Neural Net Track Reconstruction for GLAST

Brandon A. Allgood
UC Santa Cruz

in collaboration with
Bill Atwood

What are we after?

- We are after findable tracks, defined as:

- Must have at least one degree of freedom.
  - N points to fit parameters.
  - at least 1 more to test hypothesis.

- Must have 3 hits in consecutive layers in a single tower.
  - Defines a minimum pitch angle.

- Must have a sagitta (bend) less than that set by detection limits.

What are the requirements?

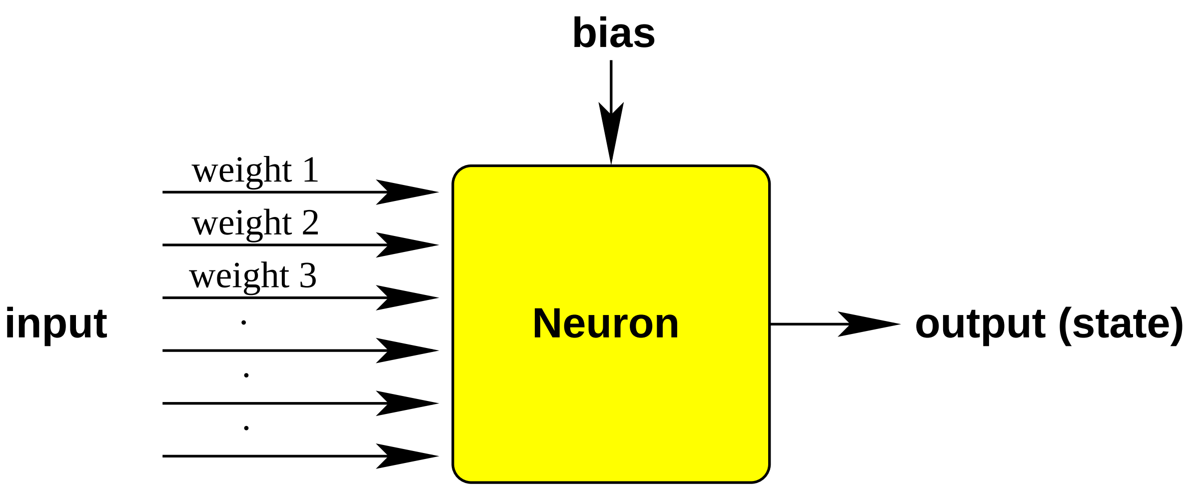
- Miss less than 1 of 10^3 findable tracks.

- Should be intrinsically 3-dimensional.

- Must apply **Globally** to the event.

What is a Neural Net?

A Neural Network consists of processing units, called neurons, and connections between them.



•The state of a neuron depends on the input neurons state and the strength (weight) of these connections.

•The dynamics of a neural net is given by:

$$S_i = sign\left(\sum_j T_{ij} S_j + Bias\right) \qquad (1)$$

where T_{ij} is the weight and S_j is the state of the input.

Neural Network Abstract

•A neural network is very similar to a spin glass, where the state of a spin is either 1 (spin up) or -1 (spin down).

•A neuron has two states as well active (1) and inactive (0).

•In both cases interactions can be long range.

•The dynamics (eq. 1) or local update rule evolves the system to a local minimum of the energy:

$$E(S) = -\frac{1}{2}\sum_{ij} T_{ij} S_i S_j \qquad (2)$$

•Functionally there are two ways to evolve the system: (both are possible using our code)

•The Simulated Annealing approach where the neurons take on states of 0 or 1.
•The Mean Field Theory approach where the neurons take on states in the range (0,1).

Mean Field Theory

•In mean field theory, one uses the thermal average of S_i.

$$V_i \equiv < S_i >_T = \left[\frac{1}{2} + \tanh\left(-\frac{1}{T}\frac{\partial E}{\partial V_i}\right)\right] \tag{3}$$

and the energy becomes equivalently,

$$E = -\frac{1}{2}\sum_{ij} T_{ij} V_i V_j \tag{4}$$

•V_i can take on any value between 0 and 1.

•T is the temperature of the system.

