Spiking Correlation Matrix Memory

Preliminary Note #1

D Kustrin

Sunday, 7th of August 1999

Abstract

Correlation Matrix Memories have been successfully applied at York to a plethora of academic and industrial problems with great success (yawn); their properties and capabilities are well understood and recently it has been suggested that a spiking architecture could be constructed that will have all the benefits of CMMs (that is you Jim!). This note describes such an architecture. The SCMM model is based completely on the spike coincidence and post-synaptic potential additive properties.

1 Introduction

The Spiking Correlation Matrix Memory (SCMM) is a spiking associative memory based on correlation matrix memories. It uses time-to-first-spike code to encode weighed sum results and either pre-somatic inhibition or coincidence detection as thresholding mechanisms.

1.1 Time-to-First-Spike Code

Historically speaking, the most popular and widely studied coding strategy has been frequency coding in which values are encoded by the firing rates. Physiological data, on the other hand, suggests that although frequency coding is present in systems in many cases it cannot account for apparent speed with which biological systems respond to stimuli (Thorpe et al. 1996). Given the observed high temporal precision with which single spikes are produced (see eg. Abeles et al. 1993) it has been suggested that the timing of single spikes may be used as a coding strategy (Thorpe et al. 1996). This strategy has been formally analysed by Maass (1999) who has also formalised weighed sums and coincidence detection techniques utilising this strategy. Both of these techniques are used in construction of the Spiking Correlation Matrix Memory.

1.2 Spike Response Model

SCMM model description is based on the standard spike response model (Gerstner 1999). Each neuron i is described by a state variable u_i which is a linear superposition of it's own post-output negative contribution and the contribution from it's pre-synaptic neurons:

$$u_i(t) = \sum_{t_i^{(f)} \in \mathcal{F}_i} \eta_i(t - t_i^{(f)}) + \sum_{j \in \Gamma_i} \sum_{t_j^{(f)} \in \mathcal{F}_j} w_{ij} \epsilon_{ij}(t - t_j^{(f)}) \qquad (1)$$

where $t_i^{(f)}$ it the time when the neuron i fires an output spike, Γ_i is the set of all pre-synaptic neurons for neuron i, w_{ij} is the weight between neurons i and j, η and ϵ are kernels describing self and pre-synaptic neuron contributions respectively and $\mathcal{F}_i = \{t | u_i(t) = \vartheta \wedge \dot{u}_i(t) > 0\}$ is the set of all firing times for neuron i. The self-contribution kernel, η, is usually modelled by the following equation:

$$\eta_i(s) = -\vartheta \exp\left(-\frac{s}{\tau}\right) \mathcal{H}(s) \tag{2}$$

with ϑ being the value the state will have at the moment of firing (ie $u_i(t) = -\vartheta$), τ a time constant and $\mathcal{H}(s)$ the Heaviside function. The response to pre-synaptic spikes is usually modelled with the following equation:

$$\epsilon_{ij}(s) = \left[\exp\left(-\frac{s - \Delta^{AX}}{\tau_m}\right) - \exp\left(-\frac{s - \Delta^{AX}}{\tau_s}\right)\right] \mathcal{H}\left(s - \Delta^{AX}\right) \tag{3}$$

where τ_s and τ_m are time constants and Δ^{AX} is the axonal transmission delay. This formulation indicates that all pre-synaptic spike inputs are summed (albeit in a nonlinear fashion) and that the timing of the incoming spikes will have a large impact on the nonlinearity of the sum.

Although relatively simplistic, neuron response model outlined above is a quite good approximation to the actual neuron behaviour. Additionally it has been shown that it is in fact a more general model then integrate-and-fire model which is it's special case (Gerstner 1999). This note uses neuron response model as described above to describe spiking CMMs.

1.3 Summing Process

Using the spike response model it is easy to see all inputs contributions presented to a neuron will be weighed and summed and that this summed value (moderated by neuron's own spiking history) will be thresholded for generation of output spikes. This capability can be used to generate not only just binary fire-spike/not-fire-spike but to also encode the sum value in the timing of the output spike. The maximum response of a single neuron occurs at ϵ kernel, which occurs at $\max \epsilon_{ij}(s) = (\tau_s/\tau_m)^{-\tau_s/\tau_s - \tau_m} - (\tau_s/\tau_m)^{-\tau_m/\tau_s - \tau_m}$ which is usually the threshold level. If n spikes are received simultaneously some preset threshold value ϑ will occur at

$$S_n(s) = \Delta^{AX} + 2\tau_s \ln(2) - 2\tau_s \ln\left(1 \pm \frac{\sqrt{n^2 - 4n\vartheta}}{n}\right) \tag{4}$$

with the first occurrence being at the positive sum. When $\tau_m = 2\tau_s$ the maximum for a single spike will occur at $S_1 = 2\ln(2)\tau_s$. Equation 4 shows that distances nonlinearly encode summed values (fig. 1).

