

Genetic Algorithms and Neural Networks as Tools in Particle Physics

Rick Field - University of Florida

Tevatron University - May 21, 1998

- Describe a minimization technique modeled after genetics and evolution (**Genetic Algorithms**).
- Show an example of the use of a genetic algorithm in Particle Physics (**six-dimensional linear cuts**).
- Describe the connection between **Genetic Algorithms** and **Neural Networks**.

Min-Max Problem

Let \mathbf{M} be a map from the data space $\{\mathbf{D}\}$ to the real numbers \mathbf{R} . The map \mathbf{M} depends on $\{\mathbf{D}\}$ and on a set of parameters $\{\mathbf{P}\}$. We would like to find a particular set of parameters $\{\mathbf{P}_0\}$ in $\{\mathbf{P}\}$ such that

$$M(P_0, D) \leq M(P_i, D)$$

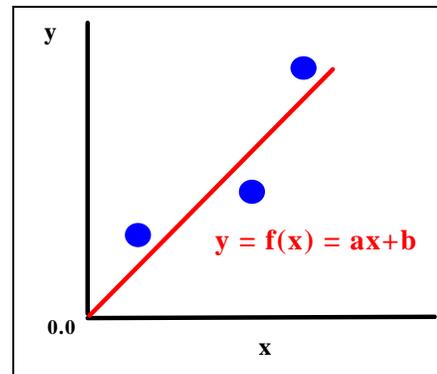
for all $\{\mathbf{P}_i\}$ in $\{\mathbf{P}\}$.

Example (Linear Regression):

The set of data

$$\{\mathbf{D}\} = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}.$$

The parameter space $\{\mathbf{P}\} = \{\mathbf{a}, \mathbf{b}\}$.



The map \mathbf{M} is

$$M(a, b, \{D\}) = \mathbf{c}^2 = \frac{1}{N} \sum_{i=1}^N (y_i - f(x_i))^2$$

Local Algorithms (gradient descent):

- Fast
- Converge to local minimum
- Require a topology on $\{\mathbf{P}\}$

Global Algorithms (enumeration, random):

- Very Slow!

Genetic Algorithms (model of genetics & evolution):

- Combine the good features of both

Linear-Cuts in Collider Phenomenology

Find the set of left L_i and right R_i cuts on the N observables O_1, O_2, \dots, O_N such that the requirement that $L_i < O_i < R_i$ maximizes the **enhancement** times the **efficiency**, where

efficiency = % signal surviving the cuts

enhancement = (% signal / % background) surviving the cuts

The set of data

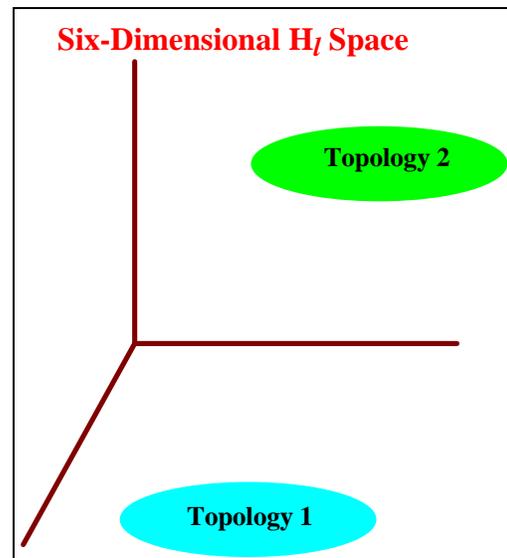
$\{D\} = \{5,000 \text{ signal and } 5,000 \text{ background events}\}$.

The parameter space $\{P\} = \{L_i, R_i\}$.

The map M is

$$M(\{P\}, \{D\}) = \text{efficiency} * \text{enhancement}$$

The observables are the first six **modified Fox-Wolfram moments**, H_1, H_2, \dots, H_6 constructed from the calorimeter cells directly.