- •The higher T, the more likely it is to escape local minima.
- •The higher T, the slower the system converges.

•Using this approximation, the process of finding the minimum energy is much faster than in the discrete case (active/not-active).

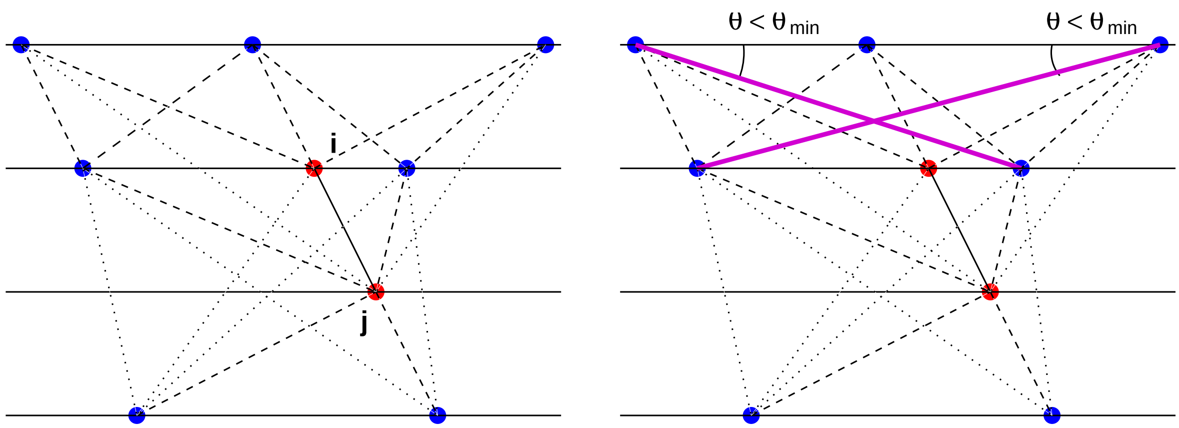
Neuron Construction

Neurons are constructed of two hits and therefore denoted by 2 indices.

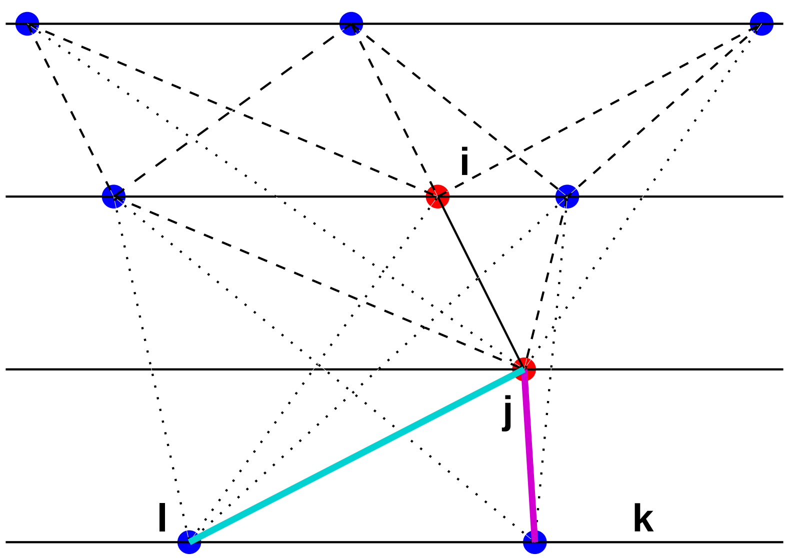
•The total number of all pairs is **HUGE**. We therefore must reduce the number of neurons (track segments).

•This done using criteria from the definition of a findable track.

•Neurons are only connected if:
- •The layer separation of the hits is less than a Max Layer Separation.
- •The neuron forms an angle of greater than θ_{min} with the horizontal plane.



