# Quantum Annealing for classification of stop events

*Dr. Pedrame Bargassa*
*IT – Lisbon*
*Mr. Artur Cordeiro Oudot Choi*
*Mr. Timothée Cabos*
*Mr. Timothée Hessel*
*Sorbonne Université*

*22/09/2021*

# Outline

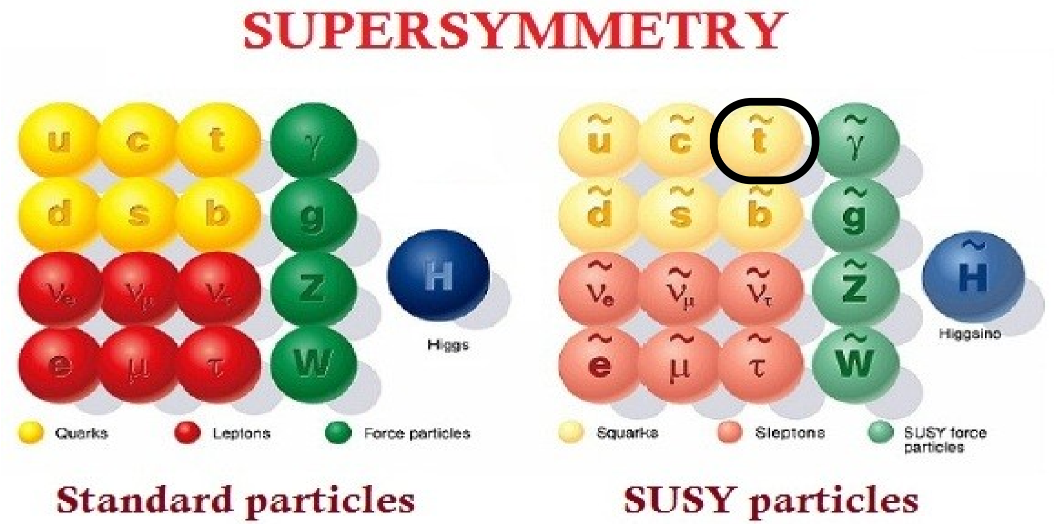
- Introduction on supersymmetric top quark
- Classification with Quantum Annealing:
  - Formulation ($H_{Ising}$) & principle
  - Implementation in hardware
- Approaches & Results
- Conclusion & Prospects

# Motivation & Scope

- Improvement of classification versus classical MVA tools
  - Quantum Annealing (QA)
  - Tool: D-Wave: 2000Q annealer / Chimera graph
- Stage in HEP: Standard Model (SM) vs Beyond-SM
  - Susymmetric top decaying in 4 bodies

- This work has 2 "points of reference":
  - QAML-Z: Quantum-Annealing based ML, PRA 102, 062405
  - Search of stop: BDT based search, JHEP 09 (2018) 065

# *Stop & its 4-body decays:*




***Brief phenomenological motivation***

***& experimental picture***

**MSSM lagrangian with soft breaking terms :**

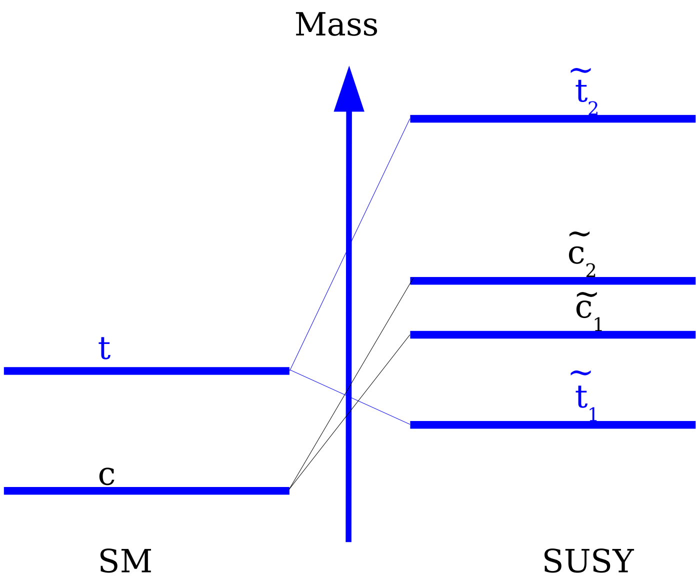
Quark left- & -right superpartners (scalars) can strongly mix to form mass eigenstates :

$$M^2_{\tilde{q}} = \begin{pmatrix} \tilde{M}_Q^2 + M_Q^2 + M_Z^2\left(\frac{1}{2} - \frac{2}{3}\sin^2\theta_W\right)\cos 2\beta & M_Q\left(A_T + \frac{\mu}{\tan\beta}\right) \\ M_Q\left(A_T + \frac{\mu}{\tan\beta}\right) & \tilde{M}_U^2 + M_Q^2 + \frac{2}{3}M_Z^2 \sin^2\theta_W \cos 2\beta \end{pmatrix}$$

"Up" squarks

$A_T$ : Tri-linear (stop) mixing term

$M_Q$ = SM quark mass




**Mass difference of quark superpartners:**

Proportional to $M_Q = M_t$ :

Strong mixing in the stops $\tilde{t}_{1,2}$ sector

$\longrightarrow$ $\boxed{\tilde{t}_1 \text{ might be the lightest squark}}$

# Stop & Cold Dark Matter

Lightest Neutralino $\tilde{\chi}^0_1$ stable: Natural candidate for Cold Dark Matter

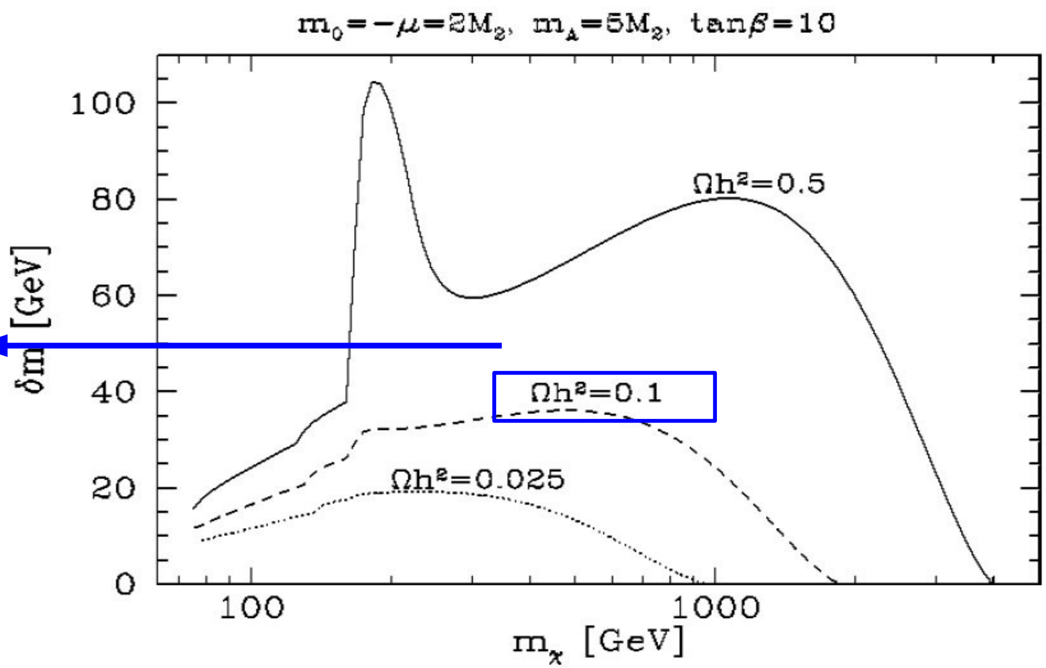
Observed $\mathbf{\Omega_{CDM} h^2 = 0.111 \pm 0.006}$ **@ 95% CL** (WMAP) well explained

**IF $\delta$m = m($\tilde{P}$) - m($\tilde{\chi}^0_1$) small**: Co-annihilations dominate

- $\tilde{\chi}^0_1\,\tilde{t}_1 \rightarrow tg,\ tH^0_i,\ bH^+$
- $\tilde{t}_1\,\tilde{t}^{(*)}_1 \rightarrow t\bar{t},\ gg,\ H^0_iH^0_j,\ H^-H^+,\ b\bar{b}$

$\Delta m = m(\tilde{P}) - m(\tilde{\chi}^0_1) \leq 50\ \mathrm{GeV}/c^2$

...Soft decay products




Boehm et al., PRD 62, 35012

**Exciting for HEP in view of Cosmology Data:**

***Is stop/sbottom/stau degenerate with LSP ?***

# $\tilde{t}_1 \to b\, f\, f'\, \tilde{\chi}^0_1$

**Final state / Event signature:**

1 lepton

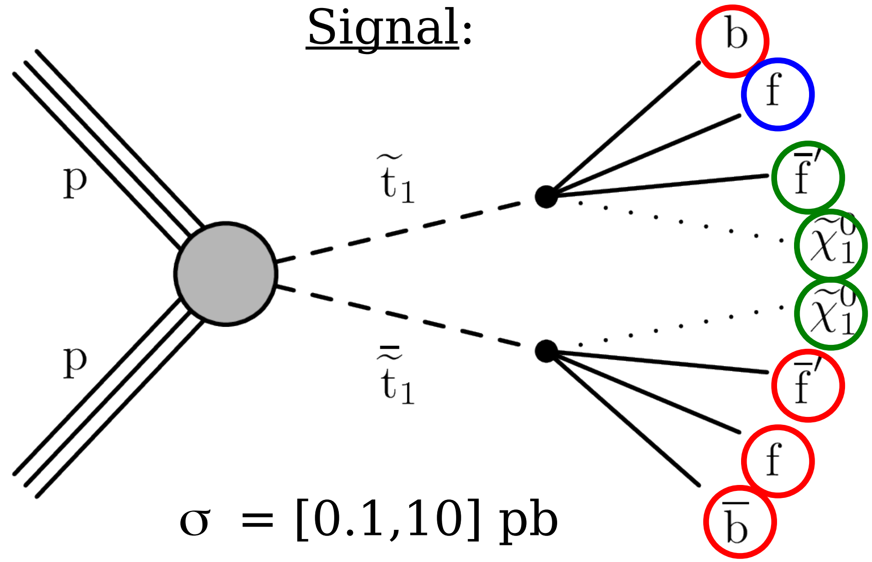
At least 1 jet

Missing transverse energy

**Preselection:**

- $p_T(j1) > 110$. 5, $3.5 < p_T(e,\mu) < 30$
- MET > 280   $H_T > 200$

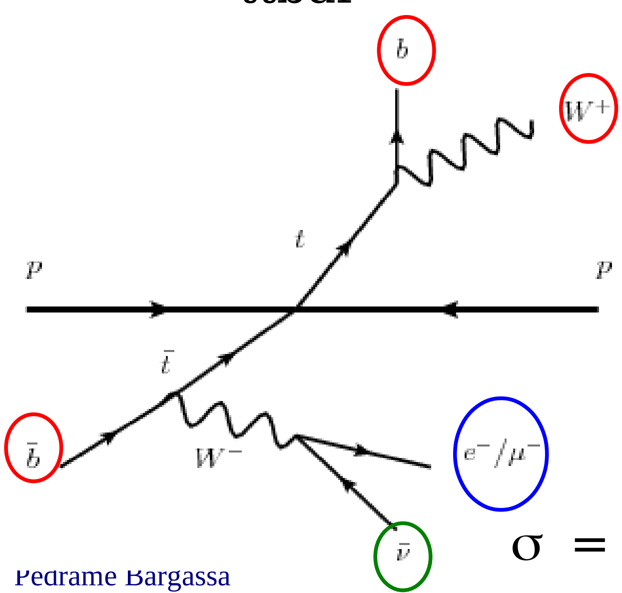
Signal:




$\sigma$ = [0.1,10] pb

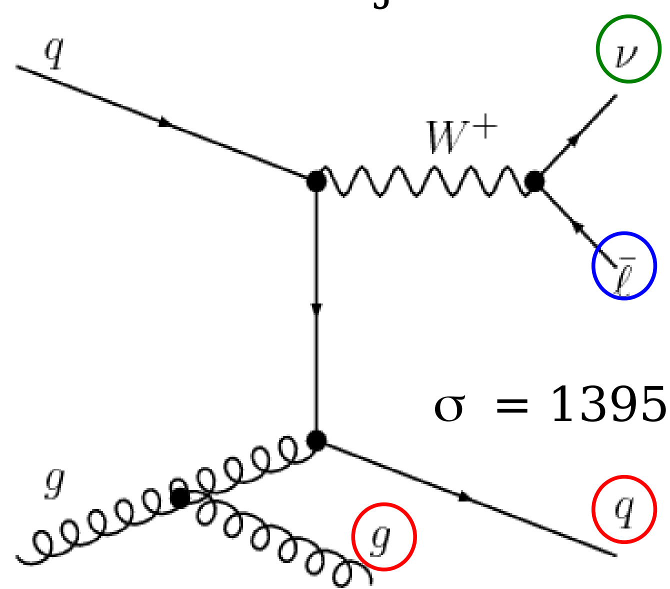
Main backgrounds:

ttbar



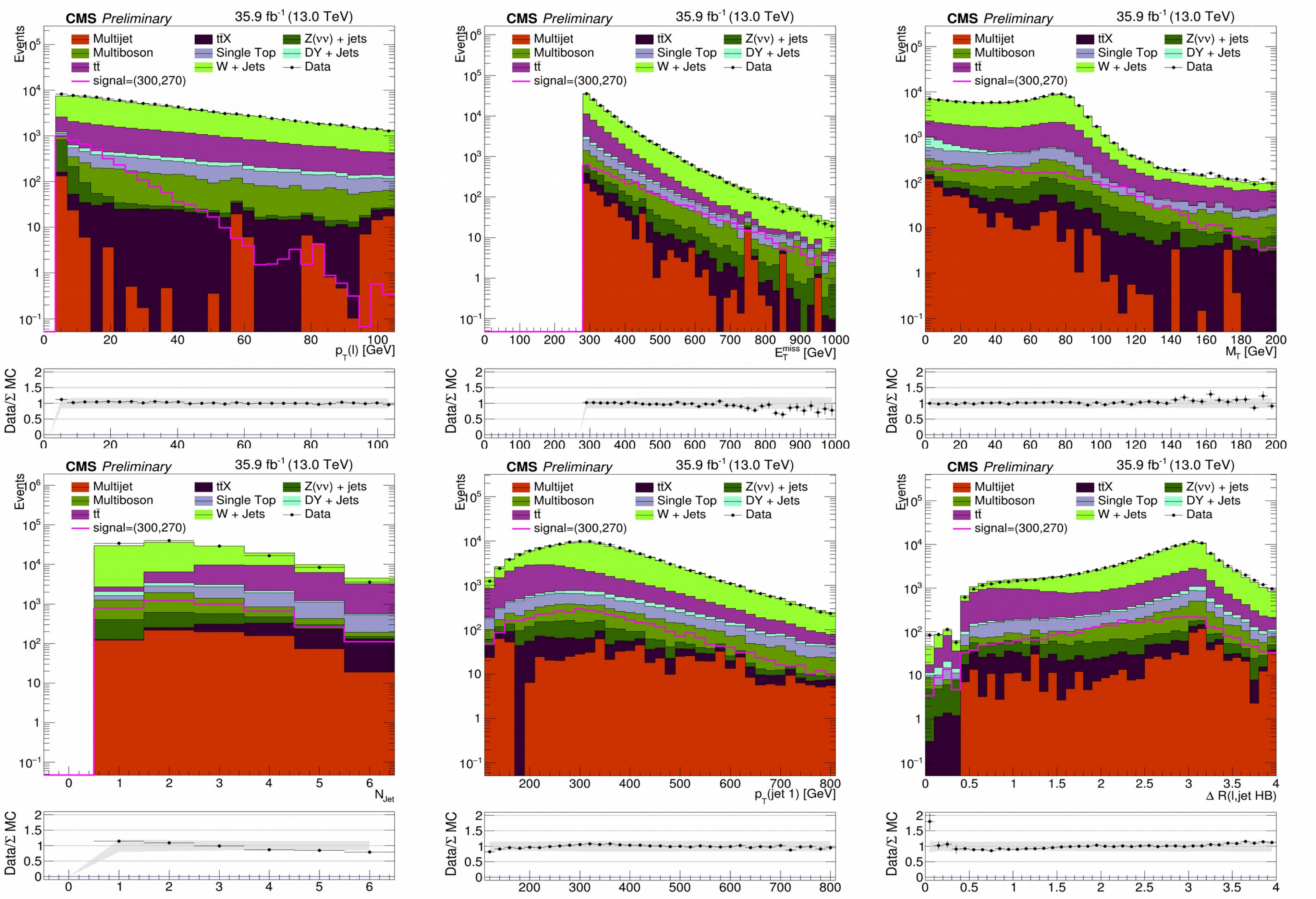

$\sigma$ = 835 pb

W+jets




$\sigma$ = 1395 pb

# **Classification challenge:** abundance & overlap



# *Quantum Annealing: Zooming & Augmenting*

# **Classification:** weak & strong classifiers

Let $x$ be the vector of input/discriminating variables

**First build a set of weak classifiers $c_i(x)$**, one for each variable

- Made by user
- $c_i(x)$: + / - for S / B
- $\nu$": combination of variable $\nu$ & its 70$^{th}$ & 30$^{th}$ percentiles
- $\nu_{+1}$: 10$^{th}$ Signal percentile
- $\nu_{-1}$: 90$^{th}$ Background percentile
- Reflective of S-B separation in $x$

$$c(\nu) = \begin{cases} +1 & \text{if } \nu_{+1} < \nu''(\nu) \\ \dfrac{\nu''(\nu)}{\nu_{+1}} & \text{if } 0 < \nu''(\nu) < \nu_{+1} \\ \dfrac{\nu''(\nu)}{|\nu_{-1}|} & \text{if } \nu_{-1} < \nu''(\nu) < 0 \\ -1 & \text{if } \nu''(\nu) < \nu_{-1} \end{cases}$$

**Then build a strong classifier** with all weak classifiers $c_i(x)$

$$R_{\vec{w}}(\vec{x}) = \sum_i w_i c_i(\vec{x})$$

$w_i = \{0,1\}$: problem to be solved by QA

$y_\tau$: {-1,+1} event-by-event ($\tau$) label for {B,S}, provided by user

# **Classification:** metric to minimize

Minimize an euclidean distance:
Quantified comparison between the labeling (y) & the classification ($R_w$) of events ($\tau$)

$$\delta(\vec{w}) = \|\vec{y} - \vec{R}_{\vec{w}}\|^2 \qquad \vec{w}_{\min} = \mathrm{argmin}_{\{\vec{w}\}}\delta(\vec{w})$$

Labeling, based on simulation:
we know what is S & B

Prediction, based on QA:
we classify S & B

**Optimization is expressed in the optimal set of spins minimizing Energy of an Ising Hamiltonian:**

$$H_{\text{Ising}} = \sum_{i<j}^{N} J_{ij} \boxed{s_i s_j} + \sum_{i=1}^{N} h_i \boxed{s_i}$$

$$J_{ij} = C_{ij}/4$$
$$h_i = \frac{1}{2}\left(\lambda - 2C_i + \sum_j C_{ij}\right)$$

**Strong classifier:** $$R(\boldsymbol{x}) = \sum_i \boxed{s_i^g} \boxed{c_i(\boldsymbol{x})} \in [-1, 1]$$

$c_i$(x): quasi-discrete: +-1/$N_{var}$. $s_i$: discrete → weak-classifiers turned on/off

# **Classification:** 1) Zoming

$$\boxed{s_i \quad \longrightarrow \quad \mu_i(t) + s_i \cdot \sigma(t) = \mu_i(t+1),}$$

- $\mu_i(t)$ is the mean value of qubit $i$ at time $t$. We have: $\forall i\ \mu_i(0) = 0$.
- $\sigma(t)$ is the search width at each annealing iteration $t$. We have: $\sigma(t) = b^t$ where $b = 1/2$ and $t \in [0, T-1]$.

Effectively shifts & narrows the region of search in the space of spins.
Updates vector $\mu_i$ collected @ final iteration (T-1) to form the strong classifier:

$$R(x_\tau) \quad = \quad \sum_{i=1}^{N} \mu_i(T-1) c_i(x_\tau),$$ → Use of weak-classifiers $c_i$ is now continuous

- Zooming increases over-fitting. 2 step randomization procedure to regularize the iterative process:
  - @ each iteration: if $E_i(t+1)>E_i(t)$: spin flip $s_i \rightarrow -s_i$ with probability $p_f(t)=\{0.16,0.08,0.04,0.02,0.01,...,0.01\}$
  - Subsequently, all qubits are uniformly randomly flipped: $s_i \rightarrow -s_i$ with probability $q_f(t)=\{0.08,0.04,0.02,0.01,0.005,...\}<p_f(t)$

# **Classification:** 2) Augmentation

Weak-classifier more "informative" about discriminating distributions

$$\boxed{c_{il}(x_\tau) \quad = \quad \frac{sgn(h_i(x_\tau) + \delta l)}{N},}$$

where $l \in \mathbb{Z}$ is the offset: $-A \leq l \leq A$, and $\delta$ is the step size. While the value $c_i$ of the old classifier has only a binary outcome for $h_i$, the new classifiers $c_{il}$ has similar but $(2l+1)$ different outcomes depending on the very distribution of $h_i$. We therefore have a better discrimination because a more continuous, thus more precise representation of the spectrum of $h_i$ with $c_{il}$ than with $c_i$.

