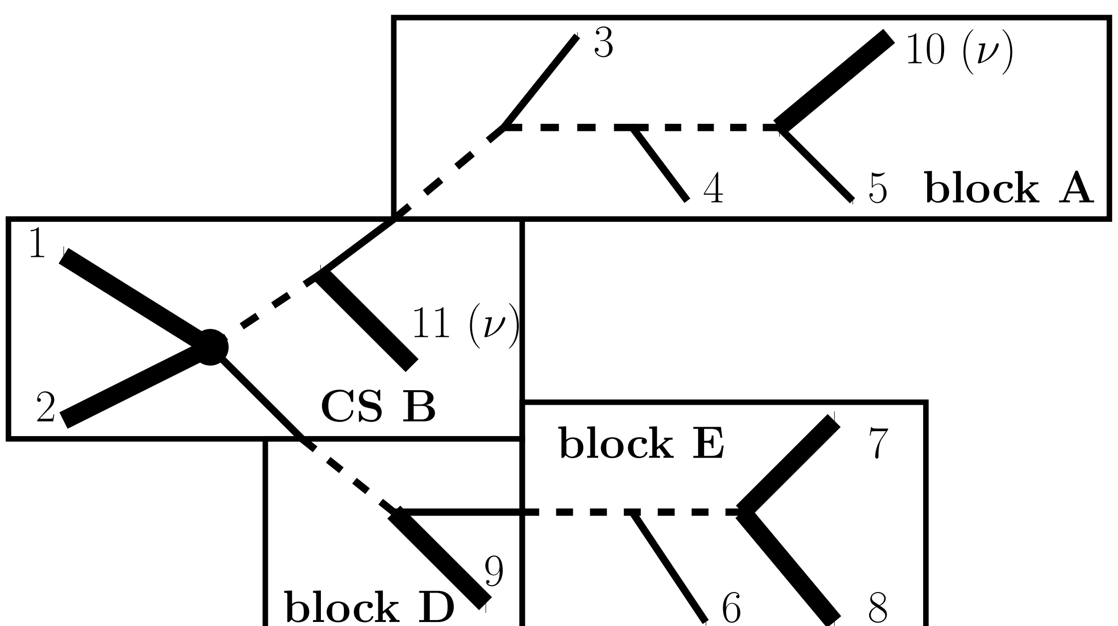
New phase-space mappings

adaptive MC integration can be used for the computation of the weights, as we know where the "peaks" lie:





for a given decay chain and a given transfer function, one needs to construct new parametrizations of the phase-space measure

