MadWeight

automatic event reweighting with matrix elements

Olivier Mattelaer

Université Catholique de Louvain

Pierre Artoisenet: UCL-CP3

Fabio Maltoni: UCL-CP3

Vincent Lemaître: UCL-CP3

Motivation and plan

- motivation : method to maximize the information that you can extract from a sample of events : matrix element method
  - test theoretical hypothesis
  - need a good understanding of the detector
  - we can extract mass, spin, cross section,...
- plan
  - weighting experimental events
  - MadWeight : automatic computation of the weights
  - Examples of application

Weighting experimental events

- matrix element method : weighting events

$$P(\boldsymbol{x}, \alpha) = \qquad |M_\alpha|^2(\boldsymbol{x})$$

where

- $|M_\alpha|^2$ is the squared matrix element

Weighting experimental events

- matrix element method : weighting events

$$P(\boldsymbol{x}, \alpha) = \qquad |M_\alpha|^2(\boldsymbol{y}) W(\boldsymbol{x}, \boldsymbol{y})$$

where

- $|M_\alpha|^2$ is the squared matrix element
- $W(\boldsymbol{x}, \boldsymbol{y})$ is the resolution function
  - $\boldsymbol{x}$: experimental measurements
  - $\boldsymbol{y}$: partonic momenta

Weighting experimental events

- matrix element method : weighting events

$$P(\boldsymbol{x}, \alpha) = \frac{1}{\sigma} \int d\phi(\boldsymbol{y}) \qquad |M_\alpha|^2(\boldsymbol{y}) W(\boldsymbol{x}, \boldsymbol{y})$$

where

- $|M_\alpha|^2$ is the squared matrix element
- $W(\boldsymbol{x}, \boldsymbol{y})$ is the resolution function
  - $\boldsymbol{x}$: experimental measurements
  - $\boldsymbol{y}$: partonic momenta
- $d\phi(\boldsymbol{y})$ is the partonic phase-space measure

Weighting experimental events

- matrix element method : weighting events

$$P(\boldsymbol{x},\alpha) = \frac{1}{\sigma}\int d\phi(\boldsymbol{y}) dw_1 dw_2 f_1(w_1) f_2(w_2) |M_\alpha|^2(\boldsymbol{y}) W(\boldsymbol{x},\boldsymbol{y})$$

where

- $|M_\alpha|^2$ is the squared matrix element
- $W(\boldsymbol{x},\boldsymbol{y})$ is the resolution function
  - $\boldsymbol{x}$: experimental measurements
  - $\boldsymbol{y}$: partonic momenta
- $d\phi(\boldsymbol{y})$ is the partonic phase-space measure
- $f_1(w_1), f_2(w_2)$ are the Parton Distribution Functions

How to evaluate the weight ?

- matrix element method : weighting events

$$P(\boldsymbol{x},\alpha) = \frac{1}{\sigma}\int d\phi(\boldsymbol{y}) dw_1 dw_2 f_1(w_1) f_2(w_2) |M_\alpha|^2(\boldsymbol{y}) W(\boldsymbol{x},\boldsymbol{y})$$

How to evaluate the weight ?

- matrix element method : weighting events

$$P(\boldsymbol{x},\alpha) = \frac{1}{\sigma}\int d\phi(\boldsymbol{y})dw_1dw_2f_1(w_1)f_2(w_2)|M_\alpha|^2(\boldsymbol{y})W(\boldsymbol{x},\boldsymbol{y})$$

- transfer functions : experimental extraction

How to evaluate the weight ?

- matrix element method : weighting events

$$P(\boldsymbol{x},\alpha) = \frac{1}{\sigma}\int d\phi(\boldsymbol{y}) dw_1 dw_2 f_1(w_1) f_2(w_2) |M_\alpha|^2(\boldsymbol{y}) W(\boldsymbol{x},\boldsymbol{y})$$

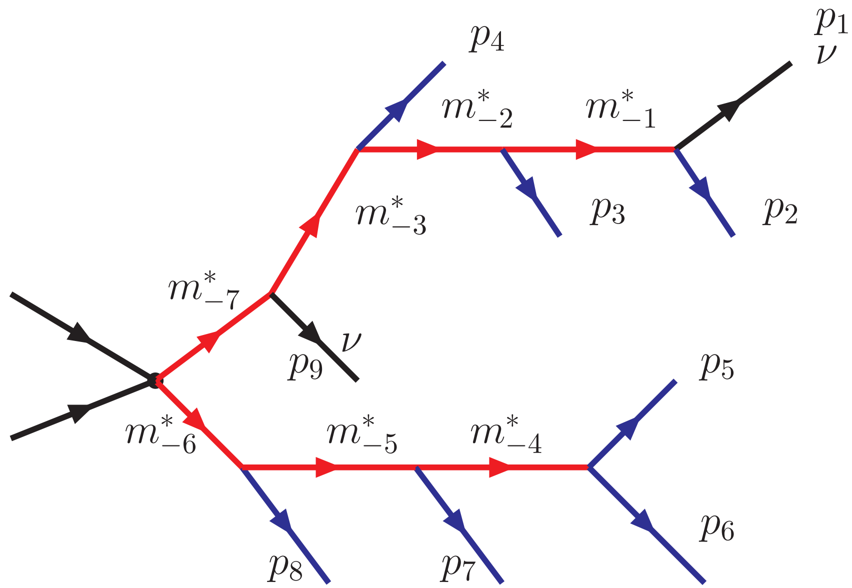
- transfer functions : experimental extraction
- numerical integration : very difficult due to the *structure in peaks* of the integrand

$|M_\alpha(\boldsymbol{y})|^2$: propagators

$$W(\boldsymbol{x},\boldsymbol{y}) \approx \prod_i \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x_i-y_i)^2}{2\sigma_i^2}}$$

Decay chain example

let us consider a specific example of decay chain :

