

Evolving Event Reconstruction via Probability Matching for the W boson Mass Measurement

Won Sang Cho (CTPU@IBS, Daejeon, Korea)

in collaboration with

Kiwoon Choi

Sung-Hak Im

Chan Beom Park

Physics Challenges in the face of LHC-14

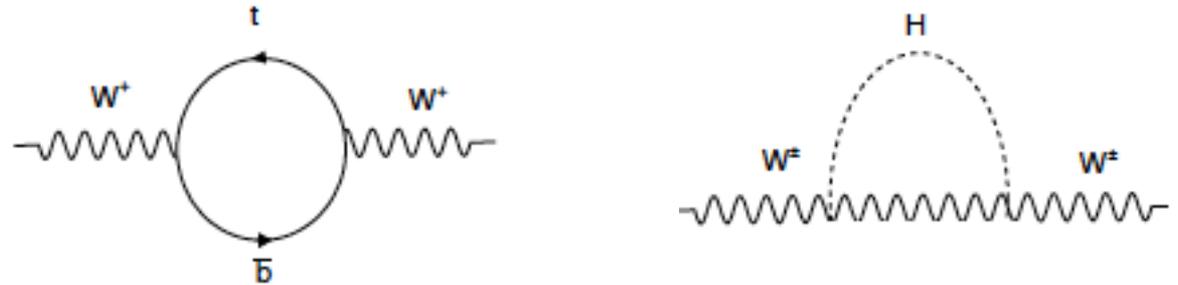
IFT, Madrid, 15 – 26 September 2014

Precise W mass measurement – Motivation

- In the Electroweak theory of the SM, W mass is predictable

$$M_W = \sqrt{\frac{\pi\alpha}{\sqrt{2}G_F \sin^2 \theta_W} \frac{1}{\sqrt{1 - \Delta r}}}$$

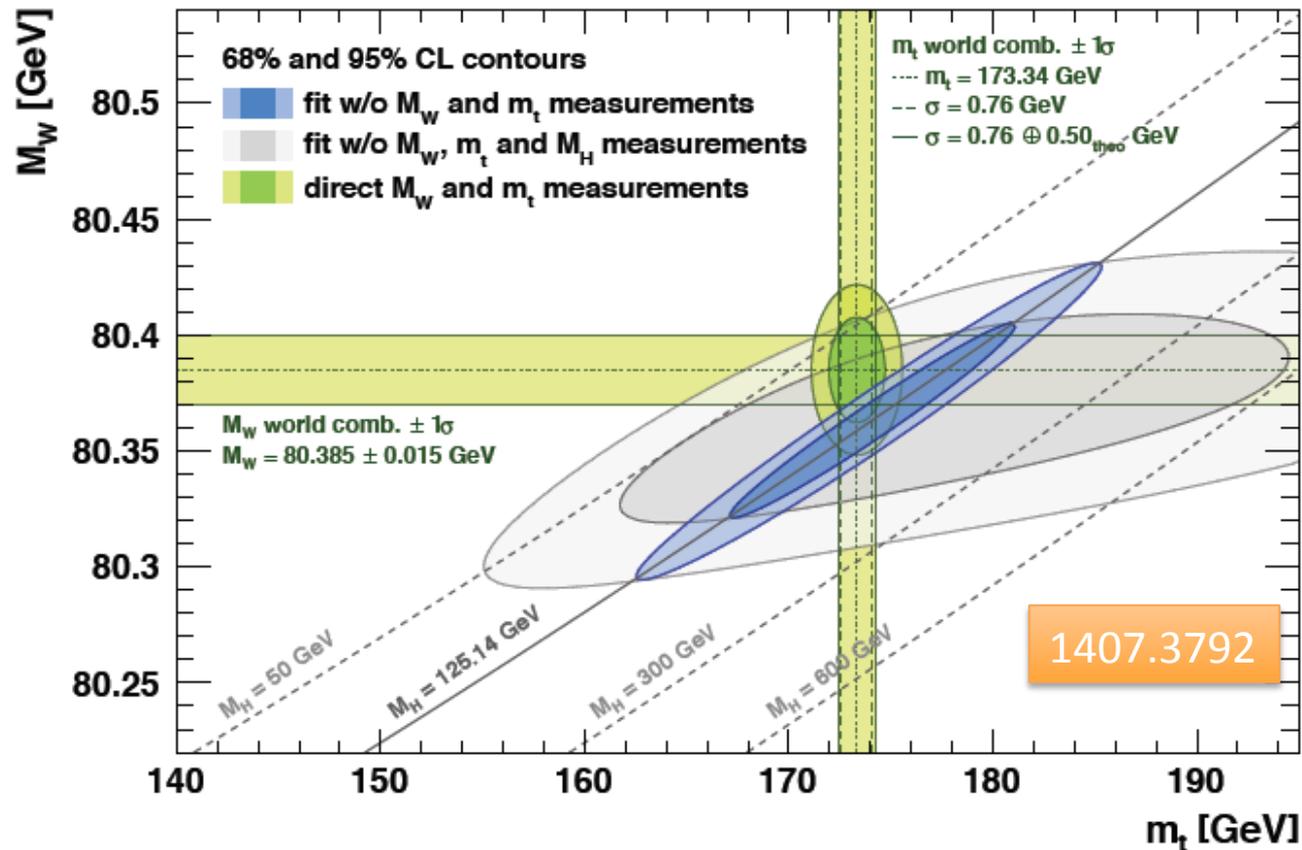
One loop radiative correction Δr to W mass depends quadratically on the top mass, and logarithmically on Higgs mass.



Measuring M_W and M_t , one can extract the information on M_H , vice versa.

Some new physics can contribute on the M_W

With the discovery of the Higgs and direct Higgs mass measurement..



