Preserving physically important variables in optimal event selections

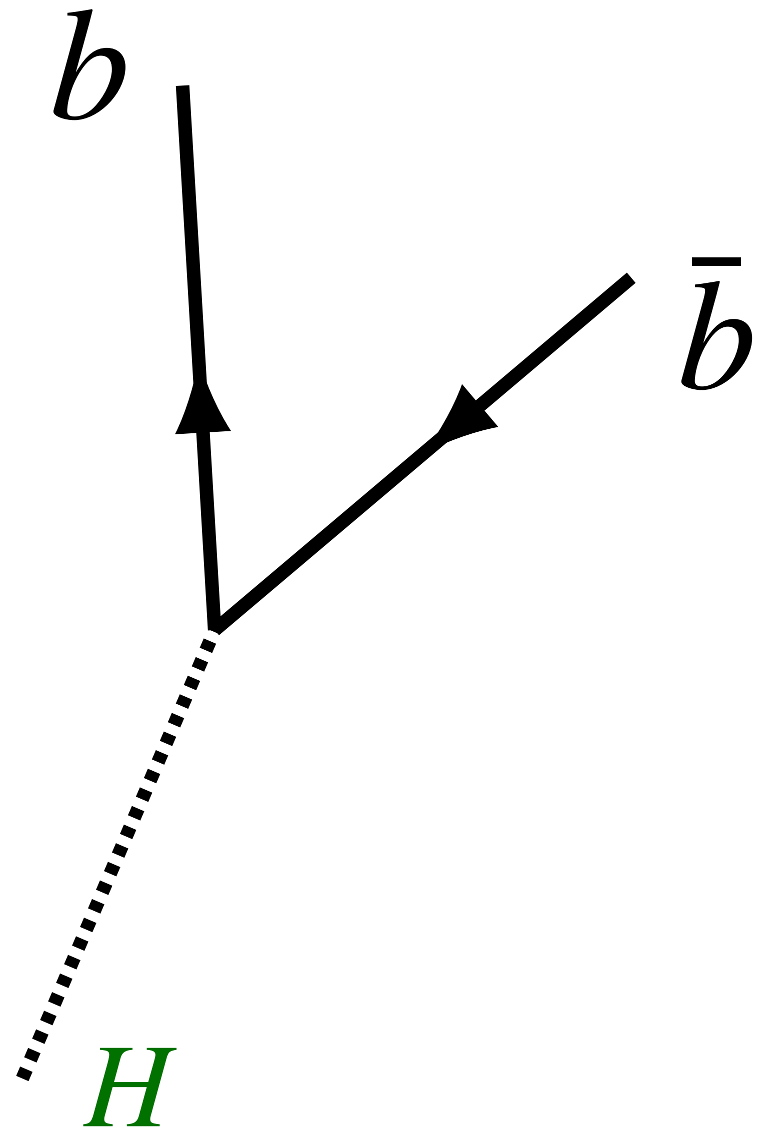
Philipp Windischhofer, Miha Zgubič, Daniela Bortoletto
philipp.windischhofer@physics.ox.ac.uk

University of Oxford

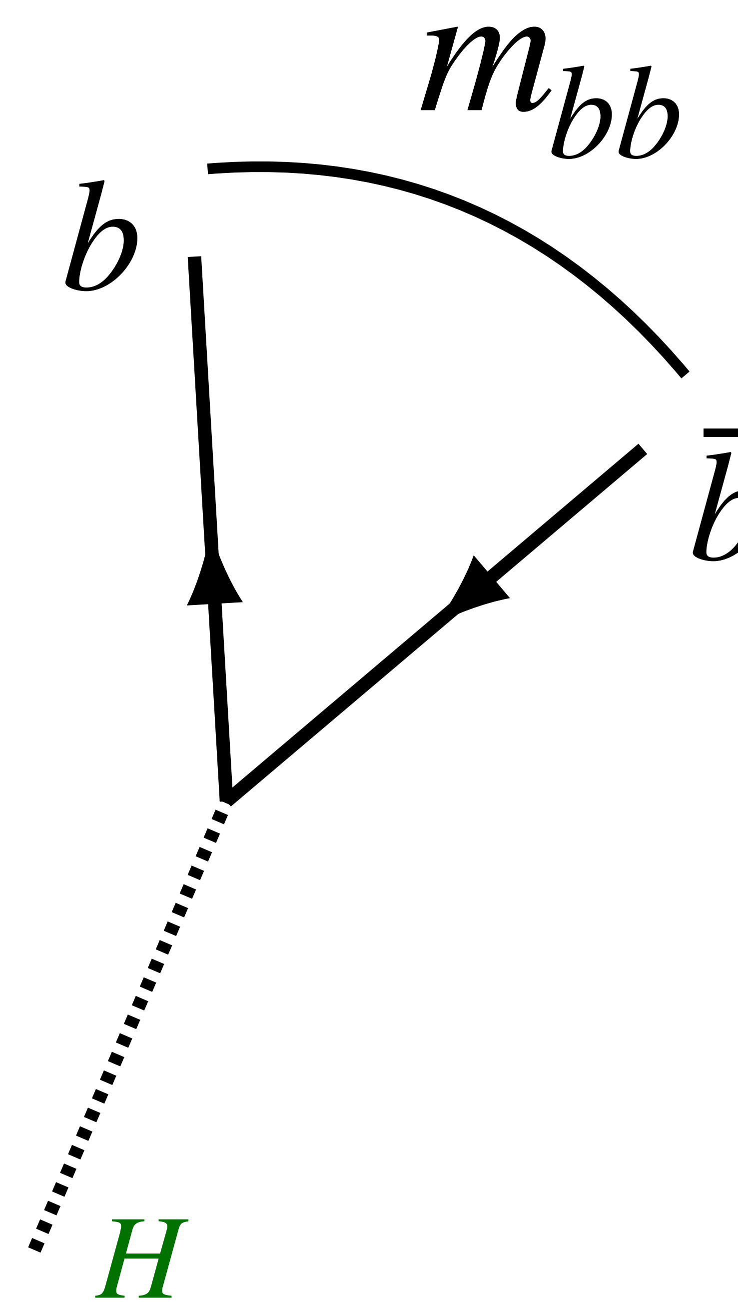
Higgs Couplings
October 1st, 2019
Oxford, United Kingdom

Preserving physically important variables in optimal event selections

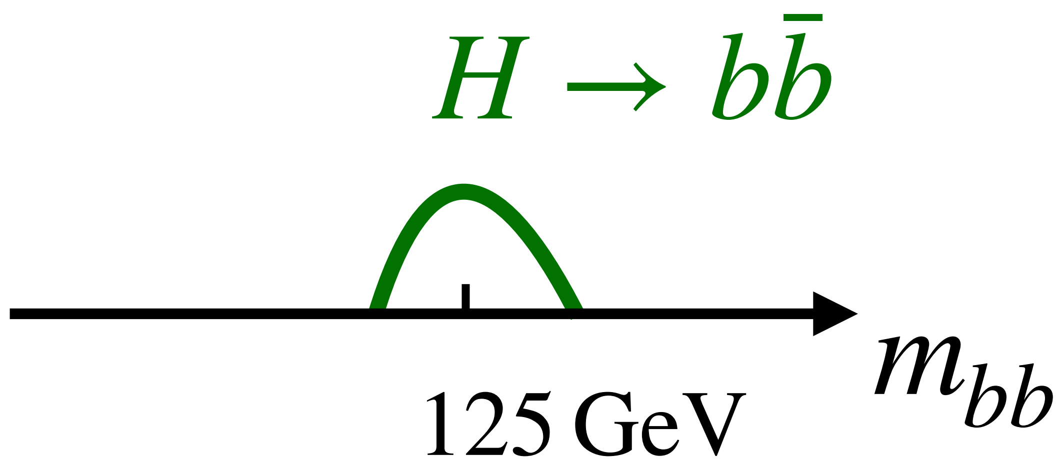
Preserving physically important variables in optimal event selections

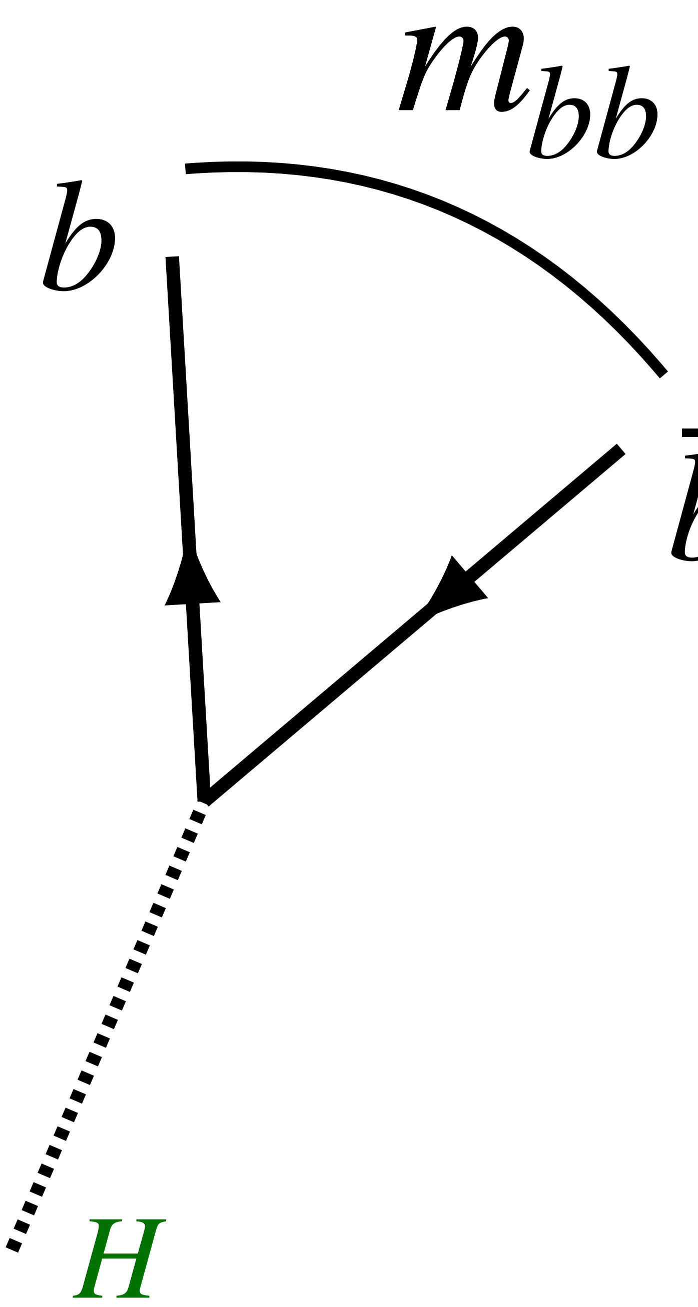


Preserving physically important variables in optimal event selections



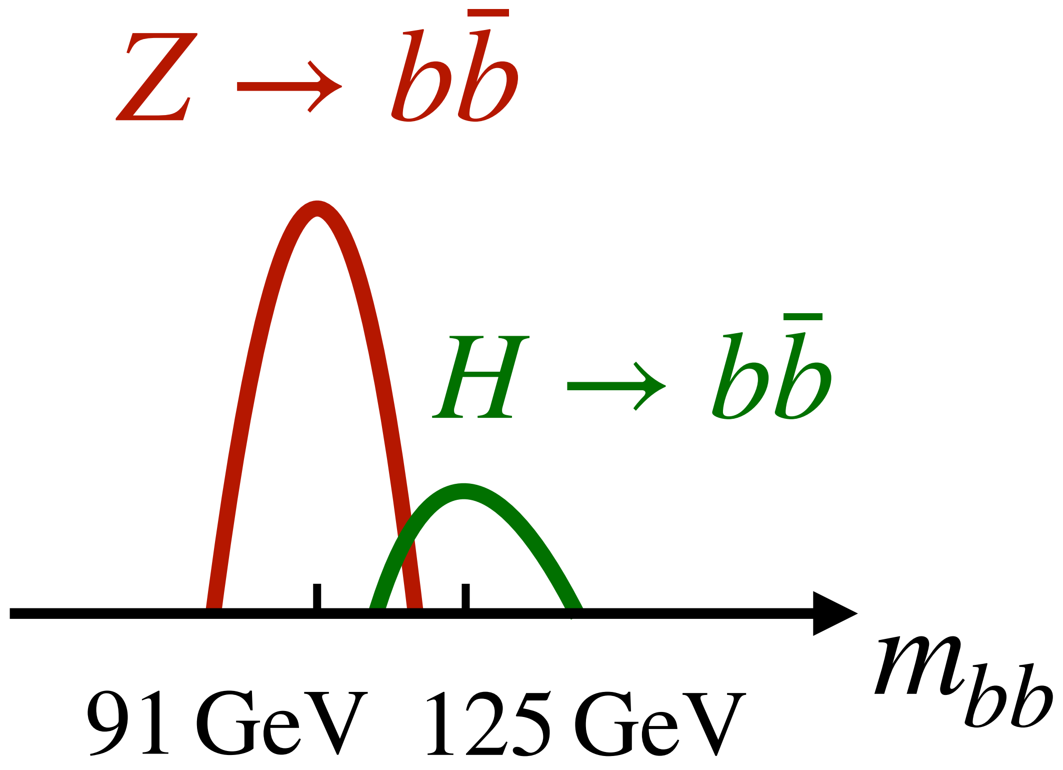
Preserving **physically** important variables in optimal event selections

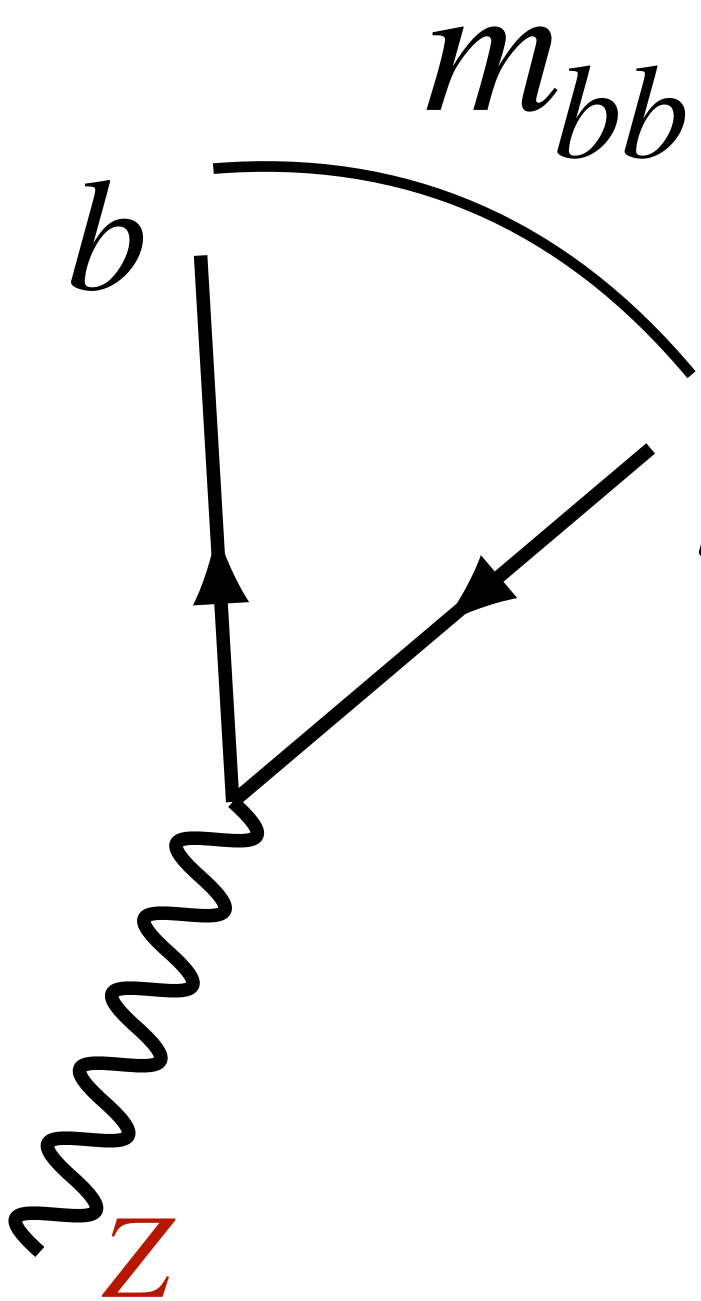




$\Longrightarrow \exists$ **physical state**

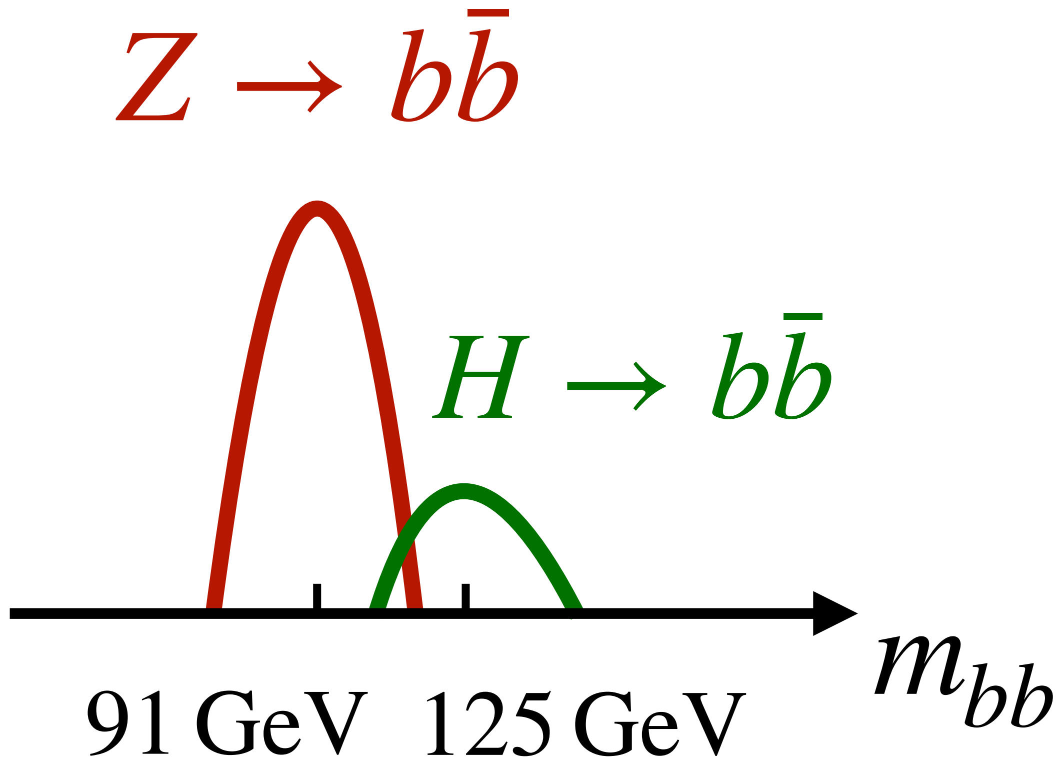
Preserving physically **important variables** in optimal event selections