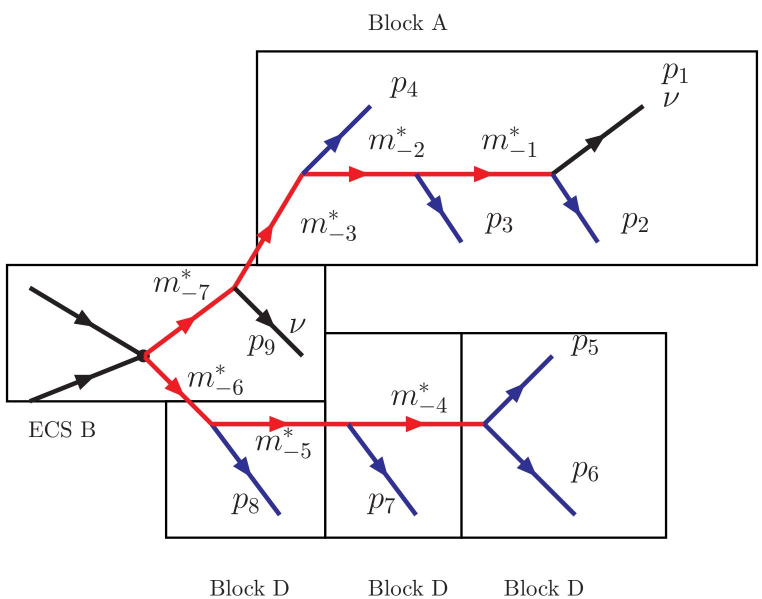


- peaks in $|M_\alpha(\boldsymbol{y})|^2$ controlled by $m^*_{-1}, \ldots, m^*_{-7}$ (7 variables)
- peaks in $W(\boldsymbol{x}, \boldsymbol{y})$ controlled by

$\theta_i, \phi_i, |p_i|^2 \quad i \in \{2, 3, 4, 5, 6, 7, 8\}$ (21 variables)

Decay chain example

- MadWeight decomposes it into blocks corresponding to local change of variables



final parametrization :

$$d\phi = d|\boldsymbol{p}_2|d|\boldsymbol{p}_3|d|\boldsymbol{p}_4|d|\boldsymbol{p}_6| \prod_{i=2}^{8} d\theta_i d\phi_i \prod_{j=1}^{7} dm^{*2}_{-j} \times J$$

MadWeight code

in general in MadWeight algorithm,

- the phase-space is splitted into *blocks*, each of them is associated to a specific local change of variables
- we only consider analytic changes of variables
- 12 different changes of variables are available
- the decomposition into blocks depends on the topology, on the widths of the Breit-Wigner distributions, and on the shape of the resolution function $\longrightarrow$ MadWeight

Examples of Application (I)

- measurement of the top-quark mass in semi-leptonic channel

Examples of Application (I)

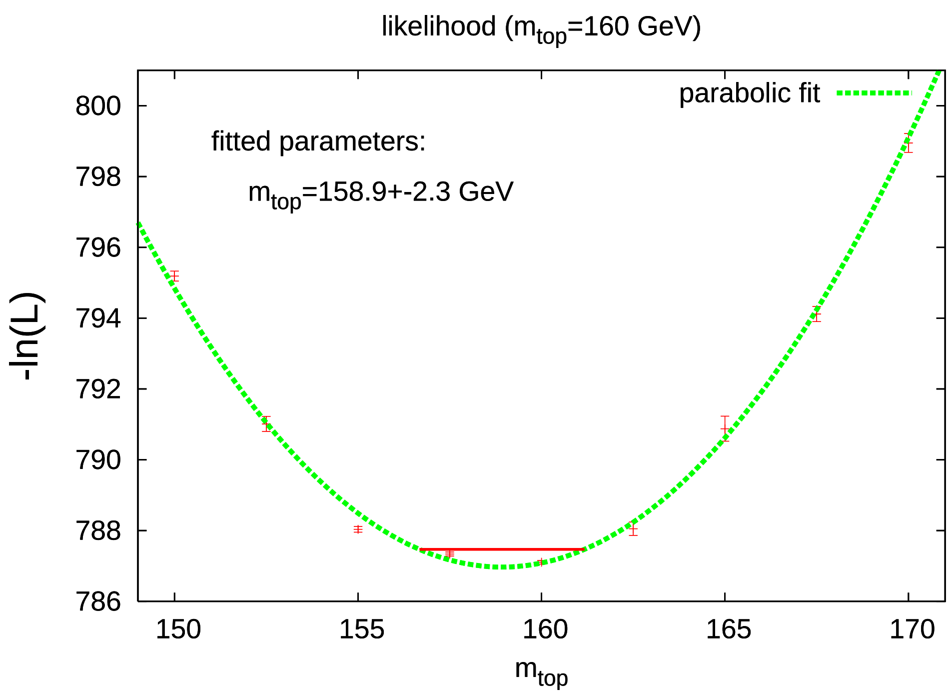
- measurement of the top-quark mass in semi-leptonic channel
- 20 Monte Carlo events (MadGraph/Pythia/PGS)

Examples of Application (I)

- measurement of the top-quark mass in semi-leptonic channel
- 20 Monte Carlo events (MadGraph/Pythia/PGS)
- $L(m_t) = e^{-N\int P(\boldsymbol{x}, m_t) d\boldsymbol{x}} \prod_{i=1}^{N} P(\boldsymbol{x}_i; m_t)$

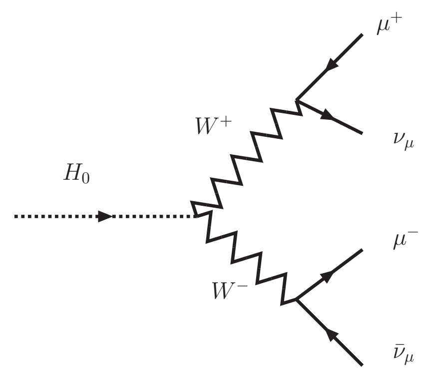
Examples of Application (I)

- measurement of the top-quark mass in semi-leptonic channel
- 20 Monte Carlo events (MadGraph/Pythia/PGS)
- $L(m_t) = e^{-N \int P(\boldsymbol{x}, m_t) d\boldsymbol{x}} \prod_{i=1}^{N} P(\boldsymbol{x}_i; m_t)$
- input : $m_{top} = 160$ GeV, output : $m_{top} = 158.9 \pm 2.3$ GeV



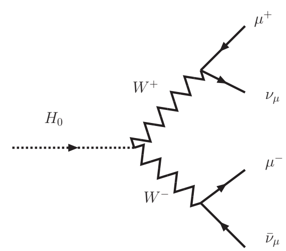
Example of Application (II)