1.4 Coincidence Detection

The spike response model's equations suggest that timing of the incoming spikes will have a great impact on a neuron's state. This is manifested by the fact that summing is nonlinear (as shown above) which results in greatest state increase when incoming spikes are coincidental and lesser impact by spikes that are further apart in time Abeles (1982). This effect can be used to detect (relative) equality of spike encoded data. Consider eq. 3. From the equation it is clear that for some preset values of the timing constants τ_s and τ_m, it is possible to find value for the threshold ϑ which will allow detection of simultaneity of arrival of two or more spikes with arbitrary error (temporal distance).

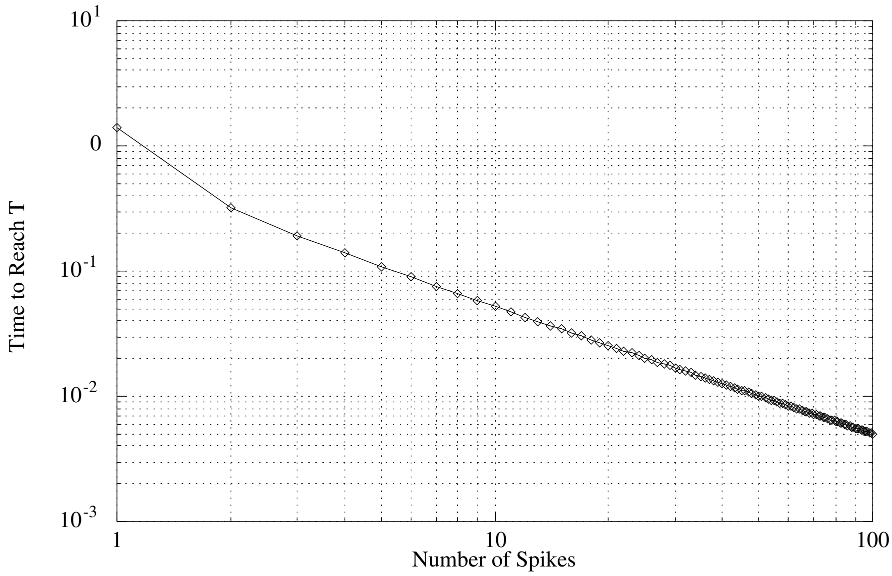

Figure 1: Distances

For example, if values are set at $\tau_s = 5.0$, $\tau_m = 10.0$ and $\vartheta = 0.3$, the temporal distance between two spikes would have to be below (approx) 22.48 time units for their occurrence to be detected as being coincidental, for $\vartheta = 0.4$ the distance is reduced to approx. 10.82 and for $\vartheta = 0.5$ the distance is zero. These three example are shown in fig. 2. Coincidence detection is a very powerful technique since it allows equality testing for temporally encoded numbers.

2 Spiking Correlation Matrix Memory

Spiking Correlation Matrix Memory utilises temporal coding, weighed summing and (in the case of Willshaw thresholding) coincidence detection to implements standard correlation matrix memory associative memory.

Correlation Matrix Memories (CMMs) are feed-forward static binary associative neural networks equivalent to a single layer fully connected feed–forward perceptron with binary weights. CMMs can operate in either auto or hetero–associative modes. Since they are quite simple they have been extensively studied and applied in various real–world problems including symbolic reasoning (Kustrin and Austin 1999), molecular structure matching (Turner and Austin 1997) and financial forecasting (Kustrin 1998). CMMs are prototypical associative memories with very simple and understandable training and recall algorithms. Given a set of binary input–output pairs represented in form

$$I_1 \rightarrow O_1, \quad I_2 \rightarrow O_2 \quad \ldots \quad I_n \rightarrow O_n \tag{5}$$

training can be expressed as a simple binary disjunction of inner products of the input–output column vectors (a form of Hebbian learning)