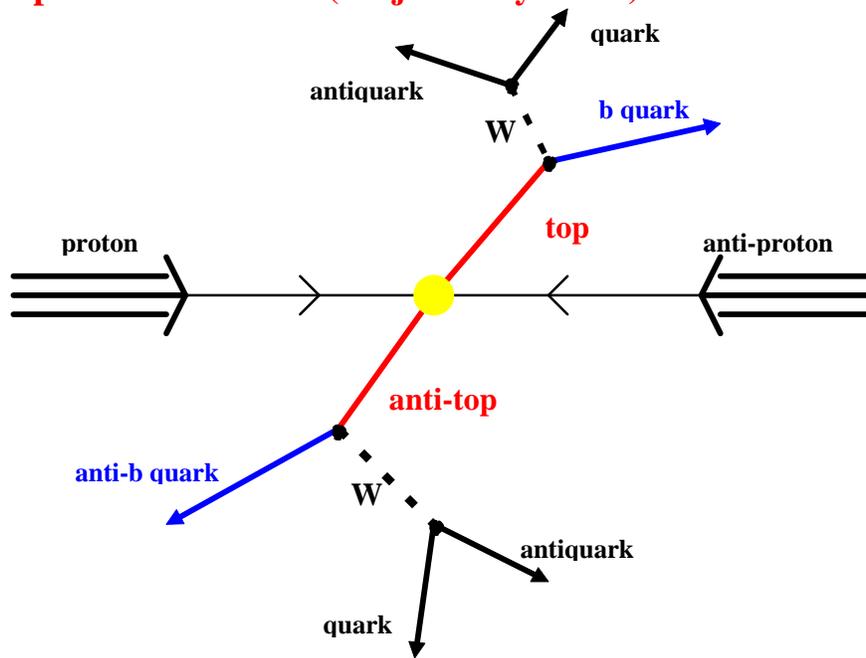


Find the region of six-dimensional H_T -space that optimizes signal over background for the six-jet decay of top-quark pairs.

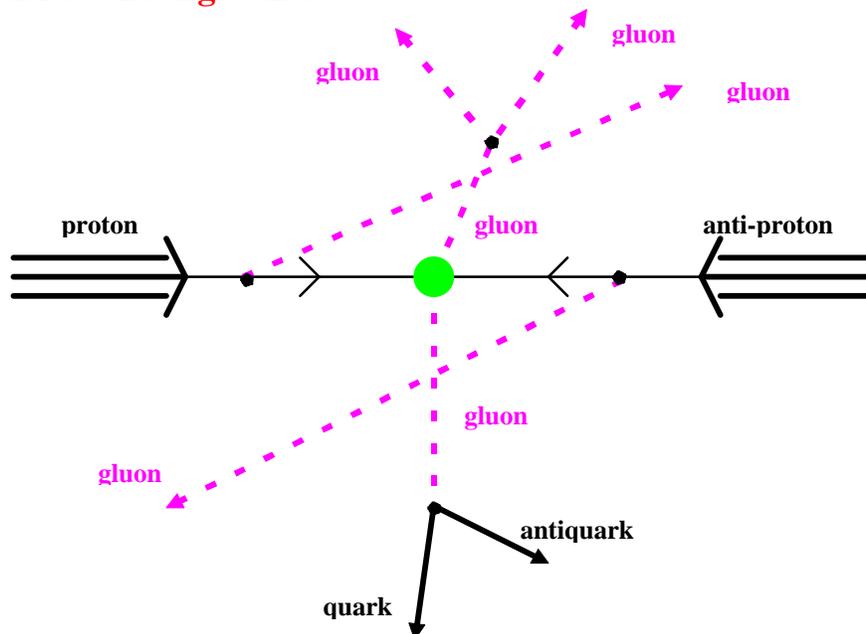
Signal & Background

Attempt to isolate the **six-jet decay mode** of top-pair production from the **QCD background** using only the **event topology**. Of course, **b-quark tagging** will improve on whatever can be accomplished using topology.