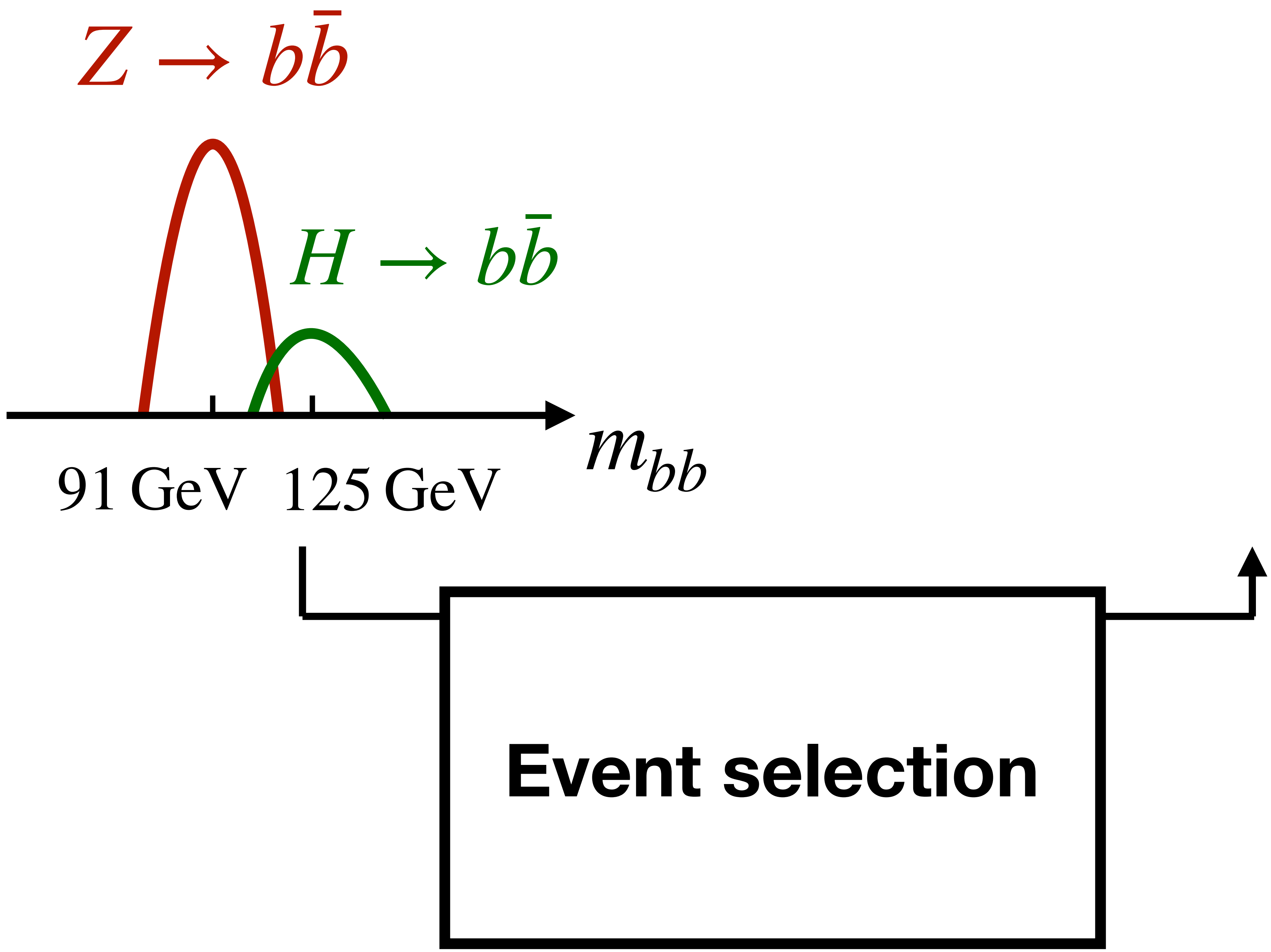


$\Longrightarrow$ **Signal (S) / background (B) discrimination**

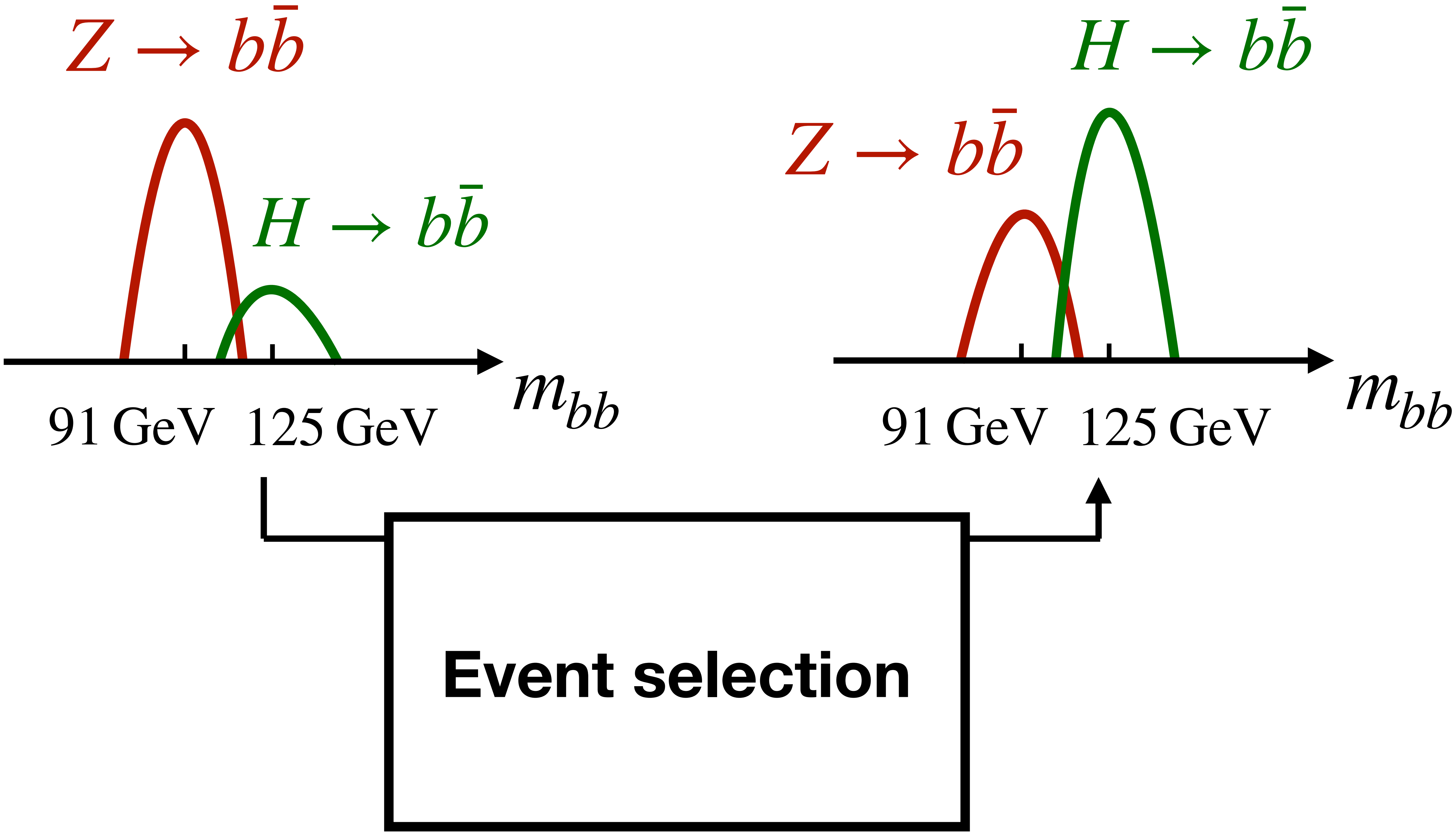
Preserving physically important variables in optimal **event selections**



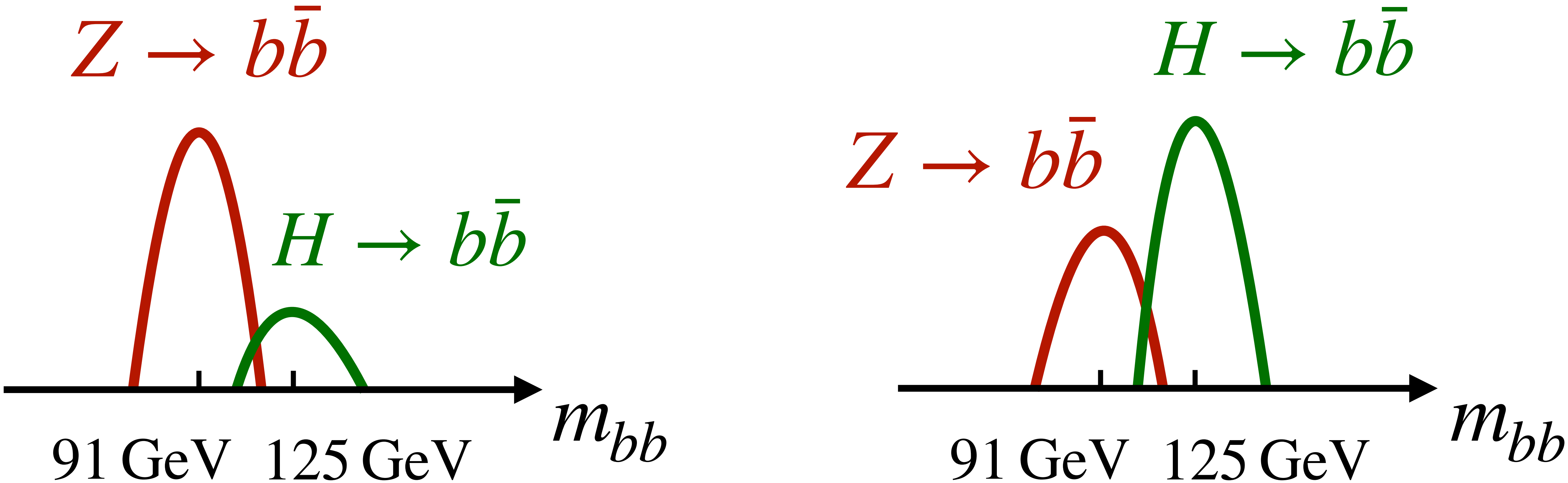
Preserving physically important variables in optimal **event selections**



Preserving physically important variables in optimal **event selections**



Preserving physically important variables in **optimal event selections**

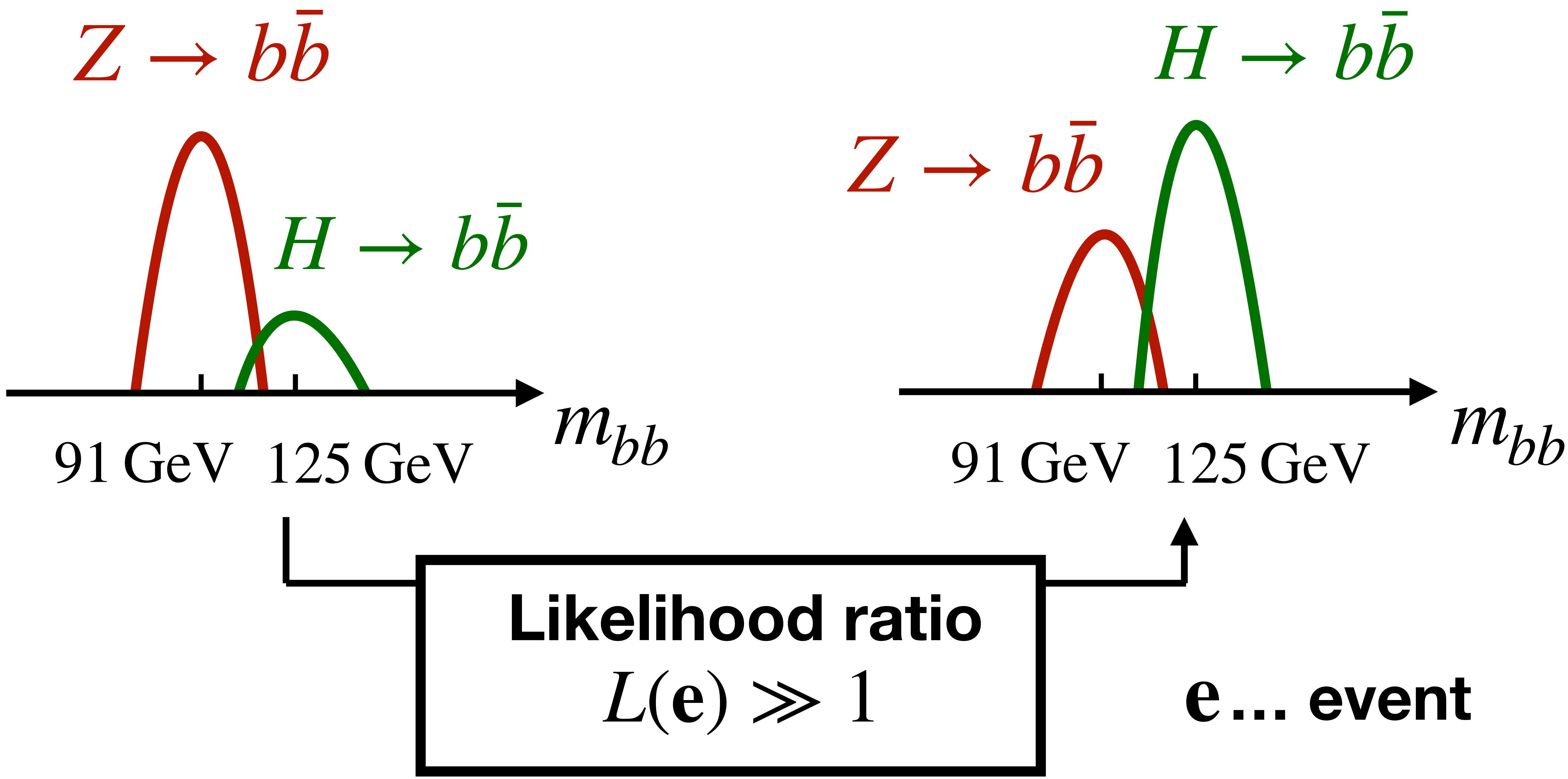


Likelihood ratio

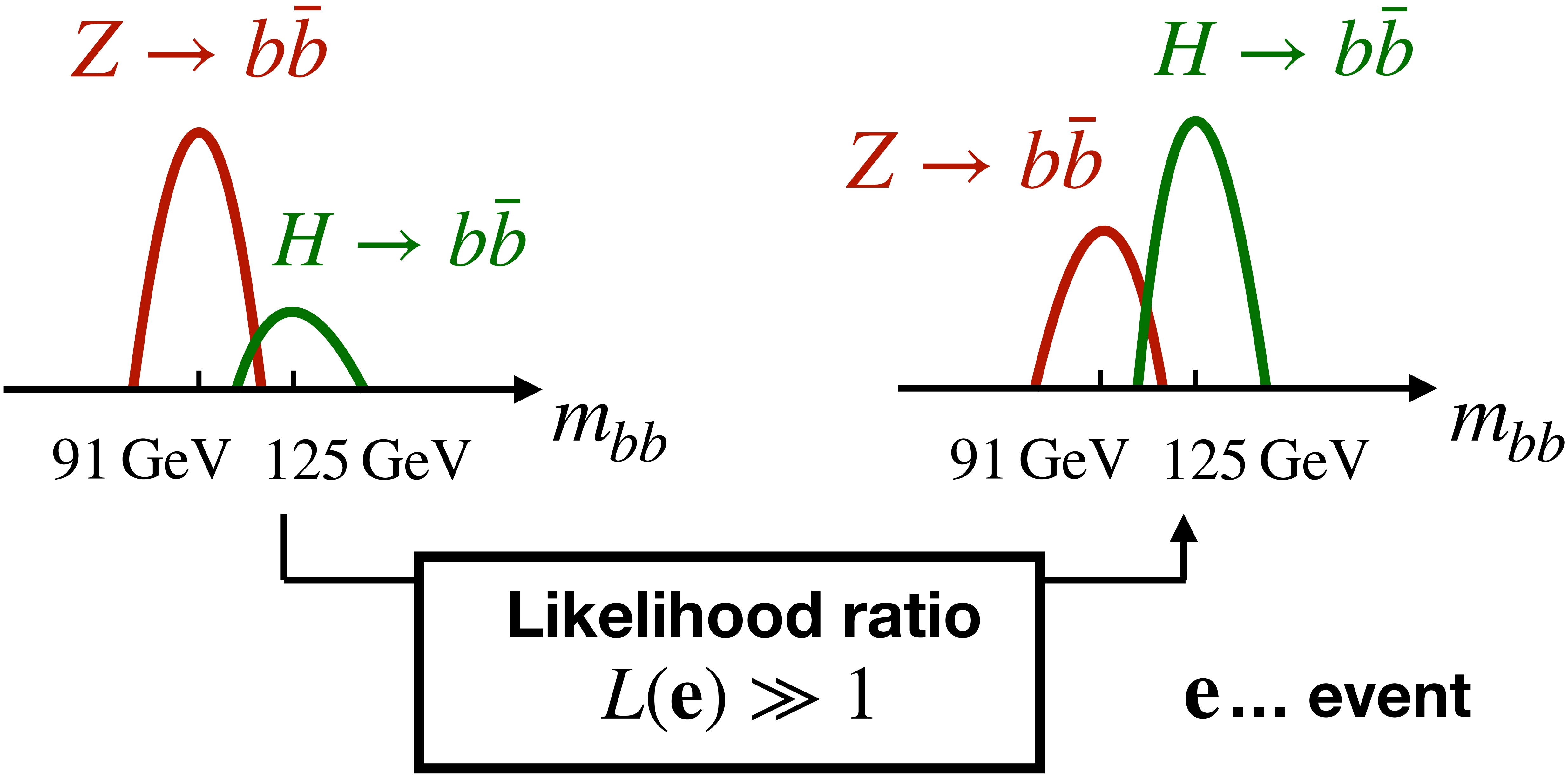
$$L(\mathbf{e}) = \frac{p_S(\mathbf{e})}{p_B(\mathbf{e})} \gg 1$$