$c_{il}(x) = \{-1,-1,\ldots,+1,+1,+1,+1\} / N_{var}$

Final Hamiltonian encoded on quantum annealer & optimized iteratively:

$$H(t) = \sum_{I=1}^{N(2A+1)} \left( -C_I + \sum_{J=1}^{N(2A+1)} \mu_J(t) C_{IJ} \right) \sigma(t) s_I + \frac{1}{2} \sum_{I=1}^{N(2A+1)} \sum_{J \neq I}^{N(2A+1)} C_{IJ} \sigma^2(t) s_I s_J$$

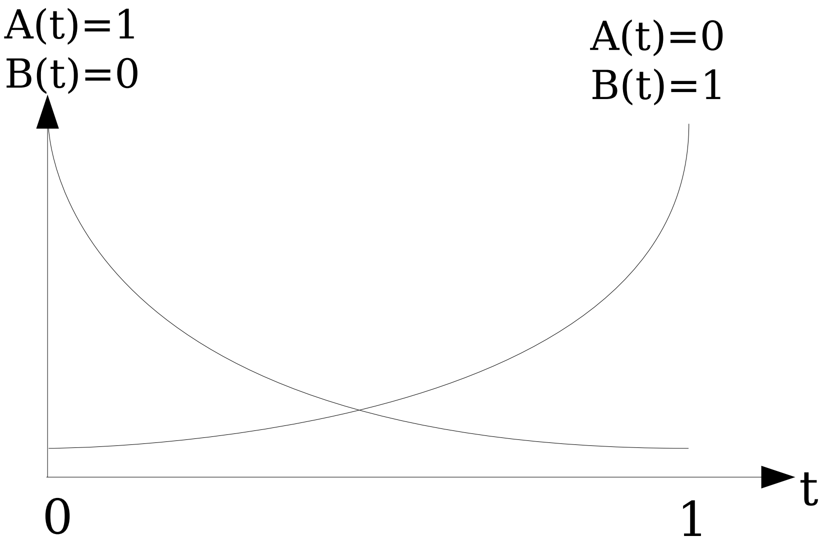
$$C_I = \sum_{\tau=1}^{S} c_{il}(x_\tau) y_\tau, \quad C_{IJ} = \sum_{\tau=1}^{S} c_{il}(x_\tau) c_{jl'}(x_\tau)$$

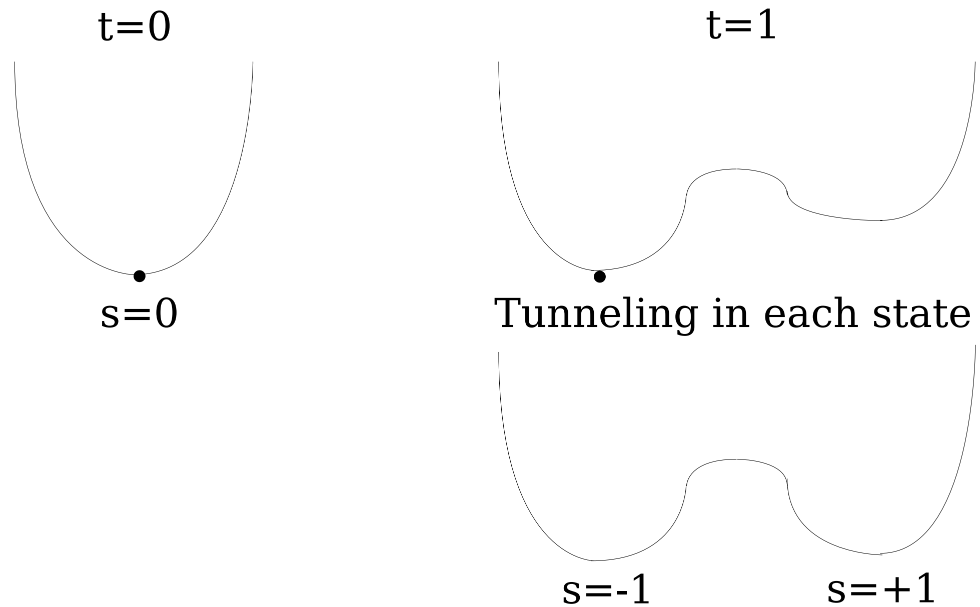
# **Classification:** annealing

QA: spin $\mathbf{s_i}$ of $H_{Ising}$ is interpreted as operator $\sigma_i^z$ on $i^{th}$ cubit in a system of N cubits

$$H_{\text{Ising}} = \sum_{i<j}^{N} J_{ij} s_i s_j + \sum_{i=1}^{N} h_i s_i$$

| | | |
|---|---|---|
| $H(t) = A(t)\ .\ \Sigma_i\ \sigma_i^x$ | $\pm$ | $B(t)\ .\ \{\Sigma_{ij}\ J_{ij}\sigma_i^z\sigma_j^z + \Sigma_i\ h_i\sigma_i^z \pm 1\}$ |
| Driver Hamiltonian<br>Transverse field | | Problem Hamiltonian: Biases<br>bought by the problem: $h_i$, $J_{ij}$: inputs |




If at t=x lowest E isn't reached:

→ increase ta ta= {1, 5, 20, 100, 2000} μs

→ pause annealing to allow the search by transverse field

# **Implementation:** $H_{Ising}$ & …hardware

Logical Hamiltonian:

$H_{Ising}$ supposes that 1 spin can be connected to all spins: coupler $J_{ij}$

$$H_{\text{Ising}} = \sum_{i<j}^{N} J_{ij} s_i s_j + \sum_{i=1}^{N} h_i s_i$$

This isn't (yet ?) feasible in HW

**Embedding:** Strongly couple a chain of qubits so that it represents one of the spins $s_i$ in HW

→ HW $H_{Ising}$ = Effective H with its limitations

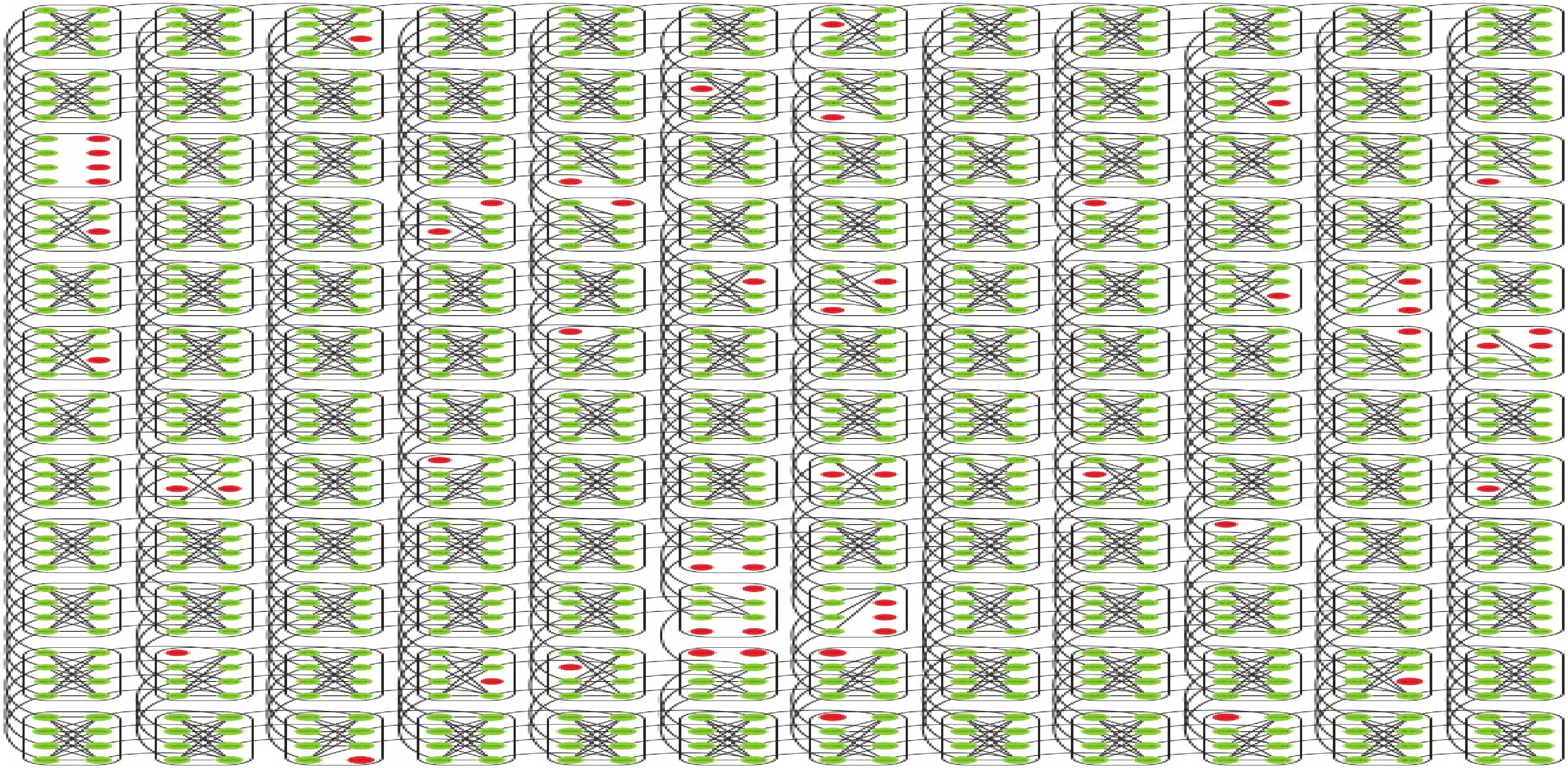
Any limitation in embedding means a non-correspondance between {logical classification problem} and {its implementation in the HW}

**Variable fixing:** Option to reduce the size of the Ising model to be encoded on the annealer

Use the polynomial-time variable fixing scheme of the D-Wave API: classical procedure to fix the value of a portion of the input variables to values that have a high probability of being optimal