Signal: Top-Pair Production (six-jet decay mode)



QCD Multi-Jet Background:



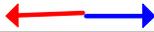
Fox-Wolfram Moments (1978)

Event Shapes in e^+e^- Annihilations:

$$H_\ell = \frac{4p}{2\ell + 1} \sum_{m=-\ell}^{\ell} \left| \sum_i^N Y_l^m(\Omega_i) \frac{p_i}{E_{tot}} \right|^2$$

where $Y_{m\ell}$ are the spherical harmonics and $\Omega_i = (\cos\mathbf{q}_i, \mathbf{f}_i)$ is the angular position of the i -th particle and \mathbf{N} is the total **number of particles** in the event. The moments H_ℓ are **rotationally invariant** and **range from 0 to 1**.

They characterize the topology of the event.

Name	Topology	H_1	H_2	H_3	H_4
"one-jet"		1.0	1.0	1.0	1.0
"two-jet"		0.0	1.0	0.0	1.0
"three-jet"		0.0	0.25	0.625	0.141
"four-jet"		0.0	0.25	0.0	0.687
"sphere"		0.0	0.0	0.0	0.0

Modified Fox-Wolfram Moments - Event Shapes in Hadron-Hadron Collisions:

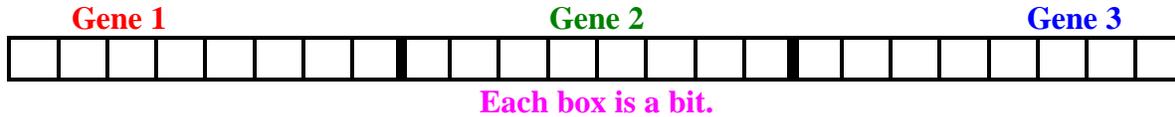
$$H_\ell = \frac{4p}{2\ell + 1} \sum_{m=-\ell}^{\ell} \left| \sum_i^N Y_l^m(\Omega_i) \frac{E_{Ti}}{E_T^{tot}} \right|^2$$

Applied to Jets - E_{Ti} is the transverse energy of the i -th **jet** and $\Omega_i = (\cos\mathbf{q}_i, \mathbf{f}_i)$ is the angular position of the i -th jet and \mathbf{N} is the total **number of jets** in the event (with $E_T > E_T(\text{cut})$).

Applied to Calorimeter Cells - E_{Ti} is the transverse energy of the i -th **cell** and $\Omega_i = (\cos\mathbf{q}_i, \mathbf{f}_i)$ is the angular position of the i -th cell and \mathbf{N} is the total **number of cells** in the calorimeter ($E_T > E_T(\text{cut})$).

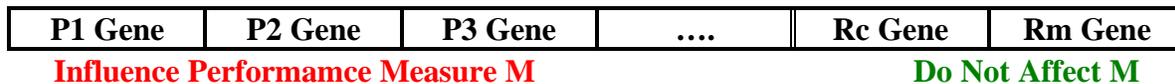
Genes & DNA

"Individuals" are characterized by their DNA which is composed of a string of genes. Numbers are represented in the computer by **N bytes** (1 byte = 8 bits) which we call a **gene**. The DNA consists of a string of genes. We use $N = 2$ but for illustration I will take $N = 1$.



Each individual carries one gene for each of the parameters in the parameter space **P** plus two extra (one for the crossover rate **R_c** and one for the mutations rate **R_m** for that individual). Also each individual has a performance measure **M** (the quantity we are trying to maximize).

DNA



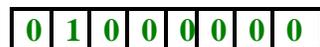
Example: Linear Cuts of H_i's:

12 parameters

- 6 left cuts of $H_1, H_2, \dots, H_6 = L_1, L_2, \dots, L_6$
- 6 right cuts of $H_1, H_2, \dots, H_6 = R_1, R_2, \dots, R_6$

DNA = 14 genes

Since L_i and R_i range from zero to one, we multiply them by **255**, so the **gene** for, for example, **L₁=0.251** looks like



The measure **M** is the **enhancement times the efficiency**.