$$M = \bigvee_{i=1}^{n} O_i I_i^{\mathrm{T}} \tag{6}$$

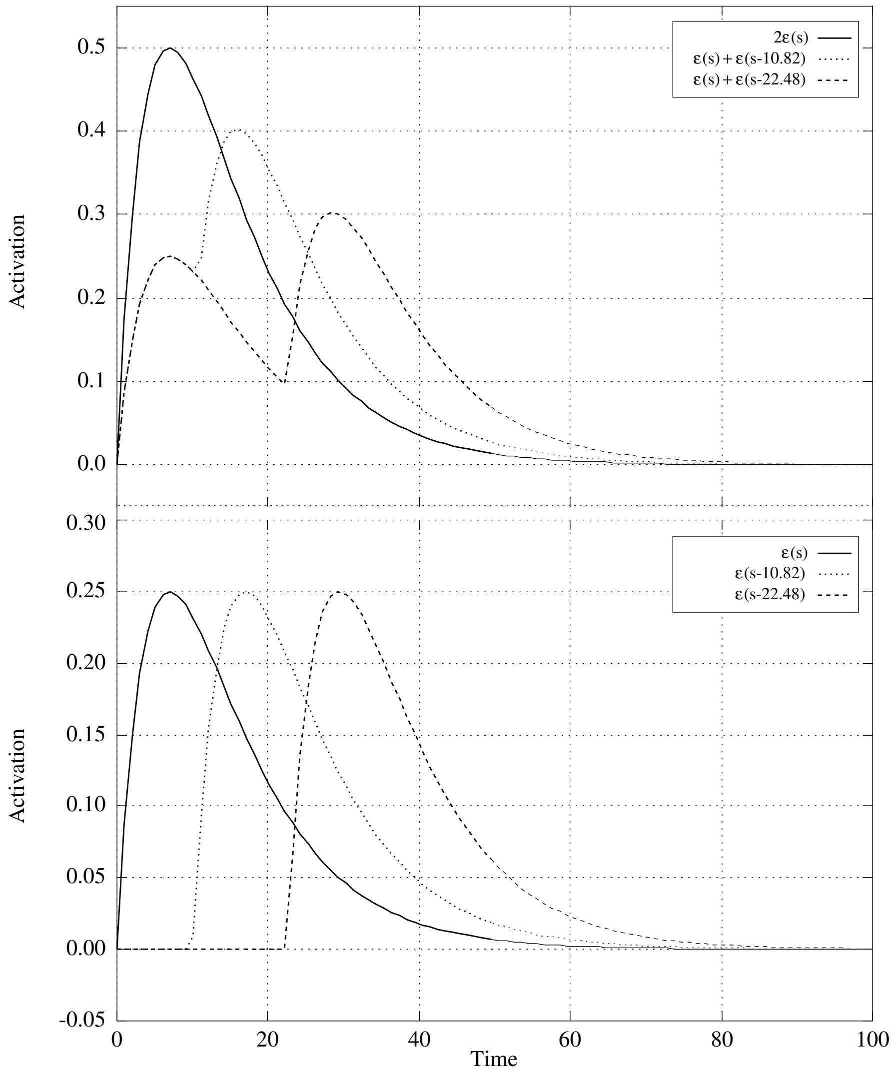

Figure 2: Example superpositions of excitatory post-synaptic potentials (top) and their components (bottom) for $\tau_m = 10.0$ and $\tau_s = 5.0$.

where $\bigvee$ is the binary disjunction and a^{T} is the transpose of the vector a. This equation can also be expressed in the *update form*:

$$\begin{aligned} M^0 &= 0 \\ M^i &= M^{i-1} \vee O_i I_i^{\mathrm{T}} \end{aligned} \tag{7}$$

which is clearly a version of the delta rule originally developed by Widrow (Widrow and Hoff 1960). Training can be viewed as construction of a mapping from the input to the output vector space and hence the CMM can be viewed as a simple mapping operator. In this framework, recall becomes simple projection of the input pattern from the input vector space to the output space:

$$O(I) = \Theta\left(MI\right) \tag{8}$$

where Θ is some threshold (transfer) function and I is some input. The need for the threshold function is intuitive: unless the input is orthogonal and of unit length, the output vector would not be a binary vector, but would contain output multiplied with the square of the input vector length.