# **Implementation:** DW chimera

2048 superconducting junction flux qubits arranged into a grid, with 5600 couplers between the qubits



$N(J_{ij})=[N_v.(N_v\text{-}1)]/2$ w $N_v=N_{var}.(2A+1)$: can reach 8646 for A=5 and $N_{var}=12$

Chimera graph doesn't necessarily have enough links:

Proceed with $[(1\text{-}$Cutoff$)\ .\ N_J]$ links: reduce the size of model to be encoded

# **Implementation:** DW chimera

Final degree of freedom: the chain strength JF
If the strength of the couplers in the ferromagnetic chains making up the logical units is 1, then the maximum strength of any other coupler is 1/JF

There is a competition between {the chain needing to behave as a single large qubit} and {the problem Hamiltonian needing to drive the dynamics}:

→ If JF is very large, the chains will “freeze out” long before the logical problem, i.e., the chains will be far stronger than the problem early on, and the transverse field terms will be unable to induce the large multi-qubit flipping events necessary to explore the logical problem space

→ If JF is very weak, the chains will be broken (i.e., develop a kink or domain wall) by tension induced by the problem, or by thermal excitations, so the system will generally not find very good solutions

→ Measure of the qubit chains performed through a majority vote (across different samplings): possibly leading to the collection of non-optimal sets of solution spins, thus to a possible loss of discriminating information

Ideally: we want the chains & the logical problem to freeze at the same time

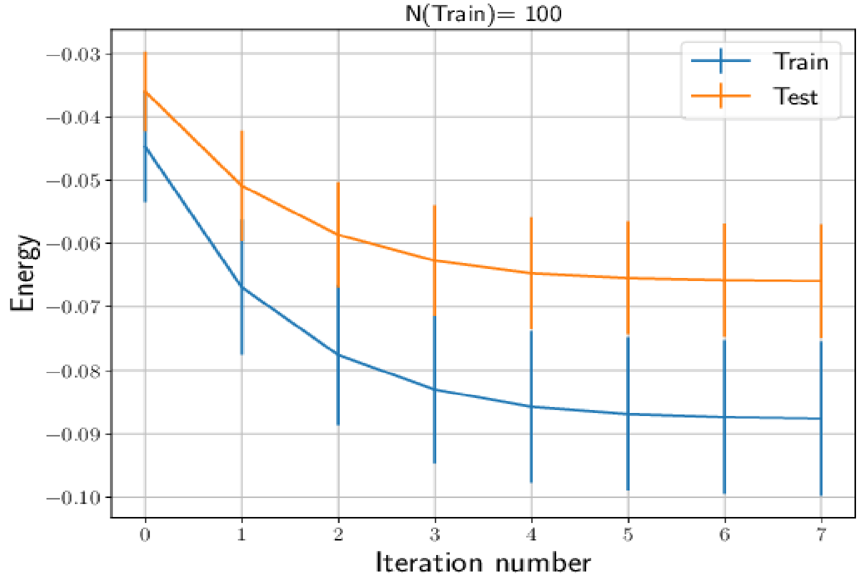
# **Implementation:** DW annealing specifics

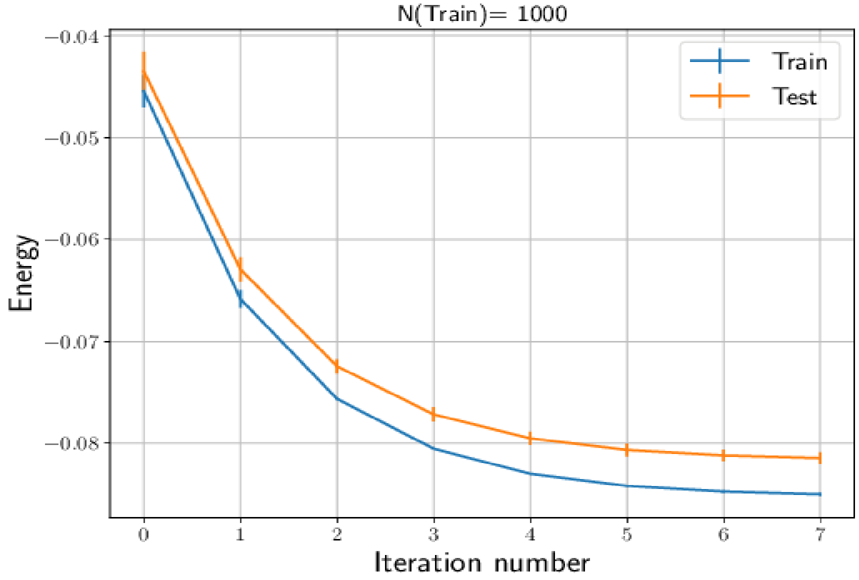
- **T(annealing) = 20 μs**
- **N(iterations) = 8**
  - **{Ising model → annealer} encoding**: reduce random errors on the $h_i$ & $J_{ij}$ by randomizing the encoding via sign flips, annealing over g gauges where g[8] ={50,10,10,...,10}
    - For each gauge: sample annealing result 200 times & measure qubit chains with a majority vote
  - **Chain strength:** ratio JF = Coupling within each chain / Largest coupling in Hamiltonian. JF[8]={3.0,1.0,0.5,0.2,0.1,0.1,...}
  - **Selection of excited states:** 2 criteria ANDed:
    - $\{E_{ground}<0: E<(1-d)E_{ground}\}$ OR $\{E_{ground}>0: E<(1+d)E_{ground}\}$
      - d[8]={0.08,0.04,0.02,0.01,...}
    - maximum total number of excited states to be selected: $n_e$
      - $N_e$[8]={1,1,...,1}
- **N(events) = 50k**

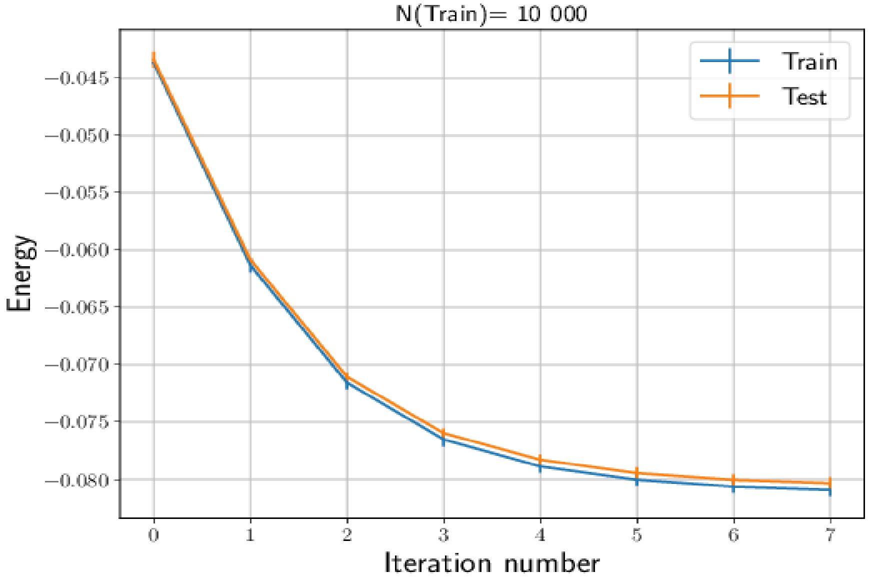
**Handles to reduce the size of the Ising model to be encoded on the annealer**

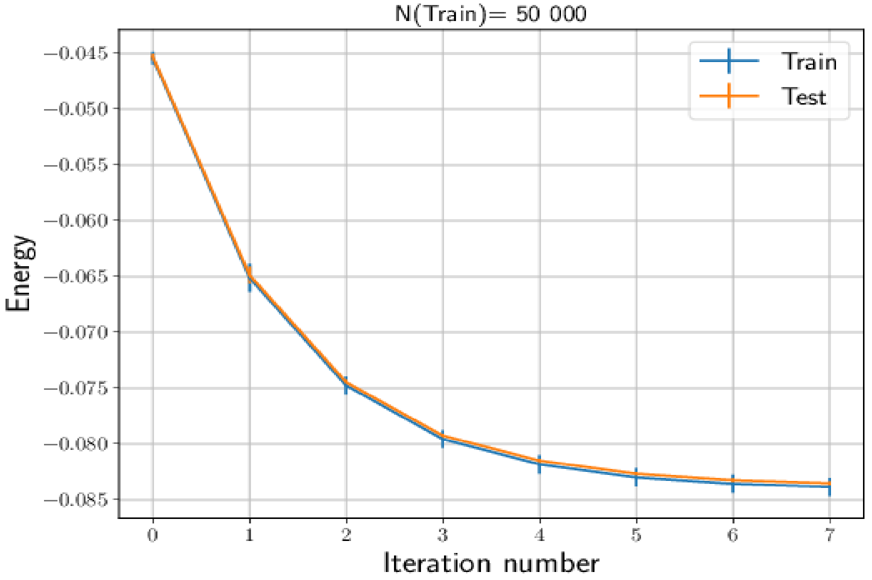
- **Cutoff C**
- **Fixing variables**: On/Off

# **Annealing**: at work



***Experimental considerations***

# **Approach:** broad lines

**1/ Use same input variables as in the BDT based search of stop**:

- Lepton: $p_T$, $\eta$, Q
- MET, $M_T$
- $p_T(j1)$, $H_T = \Sigma_i\, p_T(j_i)$, N(jet)
- b-quark: btag discriminant, N(btag), $p_T(b)$, $\Delta R(l,b)$

**to build weak-classifiers**, fed to the QA

**2/ Obtain strong classifier R through the QA machine of DW**

**3/ Assess the performance of each R through a Figure of Merit:**

Cut-scan the strong classifier for S & B: maximize FOM

$$\mathrm{FOM} = \sqrt{2\left((S+B)\ln\left[\frac{(S+B)\cdot(B+\sigma_B^2)}{B^2+(S+B)\cdot\sigma_B^2}\right] - \left(\frac{B^2}{\sigma_B^2}\right)\ln\left[1+\frac{\sigma_B^2\cdot S}{B\cdot(B+\sigma_B^2)}\right]\right)},$$

G. Cowan, "Two developments in tests for discovery: use of weighted Monte Carlo events and an improved measure", Progress on Statistical Issues in Searches, SLAC, June 4 - 6, 2012

Compare the performance of R with the one of BDT for the benchmark signal point $(m(t_1), m(\chi^0{}_1)) = (550,520)$: FOM(BDT) = 1.44