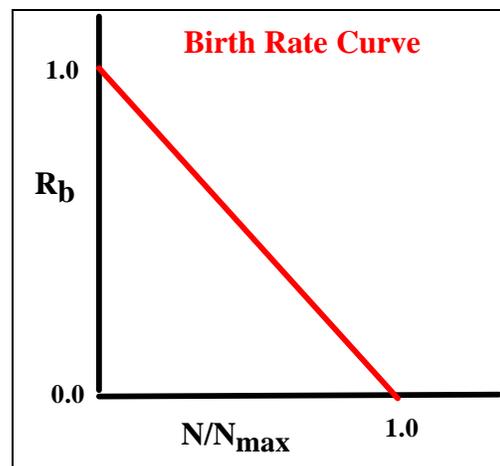
Birth and Death Rate

Maximum Population Size:

One must decide on a maximum population size, N_{\max} . In general the larger the parameter space \mathbf{P} , the larger the maximum population size should be. (We take $N_{\max} = 1,000$.)

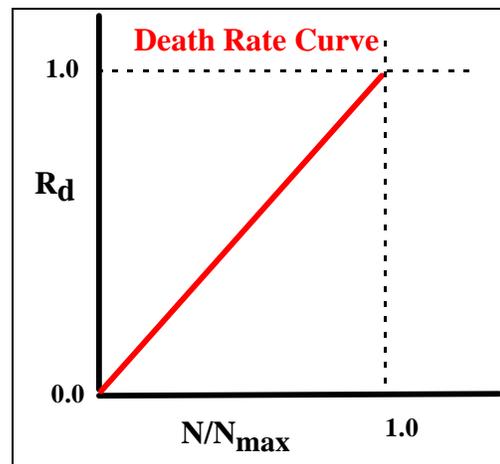
Birth Rate Curve:

The birth rate curve gives the probability, R_b , of birth (0-1) versus the relative population size, N/N_{\max} , where N is the current size and N_{\max} is the maximum size. This curve is fixed throughout time.



Death Rate Curve:

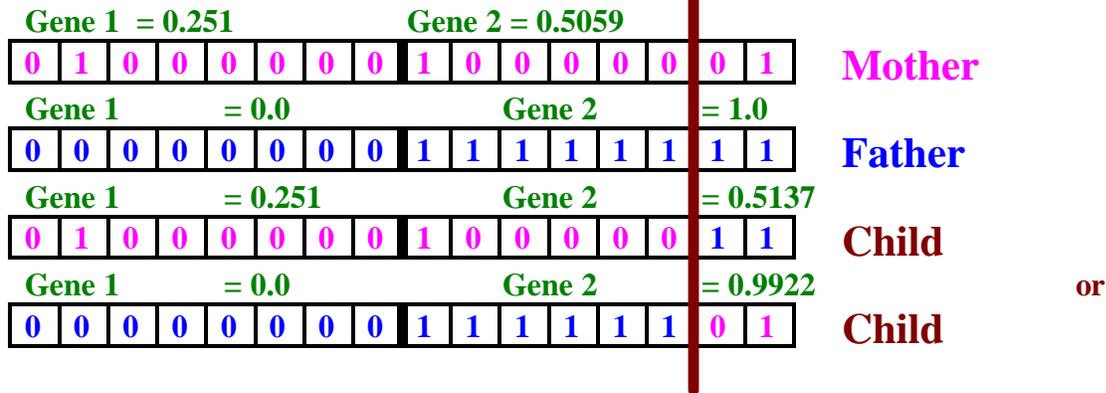
The death rate curve gives the probability, R_d , of death (0-1) from old age versus the relative population size, N/N_{\max} , where N is the current size and N_{\max} is the maximum size. This curve is fixed throughout time.