e… event

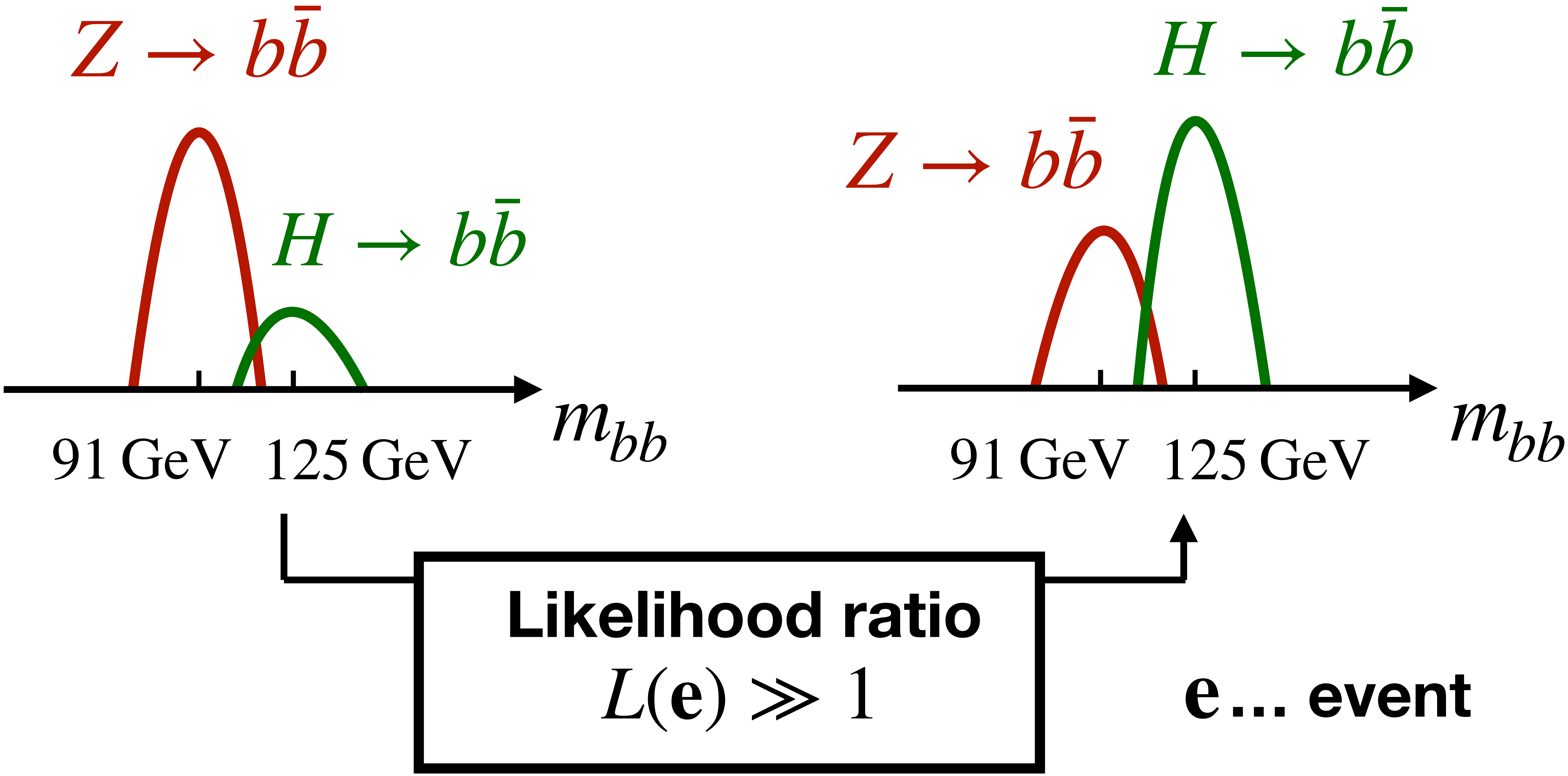
Preserving physically important variables in **optimal event selections**



~~Preserving~~ physically important variables in **optimal event selections**

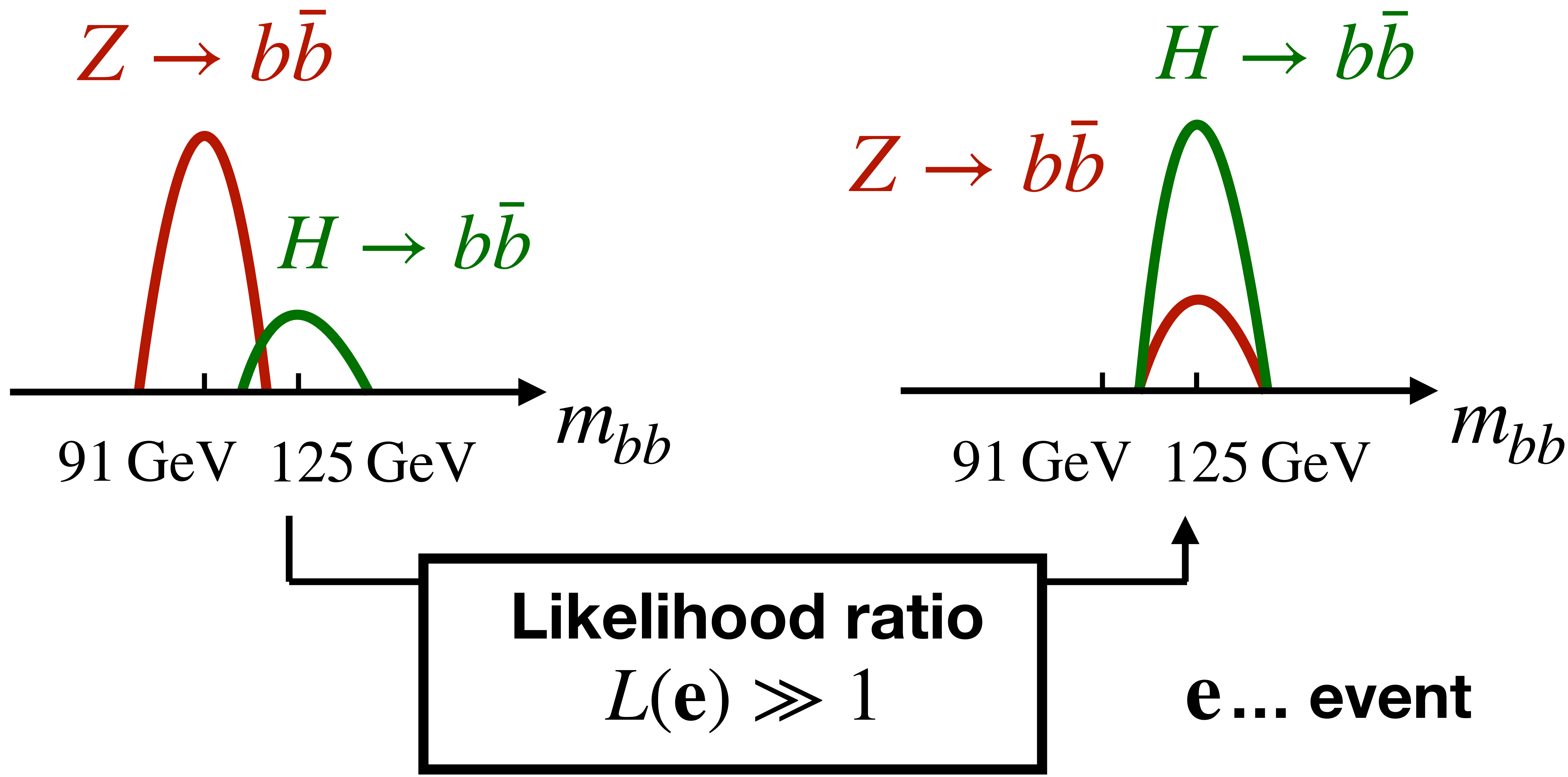


~~Preserving~~ physically important variables in **optimal event selections**



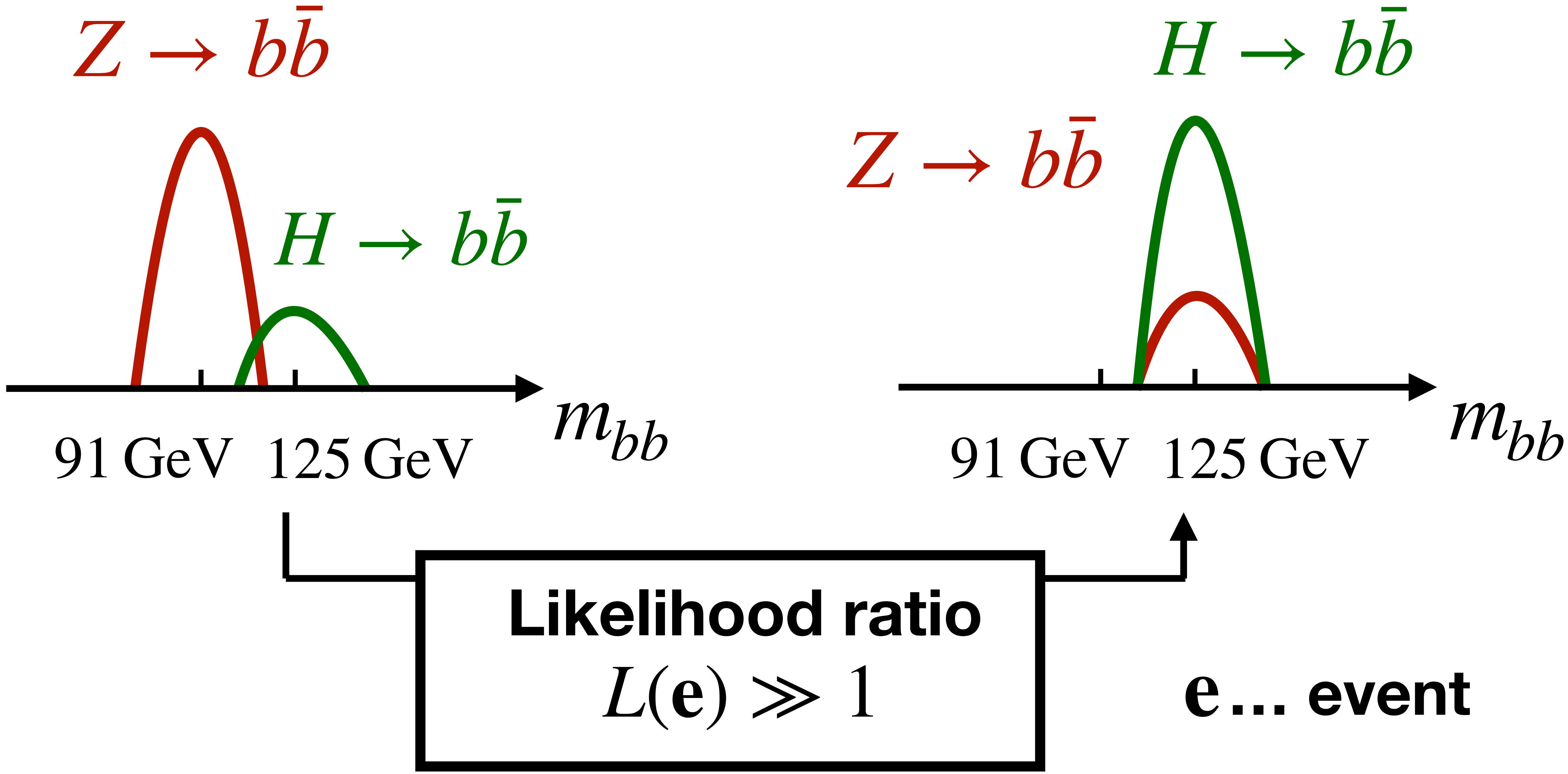
- $m_{bb}(\mathbf{e})$ and $L(\mathbf{e})$ strongly correlated $\longrightarrow$ **distortion**

~~**Preserving**~~ physically important variables in **optimal event selections**



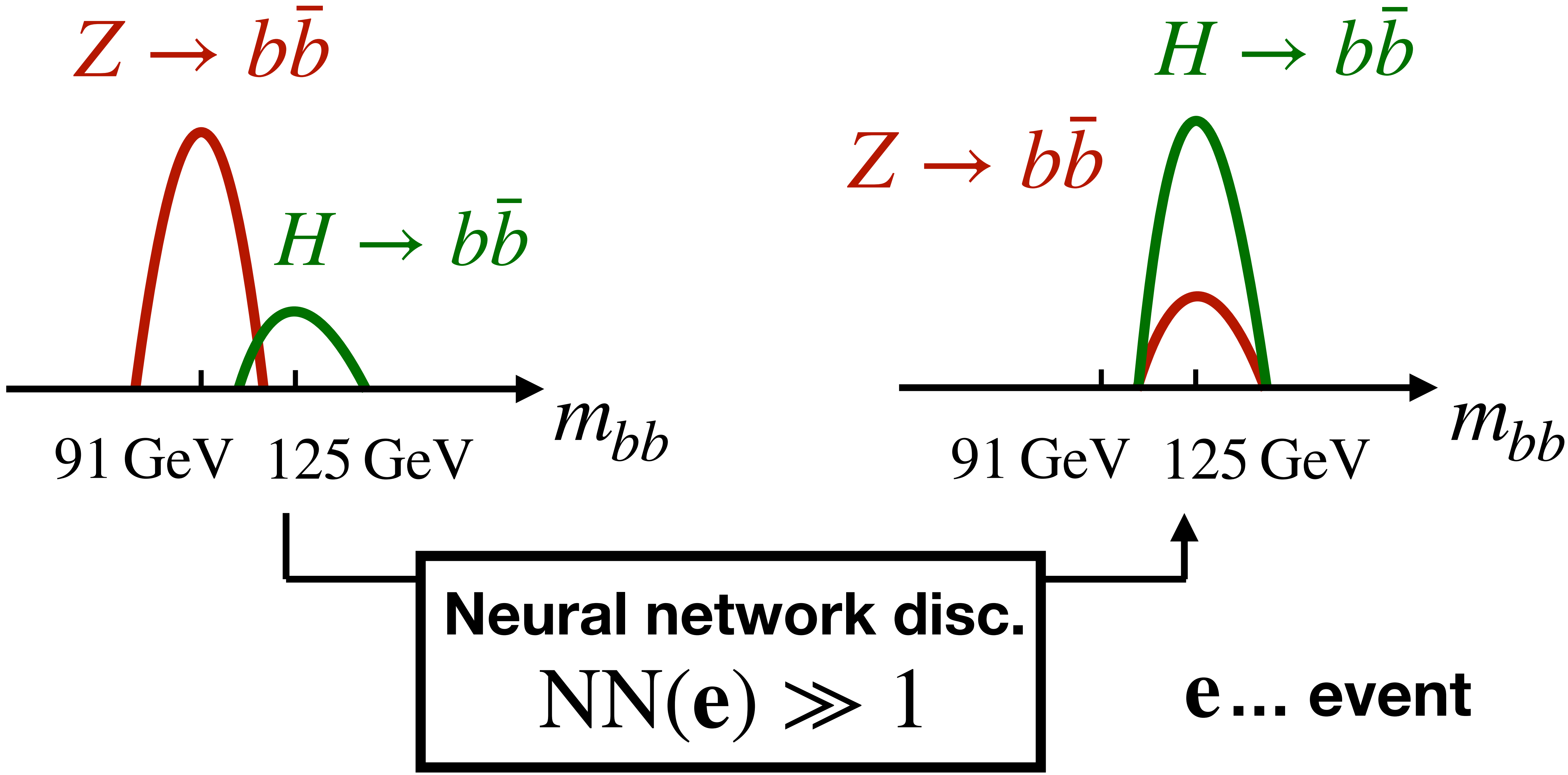
- $m_{bb}(\mathbf{e})$ and $L(\mathbf{e})$ strongly correlated $\longrightarrow$ **distortion**