- Higgs mass analysis

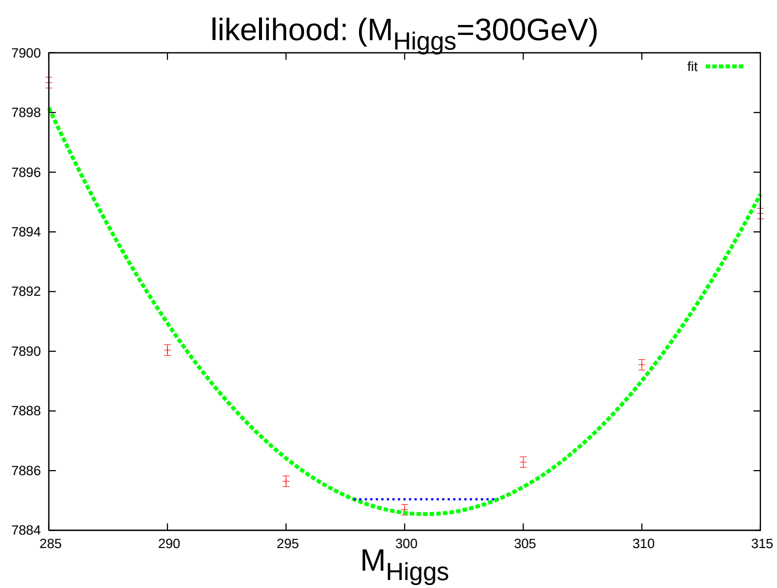


Example of Application (II)

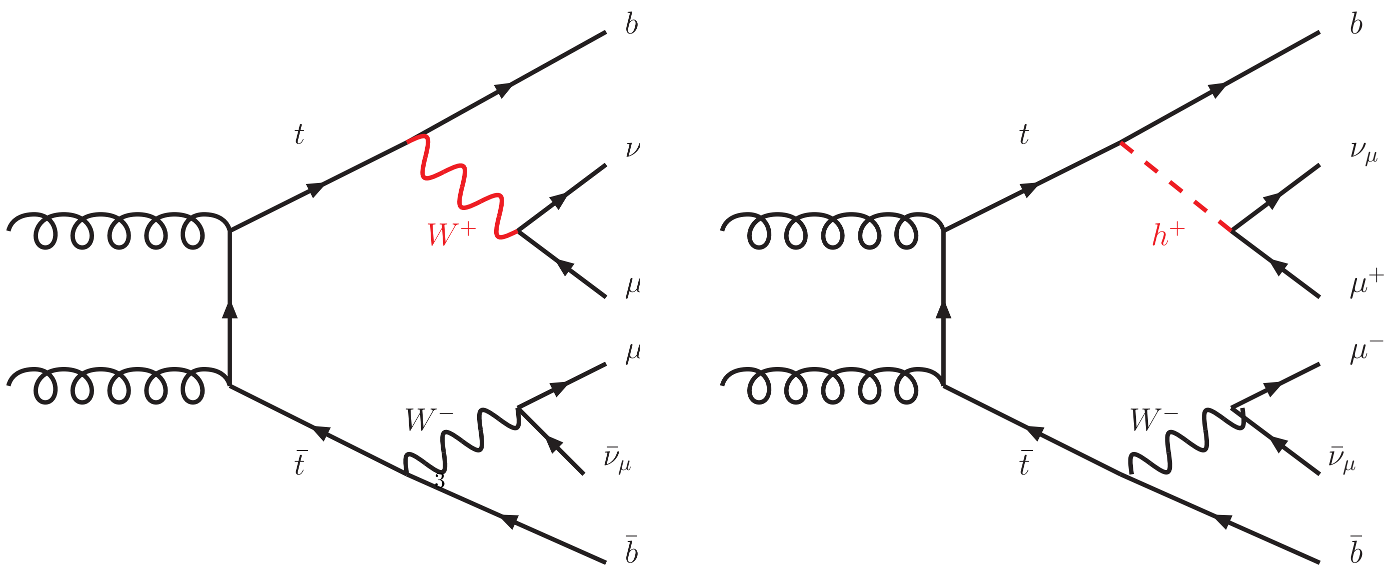
- Higgs mass analysis



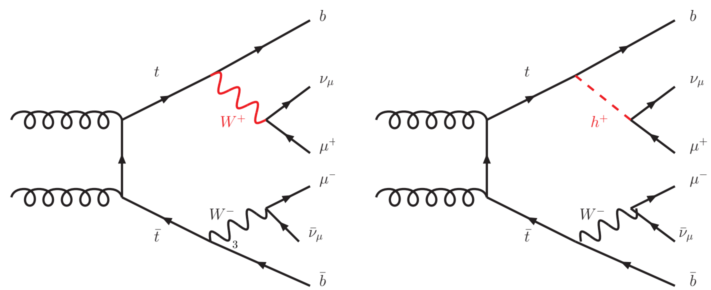
- 500 Monte Carlo events (MadGraph/Pythia/PGS)
- input : $m_{Higgs} = 300$ GeV, output : $m_{Higgs} = 300.9 \pm 3.0$ GeV



