EXTENSION OF THE HAMMING NEURAL NETWORK TO A MULTILAYER ARCHITECTURE FOR OPTICAL IMPLEMENTATION

D.R. Selviah and J.E. Midwinter

University College London, UK

INTRODUCTION

The well known Hopfield net (1) has recently been extended to a three layer architecture (2) by means of a matched filter modelling formalism (3). This enables an auto- or hetero-associative memory to be constructed optically from three holograms, one of which is formed from bipolar orthogonal patterns and fixed prior to training. However, the Hopfield net is prone to converge onto incorrect or spurious states. This tendency can be reduced by reducing the gain of the threshold on repeated iterations but no definite upper limit on the gain has yet been specified to ensure correct convergence. Moreover, a threshold with variable gain is not easy to implement optically (4,5). In this paper we examine how the less well known Hamming net (6) can be similarly extended to a three layer architecture by means of the matched filter formalism. When the convergence parameters are set below a defined upper limit the net always converges to the correct pattern (6) and so offers a definite advantage over the Hopfield net. Moreover, the Hamming convergence parameters can be made to affect the threshold offset rather than its slope and this is easier to implement optically.

THE HAMMING NET

The Hamming net is well described by Lippmann (6). It essentially consists of two cascaded subnets. The first is an interconnection net which performs the inner product of the N-bit input code with each of the M memorised N-bit vectors and outputs these inner product magnitudes in each of M corresponding channels. The M analogue magnitudes pass through an asymmetric threshold, T_1, (Fig.1 with θ=N/2) which allows the input magnitude through unchanged if it is greater than some threshold, θ, otherwise it is set to zero. This asymmetric threshold avoids the possibility of converging onto the inverse of a memorised code (such as can happen in the Hopfield net due to the symmetric thresholds). The second subnet (or MAXNET) has to select which of the inner product magnitudes greater than N/2 is the largest, and to indicate this by only giving a non-zero output in the corresponding channel. It does this by passing each magnitude to a corresponding threshold, T_2, (with θ=0) in a second layer via a weight of 1 and to all of the other thresholds in this second layer via a weight of $-\varepsilon$ (where ε is a small quanity, $\varepsilon<1/M$). The M outputs of the MAXNET are fed directly back to its M inputs and the process continues until only one output is non-zero. This will always be the closest memorised code to the input, in the Hamming sense, provided $\varepsilon<1/M$.

The first subnet is acting as a bank of matched filters (Fig.2), each storing one of the memorised codes, s, followed by gates which select the correlation peak and thresholds set to exclude signals which are obviously too low to indicate any similarity with their memorised codes. The second net is recursive and and so can be unfolded as shown in Fig.2 to give a multilayer configuration with sufficient to only give one non-zero output.

THE MATCHED FILTER MODEL OF THE HAMMING NET

In the matched filter model spatial distributions within the net are generally represented by temporal distributions for convenience of diagrammatic illustration. So, in Fig.2, the spatial distributions of correlation peak magnitudes emerging from the final set of (θ=N/2) thresholds is represented as a time sequential code consisting of bits each of which has an analogue amplitude corresponding to one of the thresholded correlation peaks.

Each of the interconnection layers within the MAXNET can be represented as the multiplication of the first analogue amplitude bit by the code $e1=(1,-\varepsilon, -\varepsilon, \ldots, -\varepsilon)$ to give a new code plus the multiplication of the second bit by the code $e_2=(-\varepsilon, 1, -\varepsilon, \ldots, -\varepsilon)$ plus all of the similar multiplications of each of the bits. This can be represented in matched filter terminology by the correlation of the input code with a monopolar code $o_1=(1, 0, 0, \ldots, 0)$ with gating to select the first bit magnitude followed by the convolution with the code $e_2=(1, -\varepsilon, -\varepsilon, \ldots, -\varepsilon)$ and so on summed over all of the channels. Fig.3 therefore shows the full matched filter representation of the Hamming net. Each correlation peak magnitude in the first subnet is <u>labelled</u> with a monopolar orthogonal code, o, given by