- Global EW fitting with $M_H(\text{exp})$ indirectly estimated M_W and $\Delta M_W \sim 8$ MeV
- Compared to the $\Delta M_t(\text{exp}) \sim 0.9$ GeV, precision on $\Delta M_W(\text{exp}) \sim 5$ MeV is required assuming equivalent contribution on Higgs mass precision.
- Both values motivate some improvements on current experimental precision on $\Delta M_W \sim 15$ MeV \rightarrow Precision M_W measurement is one of the key EW precision tests for the SM consistency.

Precision W mass measurement – Status

- World average : 80385 ± 15 MeV (sys+stat)
- CDF result (80375 MeV, 2012)

D0 (80387 MeV 2012)

Source	Uncertainty (MeV)
Lepton energy scale and resolution	7
Recoil energy scale and resolution	6
Lepton removal from recoil	2
Backgrounds	3
Experimental subtotal	10
Parton distribution functions	10
QED radiation	4
$p_T(W)$ model	5
Production subtotal	12
Total systematic uncertainty	15
W-boson event yield	12
Total uncertainty	19

Source	Uncertainty (MeV)
Electron energy calibration	16
Electron resolution model	2
Electron shower modeling	4
Electron energy loss model	4
Recoil energy scale and resolution	5
Electron efficiencies	2
Backgrounds	2
Experimental subtotal	18
Parton distribution functions	11
QED radiation	7
$p_T(W)$ model	2
Production subtotal	13
Total systematic uncertainty	22
W-boson event yield	13
Total uncertainty	26

→ Systematic uncertainty, (CDF:15, D0: 22) MeV > Statistical uncertainty, (CDF:22, D0 : 13) MeV

→ How to reduce the systematic uncertainties?

Systematic uncertainties on MW measurement

1) Experimental sources

- 1) **Lepton energy scale/energy resolution/shower model/energy loss/efficiencies**
- 2) **Backgrounds**
- 3) **Recoil model**

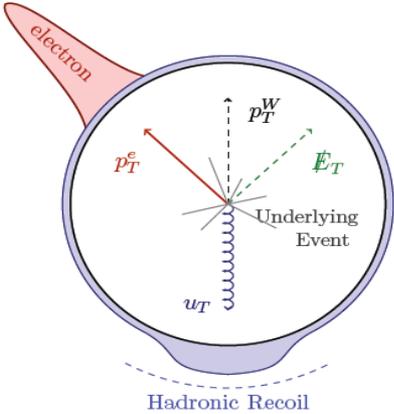
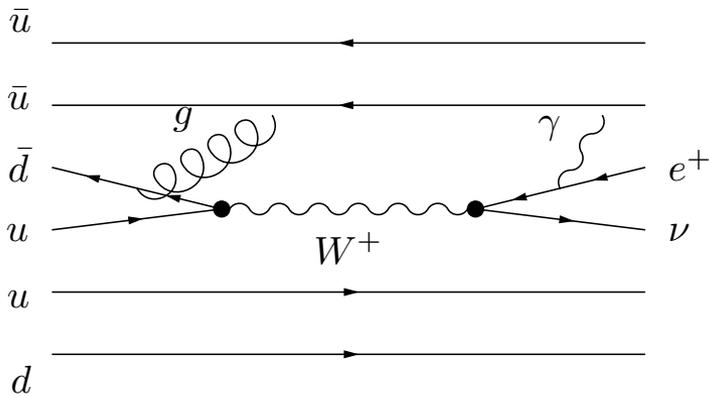
: limited by the size of the Z→ll sample => Enhanced event statistics can help.

2) W and Z production and decay models

- 1) **PDF** : Boost of the W along the longitudinal beam direction does not change a MT value. However uncertainty on the PDF affects on event acceptance rate, in particular with respect to the imposed pseudo-rapidity range, so MT distribution is also changed.
- 2) **Transverse momentum of W boson** : PT(l) distribution is very sensitive on the non-zero PT(W). MT distribution is also affected by W boson's PT.
- 3) **Photon radiation (QED)** : final state radiation from lepton

Measurement strategy and sources of systematic uncertainty

- W-boson production and decay



- Template method, using M_T , $PT(l)$ and $PT(\nu)$ variables

$$M_T = \sqrt{2P_T^l P_T^\nu (1 - \cos(\phi^l - \phi^\nu))}$$

- The distributions of the variables are unstable w.r.t the momentum profile of the mother particle, W production
 - Both M_T and $PT(l)$ distributions are not invariant under $PT(W)$ and PDF.
 - One solution is to do precise calculation and simulation on the W production (and decay) to reduce the systematic uncertainties literally.

Our approach :

Engineer the transverse mass toward the variable mostly inherits the properties of BW resonance, and try to apply the similar analysis strategy like the M_Z measurement !

Consider an 'sampled invariant mass' of W , where the missing momenta along beam direction is assigned by random variable

$$\begin{aligned}
 \tilde{M}_W(p^l, p_T^\nu; \tilde{p}_z^\nu)^2 &\equiv (p^l + \tilde{p}^\nu)^2 \quad \text{with} \quad \tilde{p}^\nu = (\sqrt{\mathbf{p}_T^{\nu 2} + \tilde{p}_z^{\nu 2}}, p_x^\nu, p_y^\nu, \tilde{p}_z^\nu), \\
 &= 2 \left[|\mathbf{p}^l| \sqrt{\mathbf{p}_T^{\nu 2} + \tilde{p}_z^{\nu 2}} - \mathbf{p}_T^l \cdot \mathbf{p}_T^\nu - p_z^l \tilde{p}_z^\nu \right], \\
 &= 2|\mathbf{p}_T^l| |\mathbf{p}_T^\nu| \left[\cosh(\eta^l - \tilde{\eta}^\nu) - \cos \phi_{l\nu} \right],
 \end{aligned}$$

In this picture, M_T is also one sampled invariant mass of W , but which is specially optimized to minimize the mass scale.

$$M_T(W) \equiv \min_{\tilde{p}_z} \tilde{M}_W \quad \rightarrow \quad \tilde{\eta}^\nu = \eta^l$$

We have no more constraint in an event by event basis. Then which information/constraints do we use for engineering toward a better solution ?