Charged Higgs : Discriminant

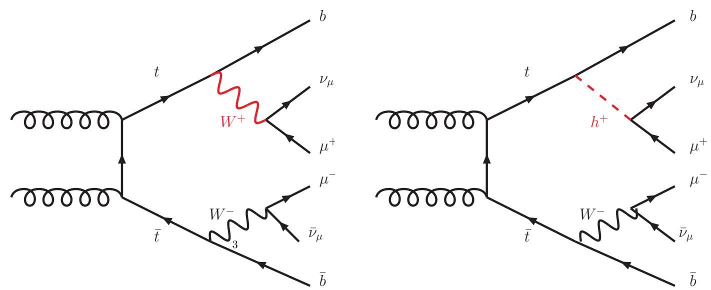


Charged Higgs : Discriminant



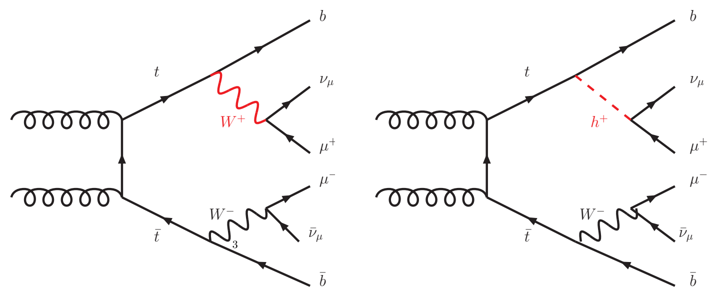
- $M_{H^+} = 100$GeV

Charged Higgs : Discriminant

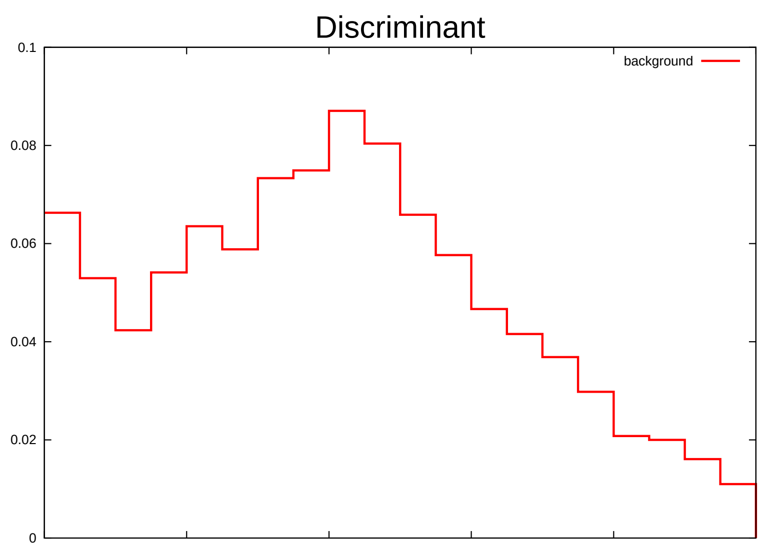


- $M_{H^+} = 100$GeV
- $d = \frac{P_S}{P_S + P_B}$

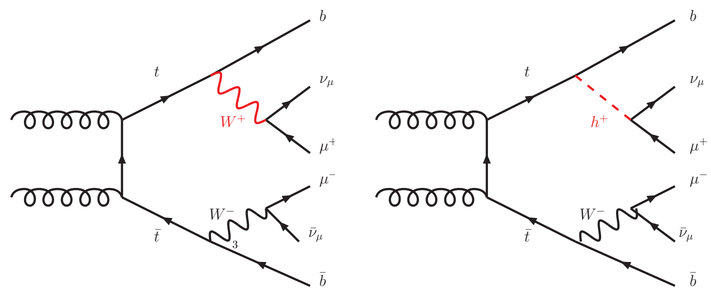
Charged Higgs : Discriminant



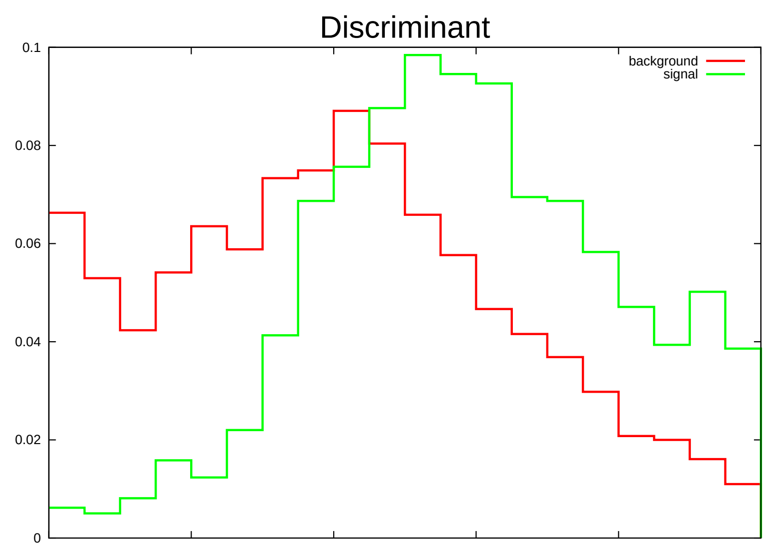
- $M_{H+} = 100$GeV
- $d = \frac{P_S}{P_S+P_B}$



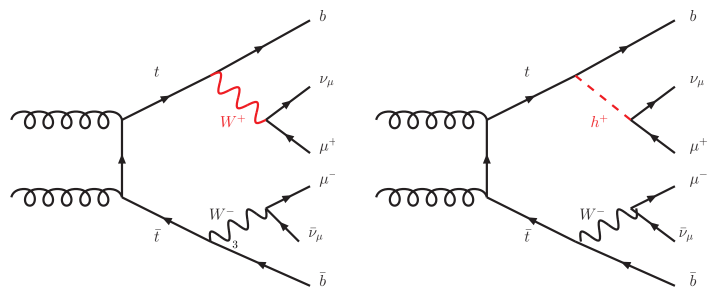
Charged Higgs : Discriminant



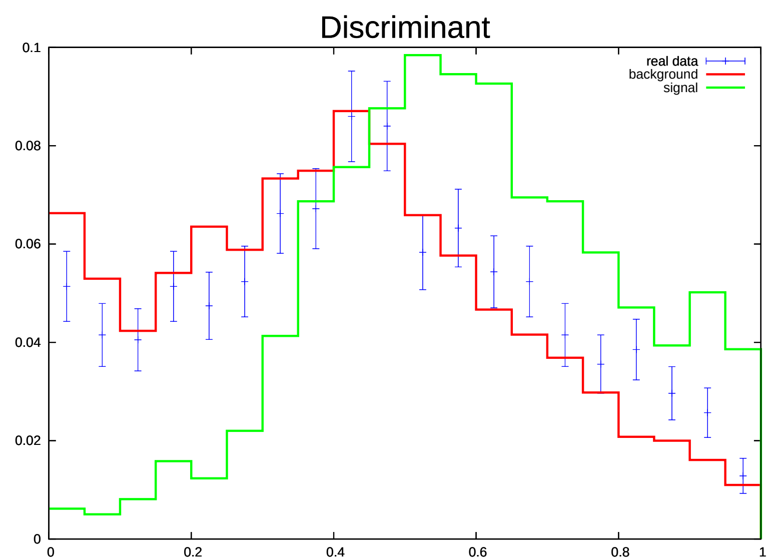
- $M_{H^+} = 100$GeV
- $d = \frac{P_S}{P_S + P_B}$