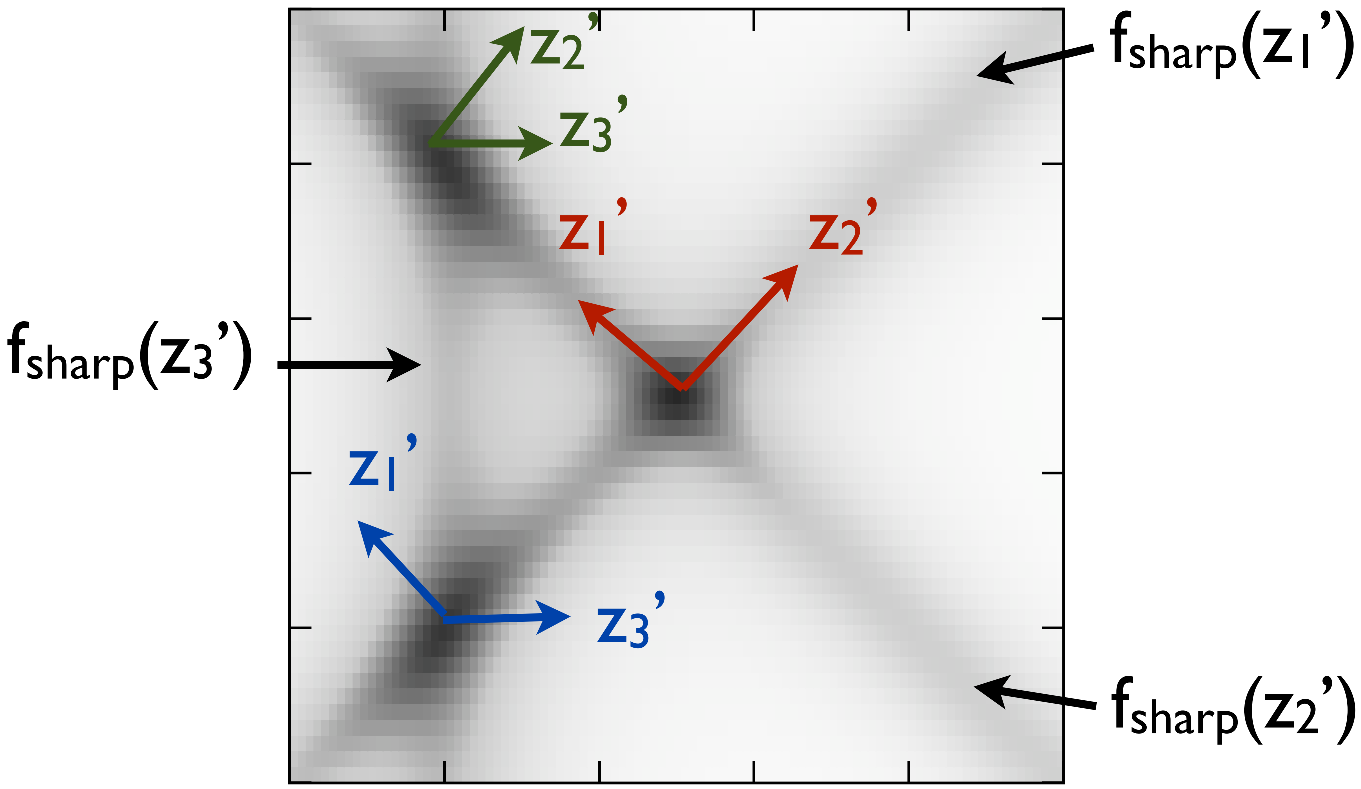


this is done automatically through the decomposition of the standard phase-space parametrization into blocks of kinematics variables subject to specific changes of variables

Multi-channel integration

for specific processes, there is no parametrization of the phase-space measure that maps all the peaks simultaneously

example: over-constrained system



decomposition into 3 channels:

$$d\phi = \beta_1 d\phi_1 + \beta_2 d\phi_2 + \beta_3 d\phi_3$$

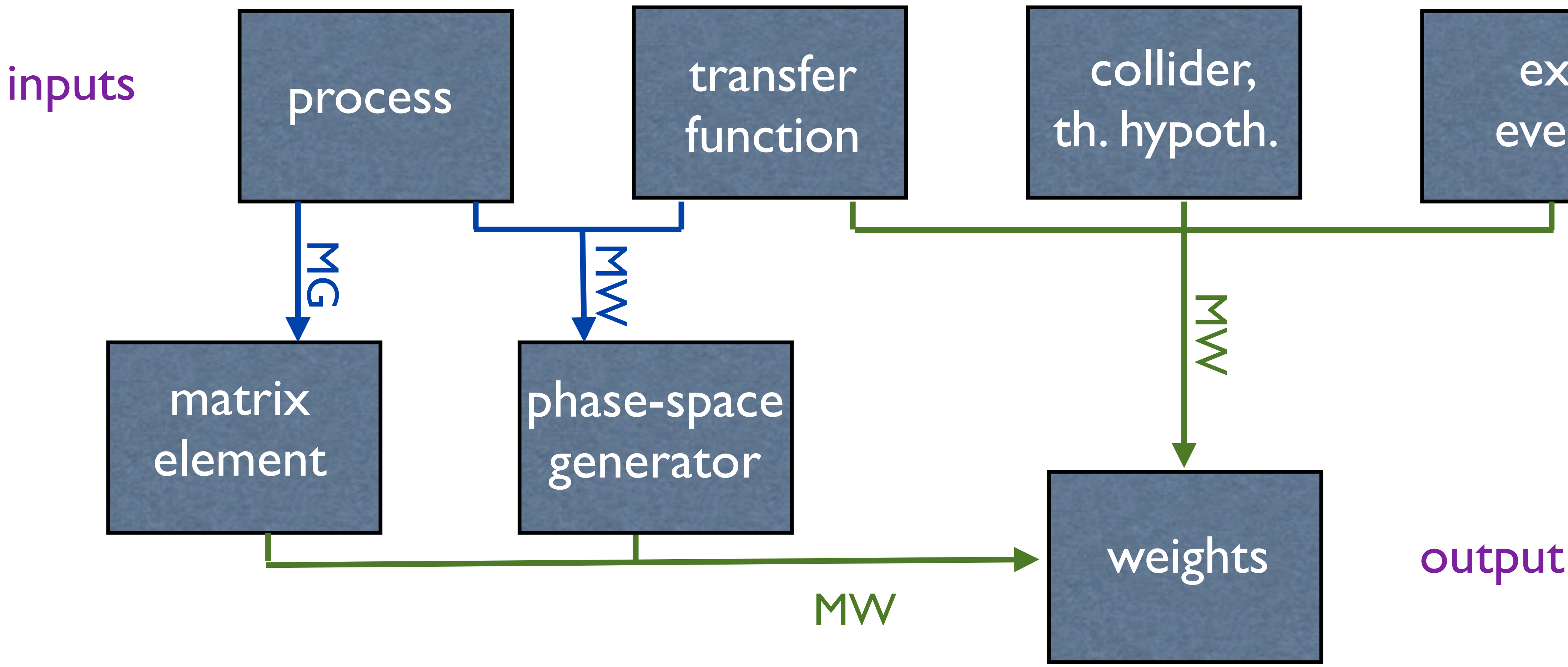
where the β 's assign the importance of each channel point by point in the phase-space:

$$\beta_1 = \frac{f_{\rm sharp}(z'_2)\, f_{\rm sharp}(z'_3)}{\sum_{i<j} f_{\rm sharp}(z'_i)\, f_{\rm sharp}(z'_j)}$$

and similarly for β_2, β_3

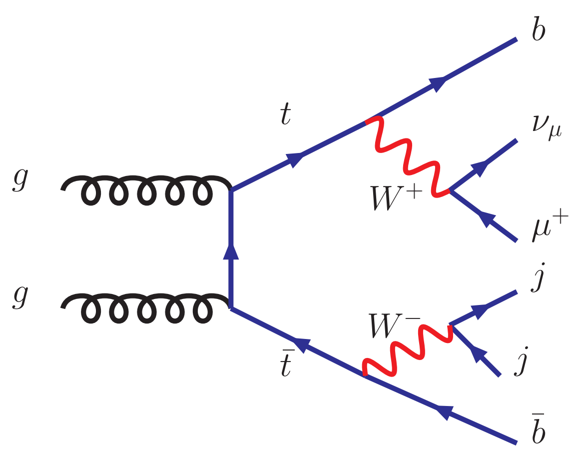
this multichannel decomposition has been generalized and automated in our algorithm to arbitrary processes