We use whole event sample in search for the solution set which most resembles the shape of physical distribution – (weighted by PDF and convoluted by resolution effect) BW resonance + modeled BG.

Engineering technology : Genetic Algorithm for the optimization in huge solution spaces

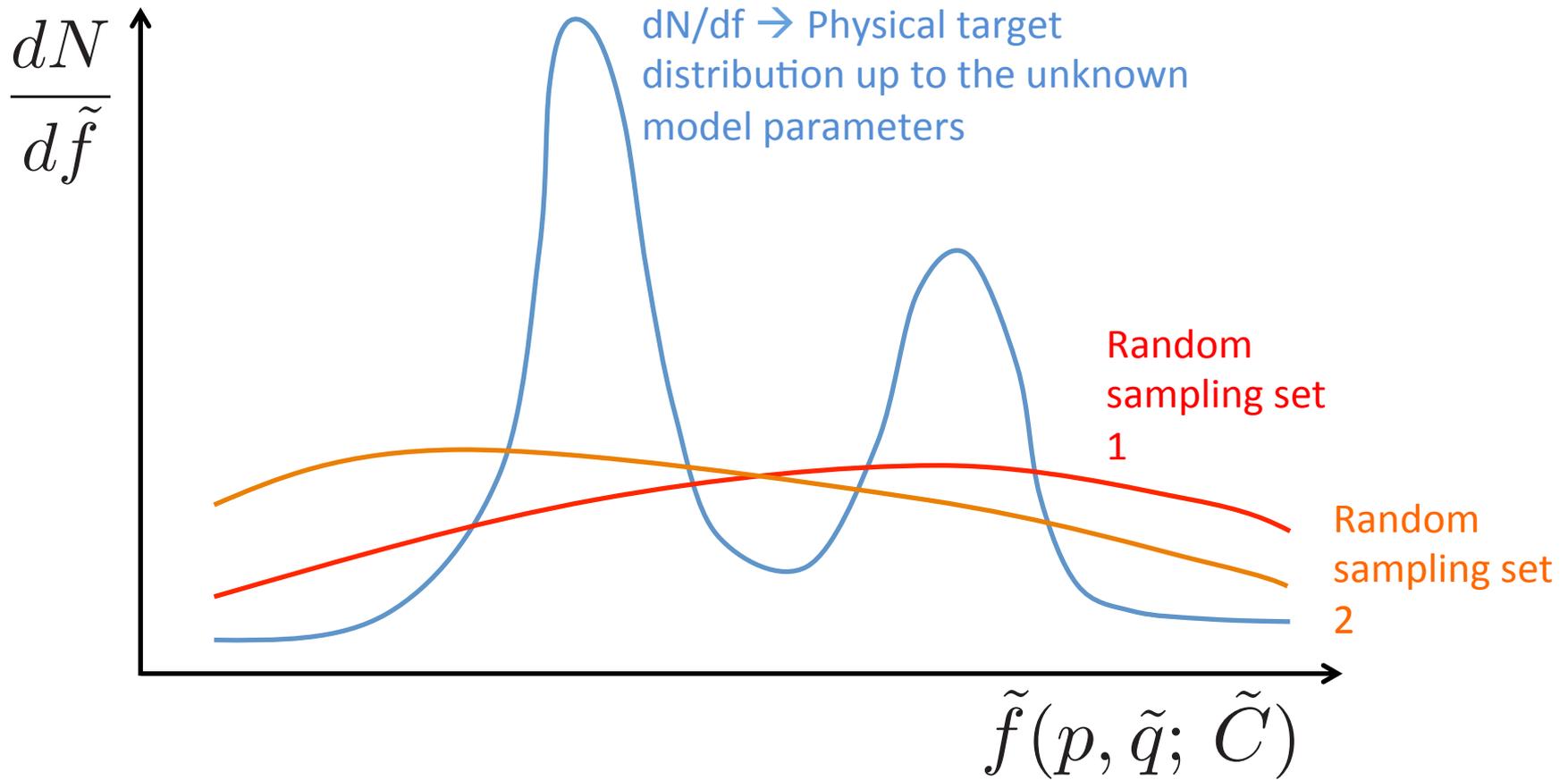
General Situation for HEP Event Reconstruction

Event ID	1	2	...	N
Known	$\{p_{vis}^1\}$	$\{p_{vis}^2\}$...	$\{p_{vis}^N\}$
Unknown	$\{q_{inv}^1\}, C^1$	$\{q_{inv}^2\}, C^2$...	$\{q_{inv}^N\}, C^N$
Observable	$f(\{p_{vis}^1\}, \{q_{inv}^1\}; C^1)$	$f(\{p_{vis}^2\}, \{q_{inv}^2\}; C^2)$...	$f(\{p_{vis}^N\}, \{q_{inv}^N\}; C^N)$

$\{p_{vis}^i\}$: set of visible momenta of the event-(i)

$\{q_{inv}^i\}$: set of unknown momenta of the event-(i)

$C^i \in \{1, \dots, N_{combi}\}$: Particle assignment ID



Dimension of the random search space :

$$\{(\{\tilde{q}^i\}, \tilde{C}^i) \mid i = 1..N_{event}\}$$

$$N_{par} = N_{event} \times (N_q + N_{combi})$$

Basic procedures for **one sampled solution** of the unknowns

1. Consider **a random sampling of the unknowns** for all event blocks.

$$\{(\{\tilde{q}^i\}, \tilde{C}^i) \mid i = 1..N_{event}\}$$

2. Select a good collider observable (non-reconstructable as q & C are unknowns) :

$$f(p, q; C)$$

3. Calculate a sampled functional value using the sampled unknown values :

$$\tilde{f}(p, \tilde{q}; \tilde{C})$$

4. Estimate the fitness as the distance between the sampled distribution and the physical distribution where unknown model parameters are optimized.

$$Fitness = Distance\ between\ \frac{dN}{df}(\vec{\alpha}_0 = (m_0, \Gamma_0, \dots))\ and\ \frac{dN}{d\tilde{f}}$$

The Genetic Algorithm does operations, generation by generation..