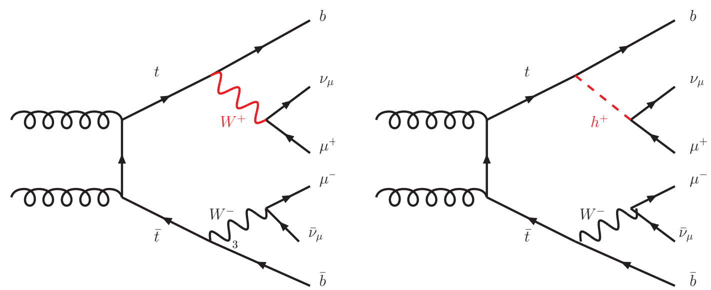
Charged Higgs : Discriminant



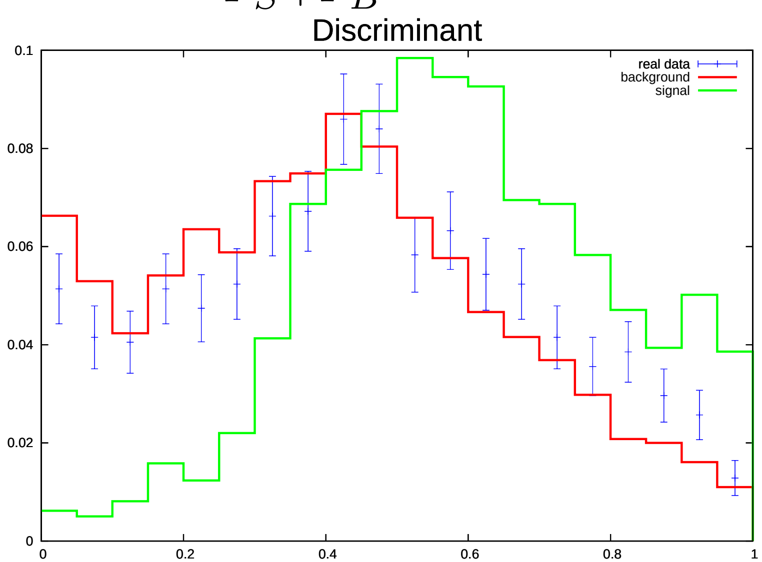
- $M_{H^+} = 100$GeV
- $d = \frac{P_S}{P_S + P_B}$



Charged Higgs : Discriminant

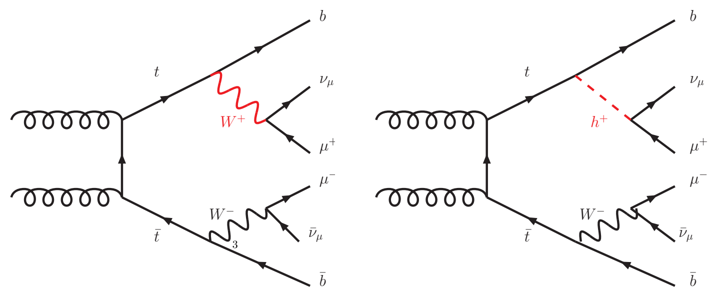


- $M_{H^+} = 100\text{GeV}$
- $d = \frac{P_S}{P_S + P_B}$

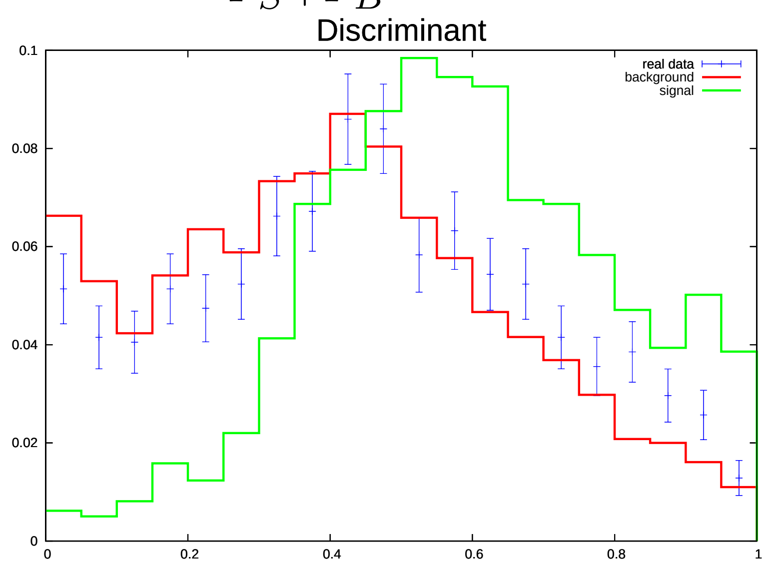


- 750 background events
- 262 signal events
- $r = 25.9\%$

Charged Higgs : Discriminant



- $M_{H^+} = 100$GeV
- $d = \frac{P_S}{P_S+P_B}$



- 750 background events
- 262 signal events
- $r = 25.9\%$
- $r_{mes} = 21 \pm 4\%$

Conclusion

- the Matrix Element method provides the best discriminator on an event-by-event basis
- both theoritical ($|M|^2$) and experimental ($\boldsymbol{x},\ W(\boldsymbol{x},\boldsymbol{y})$) information is used
- the computation of the weights requires a specific phase space generator : MadWeight
  - finds the best phase-space parametrisation(s)
  - fully automatic
  - works for "any" decay chain
- code available on madgraph.phys.ucl.ac.be (on the download page)

End

Weighting experimental events

- combine the weights into a likelihood

$$L(\alpha) = \prod_{i=1}^{N} P(\boldsymbol{x}_i; \alpha)$$

Weighting experimental events

- combine the weights into a likelihood

$$L(\alpha) = e^{-N\int P(\boldsymbol{x},\alpha)d\boldsymbol{x}} \prod_{i=1}^{N} P(\boldsymbol{x}_i;\alpha)$$