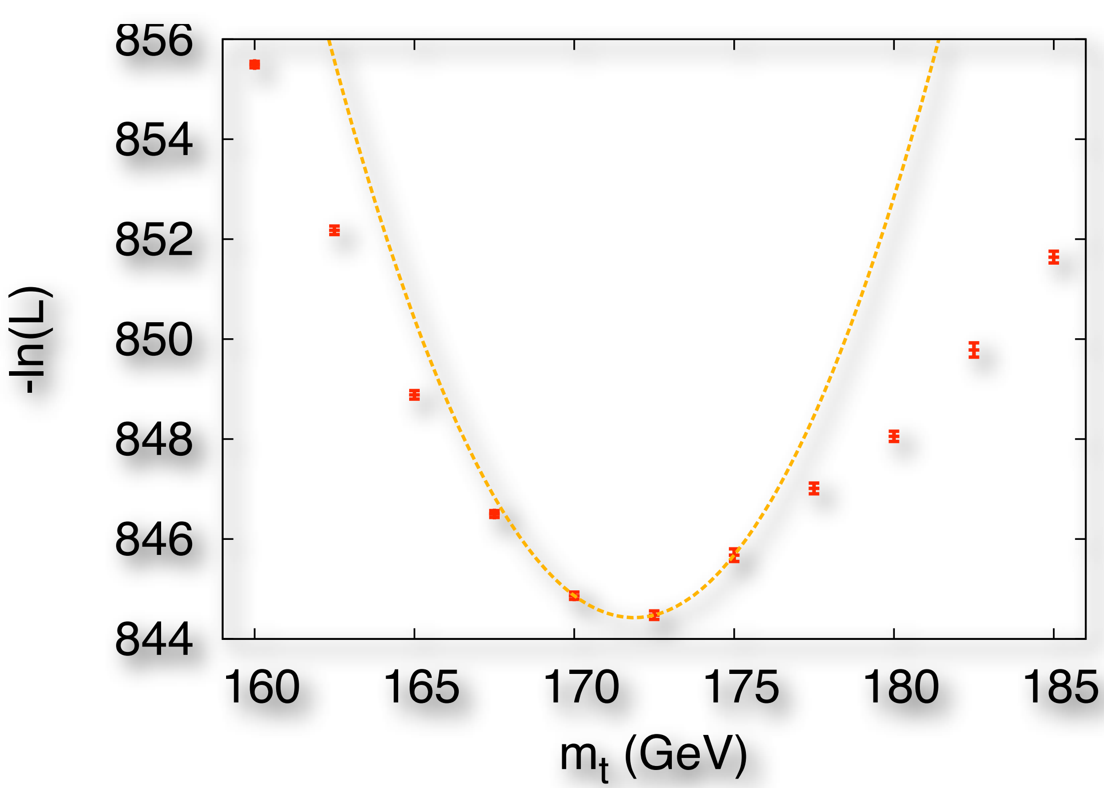
MadWeight: how does it work ?



MadWeight: how does it work ?

inputs	process	transfer function	collider, th. hypoth.	exp. events
example:	$pp \to t\bar{t}$ semi-leptonic	infinite resolution except for E_{jet}	LHC, sm, with m_t in [160 GeV, 185 GeV]	20 signal events with m_t=170 GeV





output: likelihood as a function of m_t

$$L(m_t) \propto \prod_i P(x_i|m_t)$$

Applications

MadWeight: applications

obj. # 1 (exp.): serve future measurements based on the matrix element method by providing a tool that is

- public: see http://madgraph.hep.uiuc.edu
- tested: many checks by reproducing known quantities such as volumes of integrations, total cross sections, ...
- updated with users' feedback
- suitable for improvements of either the formulation of the method itself (e.g. effects of higher order corrections in QCD) or the integration techniques

current analysis: search for a sm light Higgs at the Tevatron
(in collaboration with H. Wolf)

MadWeight: applications

obj. # 2 (exp. + th.): serve future analyses aimed at providing a better control of the potentially large systematic uncertainties (from both exp. and theoretical sources)

in particular matrix element method makes use of leading-order theory information in its fully differential form

➔ important to have a control of the impact of higher-order corrections on the method

see J. Alwall, A. Freitas, O. Mattelaer arXiv:1010.2263

MadWeight: applications

obj. # 3 (pheno): assess what is the maximum significance that can be achieved in a given analysis:

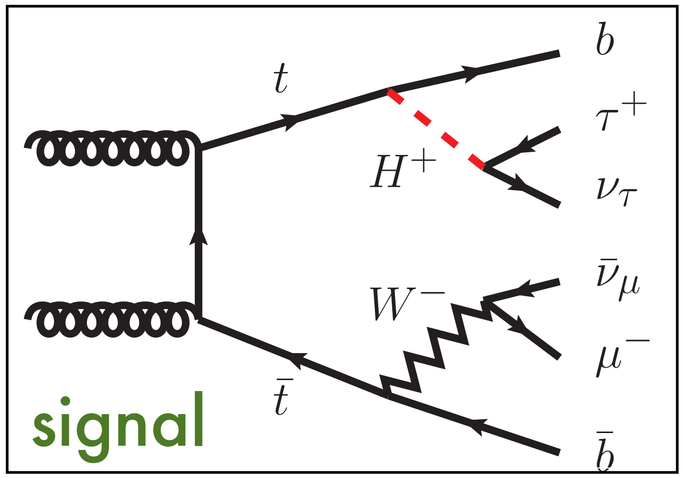
the ME method offers the possibility to optimize the discriminating power between different theoretical hypotheses, and therefore provides a way to estimate an upper-bound on the significance of a specific exp. analysis at a given luminosity

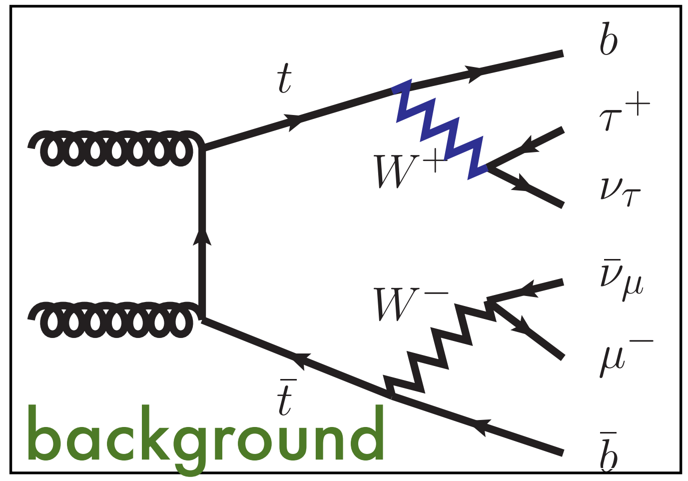
I will illustrate this with two examples

Application 1: testing spin hypotheses

Disentangling different spin hypotheses in decay chain with missing E_T

$t \to H^+ b$ vs. $t \to W^+ b$





we set

$$m_H \approx m_W$$

(keep only the spin correlation effects)

Possible discriminators:

- keeping only information from $P_T(\tau)$:

$$D(\boldsymbol{x}) = \frac{\sigma_H^{-1} \frac{d\sigma_H}{dp_{T,\tau}}}{\sigma_H^{-1} \frac{d\sigma_H}{dp_{T,\tau}} + \sigma_W^{-1} \frac{d\sigma_W}{dp_{T,\tau}}}$$

- matrix element method (keeps all information):

$$D(\boldsymbol{x}) = \frac{P_H(\boldsymbol{x})}{P_H(\boldsymbol{x}) + P_W(\boldsymbol{x})}$$

Application 1: testing spin hypotheses

Disentangling different spin hypotheses in decay chain with missing E_T

$$t \to H^+ b \quad \text{vs.} \quad t \to W^+ b$$

Analysis based on the p_T weight:

- discriminator:

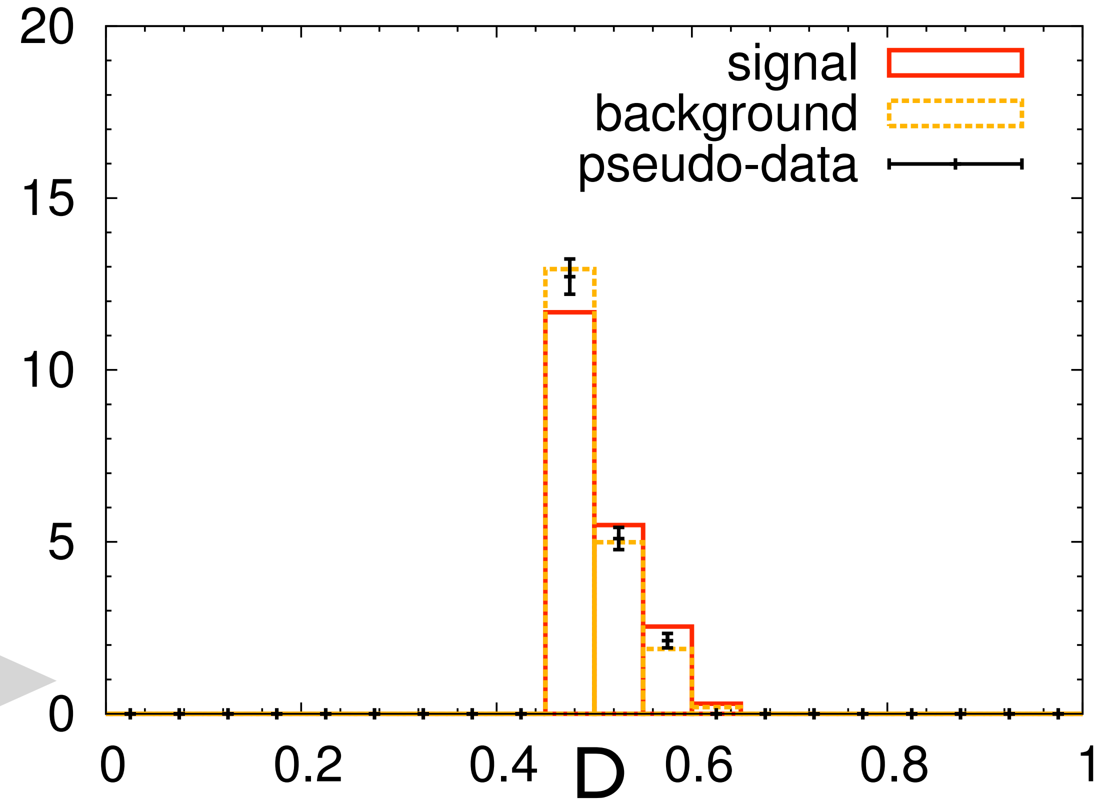
$$D(\boldsymbol{x}) = \frac{\sigma_H^{-1} \frac{d\sigma_H}{dp_{T,\tau}}}{\sigma_H^{-1} \frac{d\sigma_H}{dp_{T,\tau}} + \sigma_W^{-1} \frac{d\sigma_W}{dp_{T,\tau}}}$$