1. (Initially in the 1st generation) Consider a population of solutions (= a set of individual random solutions)
2. Estimate the fitness of all solutions
3. Sort the individuals by their ranking of the fitness
4. Select a portion of high ranked solutions, and discard others
5. Using reproduction algorithm(crossover/mutation of genes) of GA, it produces children solutions, filling the missing population.
6. Estimate the fitness of new individual solutions, and sort them again by their ranking.
7. If not converged, goto #5 again as the next generation.

The Genetic Algorithm does operations, generation by generation..

1. (Initially in the 1st generation) Consider a population of solutions (= a set of individual random solutions)
2. Estimate the fitness of all solutions
3. Sort the individuals by their ranking of the fitness
4. Select a portion of high ranked solutions, and discard others
5. Using reproduction algorithm(crossover/mutation of genes) of GA, it produces children solutions, filling the missing population.
6. Estimate the fitness of new individual solutions, and sort them again by their ranking.
7. If not converged, goto #5 again as the next generation.

→ Then, the best solution in each generation-(k), fitted into the physical target distribution which is optimized w.r.t unknown model parameters, evolves toward the true distribution,

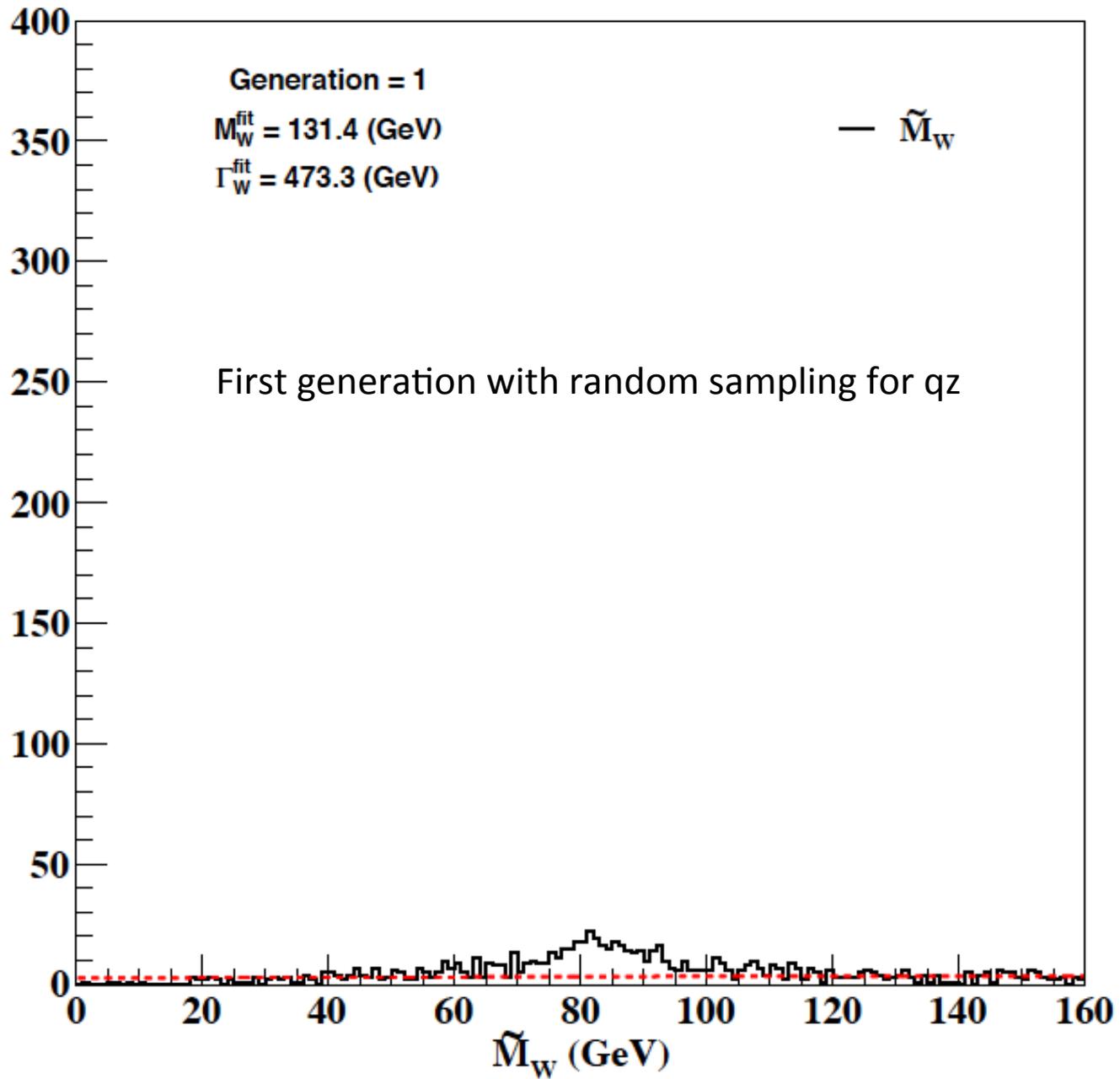
$$\lim_{k \rightarrow \infty} \frac{dN}{df}(\vec{\alpha}_k) = \frac{dN}{df}(\alpha_{true}) \sim \frac{dN}{d\tilde{f}_k}$$