# **Approach:** broad lines & treatment of data

**<u>All these</u> versus:**

- **Different variables sets**
- **Augmentation (A,δ)**
- **Cutoff & fixing variable heuristic**

- All MC simulation
- Background processes: randomized & mixed according to their abundance σ to represent SM
- Splits:
  - Background samples split in 2:
    - Big Test sample for assessing performance
    - Within QA code: split in 2: Train & Test
  - Signal:
    - All (550,520) signal point: assessing performance (FOM)
    - Within QA code: entire Δm=30, split in 2: Train & Test

<u>Uncertainty due to fluctuations of QA</u> , e.g. different embedding @ different times: We run the very same variables, with the exact same scheme 10 times, observe the maximal FOMs, and take the standard deviation

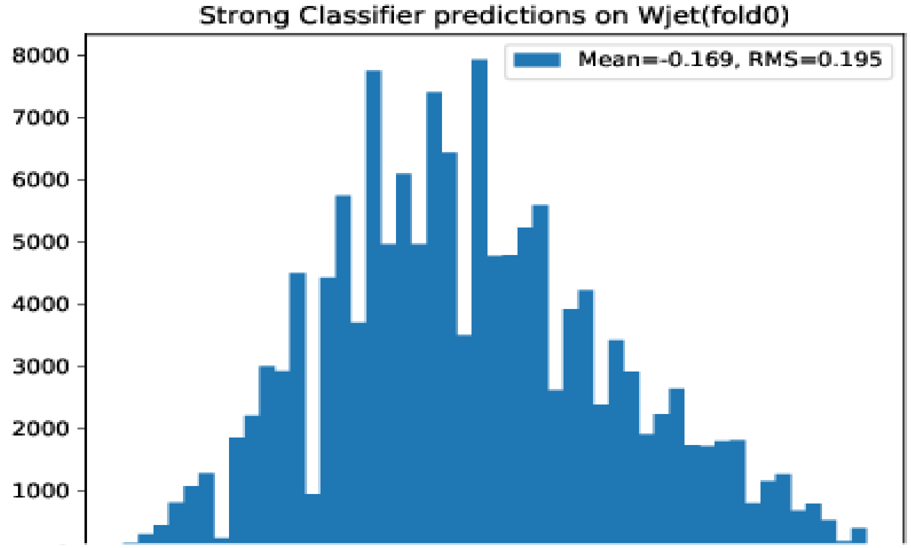
***Different settings***

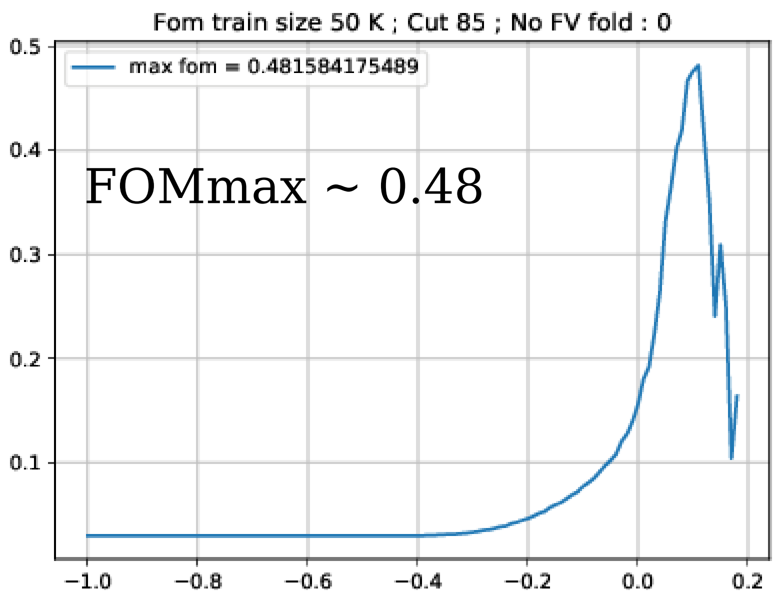
***&***

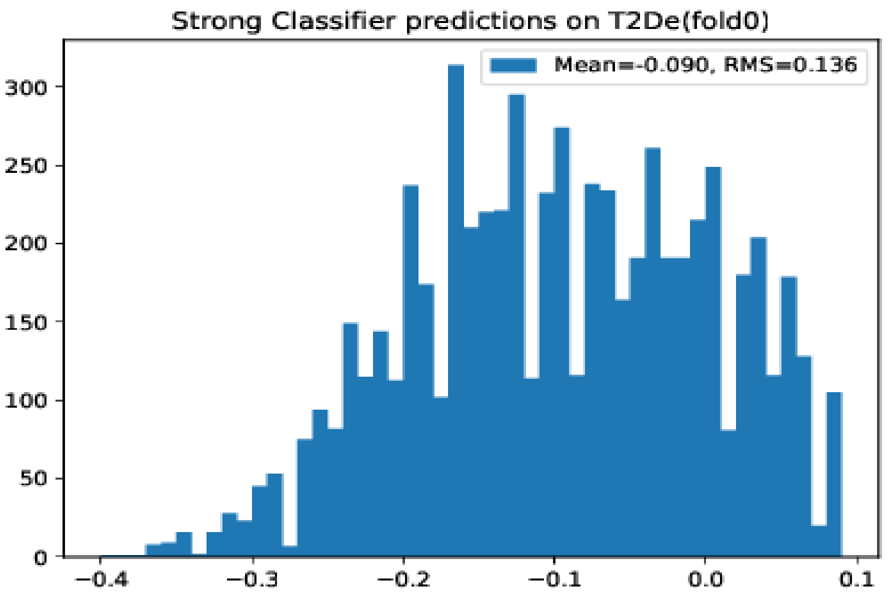
***Results***

# Approach α

Input to QA: all 12 input variables of the publication, not transformed in weak-classifiers

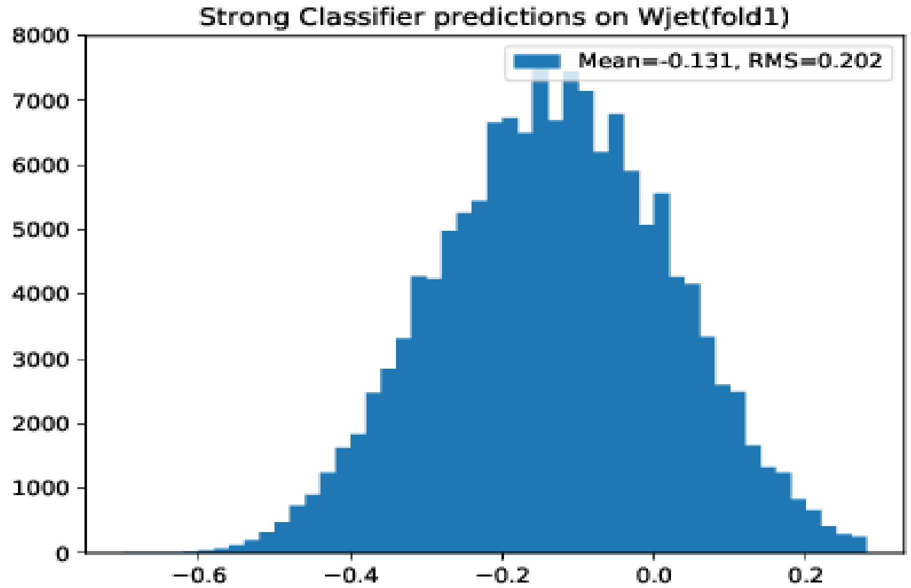


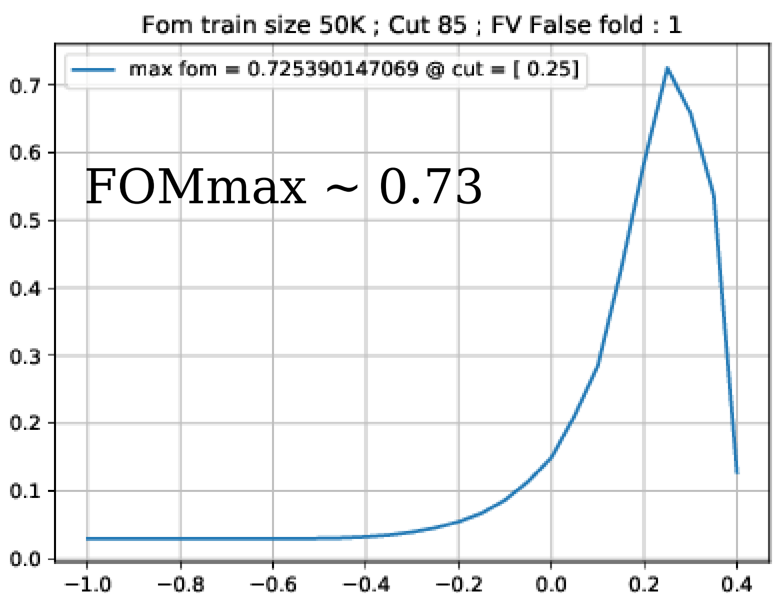



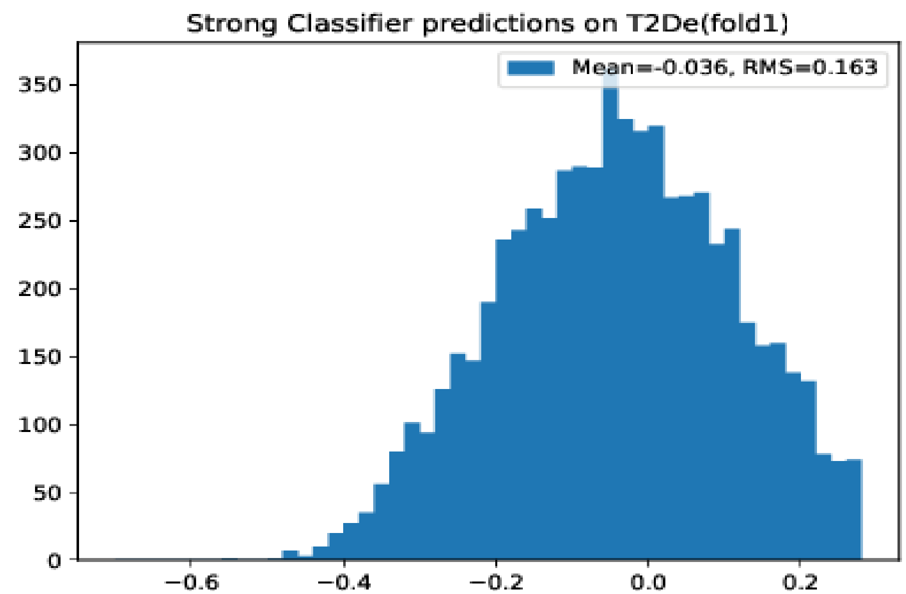



# Approach β

Input to QA: all 12 input variables of publication, transformed in weak classifiers