CMMs have always been considered as biologically plausible BLAH BLAH. Delta rule. Willshaw. Problem - not spiking — that is why we want to do it.

The spiking CMM architecture is largely dependent on the thresholding technique used. In this note two methods are considered, Willshaw threshold and the fixed value threshold. Both architectures utilise the same summing layer that computes and temporally encodes the weighed sum of inputs; Willshaw SCMM uses an additional layer that performs thresholding while fixed weight models uses pre-somatic inhibition triggered by the input.

2.1 Summing Layer

The summing layer is identical in both the Willshaw and the fixed-weight SCMMs. It utilises the additive properties of the post-synaptic potentials to perform the weighed sum of inputs. In this paper the membrane time constant τ_m is taken to be twice the synaptic time constant, $\tau_m = 2\tau_s$, which implies that the single spike maximum will be at 0.25 potential units. The firing threshold is set to be $\vartheta = 0.2$ potential units which suggests that for n simultaneous inputs a spike will be generated at time

$$S_n = \Delta^{AX} + 2\tau_s \ln\left(2\right) - 2\tau_s \ln\left(1 + \frac{\sqrt{n\left(n - \frac{4}{5}\right)}}{n}\right). \tag{9}$$

Since n spikes can contribute to the EPSP reaching the soma, the refractory response has to account for the maximal potential value that will occur after the spike is produced (at $S_{max} = 2\tau_s \ln\left(2\right)$, if refractory kernel is ignored). Since for large values of s, the PSP kernel ϵ is dominated by its membrane time constant, τ_m, the refractory kernel η should have similar temporal properties. The membrane potential part of n PSP kernels is given by

$$\epsilon^m\left(s\right) = n \exp\left(-\frac{s - \Delta^{AX}}{\tau_m}\right)$$

which has gradient

$$\dot{\epsilon}^m\left(s\right) = -\frac{n}{\tau_m} \exp\left(-\frac{s - \Delta^{AX}}{\tau_m}\right)$$

so the refractory kernel should have the form $n \exp\left(-\frac{s}{\tau_m}\right)$ shifted so that the value at the S_{max} is $-n_{max}/4$, the maximal potential for all possible n spikes. This suggests the following equation for the refractory kernel:

$$\begin{aligned} \eta(s) &= -n_{max} \exp\left(-\frac{s - S_{max}}{\tau_m}\right) \mathcal{H}(s) \\ &= -n_{max} \exp\left(-\frac{s - 2\tau_s \ln(2)}{\tau_m}\right) \mathcal{H}(s) \end{aligned} \tag{10}$$

which is similar to the standard spike-response refractory kernel, eq. 2. Since the above defined kernel is scaled by the maximum number of spikes (which is equivalent to the number of inputs to the neuron), it will have a very large refractory effect on a single spike input and the effective refractory time will be quite large. The duration of the effective refractory period is an important factor in designing SCMM since it defines a bound on the minimum value for the axonal delay from the output to input in a recursive system. Figure 3 shows an example of a single spike response and the effective refractory period for an additional single spike input. Single spike input events