Reproduction

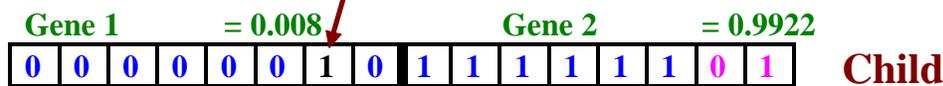
Each simulation year, depending on the population size, individuals reproduce by selecting a mate. Individuals with higher performance measure **M** have a higher probability of being selected as a mate. If the population is large, the rate of reproduction is smaller, and vice versa.

Crossover:



The children inherit certain genes from one parent and others from the other. The split position is chosen at random within the **DNA** of the parents. The child receives all the bits to the left from one parent and all the bits to the right from the other parent. The probability of a crossover is determined by **R_c**. The crossover rate is the probability of a crossover per DNA and ranges from **zero to one**. Since **R_c** is itself a gene, it is the **mothers R_c** that is used.

Mutation:

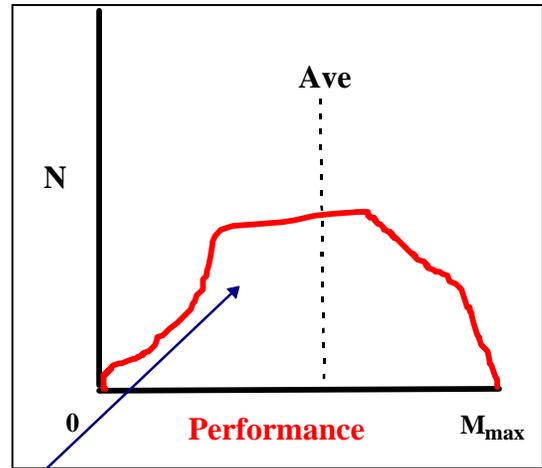


The new individual has some of its genes randomly modified. This is an extremely important factor in GA's, since this is the primary mechanism of discovering radically new solutions. The probability of a mutation is determined by the mutation rate, **R_m**. The mutation rate is the probability per bit and ranges from **R_m(min) to one**. Since **R_m** is itself a gene, it is the **mothers R_m** that is used.

Natural Selection

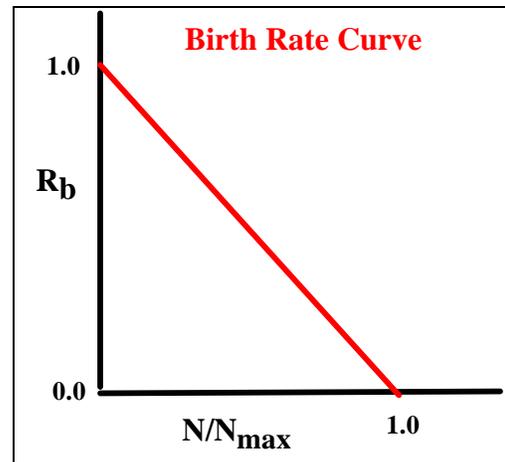
Survival of the Fittest:

At the end of each simulation year, the individuals with the **worst performance are given the highest chance of dying** and, therefore, their effect on future generations is minimized. One does not, however, exterminate the worst performers unconditionally, since good genes often require time before they lead to optimal results. At the end of each simulation year the average performance factor M_{ave} is constructed and those individuals with $M < M_{ave}$ are given a **50% chance of dying**.



Sexual Selection:

If during a simulating year it is decided that an individual will mate, **we label the individual a female**. The birth rate curve determines the probability of mating. The female then selected N_{mate} potential mates at random from the population (including herself!). She then selects the individual with the **best performance measure M** . (N_{mate} is a parameter which we take to be **10**.)