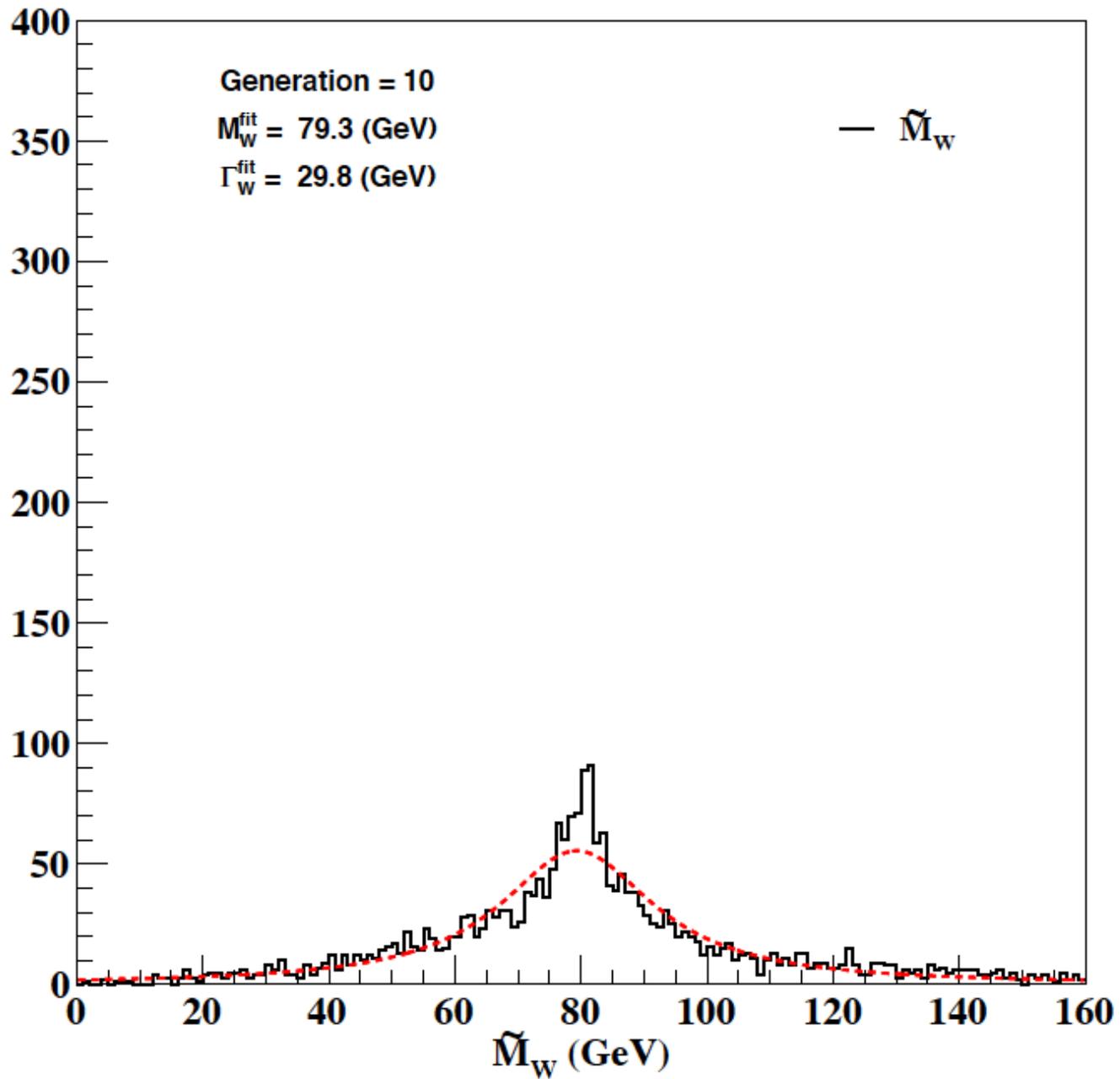
if the shape of our physical distribution possesses enough non-trivial physical information.

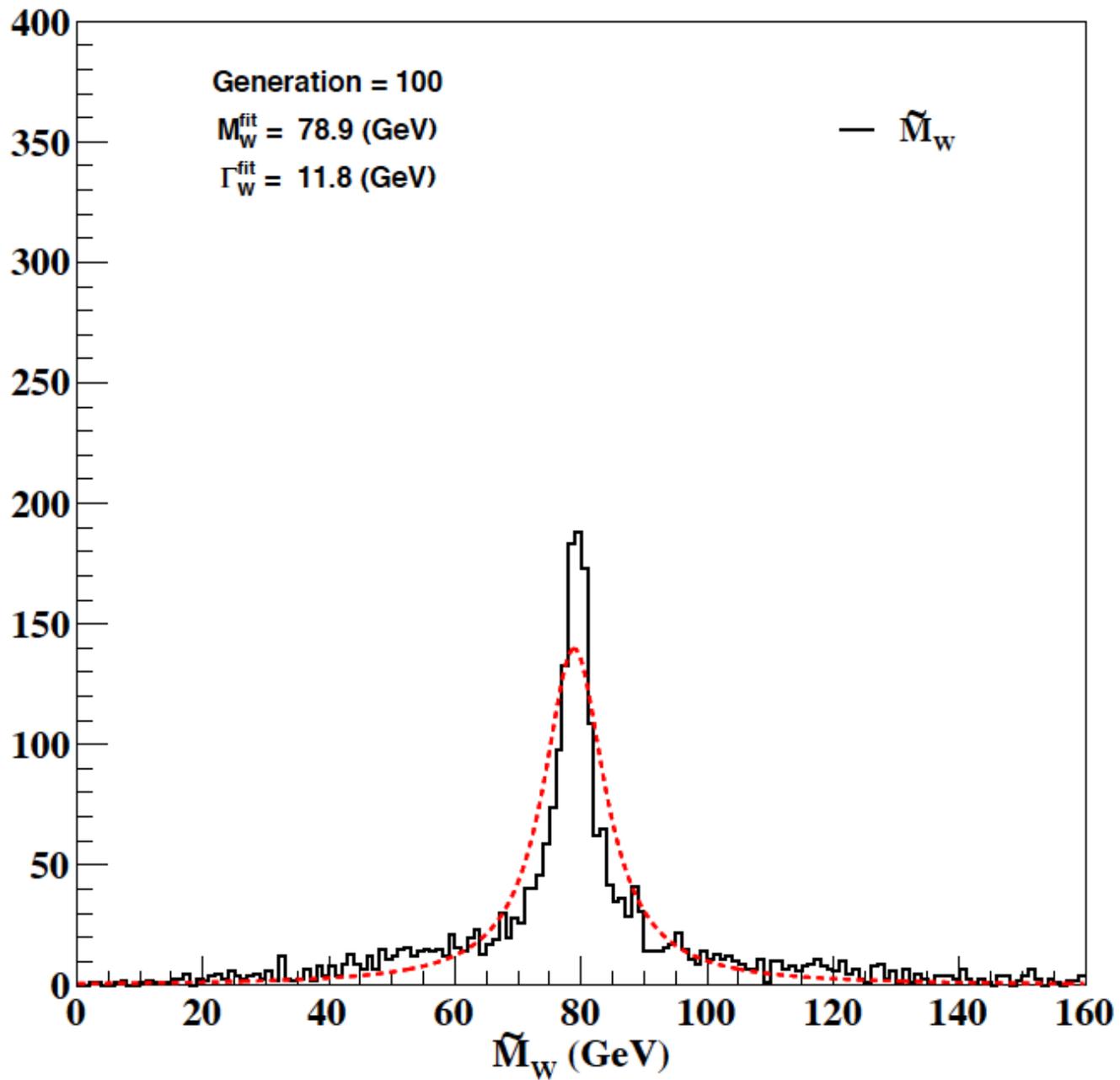
→ We apply this idea on the measurement of W boson mass, assuming pure Breit-Wigner shaped physical target distribution.