~~Preserving~~ physically important variables in **optimal event selections**



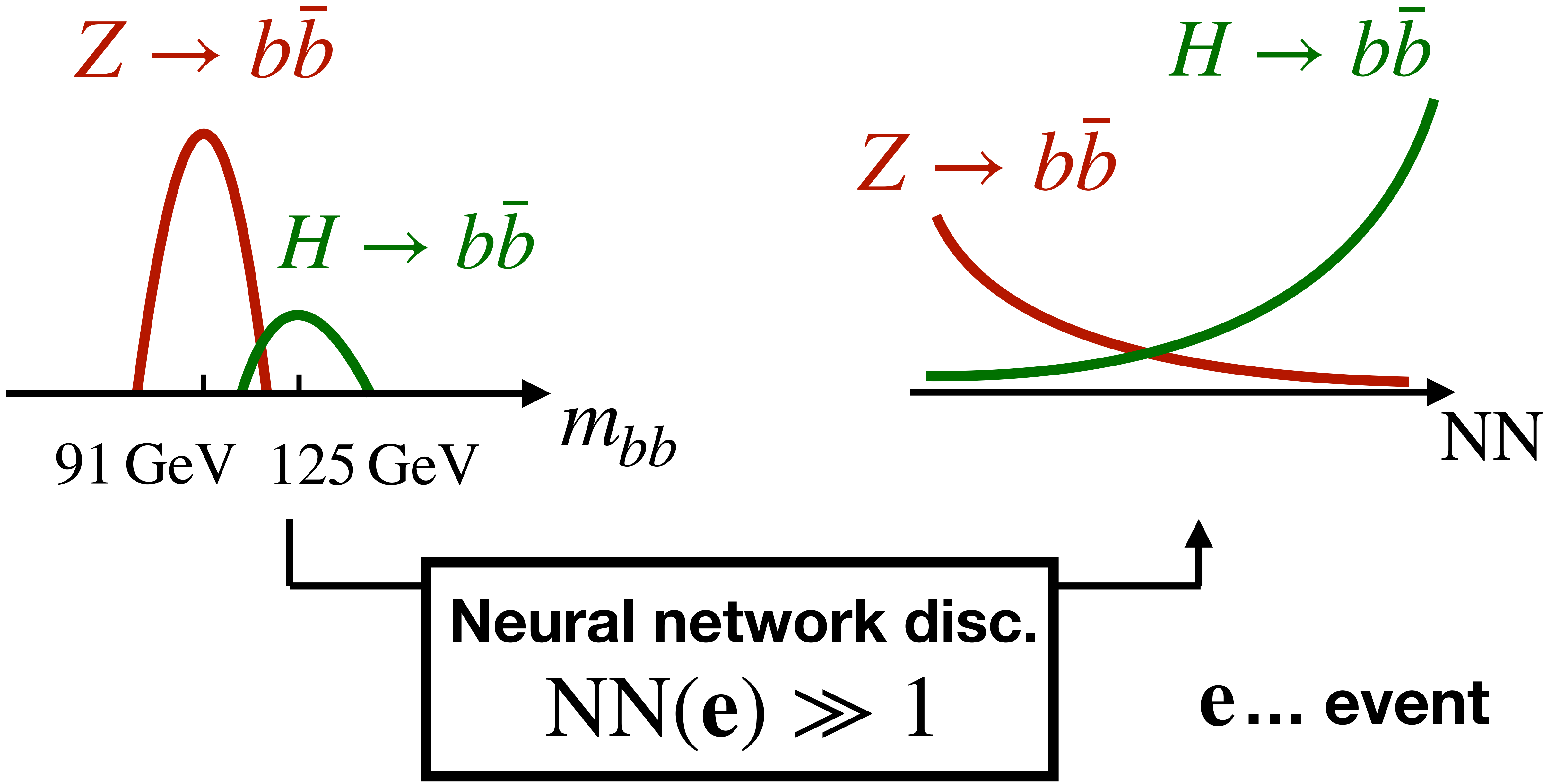
- No discrimination power left in m_{bb} $\to$ ~~**physically important**~~

~~**Preserving**~~ physically important variables in **optimal event selections**



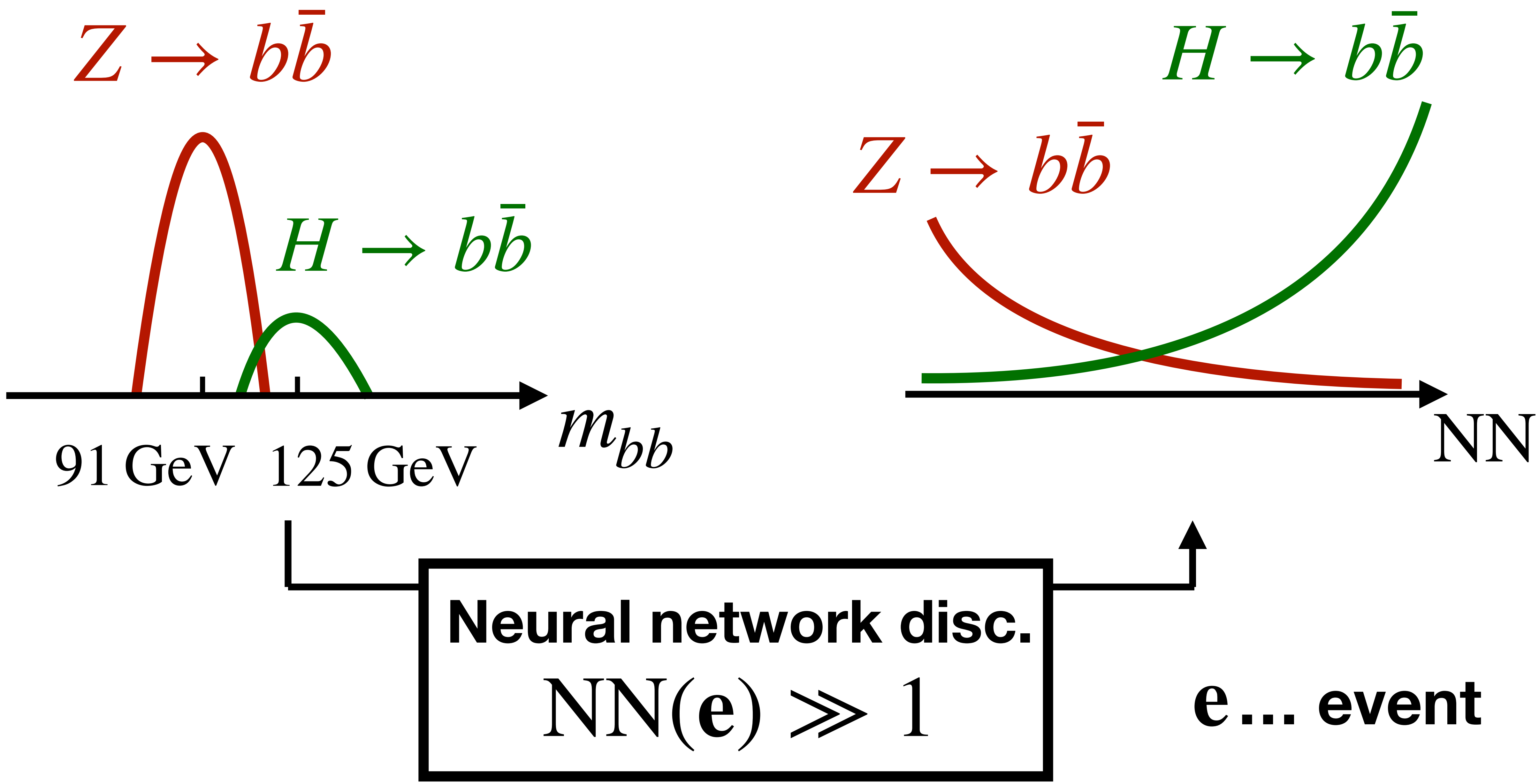
- No discrimination power left in m_{bb} $\to$ ~~**physically important**~~

~~Preserving~~ physically important variables in **optimal event selections**



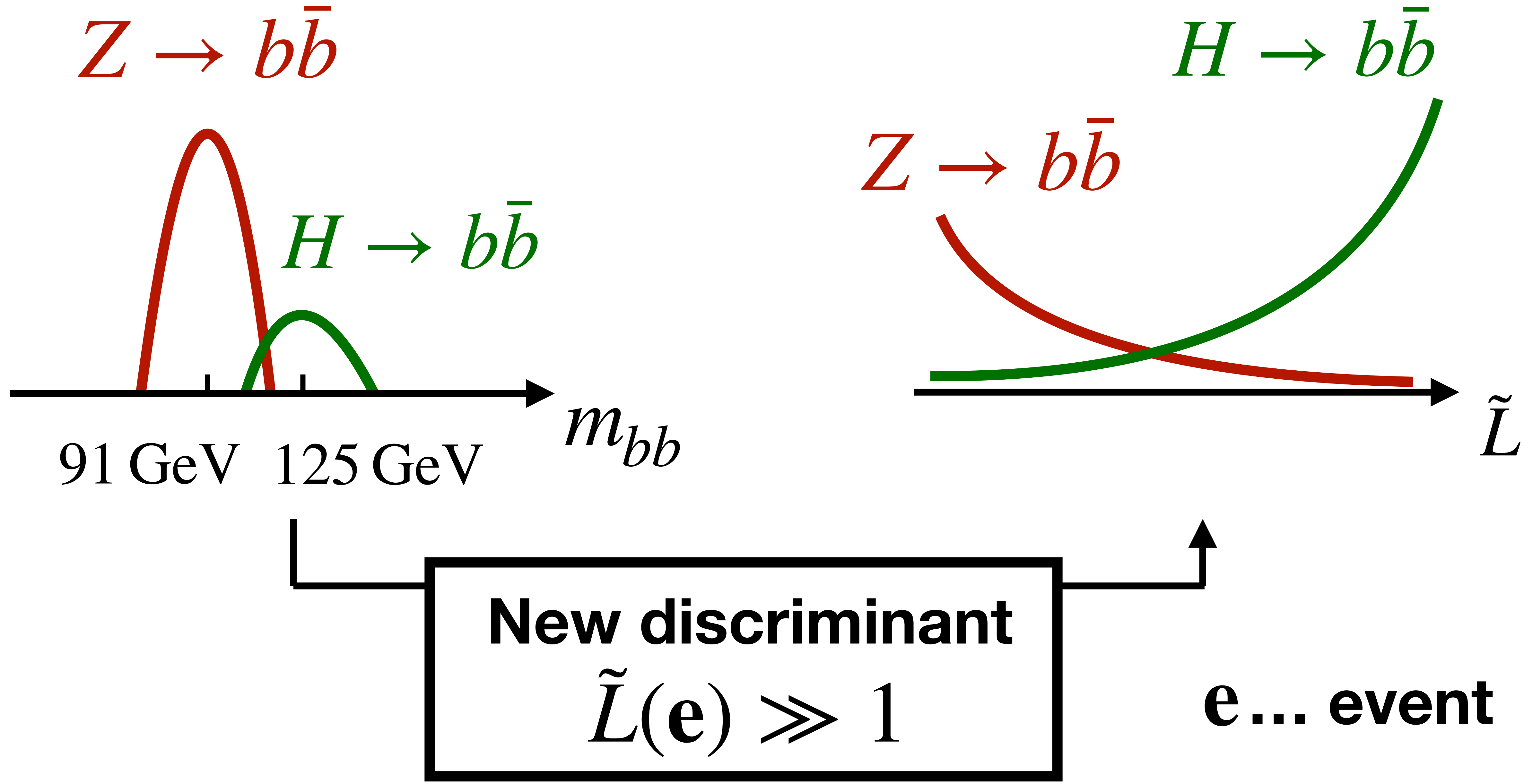
- No discrimination power left in m_{bb} $\rightarrow$ ~~**physically important**~~

~~Preserving~~ physically important variables in **optimal event selections**



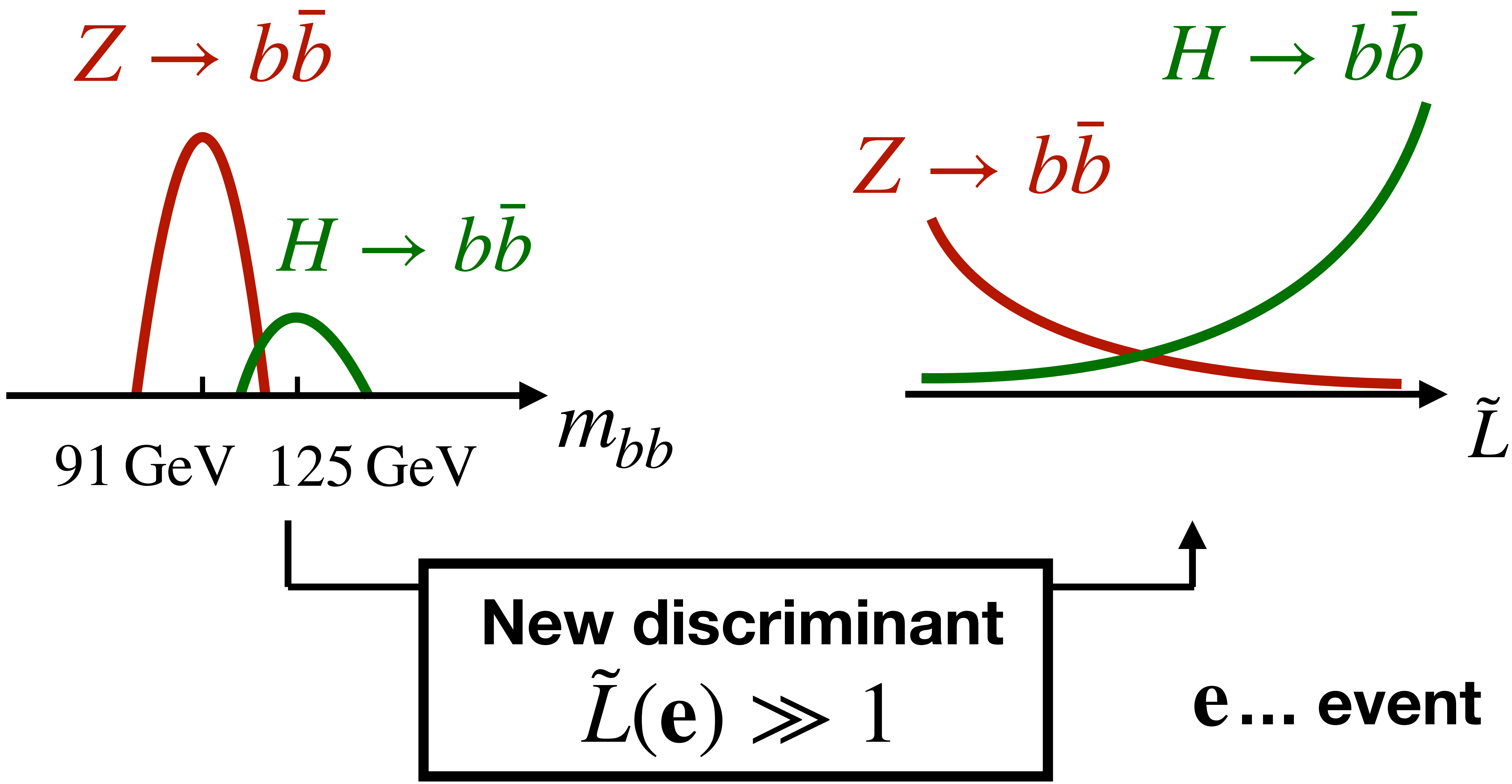
- $m_{bb}(\mathbf{e})$ vs. $\mathrm{NN}(\mathbf{e}) \leftrightarrow$ physics vs. statistics