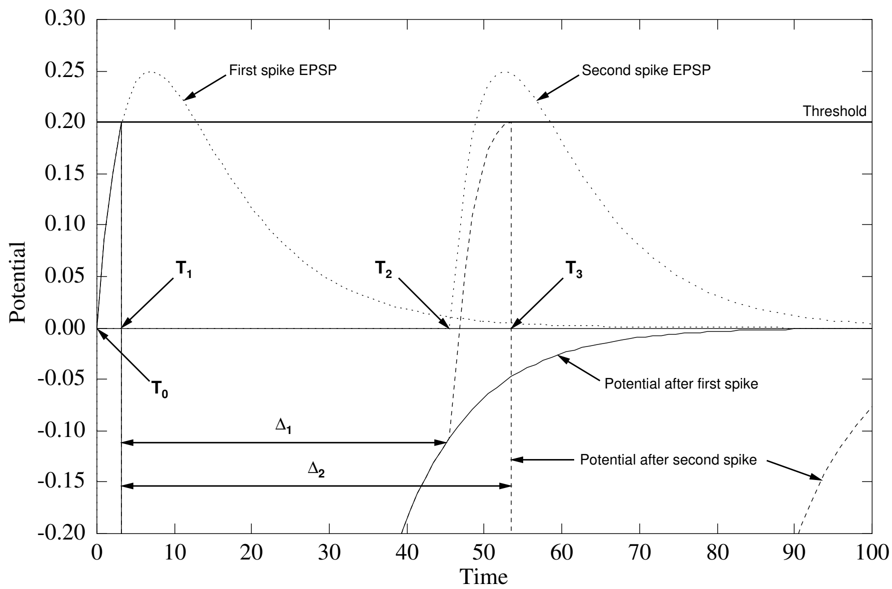

Figure 3: Post-synaptic potential graphs for single spike inputs to a four input neuron with $\vartheta = 0.2$, $\tau_s = 5.0$ and $\tau_m = 10.0$. A single spike is received by the cell at time T_0 producing the first excitatory post-synaptic potential (EPSP) which on reaching the threshold at time T_1 results in production of a spike on output and start of the refractory kernel contribution. Time Δ_1 shows the minimum time that must pass for another input spike's EPSP to be able to reach the threshold after the neuron has fired (the effective refractory time) while Δ_2 is the time between the first output spike produced by the neuron and the spike that is produced by another input spike at time T_3

will produce longest refractory periods and thus define the upper bound of the effective refractory period, Δ_r. To find this bound it is necessary to solve the following equation:

$$\epsilon(s) + \eta(s - S_1) + \epsilon(s - (S_1 + \Delta_r)) = \vartheta \tag{11}$$

for Δ_r given time constants τ_s and τ_m, the threshold ϑ, axonal delay Δ^{AX} and the number of possible inputs to the neuron, n. The solution, which will be a function of s then needs to be minimised. For example if $n_{max} = 4$, $\vartheta = 0.2$, $\tau_s = 5.0$ and $\tau_m = 10.0$ as before, Δ_r has a value of just under 36.5. This implies that the axonal delay should be at least set at this value is single spike inputs are not to be lost. In the simulations presented at the end of this note the axonal delay was set to 80 time units which is well outside effective refractory period. WHY 80?

2.2 Willshaw Threshold SCMM

Willshaw threshold is a technique that thresholds the output of a CMM at the number of bits set in the input vector (**?**). It the original paper, Willshaw argues that the method is biologically plausible. Mathematically, it can be shown that for orthogonal output vectors it is the best thresholding technique (**?**). CAN IT REALLY?

The SCMM architecture implementing the Willshaw threshold is constructed from a summing layer that sums all the weighed inputs and outputs the summed value as the time-to-next-spike with larger values having shorter temporal distances to smaller values. Additionally, and auxiliary neuron is used to calculate the total number of coincidental spikes in all the inputs to SCMM which is also produced as temporally encoded number. These two outputs are then fed to a second layer of neurons that performs coincidence detection between the weighed sum and the auxiliary neuron's output, see fig. 4. The timing constants for both summing and auxiliary neurons must be the same

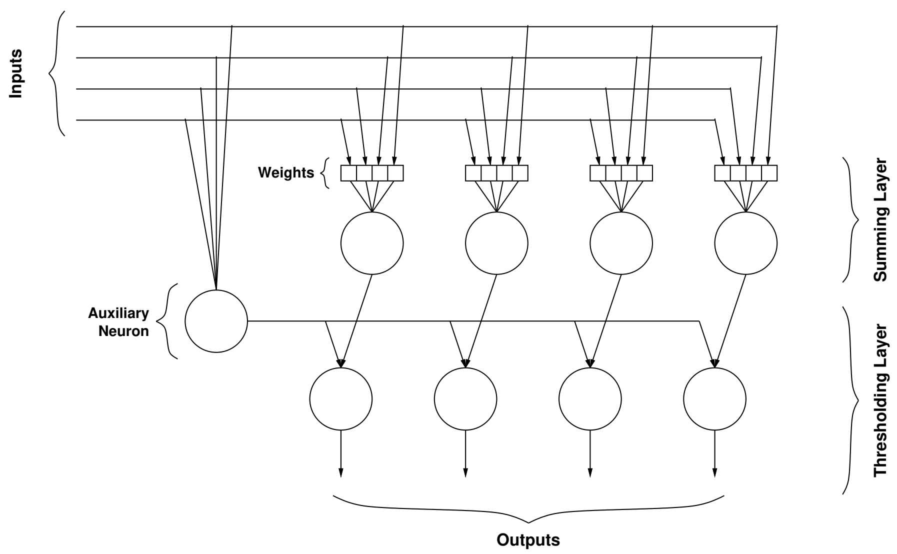

Figure 4: Willshaw threshold SCMM.

if the coincidence detection neurons are to function. For this architecture the effective refractory period will include the transmission time to the coincidence detector neurons as well so the actual axon delay will be smaller then in the case with direct recursive connections.