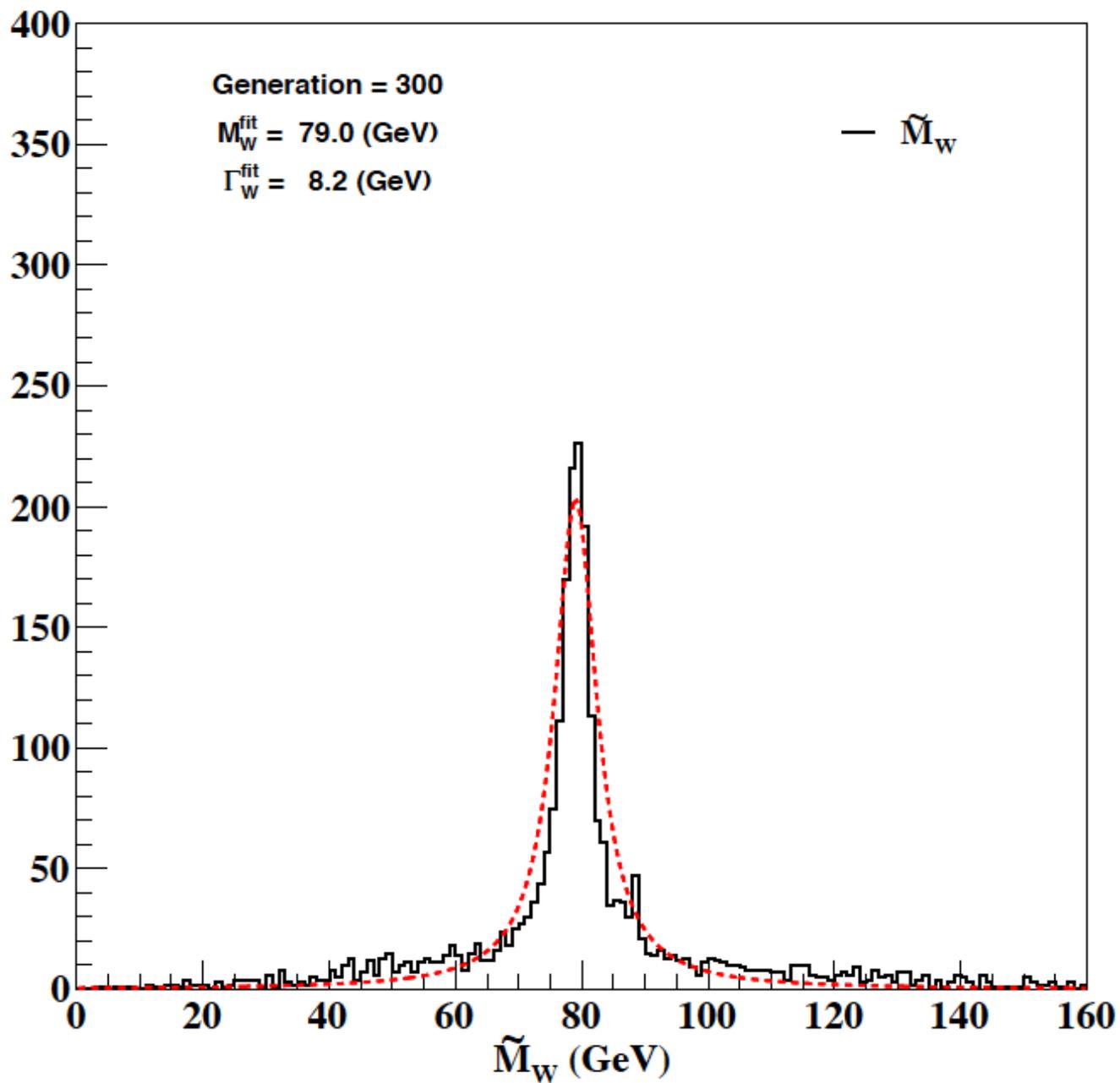
Reconstruction of the Breit-Wigner resonance peak of the W boson using genetic algorithm