Preserving physically important variables in **optimal event selections**



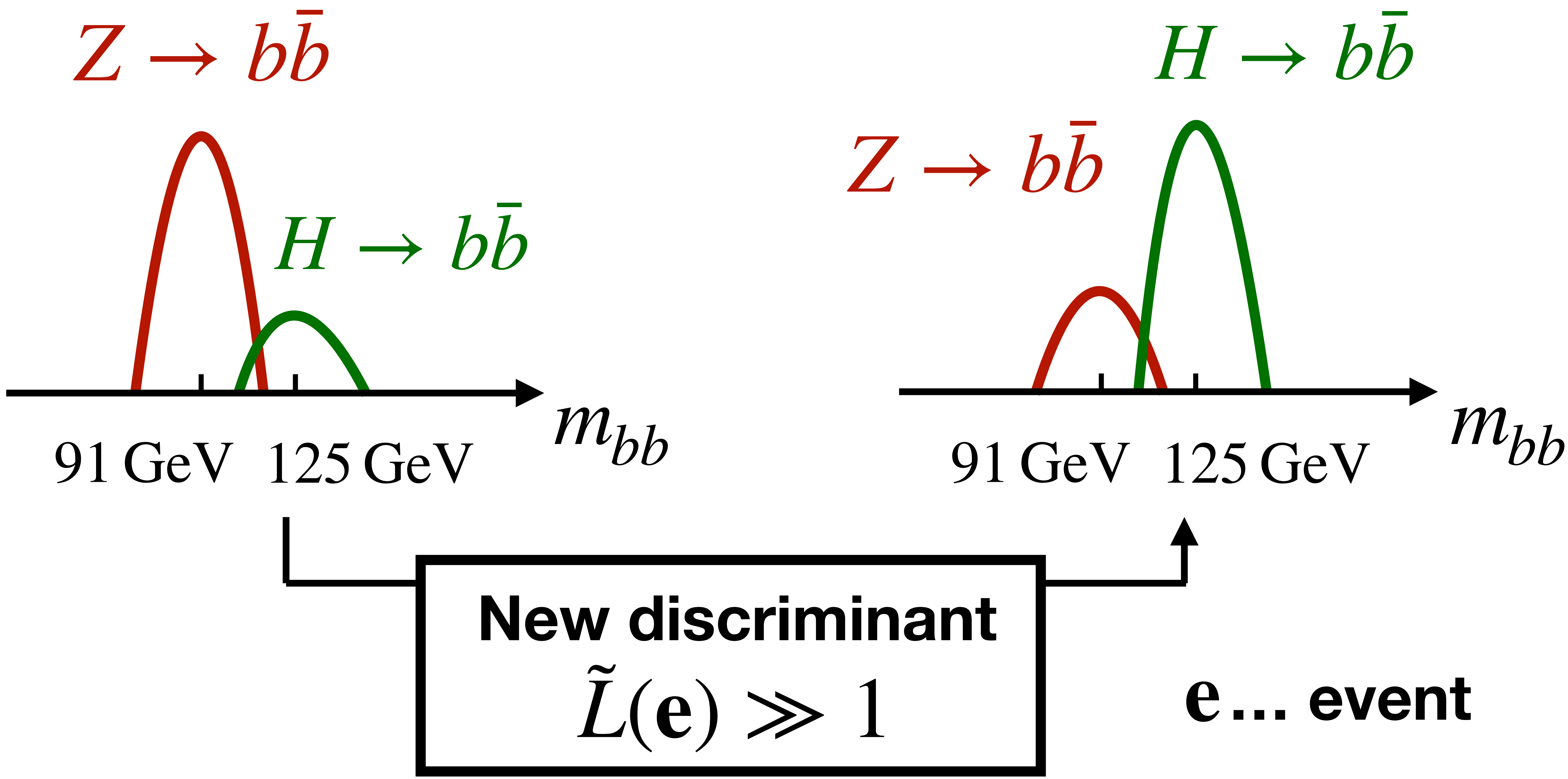
- $m_{bb}(\mathbf{e})$ vs. $\mathrm{NN}(\mathbf{e})$ $\leftrightarrow$ physics vs. statistics

Preserving physically important variables in **optimal event selections**



$\tilde{L}$ … **best discriminant that limits distortion of** m_{bb}

Preserving physically important variables in **optimal event selections**



$\tilde{L}$ … **best discriminant that limits distortion of** m_{bb}

Towards a solution

Likelihood ratio = maximally powerful discriminant

Towards a solution

Likelihood ratio = maximally powerful discriminant

$$\frac{p_S(\mathbf{e})}{p_B(\mathbf{e})} =: L(\mathbf{e})$$

Towards a solution

Likelihood ratio = maximally powerful discriminant

$$\frac{p_S(\mathbf{e})}{p_B(\mathbf{e})} =: L(\mathbf{e}) := \arg\max_f \left(\mathscr{P}[f] \right)$$

performance measure

candidate discriminant

Towards a solution

Likelihood ratio = maximally powerful discriminant

$$L(\mathbf{e}) := \arg\max_f \left(\mathscr{P}[f] \right)$$

performance measure

candidate discriminant

Towards a solution

$$L(\mathbf{e}) := \arg\max_f \left(\ \mathscr{P}[f] \ \right)$$

Towards a solution

$$L(\mathbf{e}) := \arg\max_f \left(\ \mathscr{P}[f] \ \right)$$

$\tilde{L}$: control distortion of m_{bb} by event selection

Towards a solution

$$L(\mathbf{e}) := \arg\max_f \left(\ \mathscr{P}[f] \ \right)$$

limit correlation of m_{bb} and $\tilde{L}$ $\leftrightarrow$ $\tilde{L}$: control distortion of m_{bb} by event selection

Towards a solution

$$L(\mathbf{e}) := \arg\max_f \left(\ \mathscr{P}[f] \ \right)$$

Constraint:

limit correlation of m_{bb} and $\tilde{L}$

$\leftrightarrow$

$\tilde{L}$: control distortion of m_{bb} by event selection

Towards a solution

$$L(\mathbf{e}) := \arg\max_f \left(\ \mathscr{P}[f] \ \right)$$

Constraint:

limit correlation of

m_{bb} and $\tilde{L}$

Towards a solution

$$\tilde{L}(\mathbf{e}) := \arg\max_f \left(\mathscr{P}[f] \right)$$

Constraint:

limit correlation of

m_{bb} and $\tilde{L}$

Towards a solution

$$\tilde{L}(\mathbf{e}) := \arg\max_f \Big(\ \mathscr{P}[f]$$

Constraint:

limit correlation of

m_{bb} and $\tilde{L}$

Towards a solution

$$\tilde{L}(\mathbf{e}) := \arg\max_f \Big(\mathscr{P}[f] + C(f, m_{bb}) \Big)$$

Constraint:

limit correlation of m_{bb} and $\tilde{L}$

Towards a solution

$$\tilde{L}(\mathbf{e}) := \arg\max_f \Big(\ \mathscr{P}[f] \quad + \quad C(f, m_{bb}) \ \Big)$$

↑ constraint

Constraint:

limit correlation of

m_{bb} and $\tilde{L}$

Towards a solution

$$\tilde{L}(\mathbf{e}) := \arg\max_f \Big(\mathscr{P}[f]$$

$$+ \lambda \cdot \; C(f, m_{bb}) \Big)$$

Lagrange multiplier *(constraint strength)* → λ

constraint → C

Constraint:

limit correlation of m_{bb} and $\tilde{L}$

Towards a solution

$$\tilde{L}(\mathbf{e}) := \arg\max_f \Big(\mathcal{P}[f]$$

$$- \lambda \cdot \quad \rho\left(f, m_{bb}\right) \Big)$$

Lagrange multiplier

(constraint strength)

Pearson correlation $\rightarrow 0$

Constraint:

limit correlation of

m_{bb} and $\tilde{L}$

Towards a solution

$$\tilde{L}(\mathbf{e}) := \arg\max_f \Big(\mathscr{P}[f] - \lambda \cdot \rho(f, m_{bb}) \Big)$$

Lagrange multiplier *(constraint strength)* → λ

Pearson correlation $\to 0$ → $\rho(f, m_{bb})$

$\implies \tilde{L}, m_{bb}$ linearly uncorrelated

Constraint:

limit correlation of m_{bb} and $\tilde{L}$

Towards a solution

$$\tilde{L}(\mathbf{e}) := \arg\max_f \Big(\mathscr{P}[f] - \lambda \cdot \rho(f, m_{bb}) \Big)$$

Lagrange multiplier

(constraint strength)

Pearson correlation $\to 0$

$\implies \tilde{L}, m_{bb}$ linearly uncorrelated nonlinearly related

Constraint:

limit correlation of m_{bb} and $\tilde{L}$

Towards a solution

$$\tilde{L}(\mathbf{e}) := \arg\max_f \Big(\mathcal{P}[f] - \lambda \cdot \rho(f, m_{bb}) \Big)$$

Lagrange multiplier

(constraint strength)

Pearson correlation $\rightarrow 0$

$\implies \tilde{L}, m_{bb}$ linearly uncorrelated

~~nonlinearly related~~

Constraint:

limit correlation of m_{bb} and $\tilde{L}$

Towards a solution

$$\tilde{L}(\mathbf{e}) := \arg\max_f \Big(\mathcal{P}[f] - \lambda \cdot \rho(f, m_{bb}) \Big)$$

Lagrange multiplier
(constraint strength)

Pearson correlation $\to 0$

Constraint:

limit correlation of
m_{bb} and $\tilde{L}$

Towards a solution

*http://cern.ch/go/zKv8

$$\tilde{L}(\mathbf{e}) := \arg\max_f \Big(\mathscr{P}[f] - \lambda \cdot \mathrm{MI}(f, m_{bb}) \Big)$$

Lagrange multiplier *(constraint strength)*

Mutual information* $\rightarrow 0$

Constraint:

limit correlation of m_{bb} and $\tilde{L}$

Towards a solution

* http://cern.ch/go/zKv8

$$\tilde{L}(\mathbf{e}) := \arg\max_f \left(\mathscr{P}[f] - \lambda \cdot \mathrm{MI}(f, m_{bb}) \right)$$

Lagrange multiplier
(constraint strength)

Mutual information* $\to 0$

$\implies \tilde{L}, m_{bb}$ independent

Constraint:

limit correlation of m_{bb} and $\tilde{L}$

Towards a solution

* http://cern.ch/go/zKv8

$$\tilde{L}(\mathbf{e}) := \arg\max_f \Big(\mathscr{P}[f] - \lambda \cdot \mathrm{MI}(f, m_{bb}) \Big)$$

Lagrange multiplier
(constraint strength)

Mutual information* $\to 0$

$$\implies p(\tilde{L}, m_{bb}) = p(\tilde{L}) \cdot p(m_{bb})$$

Constraint:

limit correlation of
m_{bb} and $\tilde{L}$

Towards a solution

*http://cern.ch/go/zKv8

$$\tilde{L}(\mathbf{e}) := \arg\max_f \Big(\mathscr{P}[f]$$

$$- \lambda \cdot \mathrm{MI}(f, m_{bb}) \Big)$$

Lagrange multiplier

(constraint strength)

Mutual information* $\rightarrow 0$

$$\Longrightarrow p(\tilde{L}, m_{bb}) = p(\tilde{L}) \cdot p(m_{bb})$$

Constraint:

limit correlation of m_{bb} and $\tilde{L}$ $\longleftrightarrow$ $\tilde{L}$: control distortion of m_{bb} by event selection

The final discriminant

$$\tilde{L}(\mathbf{e}) := \arg\max_f \Big(\mathscr{P}[f] - \lambda \cdot \mathbf{MI}(f, m_{bb}) \Big)$$

Into practice

$$\tilde{L}(\mathbf{e}) := \arg\max_f \Big(\ \mathscr{P}[f] \\ - \lambda \cdot \mathbf{MI}(f, m_{bb}) \ \Big)$$

Into practice

candidate discriminant: **neural network**

$$\tilde{L}(\mathbf{e}) := \arg\max_f \Big(\mathscr{P}[f] - \lambda \cdot \mathrm{MI}(f, m_{bb}) \Big)$$

Into practice

candidate discriminant: **neural network**

$$\tilde{L}(\mathbf{e}) := \arg\max_f \Big(\mathscr{P}[f] - \lambda \cdot \mathrm{MI}(f, m_{bb}) \Big)$$

numerical optimisation

gradient descent

Into practice

candidate discriminant: **neural network**

$$\tilde{L}(\mathbf{e}) := \arg\max_f \left(\mathscr{P}[f] - \lambda \cdot \mathrm{MI}(f, m_{bb}) \right)$$

numerical optimisation

gradient descent

loss function

Into practice

candidate discriminant: **neural network**

$$\tilde{L}(\mathbf{e}) := \arg\max_f \left(\mathscr{P}[f] - \lambda \cdot \mathrm{MI}(f, m_{bb}) \right)$$

numerical optimisation

gradient descent

free parameter

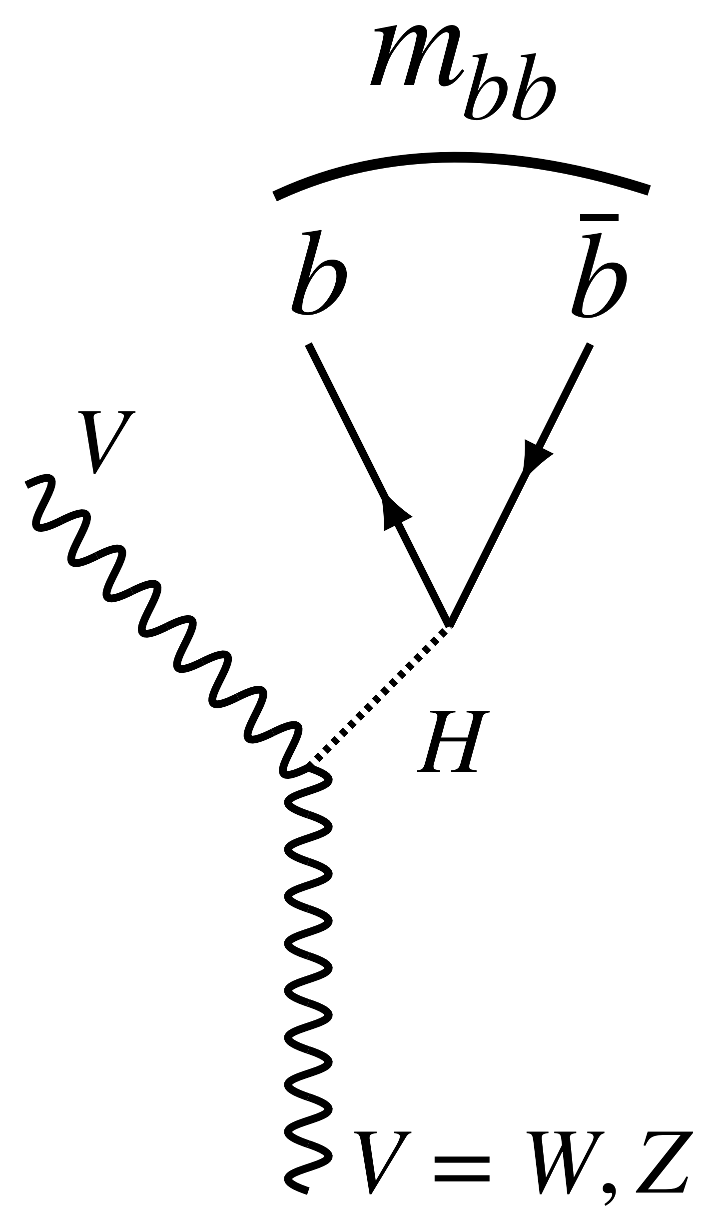
loss function

Into practice

Evaluation: toy analysis, VH/bb

Into practice

Evaluation: toy analysis, VH/bb



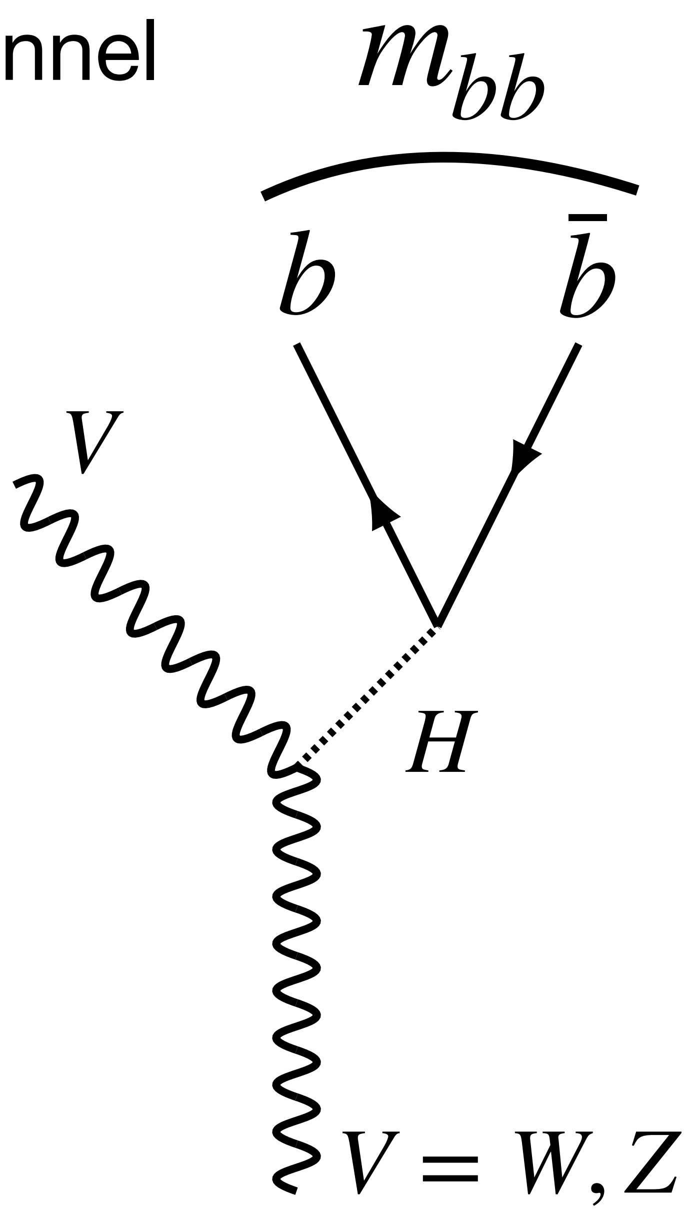
Into practice

Evaluation: **toy analysis**, VH/bb, 0-lepton channel

VH

$\hookrightarrow b\bar{b}$

$\hookrightarrow Z \to \nu\bar{\nu}, W \to \ell\nu$



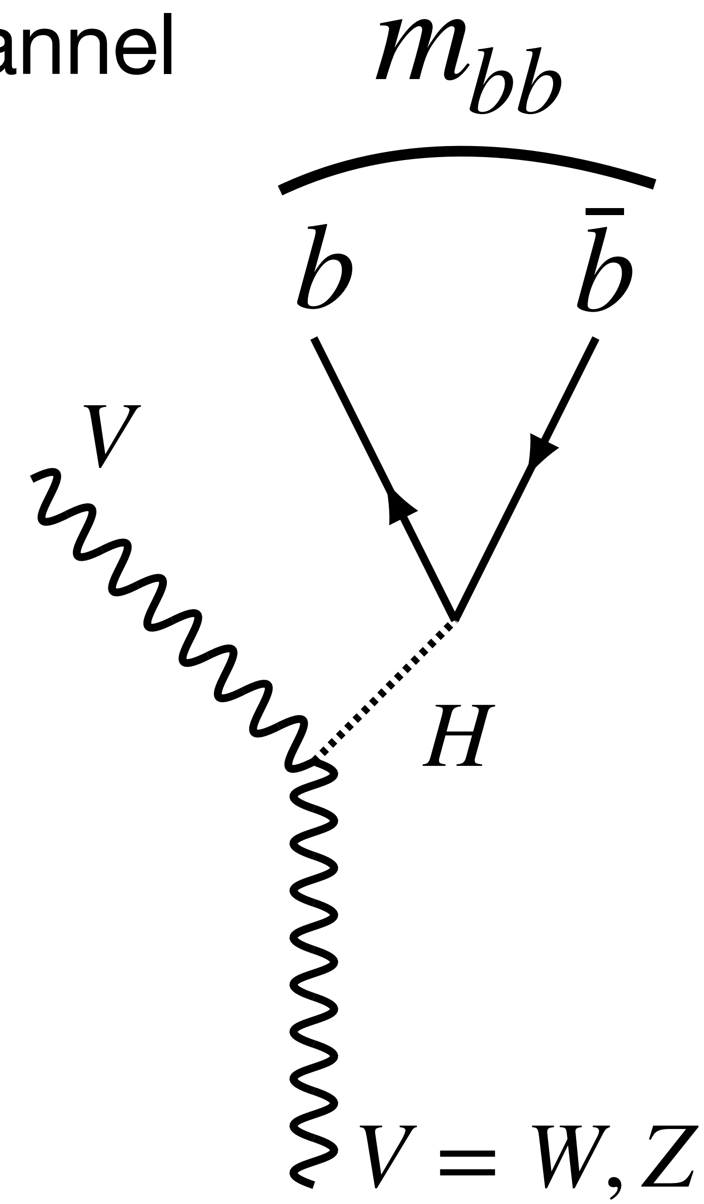
Into practice

Evaluation: toy analysis, VH/bb, 0-lepton channel

VH

$\;\;\hookrightarrow b\bar{b}$

$\hookrightarrow Z \rightarrow \nu\bar{\nu}, W \rightarrow \ell\nu$



Setup:

Backgrounds:

- W+jets, Z+jets
- ttbar
- di-boson (ZZ, WZ)

Simulation: MadGraph (LO) + Pythia8 + Delphes

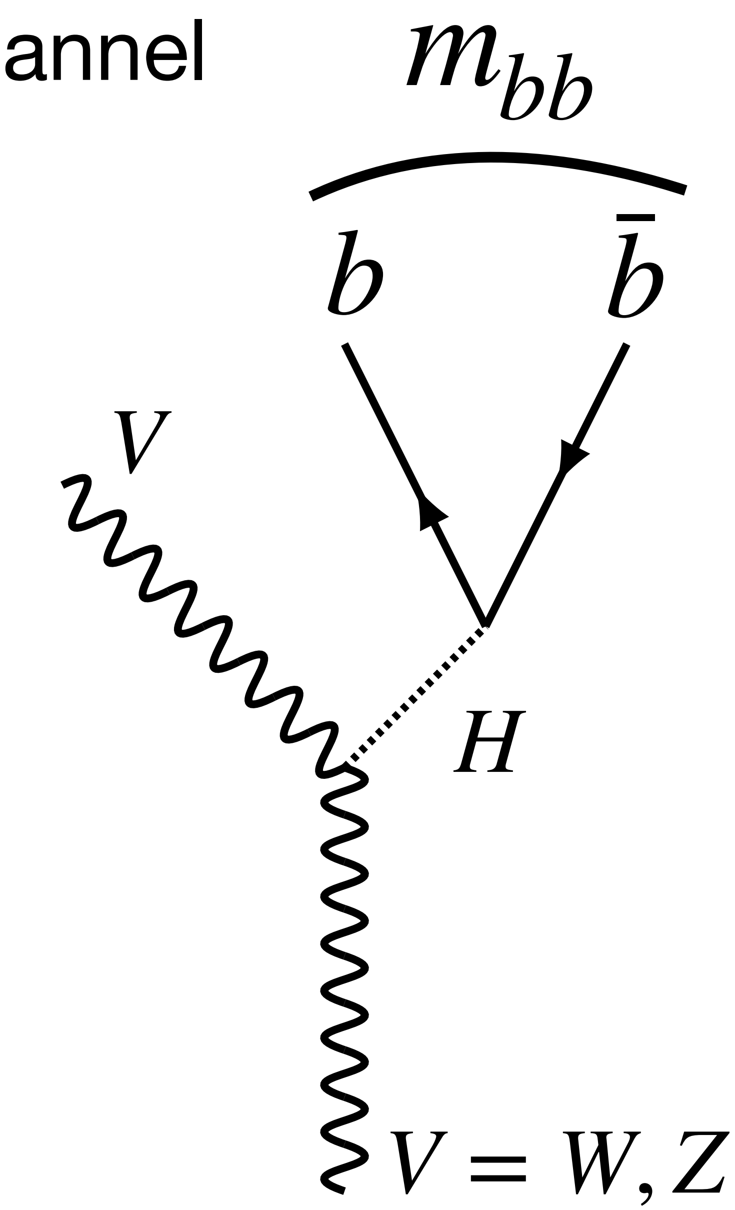
Into practice

Evaluation: toy analysis, VH/bb, 0-lepton channel

VH

$\hookrightarrow b\bar{b}$

$\hookrightarrow Z \to \nu\bar{\nu}, W \to \ell\nu$



Setup:

Backgrounds:

- W+jets, Z+jets
- ttbar
- di-boson (ZZ, WZ)

Simulation: MadGraph (LO) + Pythia8 + Delphes

Event selection: 2 or 3 jets, 2 b-tags, MET, no leptons

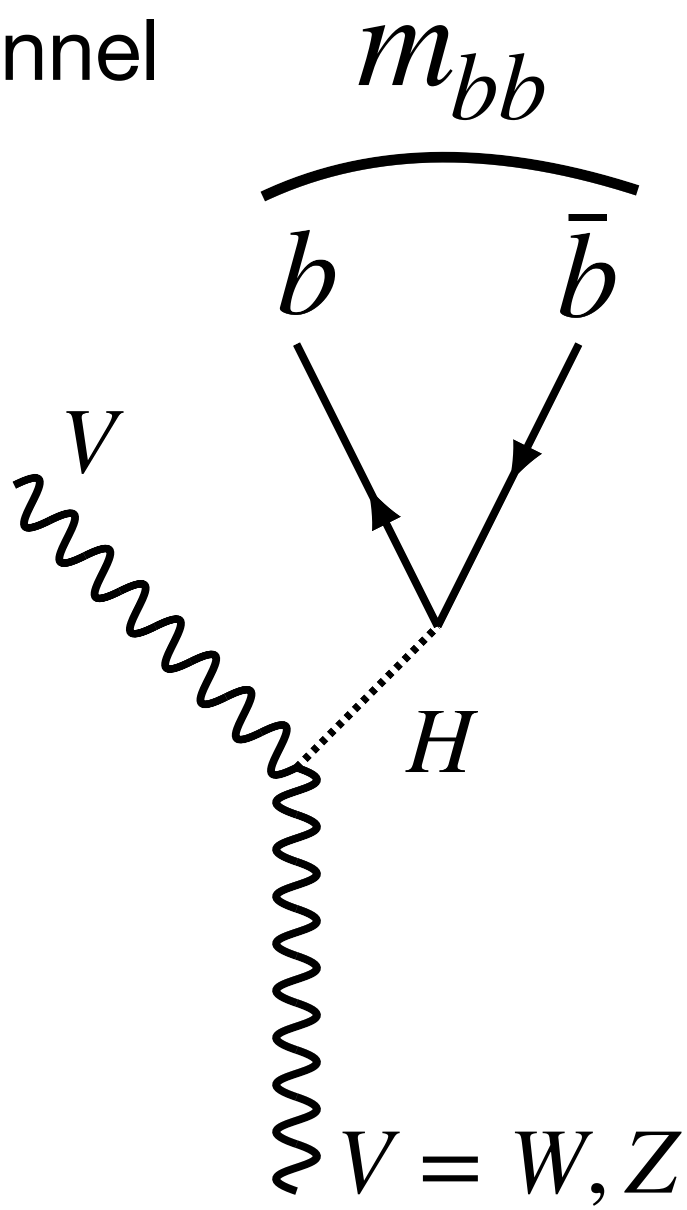
Into practice

Evaluation: toy analysis, VH/bb, 0-lepton channel

VH

$\quad \hookrightarrow b\bar{b}$

$\hookrightarrow Z \to \nu\bar{\nu}, W \to \ell\nu$



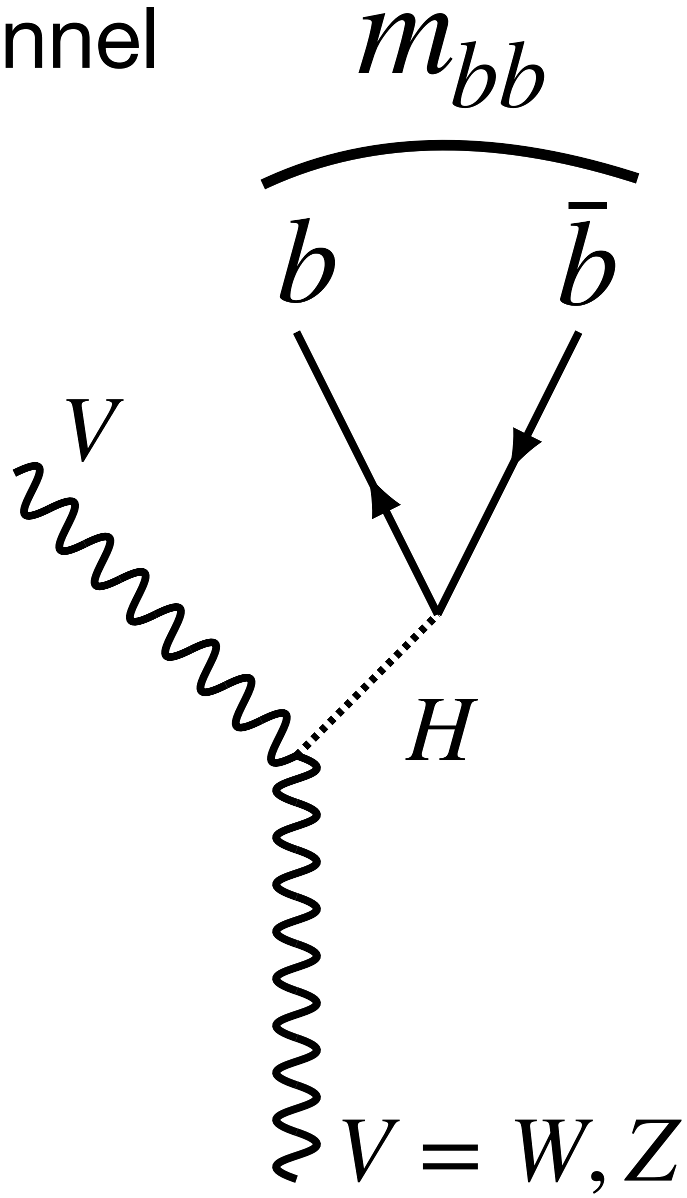
Into practice

Evaluation: **toy analysis**, VH/bb, 0-lepton channel

VH

$\quad\hookrightarrow b\bar{b}$

$\hookrightarrow Z \to \nu\bar{\nu}, W \to \ell\nu$



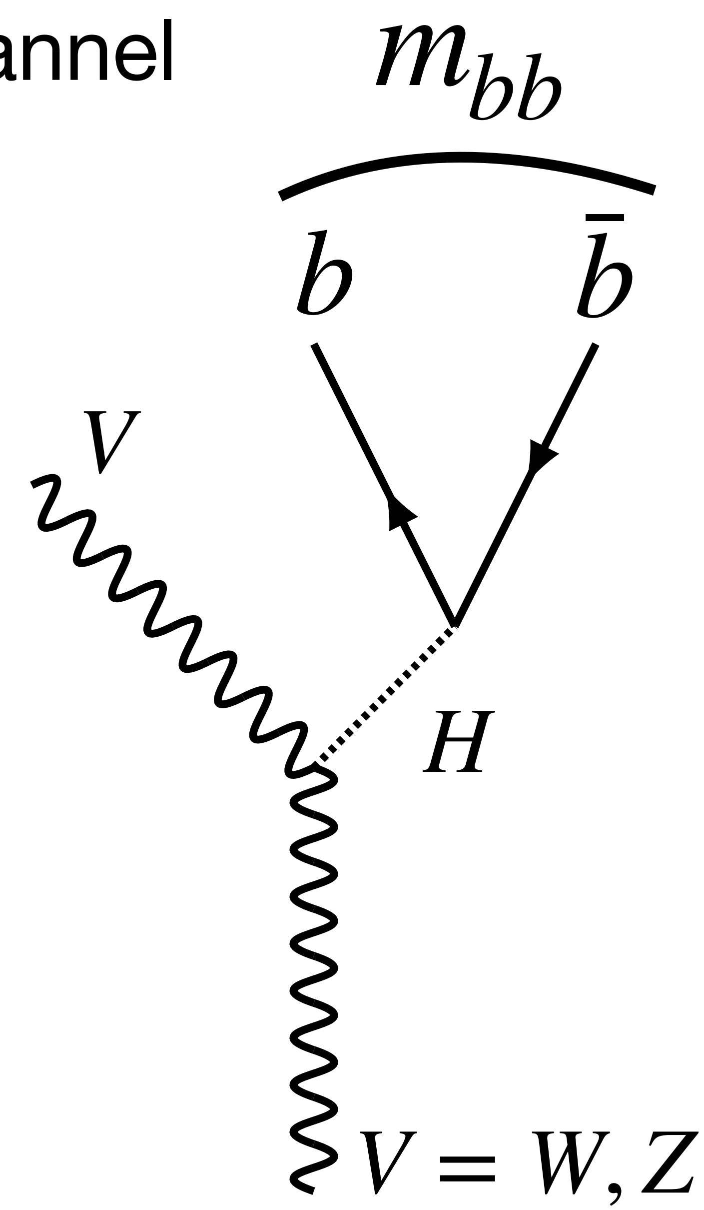
Signal regions:

	2-jet	3-jet
$\tilde{L}$ "tight"	**SR 1** m_{bb}	**SR 3** m_{bb}
$\tilde{L}$ "loose"	**SR 2** m_{bb}	**SR 4** m_{bb}

Into practice

Evaluation: **toy analysis**, VH/bb, 0-lepton channel

VH
$\quad \hookrightarrow b\bar{b}$
$\hookrightarrow Z \to \nu\bar{\nu}, W \to \ell\nu$



Signal regions:

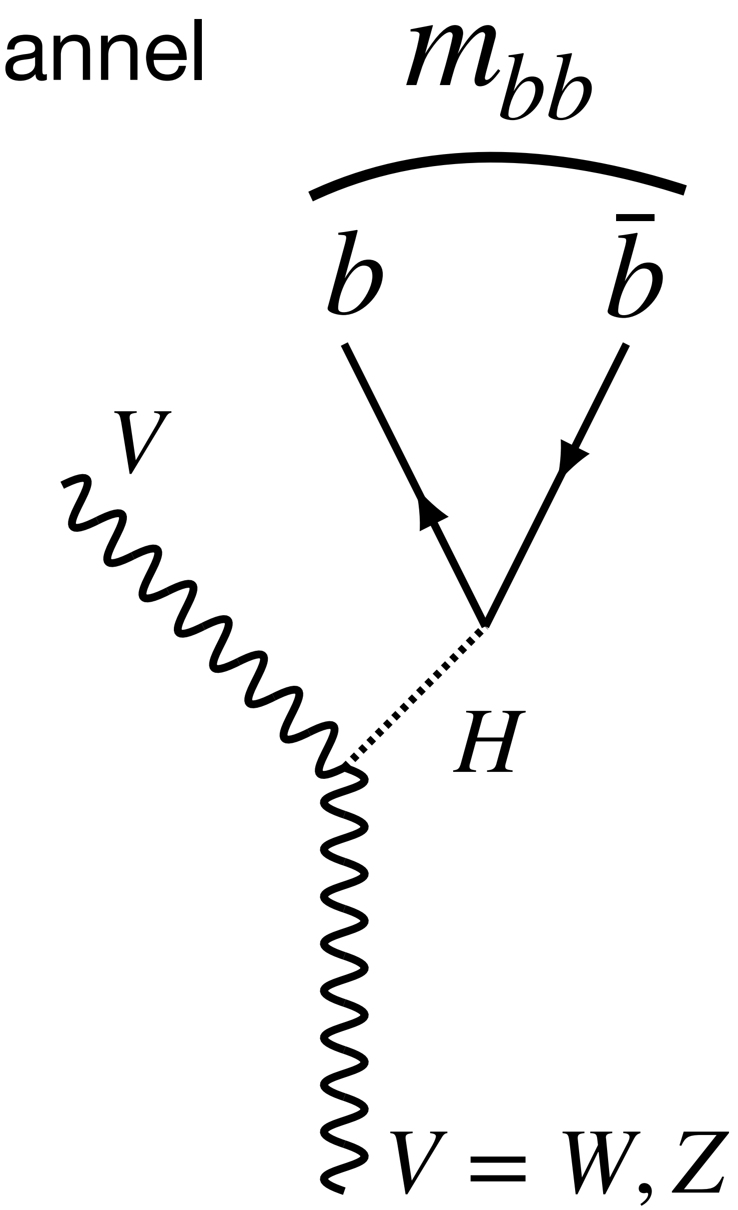
	2-jet	3-jet
$\tilde{L}$ "tight"	**SR 1** m_{bb}	**SR 3** m_{bb}
$\tilde{L}$ "loose"	**SR 2** m_{bb}	**SR 4** m_{bb}