Weighting experimental events

- combine the weights into a likelihood

$$L(\alpha) = e^{-N \int P(\boldsymbol{x},\alpha) d\boldsymbol{x}} \prod_{i=1}^{N} P(\boldsymbol{x}_i; \alpha)$$

the best estimation of α is the one that maximizes L

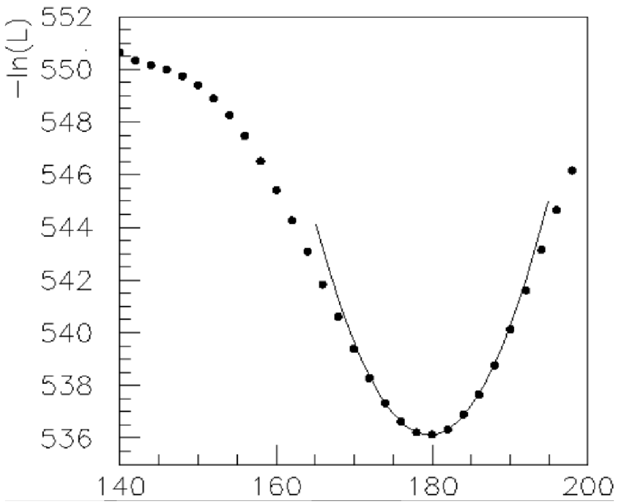
Weighting experimental events

- combine the weights into a likelihood

$$L(\alpha) = e^{-N\int P(\boldsymbol{x},\alpha)d\boldsymbol{x}} \prod_{i=1}^{N} P(\boldsymbol{x}_i;\alpha)$$

the best estimation of α is the one that maximizes L

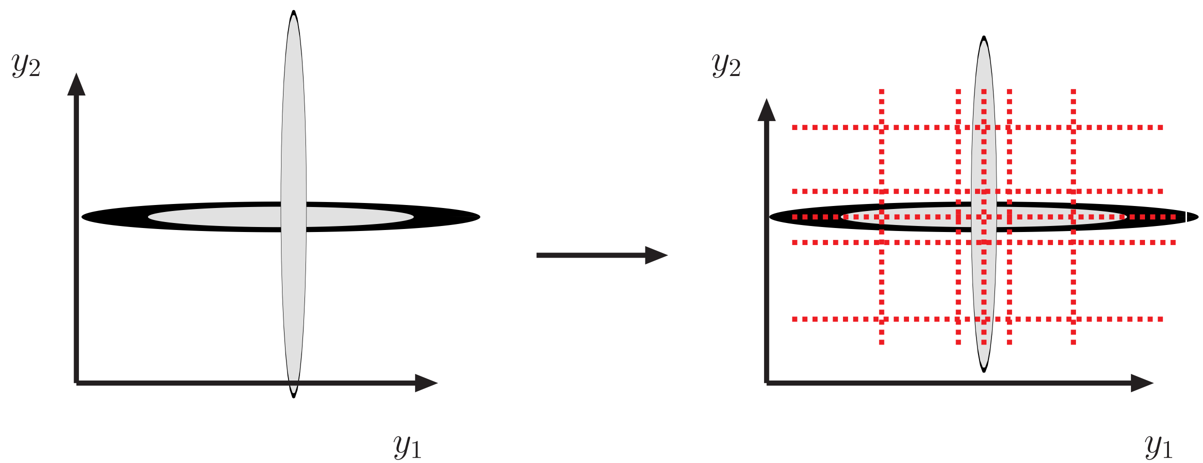
example : top-quark mass measurement from $t\bar{t} \to l^+X$ sample at DØ



- 72 events
- $M_{top} = 180.1 \pm 3.6_{stat} \pm 4.0_{sys}$ GeV
- J. Estrada : Phd dissertation, University of Rochester (2001)

Monte Carlo technics

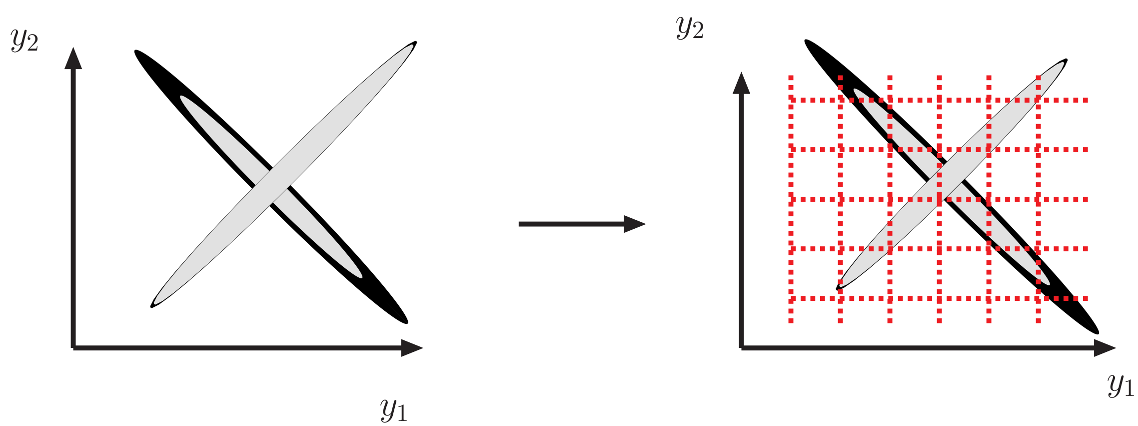
- efficiency of an adaptative MC integration :
  - case 1 : any peak is aligned along a single direction of the P-S parametrization



$\longrightarrow$ the adaptative Monte-Carlo P-S integration is very efficient

Monte Carlo technics

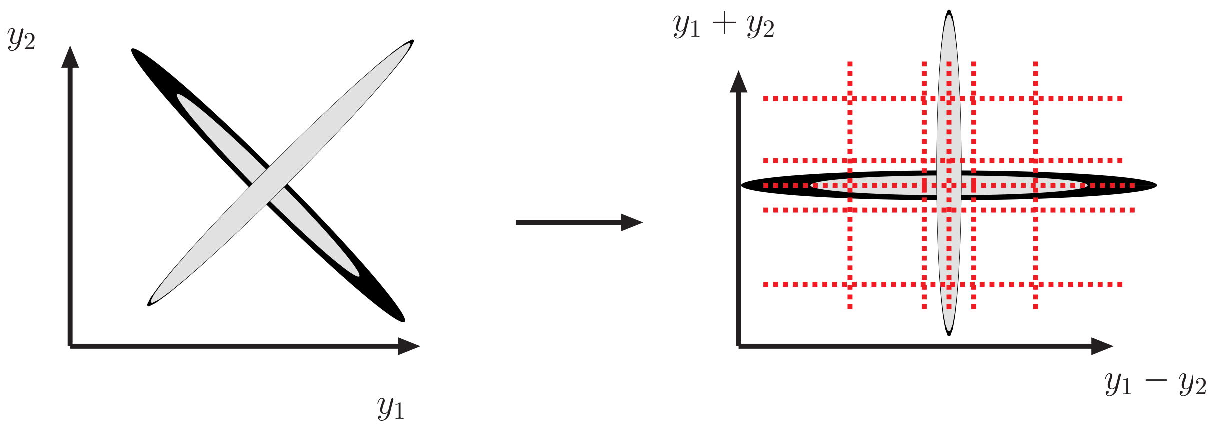
- efficiency of an adaptative MC integration :
  - case 2 : some peaks are not aligned along a single direction of the P-S parametrization