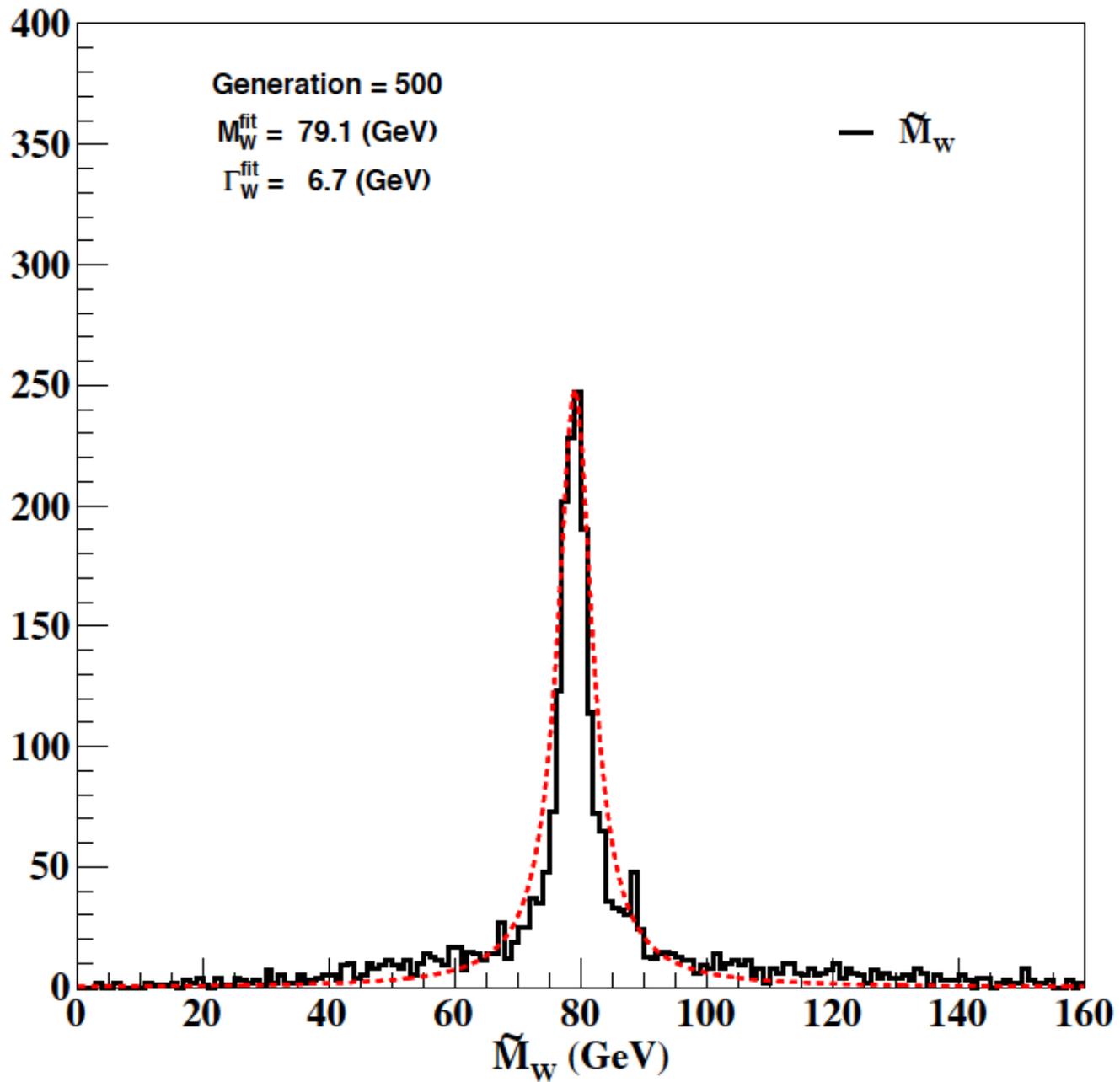
- Unknown : P_z or rapidity of neutrino
- Encoding representation for a gene : real continuous value
- Chromosome : a set of q_z , $\{q_z\}$ for whole event
- Population : randomly generated N_{pop} - chromosomes
- Fitness (of a chromosome = a solution) : calculated by various probability measures like as χ^2 , loglikelihoods, relative entropy ..
- Using the BW shape as the model of the optimized physical target distribution.