Figure of merit: Asimov fit, signal and all backgrounds floating

Into practice

Evaluation: toy analysis, VH/bb, 0-lepton channel

VH
- $\to b\bar{b}$
- $\to Z \to \nu\bar{\nu}, W \to \ell\nu$

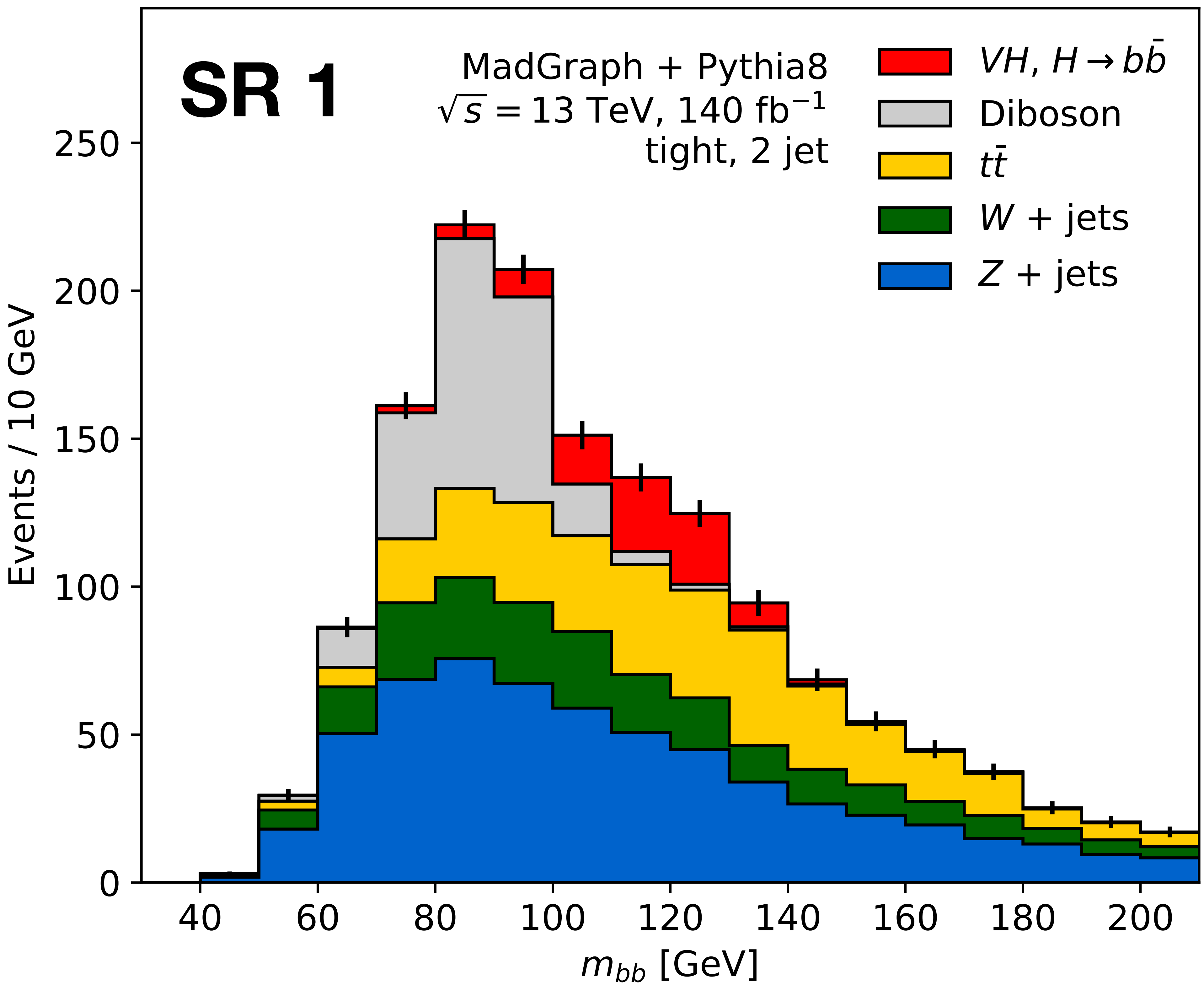


Signal regions:

	2-jet	3-jet
$\tilde{L}$ "tight"	**SR 1** m_{bb}	**SR 3** m_{bb}
$\tilde{L}$ "loose"	**SR 2** m_{bb}	**SR 4** m_{bb}

Figure of merit: Asimov fit, signal and all backgrounds floating

Into practice



Into practice

$$\tilde{L}(\mathbf{e}) = \arg\max_f \left(\mathscr{P}[f] - \lambda \cdot \mathbf{MI}(f, m_{bb}) \right)$$

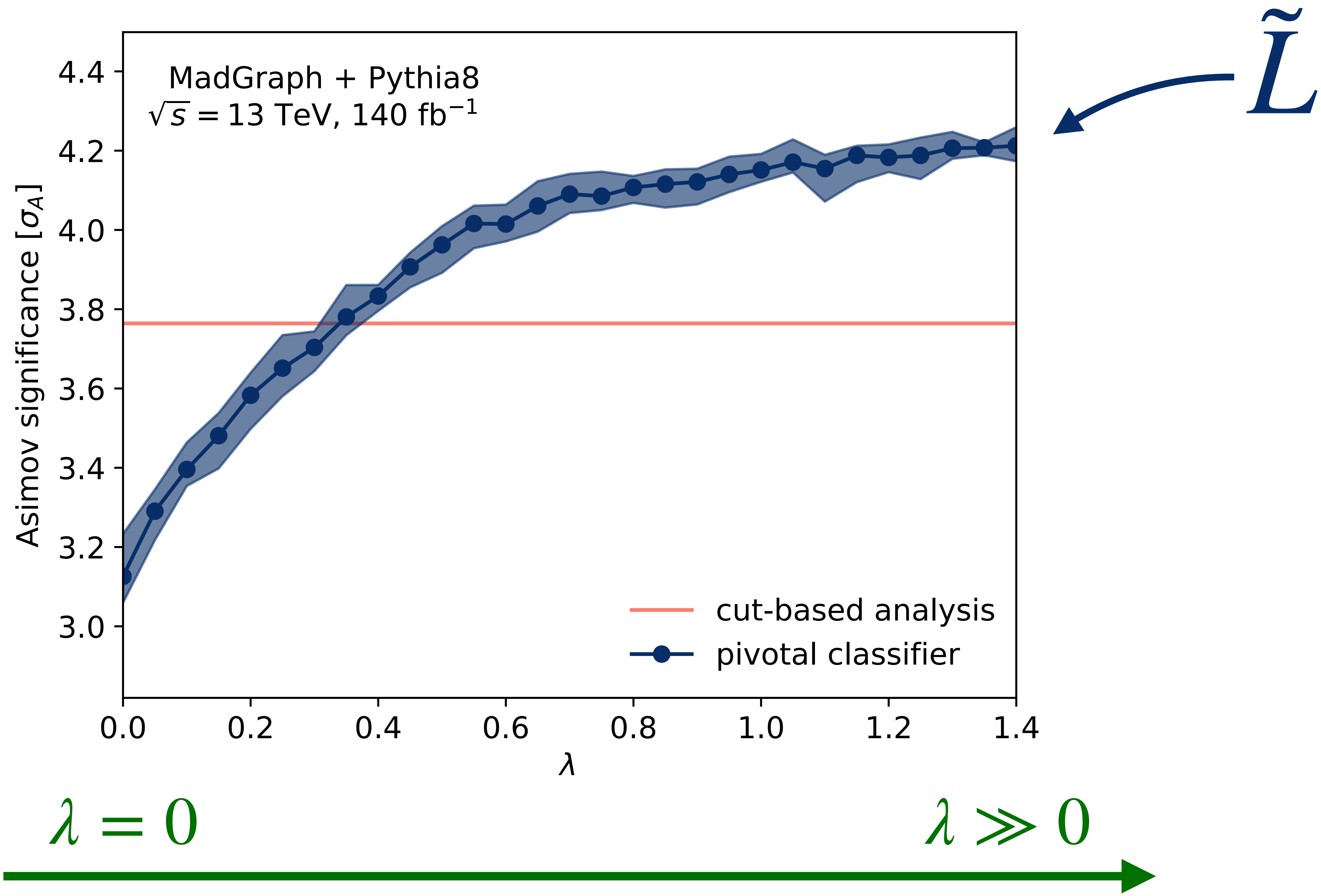
Into practice

$$\tilde{L}(\mathbf{e}) = \arg\max_f \left(\mathscr{P}[f] - \lambda \cdot \mathbf{MI}(f, m_{bb})\right)$$

$\lambda = 0$ $\lambda \gg 0$

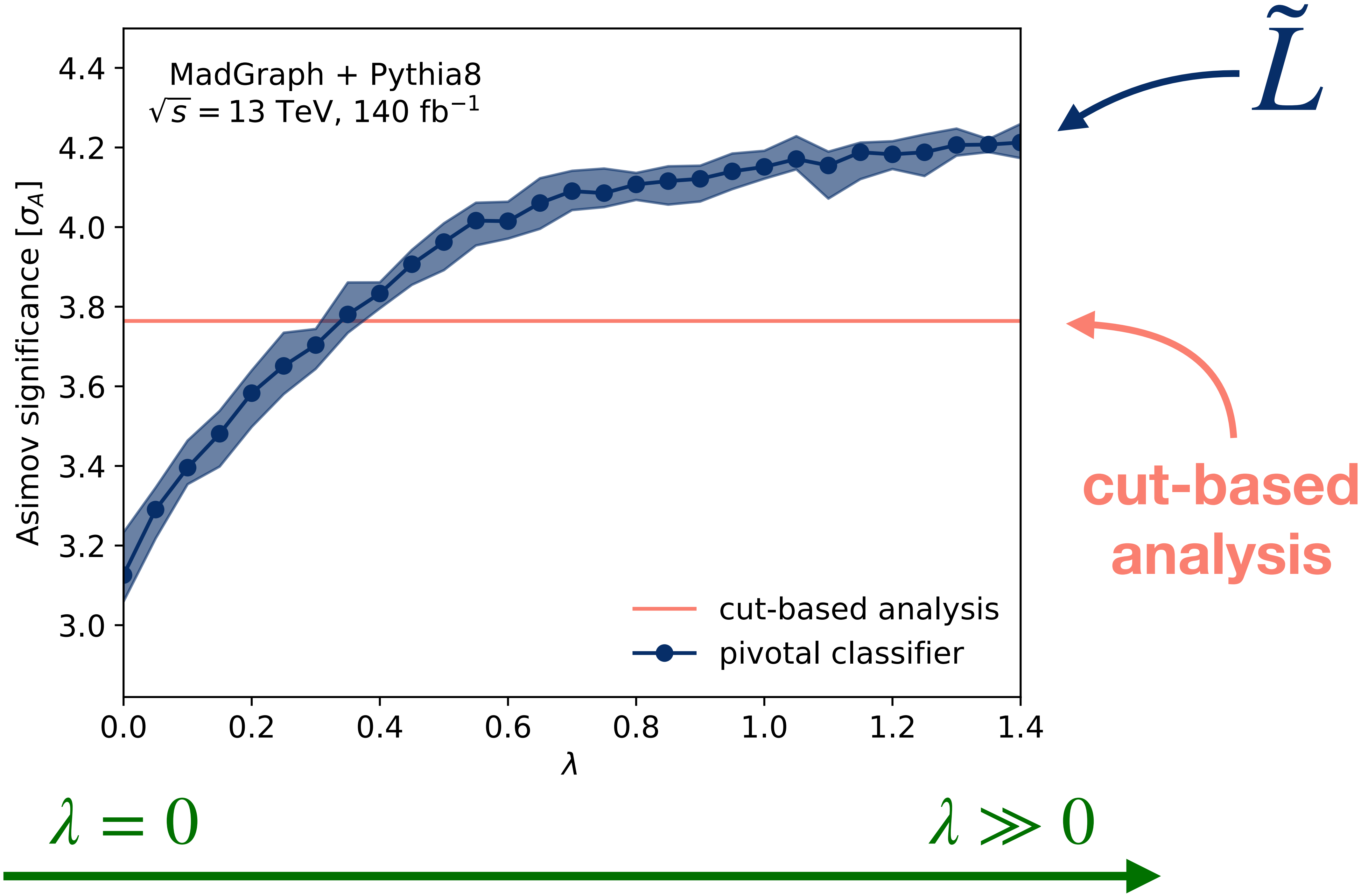
Into practice

$$\tilde{L}(\mathbf{e}) = \arg\max_f \left(\mathscr{P}[f] - \lambda \cdot \mathrm{MI}(f, m_{bb}) \right)$$



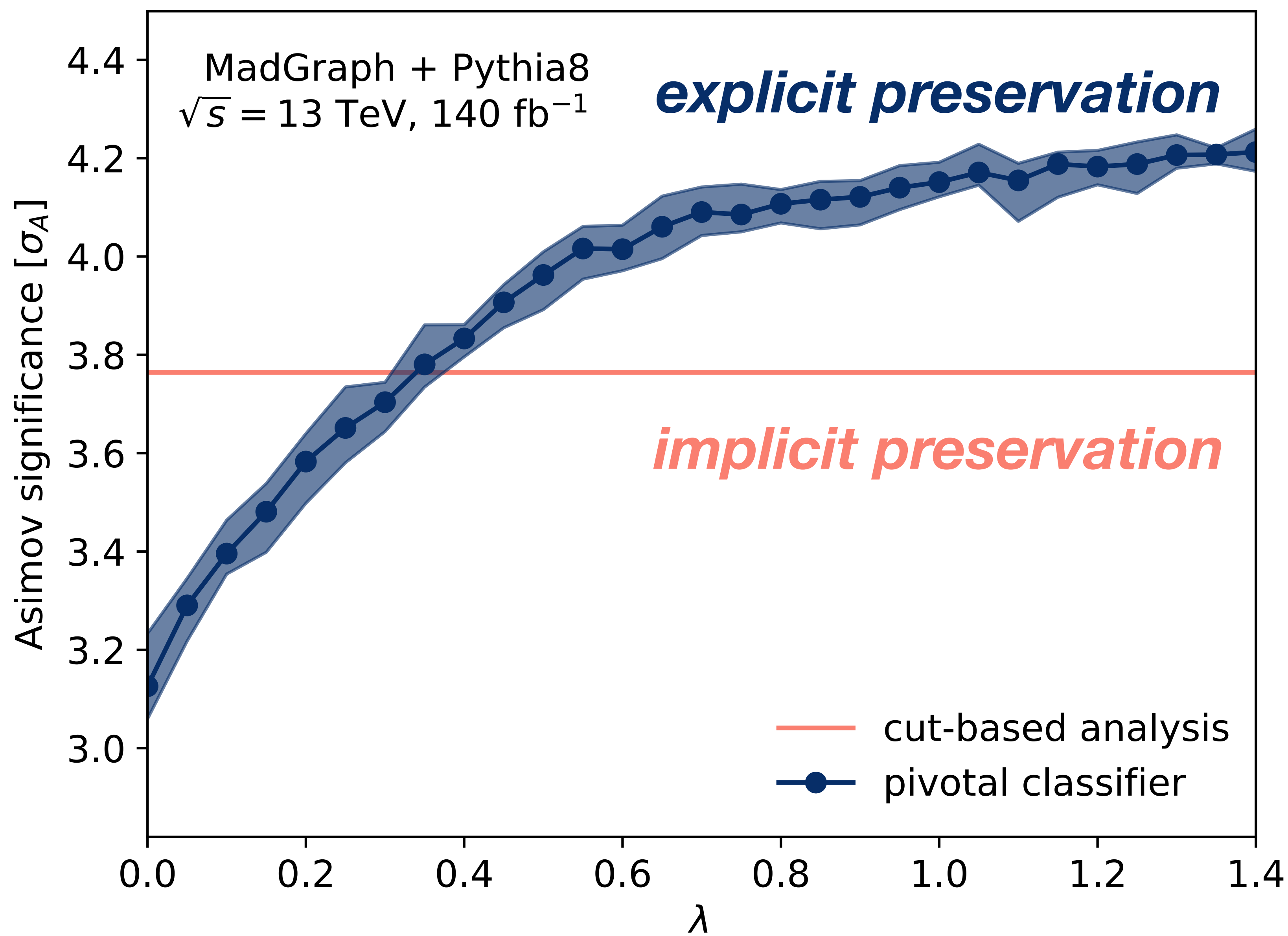
Into practice

$$\tilde{L}(\mathbf{e}) = \arg\max_f \left(\mathscr{P}[f] - \lambda \cdot \mathrm{MI}(f, m_{bb})\right)$$



Into practice

$$\tilde{L}(\mathbf{e}) = \arg\max_f \left(\mathscr{P}[f] - \lambda \cdot \mathrm{MI}(f, m_{bb}) \right)$$