$\longrightarrow$ the adaptative Monte-Carlo P-S integration converges slowly

Monte Carlo technics

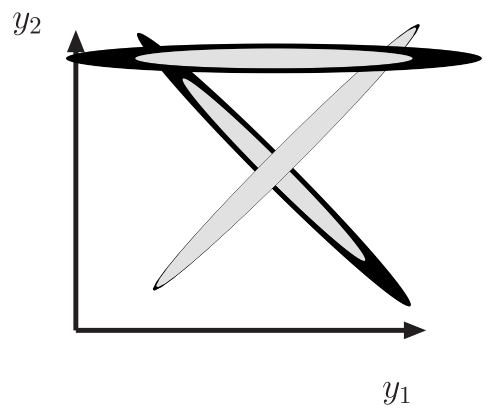
- efficiency of an adaptative MC integration :
  - possible solution : perform a change of variables



$\longrightarrow$ the adaptative Monte-Carlo P-S integration is very efficient

Monte Carlo technics

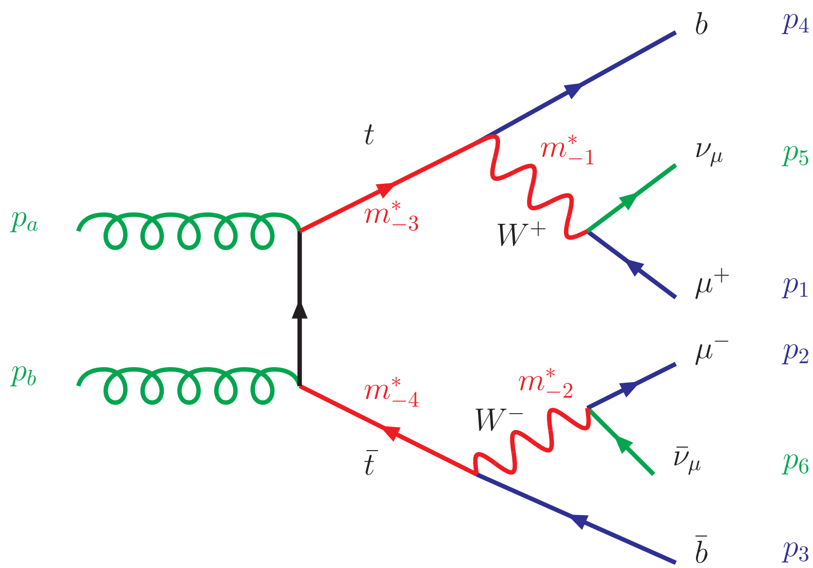
- efficiency of an adaptative MC integration :
  - case 3 : there are more peaks than phase-space variables



$\longrightarrow$ the efficiency depends of the shape, relative position, ... of the peaks

Phase-space generation

- ideal case : $t\bar{t}$



- peaks in $|M_\alpha(\boldsymbol{y})|^2$ controlled by $m^*_{-1}, \ldots, m^*_{-4}$ (4 variables)
- peaks in $W(\boldsymbol{x}, \boldsymbol{y})$ controlled by $\theta_i, \phi_i, |p_i|^2 \quad i \in \{1, 2, 3, 4\}$ (12 variables)
- $dim[d\phi] = 16$, $\rightarrow$ each peak can be aligned along a single variable of integration

Phase-space generation

which parametrization do we use ?

- natural parametrization

$$d\phi = \prod_{i=1}^{4} \frac{d^3p_i}{(2\pi)^3 2E_i} \prod_{i=5}^{6} \frac{d^3p_i}{(2\pi)^3 2E_i} dx_1 dx_2 \delta^4 \left(p_a + p_b - \sum_j p_j \right)$$

where all the peaks in $W(\boldsymbol{x}, \boldsymbol{y})$ are aligned

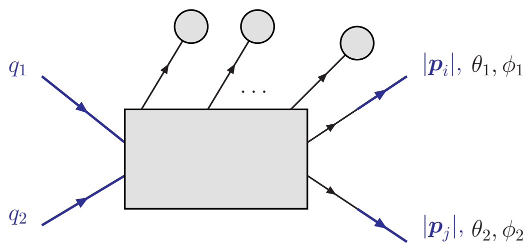
- we apply local changes of variables to reach the parametrization

$$d\phi = \prod_{i=1}^{4} d\theta_i d\phi_i d|\boldsymbol{p}_i| \prod_{j=1}^{4} dm^{*2}_{-j} \times J$$

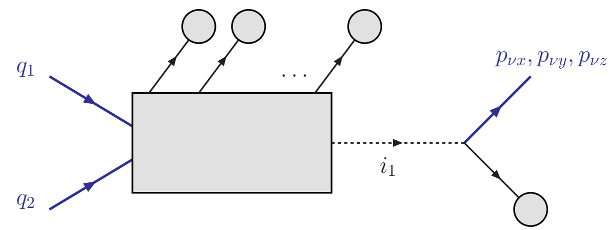
where each Breit-Wigner distribution is also aligned

MadWeight : changes of variables

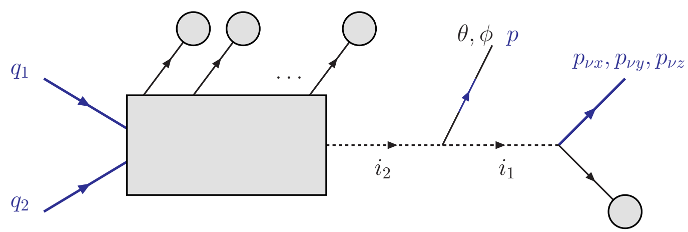
- changes of variables to restore energy momentum conservation