2.3 Fixed Weight Threshold SCMM

The fixed weight threshold simply compares the summed value with some fixed value and produces output if the summed value is greater or equal to the output value. This technique can be simply implemented by inhibitory synapses locate very close to the soma and driven by an auxiliary neuron with preset firing frequency and different temporal characteristics to the summing neurons (see fig. 5). The auxiliary neuron needs to produce it's spikes much faster then the summing neurons if it's output is to affect the inputs. Additionally, the inhibitory synapses must produce post-synaptic potentials with different temporal characteristics then the excitatory synapses. TALK ABOUT

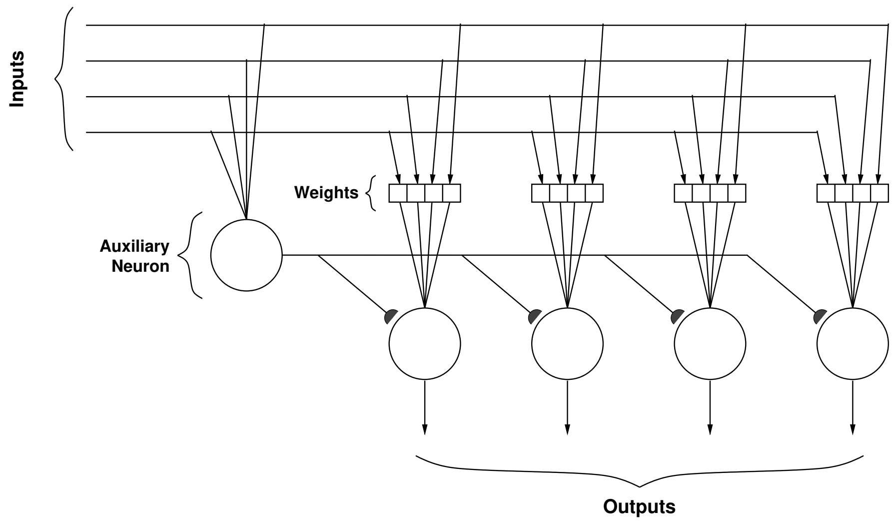

Figure 5: Fixed weight SCMM.

AUXILIARY TIME CONSTANTS HERE, I GUESS.

3 Examples

As a toy problem that will illustrate SCMM operation the following set of relations, $\{A \to B, B \to C, C \to A\}$, and their binary codings $A = [1001]$, $B = [0110]$, $C = [0011]$, $D = [0001]$ were converted into a CMM weights matrix

$$w = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}. \tag{12}$$

This weight matrix was then used in construction of both Willshaw and fixed-weight threshold SCMMs. All SCMM neurons had time constants fixed at $\tau_m = 2\tau_s$ and $\tau_s = 5.0$ with threshold at $\vartheta = 0.24$. In the Willshaw-threshold architecture the threshold layer neurons had their timing constants set at $\tau_m = 10.0$ and $\tau_s = 1.0$, to facilitate more accurate discrimination of summed outputs. The Willshaw SCMM architecture is presented in fig. 6. The architecture is recurrent and by presenting input vector corresponding to symbol A it should produce the sequence $A \to B \to C$ after which it will enter a cycle of producing vectors [1011] and [0010] due to the nature of the Willshaw method. The output of the simulation is presented in fig. 7.

4 Discussion

What is interesting?