$\tilde{L}$

cut-based analysis

$\lambda = 0$ $\lambda \gg 0$

Summary

- Event selections *distort* distributions
    - Physical observables lose interpretation

Summary

- Event selections *distort* distributions
  - Physical observables lose interpretation

- Construct discriminants that obey auxiliary constraints
  - Require independence $\longleftrightarrow$ *preserve* shapes

Summary

- Event selections *distort* distributions
  - Physical observables lose interpretation
- Construct discriminants that obey auxiliary constraints
  - Require independence $\longleftrightarrow$ *preserve* shapes
- **Ready to be used in practice**

Summary

- Event selections *distort* distributions
  - Physical observables lose interpretation

- Construct discriminants that obey auxiliary constraints
  - Require independence $\longleftrightarrow$ *preserve* shapes

- **Ready to be used in practice**

More information:

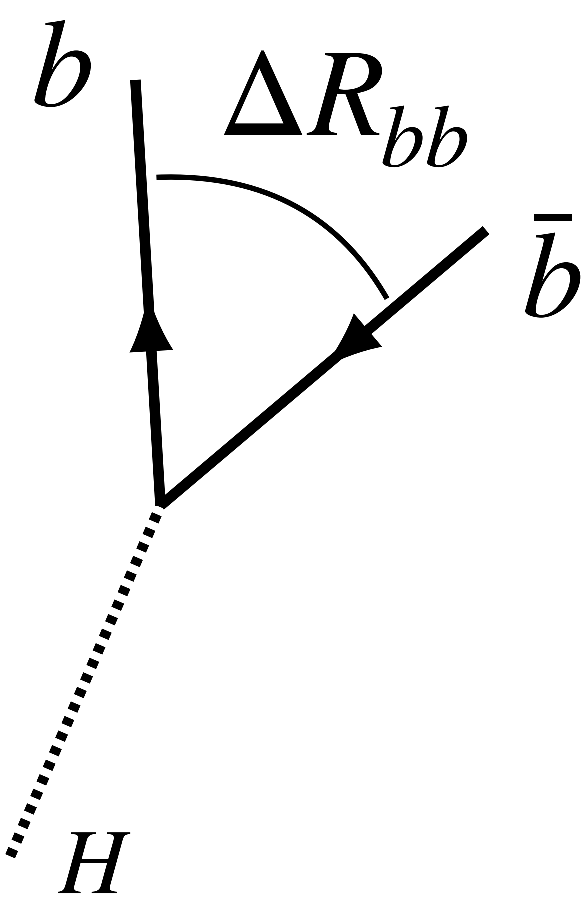
arXiv:1907.02098

Backup

Cut-based analysis

Signal regions: (inspired by a published ATLAS analysis)

	2-jet	3-jet
MET > 250 GeV $\Delta R_{bb} < 1.2$	**SR 1** m_{bb}	**SR 3** m_{bb}
250 > MET > 150 GeV $\Delta R_{bb} < 1.8$	**SR 2** m_{bb}	**SR 4** m_{bb}



Optimise cuts to maximise Asimov significance

“Pivotal classifier” analysis

Signal regions:

		2-jet	3-jet
$\tilde{L}$	“tight”	**SR 1** m_{bb}	**SR 3** m_{bb}
$\tilde{L}$	“loose”	**SR 2** m_{bb}	**SR 4** m_{bb}

- Cuts on $\tilde{L}$ designed to achieve the same signal efficiency in each region

Mutual Information

$\mathbf{f, m_{bb}}$ **independent** $\Longrightarrow$ $\mathbf{f, m_{bb}}$ **linearly uncorrelated**

Mutual Information

$$\mathrm{MI} = 0$$

$$\Downarrow$$

no nonlinear relationship

Pearson correlation

$$\rho = 0$$

$$\Downarrow$$

no linear relationship

Mutual Information

Two random variables:

$$X \qquad\qquad Y$$

Objects that characterise them:

$$p_X(x) \qquad\qquad p_Y(y)$$

$p_{(X,Y)}(x, y)$ … joint probability distribution

Mutual Information:

$$\text{MI}(X, Y) = D_{\text{KL}}\left(p_{(X,Y)} \,||\, p_X \cdot p_Y\right)$$

Kullback-Leibler divergence

"distance" between densities

Mutual Information

Kullback-Leibler divergence:

$$D_{\mathrm{KL}}(P\,||\,Q) = \int \mathrm{d}x\, p(x)\, \log\left(\frac{p(x)}{q(x)}\right)$$

"Distance" between probability distributions p and q.

Mutual information:

$$\mathrm{MI}(X,Y) = D_{\mathrm{KL}}\left(p_{(X,Y)}\,||\,p_X p_Y\right)$$

"Distance" between joint PDF and product of marginal PDFs.

$$\mathrm{MI}(X,Y) = \int \mathrm{d}x\,\mathrm{d}y\, p_{(X,Y)}(x,y)\, \log\left(\frac{p_{(X,Y)}(x,y)}{p_X(x)\,p_Y(y)}\right)$$

How to compute Mutual Information?

MI can recognise arbitrary complicated dependencies between random variables…

$$\mathrm{MI}(X, Y) = \int \mathrm{d}x\, \mathrm{d}y\, p_{(X,Y)}(x, y)\, \log\left(\frac{p_{(X,Y)}(x, y)}{p_X(x)\, p_Y(y)}\right)$$

… but requires knowledge about distributions.

MI also admits a functional definition:

$$\mathrm{MI}(X, Y) = \sup_{T \in \mathscr{F}} \langle T \rangle_{p(X,Y)} - \langle e^{T-1} \rangle_{p(X)p(Y)}$$

How to compute Mutual Information?

$$\tilde{L}(\mathbf{e}) := \arg\max_f \Big(\mathscr{P}[f] - \lambda \cdot \mathrm{MI}(f, m_{bb}) \Big)$$

neural network ($\mathscr{P}[f]$)

loss function ($\mathscr{P}[f] - \lambda \cdot \mathrm{MI}(f, m_{bb})$)

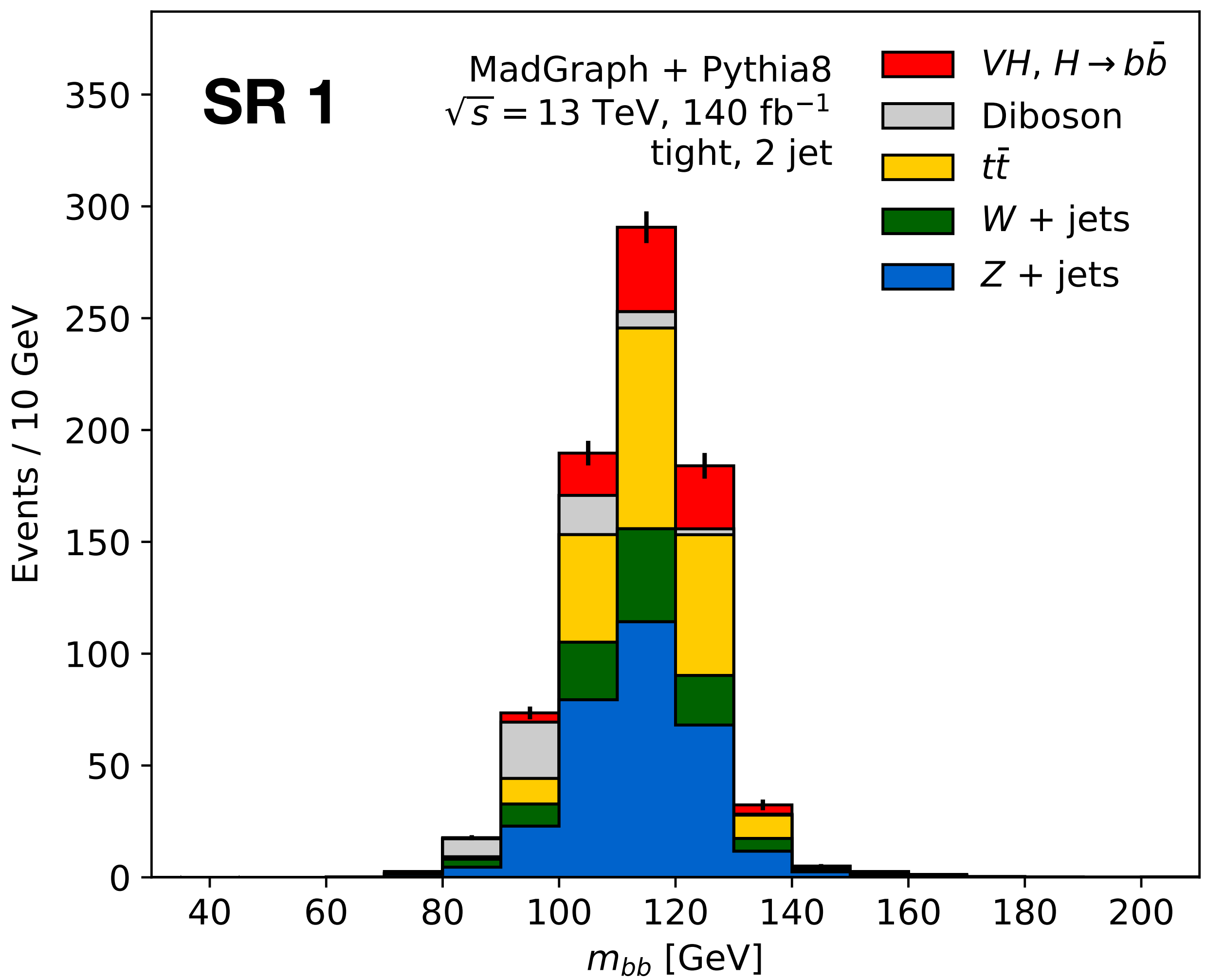
must be *differentiable* w.r.t. f

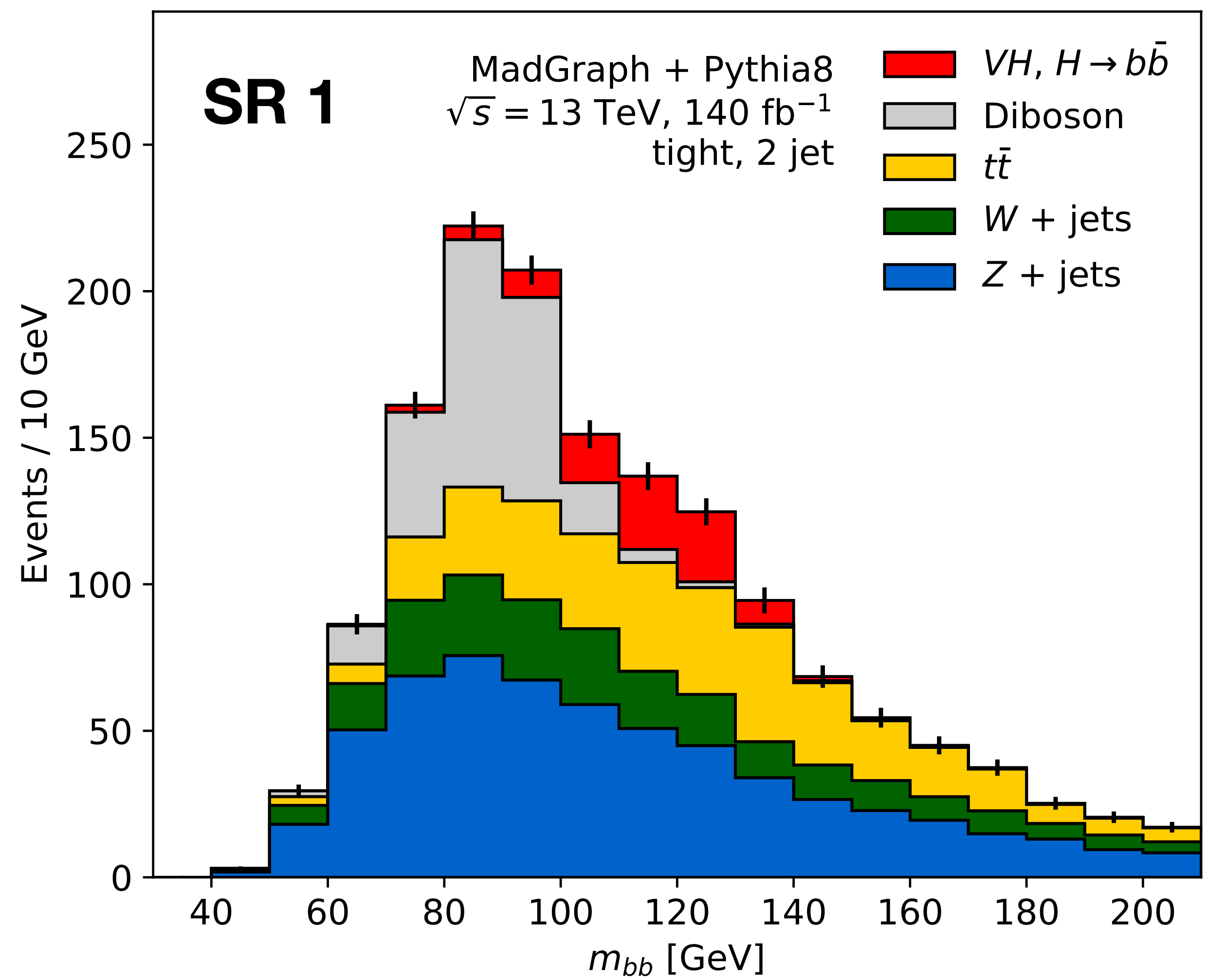
$$\mathrm{MI}(X,Y) = \sup_{T \in \mathscr{F}} \left(\langle T \rangle_{p(X,Y)} - \langle e^{T-1} \rangle_{p(X)p(Y)} \right)$$

adversarial **neural network** (T)

Into practice

$$\tilde{L}(\mathbf{e}) = \arg\max_f \left(\mathscr{P}[f] - \lambda \cdot \mathrm{MI}(f, m_{bb}) \right)$$

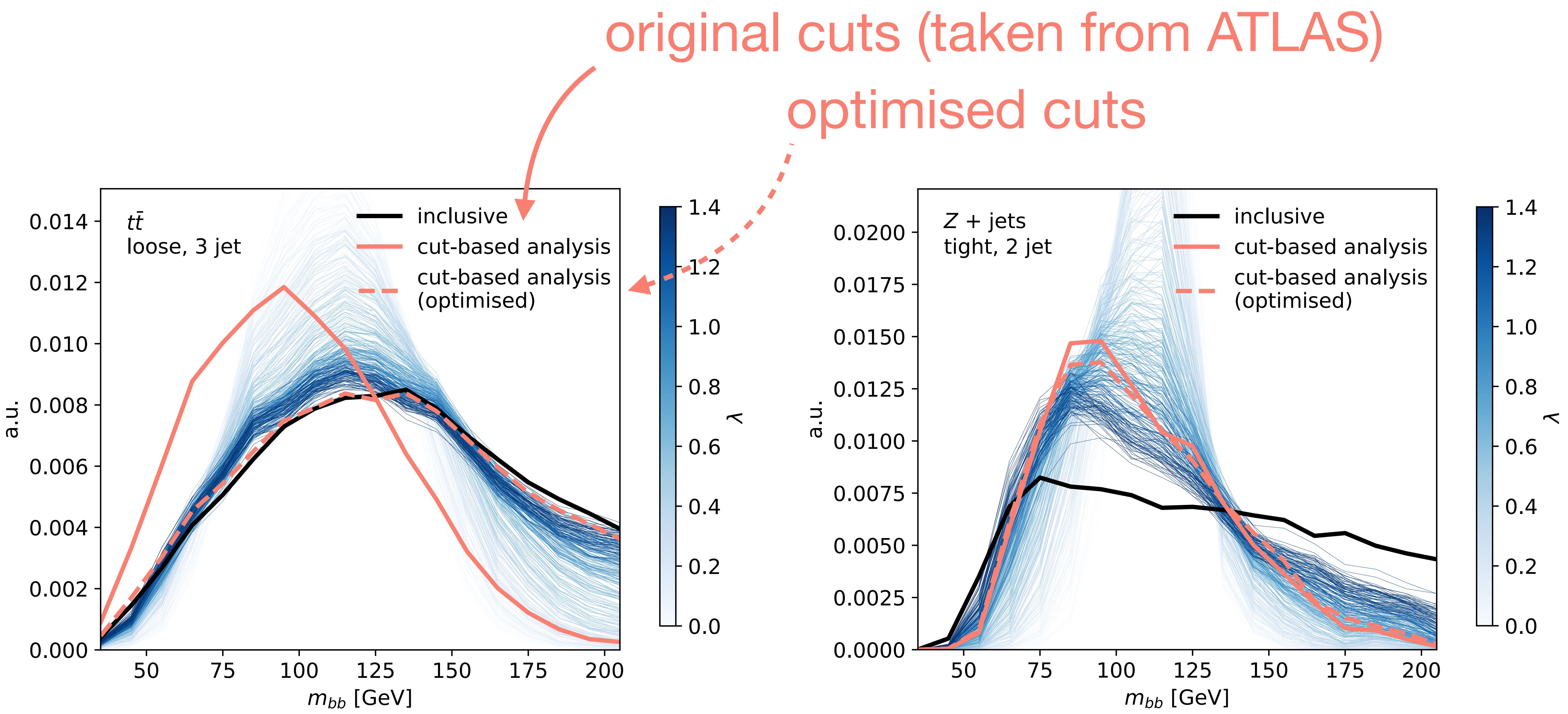




$\lambda = 0$ → $\lambda \gg 0$

More performance plots

Comparison: cut-based vs. pivotal classifier, mass shapes



More performance plots

Comparison: cut-based vs. pivotal classifier, significances