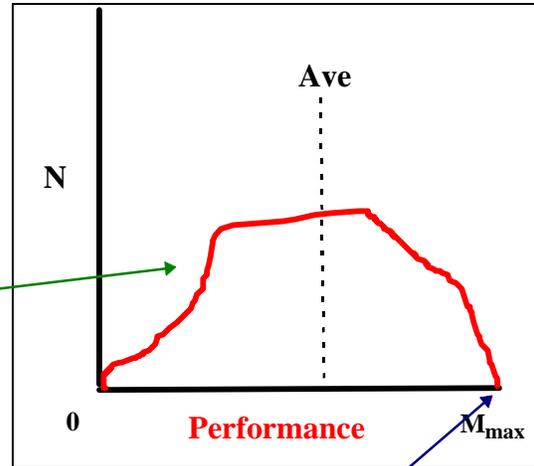


Two Ways to Die

Death by being Unfit:

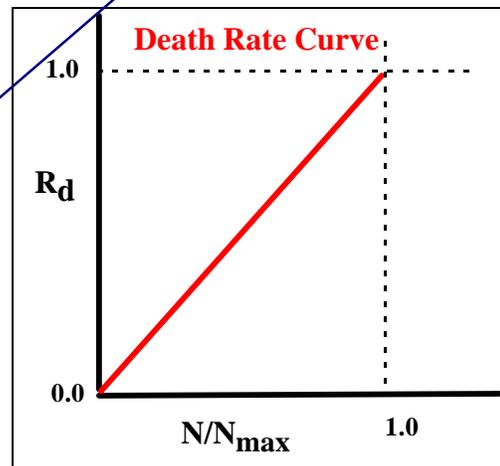
At the end of each simulation year, the individuals with the **worst performance** are given the highest chance of dying.

Individuals with $M < M_{ave}$ have a **50% chance of dying**.



Death from Old Age:

At the end of each simulation year the **death rate curve** is examined to determine the probability of death from old age. Note that the **star performer** (individual with the highest performance M) is **immune from death from old age**.



In the Beginning (year 1)

- Start at **year one with two individuals** ($N = 2$). Adam and Eve have their DNA selected at random ($N_{max} = 1,000$).

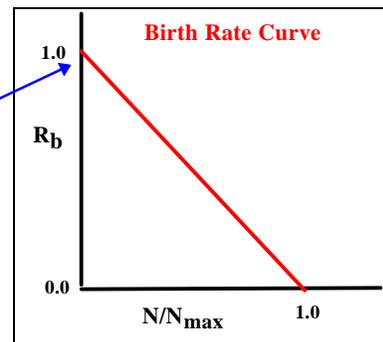
Individual 1 DNA (Adam and/or Eve)

P1 Gene	P2 Gene	P3 Gene	Rc Gene	Rm Gene
---------	---------	---------	------	---------	---------

Individual 2 DNA (Eve and/or Adam)

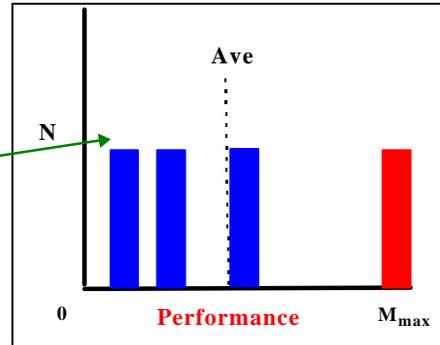
P1 Gene	P2 Gene	P3 Gene	Rc Gene	Rm Gene
---------	---------	---------	------	---------	---------

- Start with **individual 1** and check the birth rate curve at the point $N/N_{max} = 0.002$ (very high probability of giving birth!).
- If birth is chosen then individual 1 is Eve and she examines **10 mates at random** (in this case Adam and herself!) and **selects the mate with the highest performance M** .
- A child is produced based on the **mother's crossover and mutation rates**.
- Proceed to individual 2 and check the birth rate curve at the point $N/N_{max} = 0.002$ (very high probability of giving birth!).
- If birth is selected then individual 2 is also female and she examines **10 mates at random** (in this case individual 1 and herself) and **selects the mate with the highest performance M** .
- A child is produced based on the **mother's crossover and mutation rates**.
- During year 1 there will probably be **four individuals** ($N=4$).