# Accounting for correlations

Until now: we introduced the input variables of the BDT, to the QA

But: MVA tools take into account correlations, while QA: not really

Idea: introduce 1-D variables, from the 12 original, which (already) contain correlative information as additional variables to the QA

| Variable | Description |
|---|---|
| $p_T(l)$ | $p_T$ of the lepton $l$ |
| $\eta(l)$ | Pseudo-rapidity of the lepton $l$ |
| $Q(l)$ | Charge of the lepton $l$ |
| $E_{\mathrm{T}}^{\mathrm{miss}}$ | Missing transverse-energy |
| $M_{\mathrm{T}}$ | Transverse invariant mass of the ($E_{\mathrm{T}}^{\mathrm{miss}}$,$p_T(l)$) system |
| $p_T(j_1)$ | $p_T$ of the leading jet |
| $H_T$ | $\sum_i p_T(jet(i))$ |
| $N_{\mathrm{jets}}$ | Multiplicity of selected jets |
| $Disc(b)$ | b-quark tagging discriminant |
| $N(b)$ | Number of b-tagged jets |
| $p_T(b)$ | $p_T$ of the jet with highest b-discriminant |
| $\Delta R(l, b)$ | Distance between the lepton and the jet with the highest b-discriminant |

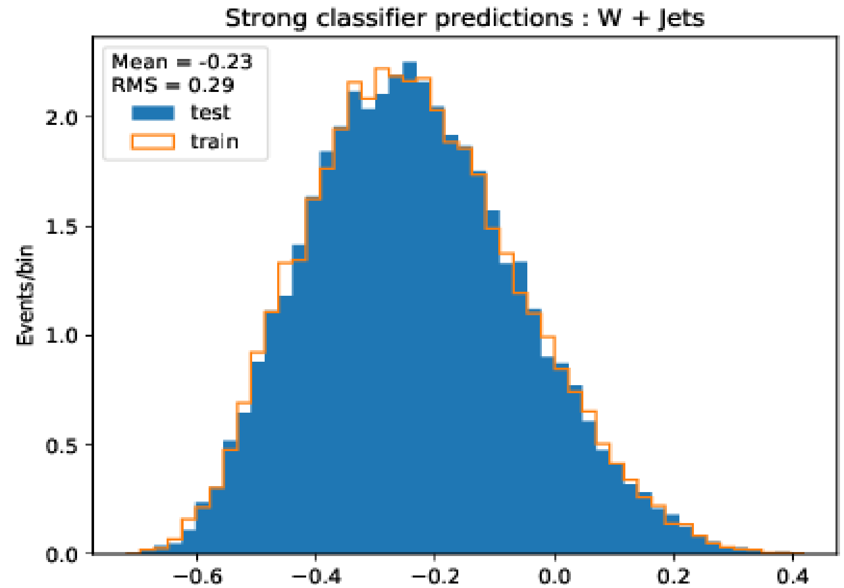
Operations guided by physics:

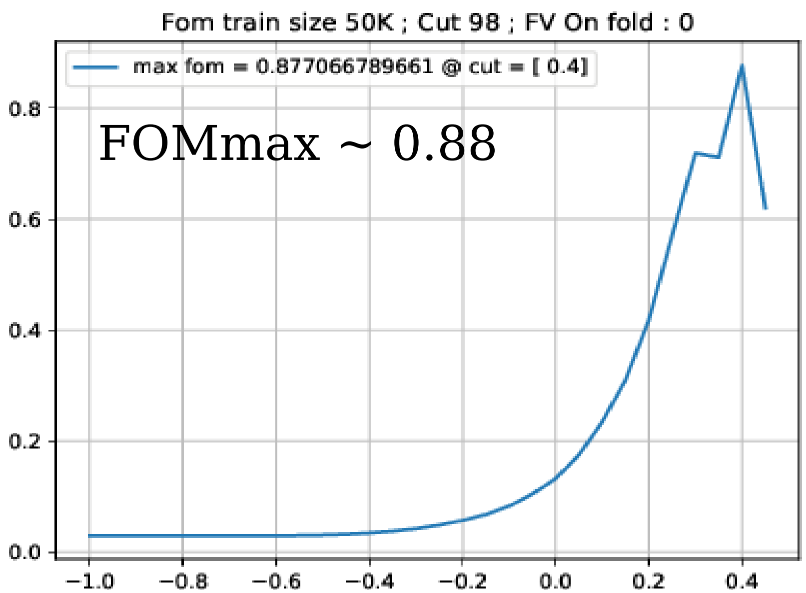
* Why a given set of variables & operations separate S from this and/or than B process: understood and checked in 2D distributions

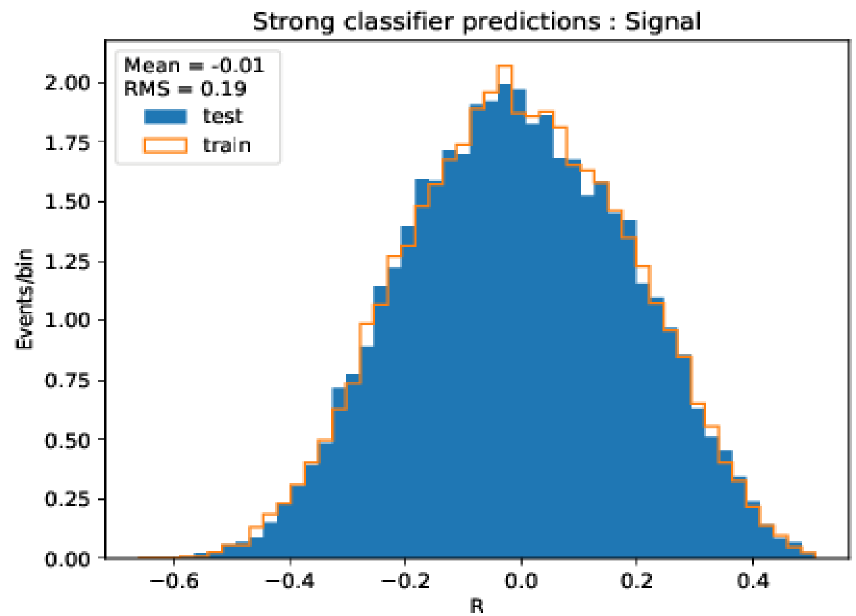
* How much: guided my FOM maximisation

# Approach A

Input to QA: add $N_V$ new variables made of physics-guided operations between pairs of the same variables: trying to capture differences of correlation between S & B. $N_V$=5

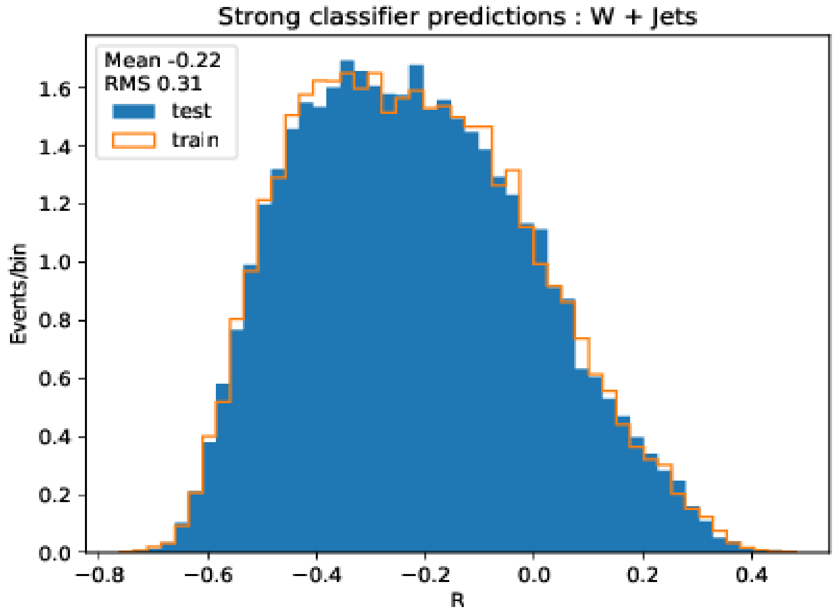


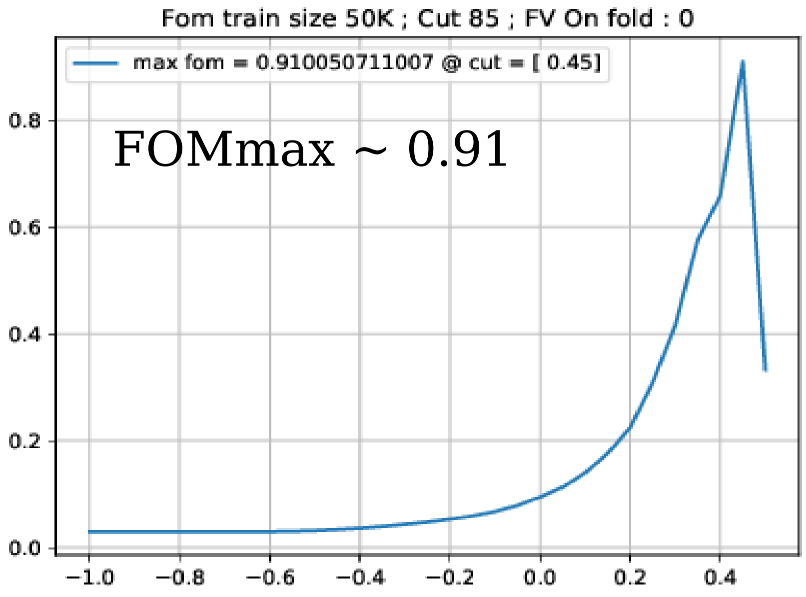



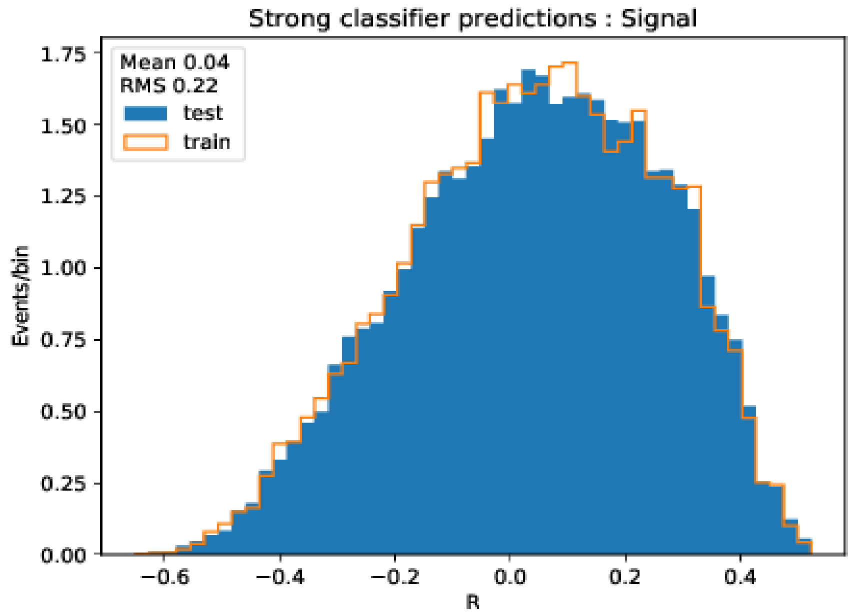



# Approach B

Input to QA: add $N_V$ new variables made of physics-guided operations between pairs of the same variables: trying to capture differences of correlation between S & B. $N_V$=9