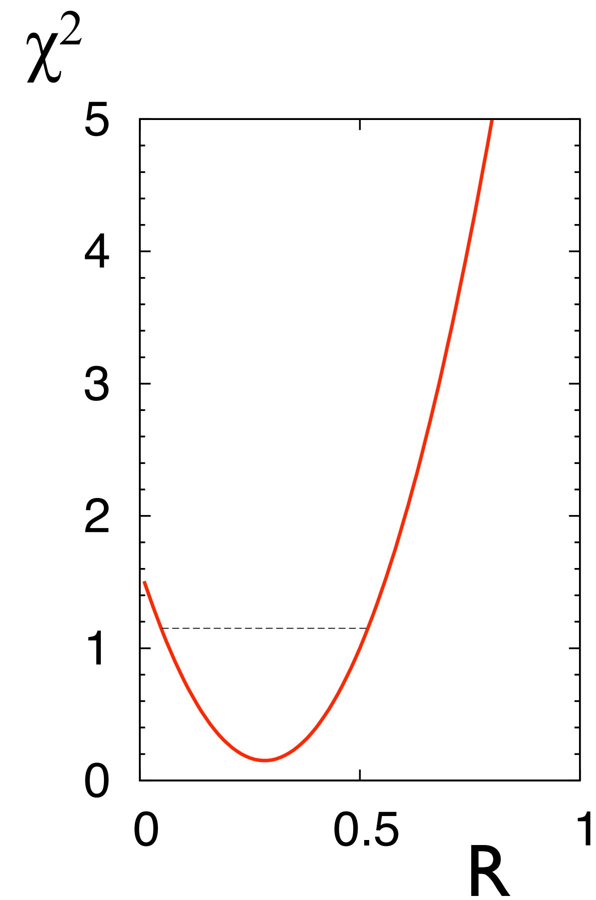
- "data": 240 signal events

760 background events

fraction of signal events:

R_{in}=24%





By fitting the event density distribution of the pseudo data by a superposition of the expected distributions for the signal and for the background, we get

reconstructed fraction of signal events (R_{out}) = 28±24%

Application I: testing spin hypotheses

Disentangling different spin hypotheses in decay chain with missing E_T

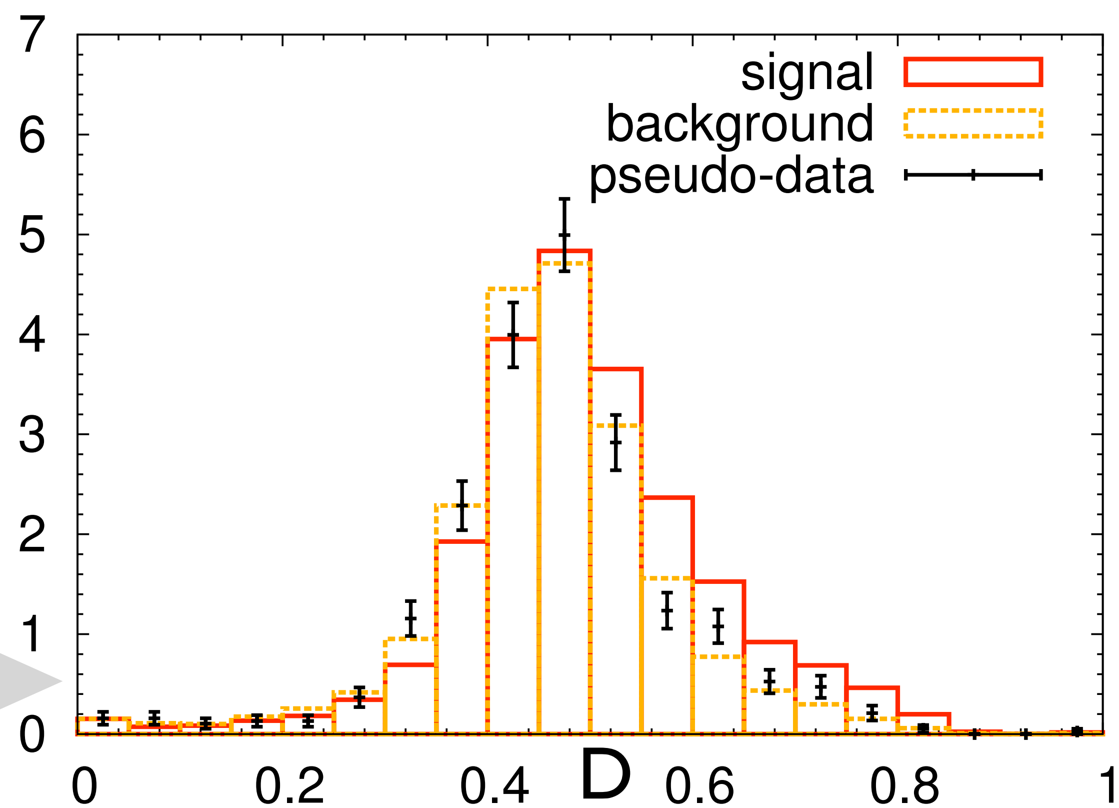
$$t \rightarrow H^+ b \qquad \text{vs.} \qquad t \rightarrow W^+ b$$