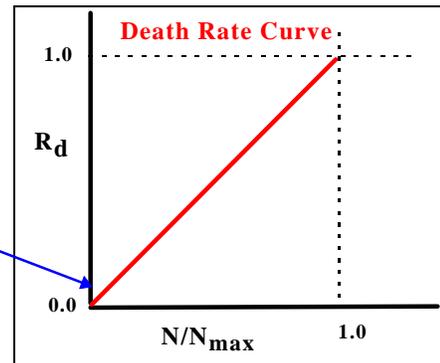


At the End of Year 1

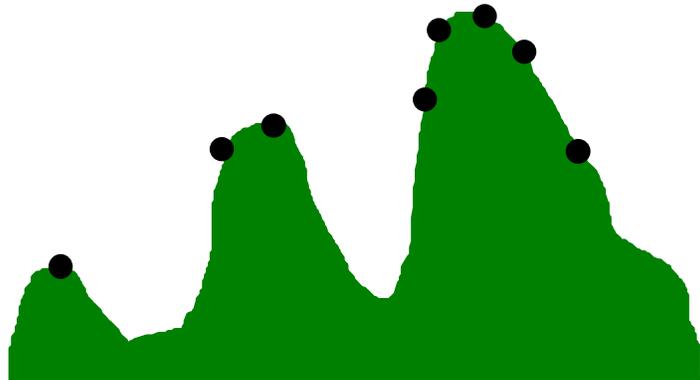
- Construct the **performance distribution** of the four individuals and assign those individuals with $M < M_{\text{ave}}$ are given a 50% chance of dying.



- Examine the **death rate curve** with $N/N_{\text{max}}=0.003$ (assuming one died in the previous step). Note that the **star performer** (individual with the highest performance M) is **immune** from death from old age.

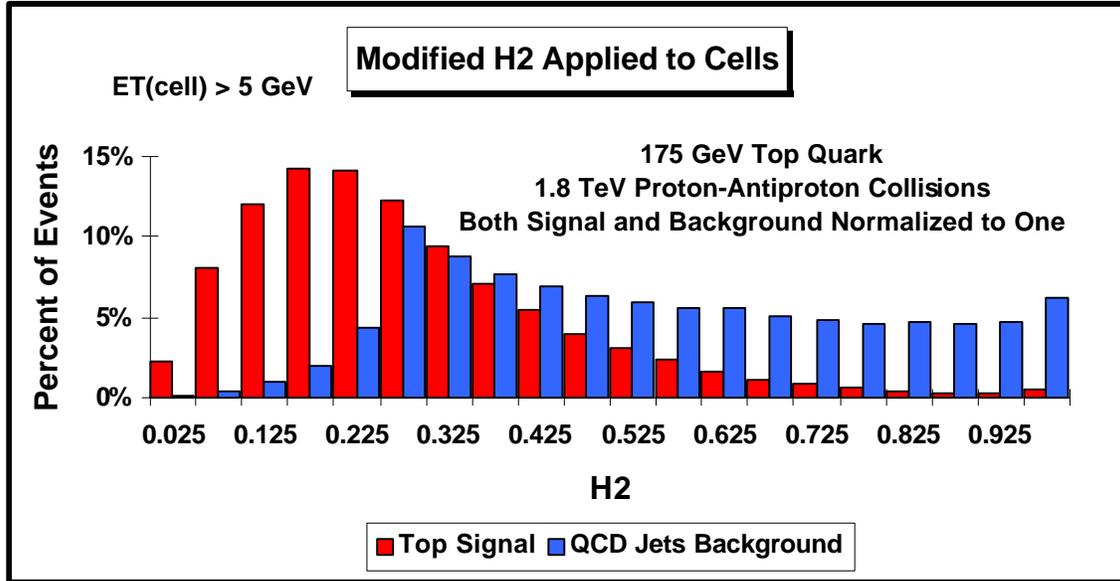


- Continue to the next simulation year with the remaining individuals (in this case probably 3).
- Population quickly grows to N_{max} and forms colonies at the tops of the mountains!



Invariant Mass Distribution

Topological Differences between Signal and Background



Multi-Jet Invariant Mass