# Comparison with BDT

| Variable set | Fixing variable | $C$ [%] | $(\delta, A)$ | FOM |
|---|---|---|---|---|
| $\alpha$ | False | 85 | (0.0250,3) | $0.48 \pm 0.03$ |
| $\beta$ | False | 85 | (0.0250,3) | $0.73 \pm 0.04$ |
| A | True | 98 | (0.0090,5) | $0.88 \pm 0.04$ |
| B | True | 85 | (0.0070,3) | $0.91 \pm 0.05$ |
| BDT | NA | NA | NA | $1.44 \pm 0.06$ |

Question: Are our so far best approaches (A & B) totally comparable to BDT ?

1/ we include the same variables

2/ we include correlative information made of the same variables

BDT: before being trained, the discriminating variables fed are decorrelated

QA: we don't decorrelate variables, whose "natural" basis might not be the optimal one for a linear classifier

So: we first pass our (different sets of) variables through Principal Component Analysis [i], then feed them to the QA

[i] "A Tutorial on Principle Component Analysis", J. Shlens, arxiv:1404.1100v1

# Comparison with BDT

| Variable set | Fixing variable | $C$ [%] | $(\delta, A)$ | FOM |
|---|---|---|---|---|
| $\alpha$ | False | 85 | (0.0250,3) | $0.48 \pm 0.03$ |
| $\beta$ | False | 85 | (0.0250,3) | $0.73 \pm 0.04$ |
| A | True | 98 | (0.0090,5) | $0.88 \pm 0.04$ |
| B | True | 85 | (0.0070,3) | $0.91 \pm 0.05$ |
| PCA(A) | False | 95 | (0.0145,3) | $1.57 \pm 0.24$ |
| PCA(B) | True | 95 | (0.0070,5) | $1.09 \pm 0.17$ |
| BDT | NA | NA | NA | $1.44 \pm 0.06$ |

- $\alpha$ vs $\beta$: effect of the weak-classifier
- $\beta$ vs B: effect of additional variables, even though A & B almost equivalent
- A vs PCA(A), B-PCA(B): effect of uncorrelating basis
- At PCA level: conceptually doing the same things than in BDT approach (JHEP 09 (2018) 065): the only difference: QA versus BDT
- PCA approaches: larger uncertainties. In PCA basis, the variables/$c_i$'s are more decorrelated. A weight variation ($\mu_i$) for one variable is more independent from another. When a $\mu_i$ fluctuates (e.g. b/c of the state of the machine), R is more sensitive to all variations of all $\mu_i$, which are more independant, hence a larger variation of R itself

# Conclusions

This is the second attempt of a **binary classification** based on **quantum annealing**
**Based on a small & well motivated variables**

- **Systematic studies versus:**
  - Number of events = {100, 1k, 10k, 50k}
  - Internal annealing parameters: Cutoff on coupling $J_{ij}$
  - Augmentation parameters ($\delta$,A) versus $N_{var}$
  - w & w/o weak-classifiers
  - Sets of variables
- **Best result achieved:**
  - Original 12 variables + (only) 5 “correlative” variables + PCA
  - QA: $1.57 \pm 0.24$ / BDT: $1.44 \pm 0.06$

    **Our QA approach can do as good, if not better than classical MVA based approach**

    This is the first time that an annealing based quantum classification is at least as good as a classical one
- Publication: in proof reading. https://arxiv.org/abs/2106.00051

# **Discussion:** Focus on QA specifics

- **Treatment of correlations:**
  - MVA tools (somehow differently) exploit differences of correlation: BDT: one can do PCA. NN: separates S & B via new basis of variables through different (layers of) nodes
  - QA: Introduce to some extent correlation among the variables through the coupler $J_{ij}$: HW chain corresponding to the coupler breakes → loss of coupling among variables
    - Should improve with better device & embedding
  - QA: Have to *warm-up* with operations among variables
- **Embedding:** {Logical problem}→{HW}: Should improve with device, new schemes ?
- **Feezing time:** What if the chains & logical problem don’t freeze out “at the same time”, i.e. reasonably close in time ? Can lead to pick-up sub-optimal set of spins, thus strong classifier
- **Broken chains:** When happens: has to use majority vote across different draws → can “dilute” acquired discriminating information ?
  - The larger number of available couplers in the machine will render each chain more stable, thus less prone to be broken

# Outlook

**Test & benchmark the same classification problem:**

**Quantum versus Classical** (*latest/greatest*)

**Within Quantum Annealing**: versus different graphs

- Following the **same $H_{Ising}$ based approach with Tensor-Network**
  - Classical but advanced tool
- Try on a more advanced quantum annealer: **Pegasus graph** of D-Wave

**Merci à**

**Artur**

**Timothée**

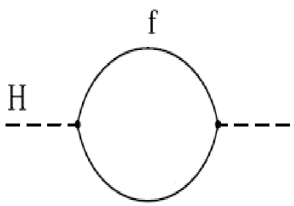
**Timothée**

***Backup***

- **Greatest discovery in HEP in last decades:**

  **Higgs boson: $m_H$ = 125 GeV/$c^2$**

- **Consider Higgs mass correction from fermionic loop:**



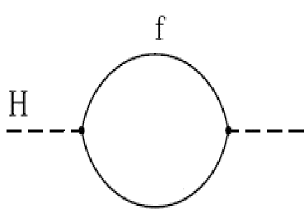

$$\Delta m_H^2 = \frac{\lambda_f^2}{16\pi^2}\cdot\left[-2\Lambda_{UV}^2 + \ldots\right]$$

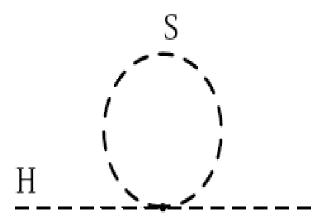
$\Lambda_{UV}$: Energy-scale at which new physics alters the Standard-Model (momentum cut-off regulating the loop-integral)

**If $\Lambda_{UV} \sim M_P$ → $\Delta m^2_H \sim O(10^{30})$ larger than $m_H$ !!!**

And all Standard-Model masses indirectly sensitive to $\Lambda_{UV}$ !!!

$$\Delta m_H^2 = \frac{\lambda_f^2}{16\pi^2}\cdot\left[-2\Lambda_{UV}^2 + \ldots\right]$$



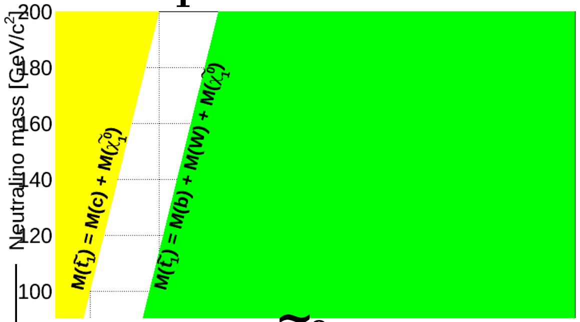

$$\Delta m_H^2 = \frac{\lambda_S}{16\pi^2}\cdot\left[\Lambda_{UV}^2 - \ldots\right]$$

$\Delta m^2_H$ quadratic divergence canceled :

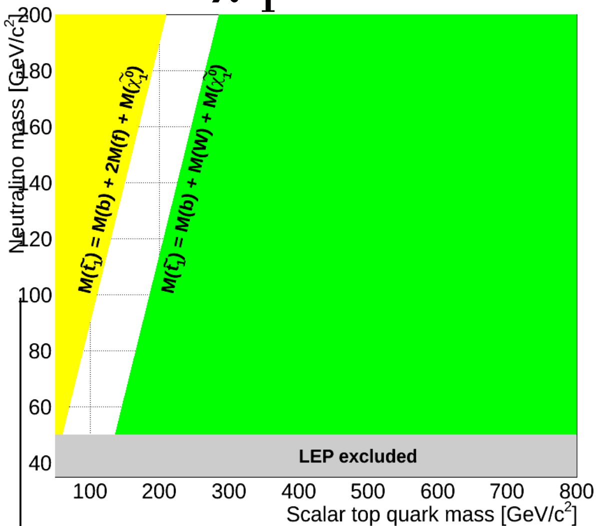
**Hierarchy problem naturally solved !**