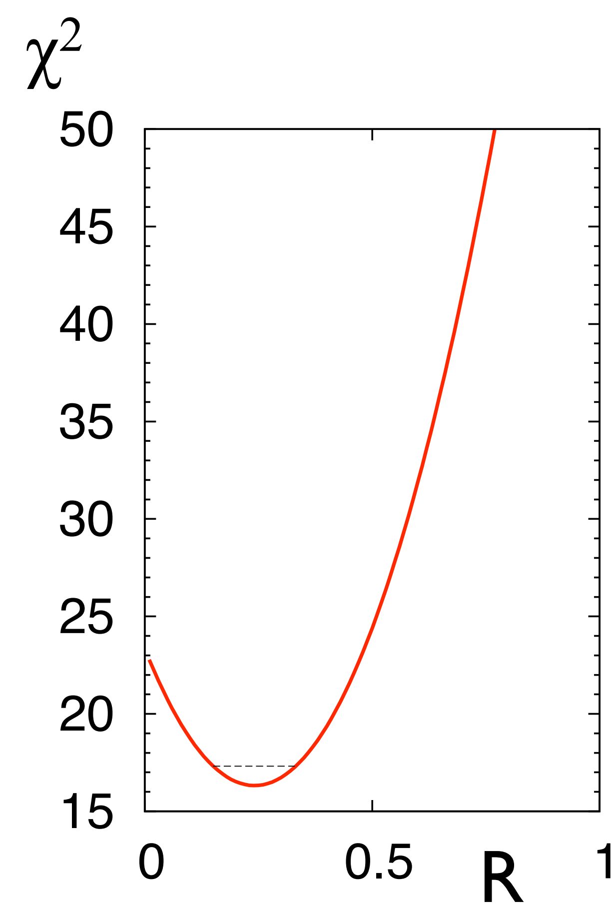
Analysis based on the matrix element weight:

- discriminator:

$$D(\boldsymbol{x}) = \frac{P_H(\boldsymbol{x})}{P_H(\boldsymbol{x}) + P_W(\boldsymbol{x})}$$

- "data": 240 signal events
760 background events
fraction of signal events:
R_{in}=24%