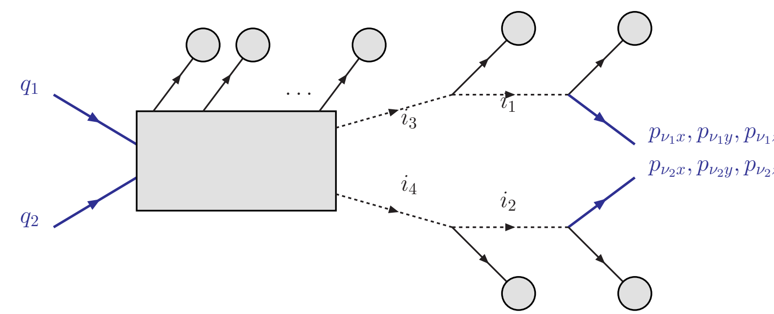
Class A



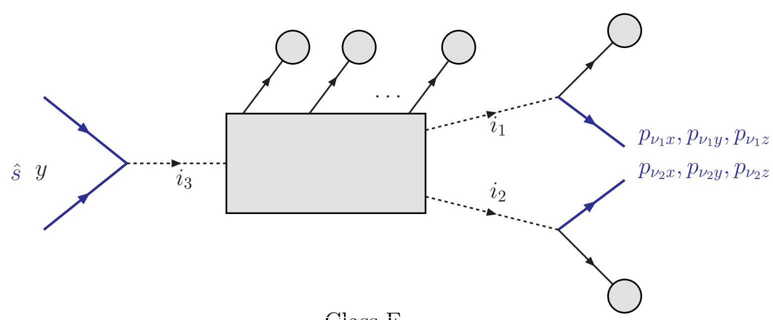
Class B



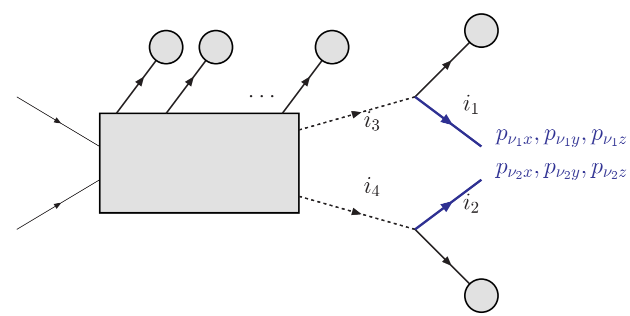
Class C



Class D



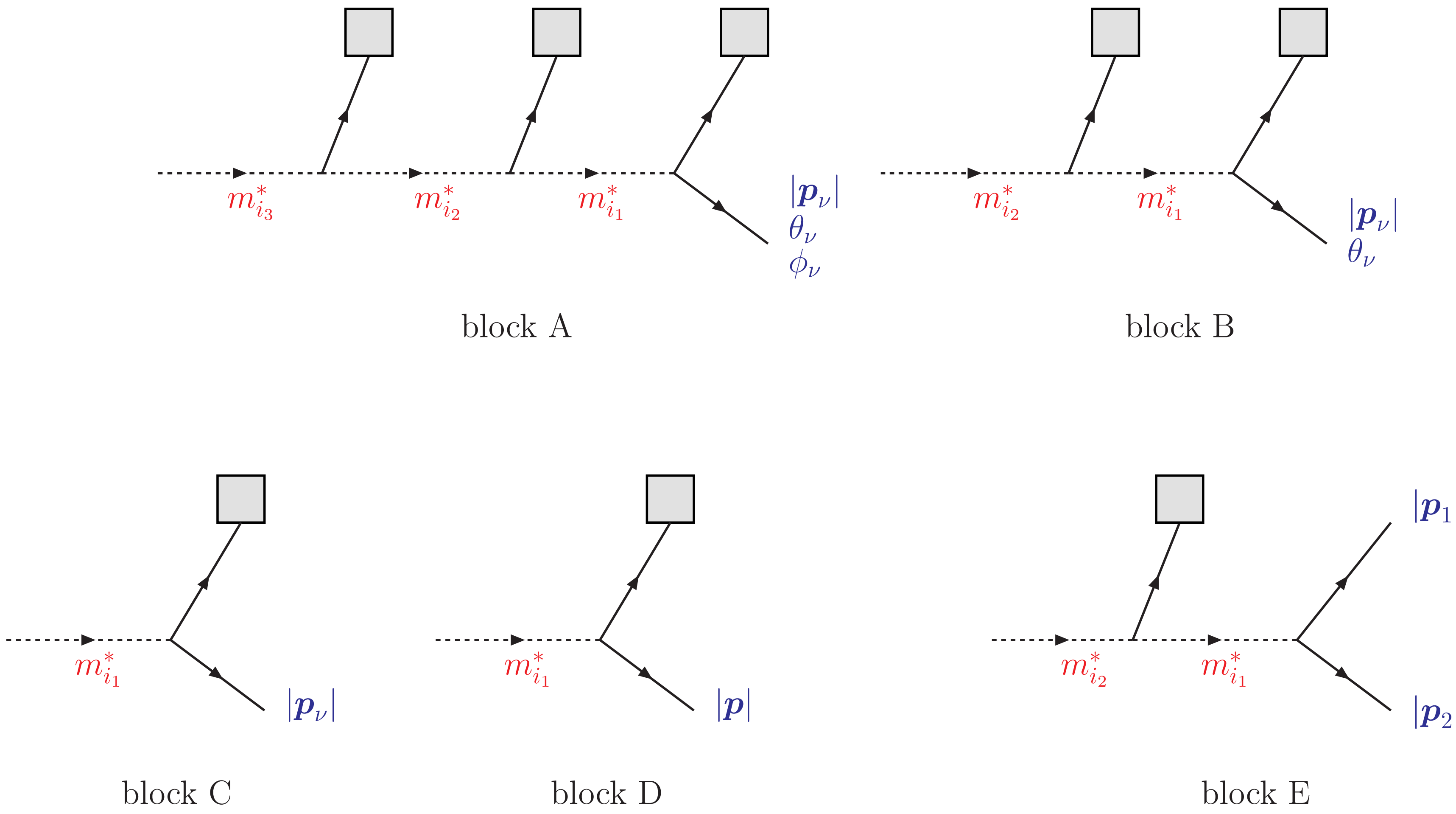
Class E



Class F

MadWeight : changes of variables

- auxiliary changes of variables :



Weighting experimental events

- advantages :
  - it takes into account the full matrix element (in particular spin-correlation effects)
  - resolution of the detector is included
  - it is particularly usefull for processes with missing particles
- drawbacks :
  - the evaluation of the weight is time-consuming compare to other methods
  - what are the systematics errors ?