# $\tilde{t}_1$ : Stop decays complementary in mass regions

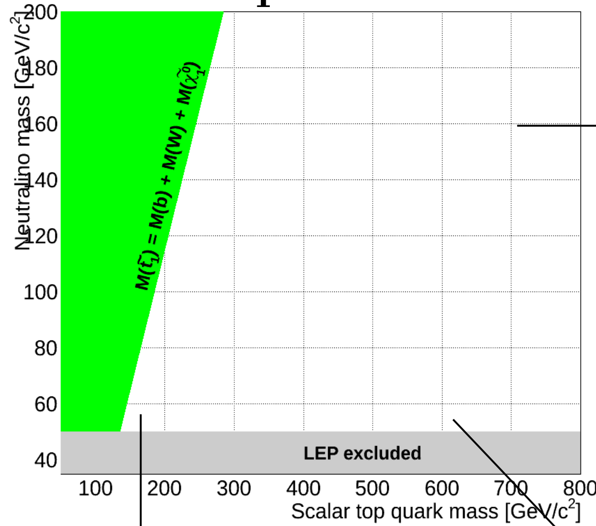
## $c\,\tilde{\chi}^0_1$




## $b\,f\,f'\,\tilde{\chi}^0_1$




## $b\,W\,\tilde{\chi}^0_1$




**$\tilde{t}_1 \to t\,\tilde{\chi}_1^0$**

$t+\tilde{\chi}_1^0 < \tilde{t}_1$

**$\tilde{t}_1 \to b\,\tilde{\chi}_1^+$**

$b+W+\tilde{\chi}_1^0 < \tilde{t}_1$

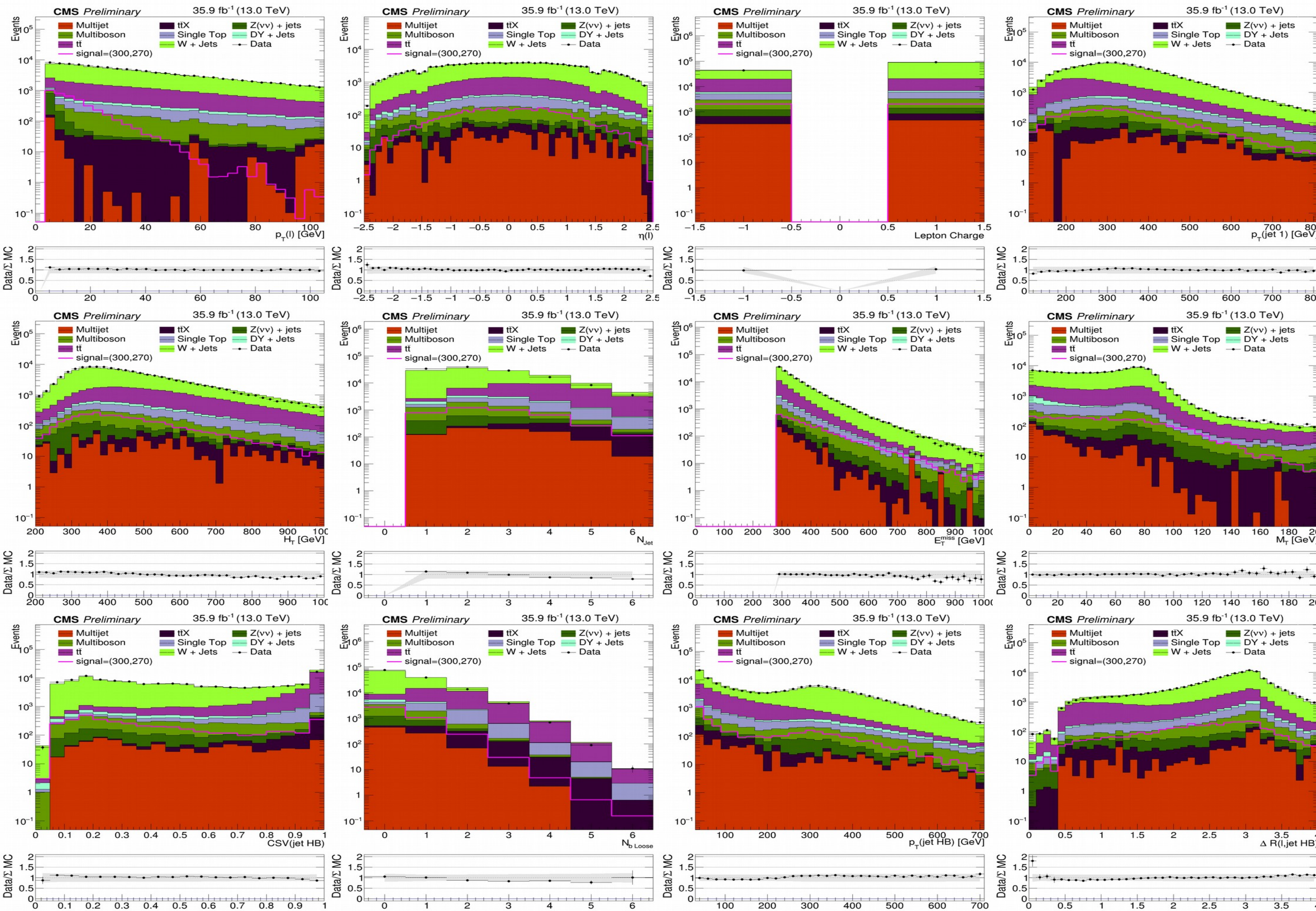
$W+\tilde{\chi}_1^0 < \tilde{\chi}^+_1 < \tilde{t}_1\text{-}b$

$c+\tilde{\chi}_1^0 < \tilde{t}_1 < b+W+\tilde{\chi}_1^0$

$b+f+f'+\tilde{\chi}_1^0 < \tilde{t}_1 < b+W+\tilde{\chi}_1^0$

**$\tilde{t}_1 \to t^*\,\tilde{\chi}_1^0$**

$b+W+\tilde{\chi}_1^0 < \tilde{t}_1 < t+\tilde{\chi}_1^0$


CMS Preliminary
35.9 fb⁻¹ (13.0 TeV)
Multijet
Multiboson
tt
signal=(300,270)
ttX
Single Top
W + Jets
Z(νν) + jets
DY + Jets
Data
Events
Data/Σ MC
p_T(l) [GeV]
η(l)
Lepton Charge
p_T(jet 1) [GeV]
H_T [GeV]
N_Jet
E_T^miss [GeV]
M_T [GeV]
CSV(jet HB)
N_b Loose
p_T(jet HB) [GeV]
Δ R(l,jet HB)

# Weak classifier construction

**What we want from a weak classifier for variable $\nu$:**

Be +/- for Signal/Background dominated region of a variable $\nu$

$\nu_{cut}$: S = 70% for $\nu < \nu_{cut}$

B < 70% for $\nu < \nu_{cut}$: S on the left

B on the right

- Center around $\nu_{cut}$: $\nu' = \nu - \nu_{cut}$
- Invert b/c we want S to be on the right/positive side: $\nu'' = -\nu'$

B > 70% for $\nu < \nu_{cut}$:

S & B on the same side of $\nu_{cut}$:

$\nu_{newcut}$: S = 30% for $\nu < \nu_{newcut}$

70% of S is above $\nu_{newcut}$

B > 30% for $\nu < \nu_{newcut}$

Most of B is below $\nu_{newcut}$

B/S on left/right: nothing to do:

$\nu'' = \nu - \nu_{newcut}$

$S(\nu'') > B(\nu'')$

B < 30% for $\nu < \nu_{newcut}$

More than 70% of B above $\nu_{newcut}$, in region where 70% of S is:

Variable not worth

$$H_{\text{Ising}} = \sum_{i<j}^{N} J_{ij} s_i s_j + \sum_{i=1}^{N} h_i s_i$$

$$J_{ij} = C_{ij}/4$$

$$h_i = \frac{1}{2}\left(\lambda - 2C_i + \sum_j C_{ij}\right)$$

$$C_{ij} = \sum_\tau c_i(\vec{x}_\tau) c_j(\vec{x}_\tau) = C_{ji} \text{ and } C_i = \sum_\tau c_i(\vec{x}_\tau) y_\tau$$

# **Annealing:** the principle

$$H_{\text{Ising}} = \sum_{i<j}^{N} J_{ij} s_i s_j + \sum_{i=1}^{N} h_i s_i$$

$H(t) = A(t) \,.\, \Sigma_i\, \sigma_i^x \quad \pm \quad B(t) \,.\, \{\Sigma_{ij} J_{ij} \sigma_i^z \sigma_j^z + \Sigma_i h_i \sigma_i^z \pm 1\}$

Driver Hamiltonian
Transverse field

Problem Hamiltonian: Biases bought by the problem: $h_i$, $J_{ij}$: inputs

<u>Adiabatic theorem</u>: If Hamiltonian is interpolated from initial H(t=0) to final H(t=ta) slowly enough compared to minimum ground-to-first-excited state gap of H(t), the system will be in the ground state of H(t=ta) with high probability, provided it was initialized in the ground state of H(t=0). One can evolve from a simple, easy to initialize H at t=0 to a complicated H with an unknown ground state at t=ta, with ta: annealing time
<u>Summary</u>: If t decreases slowly enough: able to reach **lowest E solution**

A(t), B(t): annealing schedules. Monotonically: A(t): 1 → 0 / B(t): 0 → 1

- It can be simulated (SA)
- Specific to Quantum Annealing (QA): physically implemented by a circuit, with control lines in which $e^{\pm}$ are in quantum mode (superconducting metal). One can control the tunel barrier

# **Implementation & Classification:** sampling & result

Quantum annealers are (naturally ?) run in a batch mode in which one draws many samples from a single Hamiltonian
Repeated draws for QA are fast: DW averages ~5000 samples/s under optimal conditions

→ HW solution spins after multiple reads

→ Read-out to get spins a la $H_{\text{Ising}} = \sum_{i<j}^{N} J_{ij} s_i s_j + \sum_{i=1}^{N} h_i s_i$

→ Strong classifier $R(\boldsymbol{x}) = \sum_i s_i^{\text{g}} c_i(\boldsymbol{x}) \in [-1, 1]$

It’s the strong classifier that we use similarly to the output of NN or BDT…

# **Classification:** weak & strong classifiers

1/ To avoid mismodelling of input variables in regions with low statistics (tails): distribution of each variable is truncated

Remark:
The strong classifier is, as defined and if not modified, only taking into account linear combinations of all variables via the weak classifiers:

$$R_{\vec{w}}(\vec{x}) = \sum_i w_i c_i(\vec{x})$$

So all the exploitation of differences of e.g. 2-by-2 correlation between S & B by MVA tools is a priori absent

2/ + - * / operations between input variables performed and fed as weak classifiers

→ Introduce 1- & 2-point correlations among variables
→ Help physicist to spot possibly meaningful - having physics meaning - new variables., i.e. help discern new discriminating variables so far missed by the physicist

Approach $\alpha$:

Even when input variables are in [-1,+1], it can have the bulk of distribution below or above 0, which is the criterion for being classified as S- or B-like. It's therefore possible that a majority of events are - or +, but at the same time for S **and** B -> less discrimination. i.e. less c_i's are relevant, particularly given the size $\delta l$ of the scan.

$\delta l$ isn't necessarily of the right size for one input variable, and thus doesn't scan it properly.

Approaches with weak-classifiers (>=$\beta$):

The distribution is centered around 70th percentile of S: we already know that in + values of wc, are have 70% of S and <70% of B.

$\delta l$ has the right size to scan a weak-classifier, because wc's are scaled as function of initial distribution of input variables to reflect 30th, 70th percentile -> We have more c_il's which have discriminating information than only with input variables where a lot of c_i's give the same information because most of the time being on the same side of the bulk of the distribution.