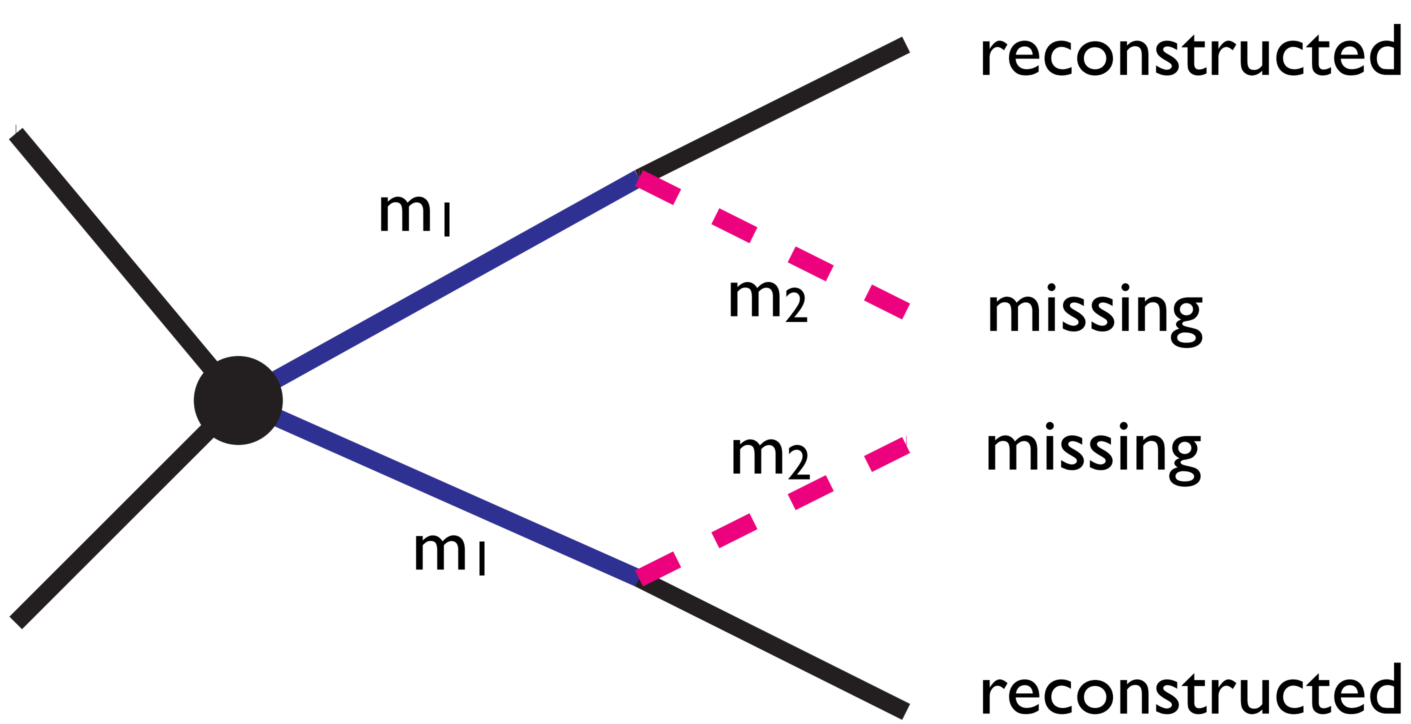


The discriminating power is substantially improved. The fit of the distribution associated with the pseudo-data gives:

reconstructed fraction of signal events (R_{out}) = 24±9%

Application II: mass reconstruction

Consider the following symmetric decay chain



assuming a pure sample of signal events, is it possible to reconstruct both m_1 and m_2 ?

for limited statistics, kinematic methods can only reconstruct the quantity $(m_1^2-m_2^2)/2m_1$

Application II: mass reconstruction

Q: assuming that the masses m_1 and m_2 are the only unknown, what is the maximum significance that can be achieved in measuring these masses at a given luminosity ?

Let us consider a specific example:

$$pp \to (\tilde{\mu}_r^+ \to \mu^+ \tilde{\chi}_1)(\tilde{\mu}_r^- \to \mu^- \tilde{\chi}_1)$$

sample of 50 events with $m_{\tilde{\mu}_r} = 150$ GeV

$m_{\tilde{\chi}_1} = 100$ GeV

$$(m^2_{\tilde{\mu}_r} - m^2_{\tilde{\chi}_1})/2m_{\tilde{\mu}_r} = 42 \text{ GeV}$$

possible discriminators:

- keeping only information from $p_T(\mu^+)$, $M(\mu^+, \mu^-)$

$$P(x|\tilde{\mu}_r, \tilde{\chi}_1) = \sigma^{-1}\frac{d\sigma}{dp_{T\mu}}(p_{T\mu}|m_{\tilde{\mu}_r}, m_{\tilde{\chi}_1}) \times \sigma^{-1}\frac{d\sigma}{dM_{\mu\mu}}(M_{\mu\mu}|m_{\tilde{\mu}_r}, m_{\tilde{\chi}_1})$$

- matrix element method (keeps all information):

$$P(x|\tilde{\mu}_r, \tilde{\chi}_1) = \text{ matrix element weight}$$

Application II: mass reconstruction

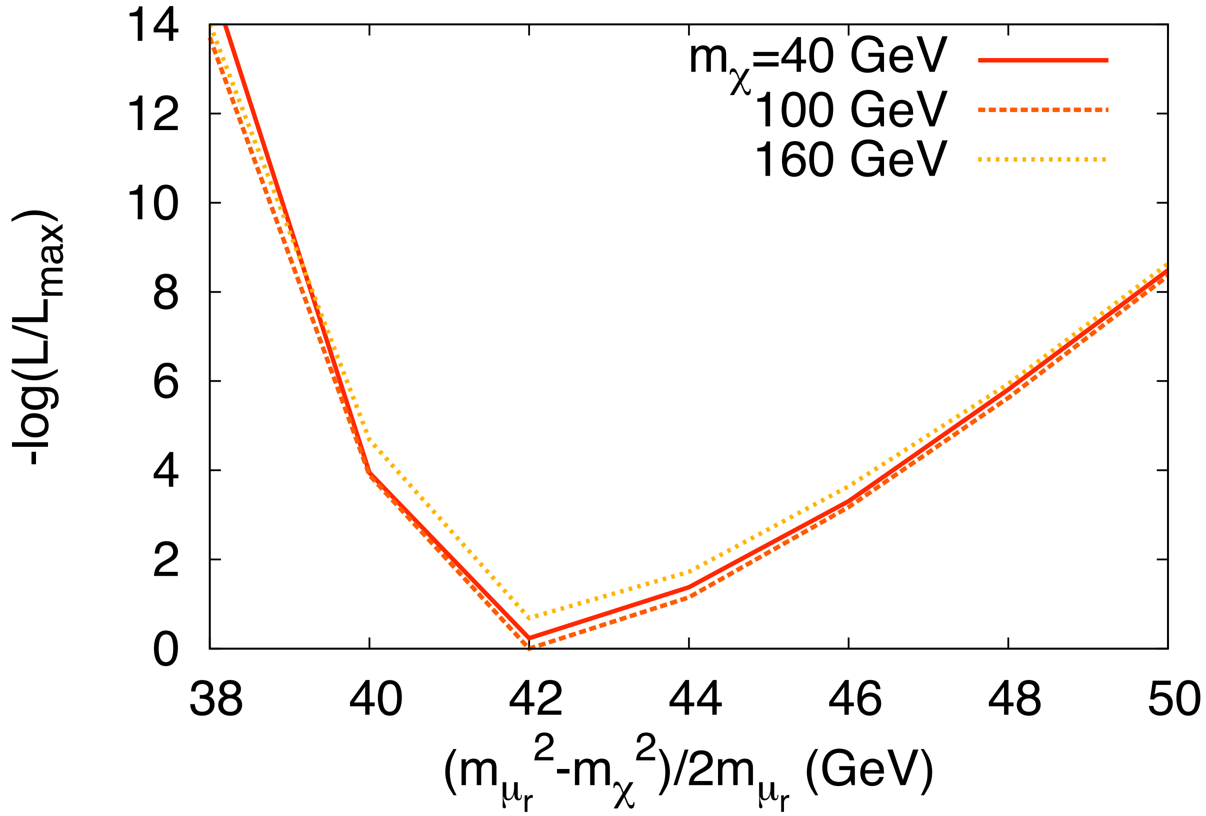
Q: assuming that the masses m_1 and m_2 are the only unknown, what is the maximum significance that can be achieved in measuring these masses at a given luminosity ?