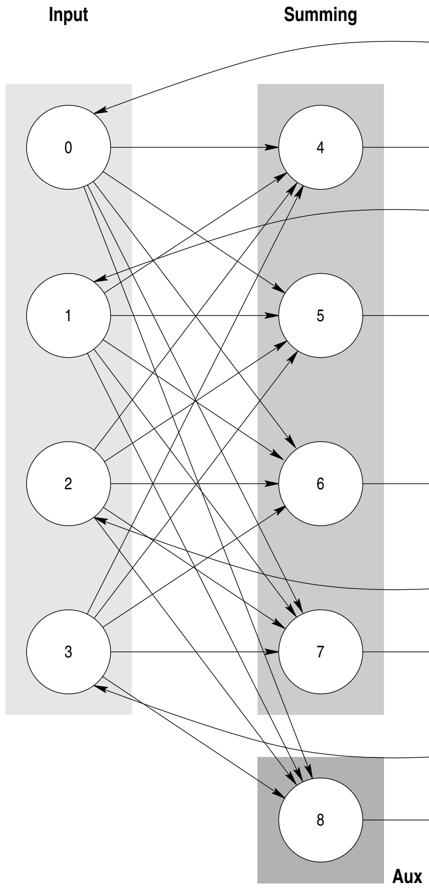

Figure 6: The Willshaw-threshold SCMM architecture used in a the example CMM. All non-thresholding had threshold at 0.25 and timing constants set at $\tau_m = 2\tau_s$ and $\tau_s = 5.0$. Thresholding neurons were set at $\vartheta = 0.5$, $\tau_m = 10.0$ and $\tau_s = 1.0$. Axonal delay for all neurons was set at 80 time units.

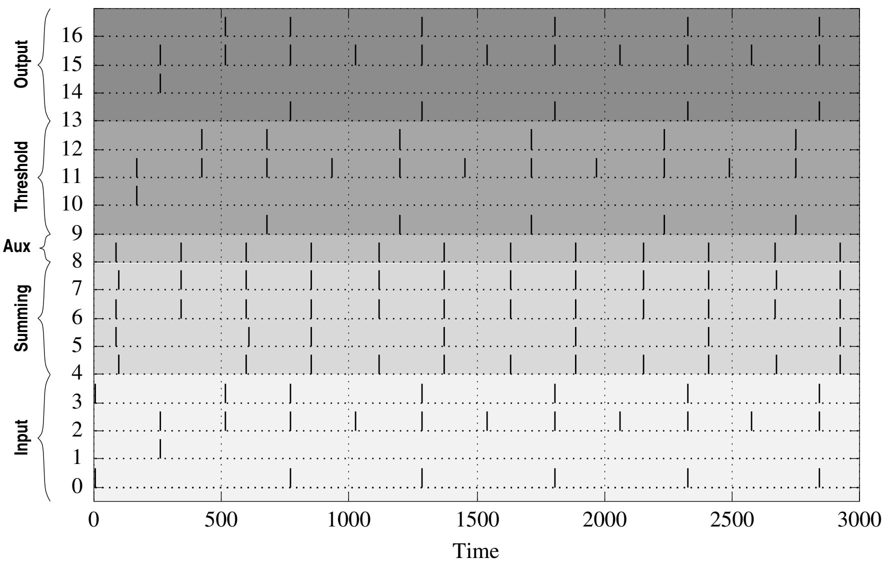

Figure 7: Spike trains for all neurons in the Willshaw-threshold SCMM used in a the example CMM.

References

Abeles, M. (1982). Role of the cortical neuron: integrator or coincidence detector? *Israel Journal of Medical Sciences 18*(1), 83–92.

Abeles, M., H. Bergman, E. Margalit, and E. Vaadia (1993). Spatiotemporal firing patterns in the frontal-cortex of behaving monkeys. *Journal of Neurophysiology 70*(4), 1629–1638.

Gerstner, W. (1999). Spiking neurons. In W. Maass and C. M. Bishop (Eds.), *Pulsed Neural Networks*, pp. 3–54. MIT Press.

Kustrin, D. (1998). *Forecasting Financial Time Series with Correlation Matrix Memories for Tactical Asset Allocation*. Ph. D. thesis, Department of Computer Science, University of York.

Kustrin, D. and J. Austin (1999, March). Connectionist graph traversal implementing a simple propositional logic. In D. Kustrin (Ed.), *Third Wieghtless Neural Networks Workshop*, York.

Maass, W. (1999). Computing with spiking neurons. In W. Maass and C. M. Bishop (Eds.), *Pulsed Neural Networks*, pp. 55–86. MIT Press.

Thorpe, S., D. Fize, and C. Marlot (1996). Speed of processing in the human visual system. *Nature 381*(6582), 520–522.

Turner, M. and J. Austin (1997). A neural network technique for chemical graph matching. In M. Niranjan (Ed.), *Proceedings of the Fifth International Conference on Artificial Neural Networks*. IEE.

Widrow, B. and M. E. Hoff (1960). Adaptive switching circuits. IRE Western Electric Show and Convention Record, part 4.