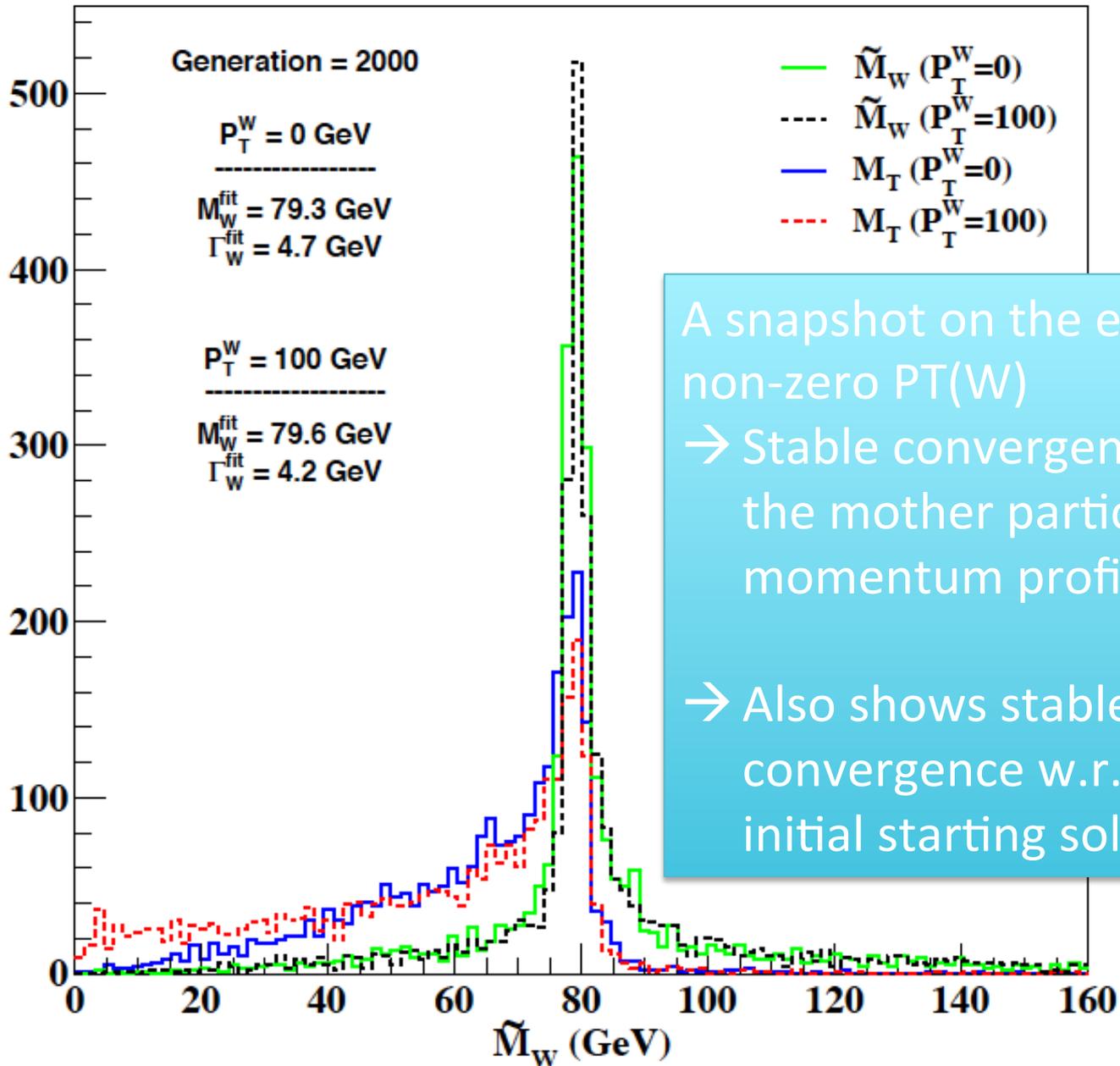












A snapshot on the effect of non-zero $P_T(W)$

- Stable convergence w.r.t the mother particles momentum profile
- Also shows stable convergence w.r.t various initial starting solution

→ The quality of evolution depends on the kind of the measure of the distance between probability distributions.

Conclusions

- We are testing various collective stochastic optimizations by employing genetic algorithms (GA), in search for the best random sequence (of HEP event unknowns) which fits the best into physical target distributions.
- In this process, the best random sequence can have good correlation with the true sequence. We demonstrated that such a good solution (as an extremum searched by GA) really exists, in the case of sampled invariant mass functions of a random sequence realizing the Breit-Wigner distribution of the M_w , collectively.

Conclusions

- In this way, we engineer our sampled invariant mass distribution to have the properties of BW resonance.
- The GA, based on the principle of Natural Selection, is found to be very efficient and powerful for implementing the (complicated) collective stochastic optimizations for HEP event reconstruction.
- Developing more efficient GA algorithm is required.

Backup

Genetic Algorithm

- Genetic algorithm is a search and optimization technique based on Darwin's Principle of Natural Selection.
- "Select the best individuals with good fitness, and Exchange good genes, and Select .."
- Optimizes a large number of (continuous/discrete) parameters with extremely complex objective function. It can rather easily jump out of a local minimum, in compared to the usual gradient-based search algorithms.
- Does not require derivative information.
- May encode the parameters so that the optimization is done with the encoded parameters

Simple GA works with the binary encoding for each variable.

$gene^i$ = *an encoded unknown value (q^i or C^i)
of an event block*

$chromosome$ =
$$\left(\begin{array}{cccc} Gene^1 & Gene^2 & \dots & Gene^{N_{event}} \\ 010101 & 101010 & \dots & 110010 \end{array} \right)$$

(= *individual*) = *a set of genes = a set of encoded unknowns*
 $\leftrightarrow \{ \tilde{q}^1 / \tilde{C}^1, \tilde{q}^2 / \tilde{C}^2, \dots, \tilde{q}^{N_{event}} / \tilde{C}^{N_{event}} \}$
 \rightarrow *A distribution of $\{ \tilde{m}(p^i, \tilde{q}^i | \tilde{C}^i) \}$*

$population$ = *set of (random)chromosomes*
 \leftrightarrow *set of (random)solutions*
 \rightarrow *larger population \rightarrow better genetic diversity*
 \rightarrow *GA can survey larger solution space.*

$fitness$ = *quality of a solution*

Flow-chart of Genetic Algorithm

1. Encode Chromosomes
 - 1.1 Binary encoding
 - 1.2 Continuous encoding
2. Generating initial population
3. Evaluation of fitness values of all chromosomes
4. Convergence test (continue, if not converged)
5. Selection and Mating
 - 5.1 Roulette Wheel Selection in proportion to fitness
6. Reproduction (Crossover, Mutation)

Reproduction operators

- ▶ Crossover : random points are chosen on the individual chromosomes and the genes are exchanged at this point. This is the first operation where the GA explores another points in the variable space.
e.g.) single point crossover

$$\begin{array}{r} \begin{array}{c} 1010101 \\ 0101010 \end{array} \left| \begin{array}{c} 0001 \\ 1110 \end{array} \right. \quad (1) \\ \rightarrow \begin{array}{c} 1010101 \\ 0101010 \end{array} \left| \begin{array}{c} 1110 \\ 0001 \end{array} \right. \quad (2) \end{array}$$