Weight Matrix



•Definition of weight matrix:

$$T_{ijk} = \frac{\cos^{\lambda} \psi_{ijk}}{d_{ij}^{\mu} + d_{jk}^{\mu}} \quad (5)$$

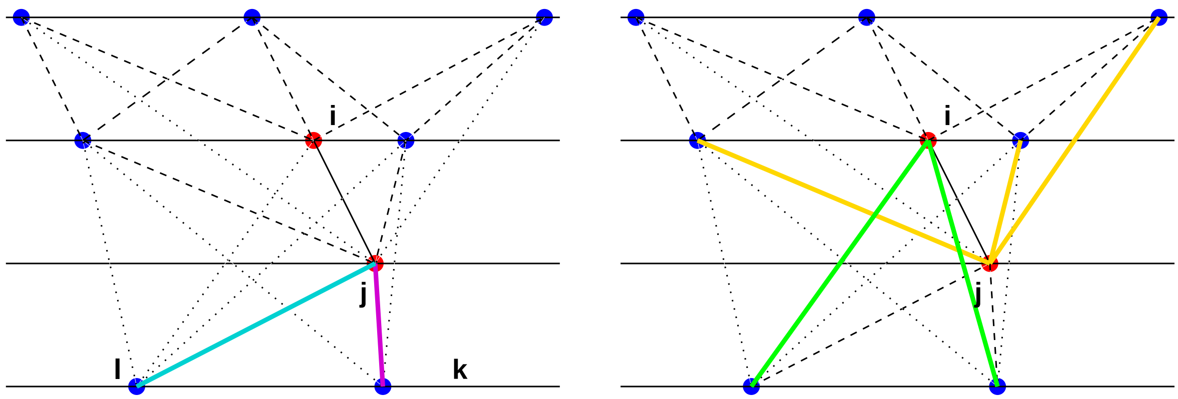
•This weight will favor straighter tracks.

•The parameter λ will be a function of the registered calorimeter energy. Higher energy $\rightarrow$ Straighter track.

•d is defined as the layer separation of the two hits (i and j). Biased against neurons which skip layers.

Expanded Update Rule

$$V_{ij} = \frac{1}{2}\left[1 + \tanh\left(\frac{c}{T}\sum_k T_{ijk}V_{jk} - \frac{\alpha}{T}\left(\sum_{(k\neq j)} V_{ik} + \sum_{(l\neq i)} V_{lj}\right) - \frac{\beta}{T}\left(\sum_{kl} V_{kl} - N_a\right)\right)\right] \quad (6)$$



•Using the parameters c, α and β we can adjust the relative importance of the three terms.

•The c term is the previously defined weight matrix.

•The α term is should graphically on the right. It has the opposite sign of the weight matrix, and biases against to many active neurons connected to one hit.

•The β term is a stimulation term. The stimulation comes from the fact that β contributions is positive if $N_a < \sum V_{kl}$.

New Energy Function

- The energy function also becomes:

$$E = -\frac{1}{2}\left[c\sum_{ijk} T_{ijk}V_{ij}V_{jk} - \alpha\left(\sum_{ijk(k\neq j)} V_{ij}V_{ik} + \sum_{ijl(l\neq i)} V_{ij}V_{lj} \right) - \beta\left(\sum_{kl} V_{kl} - N_a \right)^2 \right] \tag{7}$$

Updating Neurons

- In practice we have replaced the β term with a bias (knob) on each neuron.

- Near the end of the relaxation, the bias of neurons no longer enforced by their neighbors can be turned down, thus quickening convergence.

- We randomly queue the neurons and update each neuron at every timestep following the queue allowing the update to have an immediate.

- Convergence is determined using:

$$\frac{1}{N}\sum_{ij}\left|V_{ij}^{I} - V_{ij}^{I+1}\right| \leq criterion \tag{8}$$

where I is the iteration number.

- Once a minimum is found, T can be lowered and the system can be relaxed again in order to test the current best state of the system.

- Finally we make a cut in activation level (ie $V_i > 0.9$). The neurons which are above this cut become the hypothesis for the track fitting routine.

NN applied to Aleph

This Neural Network track finding algorithm was applied to simulations of the Aleph detector in 1990.

- They found it to be more accurate than other track recon algorithms

- They found that the behavior of the network did not depend strongly on the parameters.

- They report that it converged very quickly in most cases (6 iterations).

- They also found that the neuron setup was the largest fraction of the algorithms running time.

Implimentation