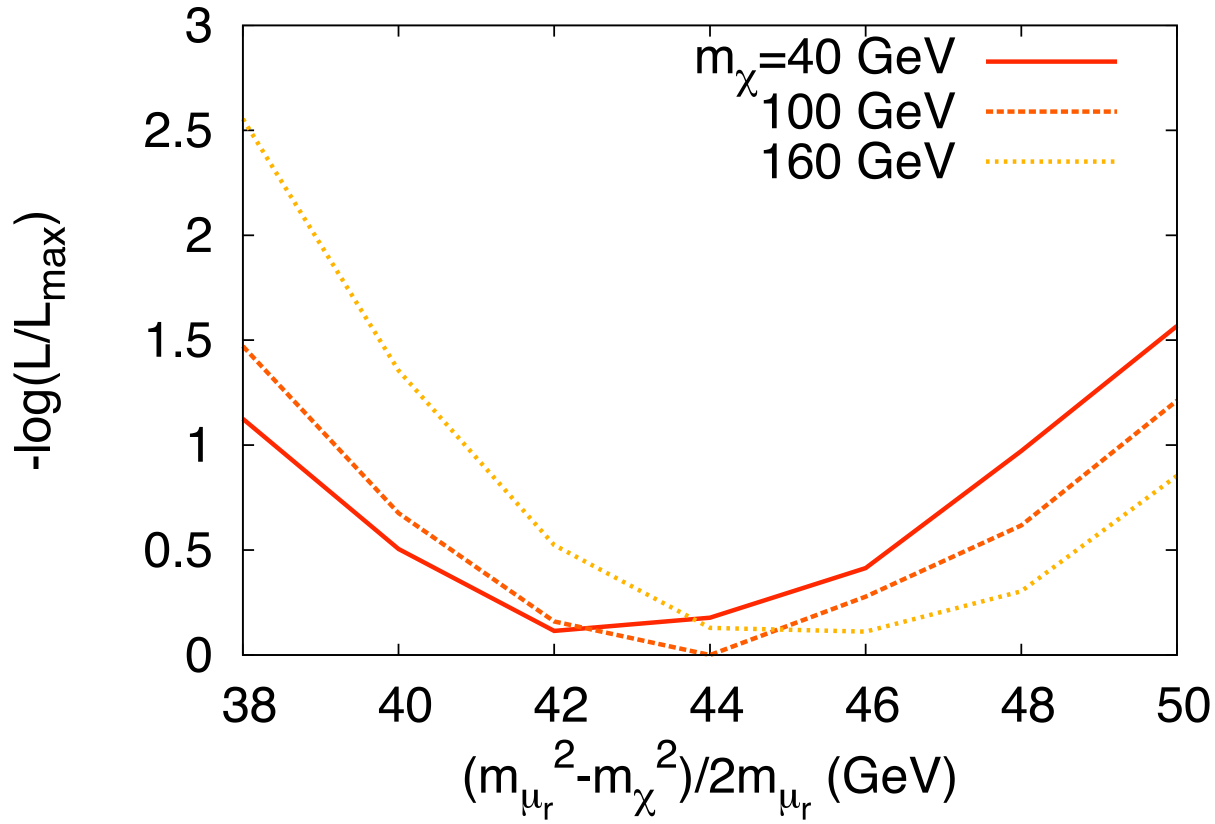
Let us consider a specific example:

$$pp \to (\tilde{\mu}_r^+ \to \mu^+ \tilde{\chi}_1)(\tilde{\mu}_r^- \to \mu^- \tilde{\chi}_1)$$

sample of 50 events with $m_{\tilde{\mu}_r} = 150$ GeV, $m_{\tilde{\chi}_1} = 100$ GeV

- matrix element method (keeps all information)
- keeping only information from $p_T(\mu^+)$, $M(\mu^+, \mu^-)$





Conclusion & perspectives

- the matrix element method is a powerful technique to maximize the significance of a specific measurement
- I presented generic algorithm to evaluate the weights appearing in the matrix element method
- the corresponding tool (MadWeight) is aimed at providing a dynamical reference framework for future analyses, that is convenient for improvements of both the method itself and the integration techniques
- directions of future investigations include
  - new measurements based on the ME method
  - a better understanding of the systematics of theoretical origin
  - study of the maximum significance that can be achieved in specific measurements