$$\sum_{\tau=1}^{N} o_{i\tau}\, o_{j\tau} = 0 \qquad i\neq j \qquad (1)$$

$$\sum_{\tau=1}^{N} o_{i\tau}\, o_{j\tau} = 1 \qquad i=j \qquad (2)$$

where the τ subscript is the bit index and the i and j subscripts are the code indices. The MAXNET then iterates to select the strongest correlation peak which is output. In this model the bit positions in the codes are represented in the time domain but they could equally well be in the spatial domain by replacing time by space as the independent variable which would be done in an optical implementation.

THREE LAYER NEURAL NET BASED ON THE HAMMING NET

A multilayer architecture can be realised by following the MAXNET by a pattern association net (2) which remaps the monopolar orthogonal code set labels back to either the original set for pattern recognition purposes or to a new set of codes, p_1 for pattern association (Fig.4).

A suitable optical element for performing weighted interconnections is a hologram formed from an input pattern and a reference pattern. These are associated in the holographic medium and correspond to a single channel of the net consisting of a correlation with the first image, and a convolution with the second. Optically, the feedforward association net following the Hamming net can be implemented by forming superimposed or space multiplexed holograms between the pairs of codes in all of the channels. The representation of a neural network in terms of matched filters is therefore a direct description of how one might form the holograms for an optical implementation.

The three layer Hamming based net can, therefore, be constructed using three holograms, the middle one being recorded and fixed before training, while the first and last are recorded during training using the orthogonal code set as the reference patterns.

HAMMING NET IMPLEMENTATION USING OPTICAL INTENSITY ENCODING

So far we have assumed that it is possible to form holograms using patterns having both positive and negative bits. This is so, provided we use phase patterns giving 0° phase for +1 bits and 180° phase for -1 bits. However, phase errors can easily occur due to misalignment and errors in holograms and other components of as little as one quarter of a wavelength ~ 0.2 μm. A better method is to use intensity encoding where the tolerances are not so tight, however, intensities can only be positive. The Hamming net has an advantage over the Hopfield net in this respect since the Hamming net uses monopolar orthogonal codes, whereas the Hopfield net uses bipolar orthogonal codes. Nevertheless, the e codes contain negative bits in the Hamming net and so would entail the use of phase patterns. One way around this has been suggested by White (7) for the Hopfield net, where the algorithm is rewritten in such a way that each hologram described above is replaced by two parallel holograms, one dealing with the positive bits and one with the negative bits. Both use positive intensity for the presence of the bit and zero intensity elsewhere. An alternative method described below for the Hamming net involves reformulating the net with the help of the matched filter formalism into a purely monopolar format.

We neglect losses due to subdivision of the signal and we assume ideal lossless components. If the quanity $+\varepsilon$ is added to each bit of the e codes in Fig.4, they become the same as the monopolar orthogonal codes in the same channel but scaled by $(1+\varepsilon)$. So the central interconnection net can be simply replaced by an amplifier with a gain of $(1+\varepsilon)$.

This addition means that each bit of the code reaching the threshold is larger by the addition of $\varepsilon.\ (A_1+A_2+A_3+\ldots+A_M)$ where the A_i are the intermediate amplitudes in each channel after the monopolar matched filters. If the MAXNET threshold is set to

$$\theta = \varepsilon . \sum_{i=1}^{M} A_i$$

$$\theta = \varepsilon . M . A_{mean}$$

where A_{mean} is the average amplitude, then the output from the threshold is just as it was before the extra addition since we are effectively subtracting ε from each of the codes by doing this. This gives a great simplification since now the transformed intermediate or MAXNET (Fig.5) consists of only two elements: an amplifier with a gain of $(1+\varepsilon)$ and a threshold whose offset, θ, decreases on successive iterations since A_{mean} decreases as some of the A_i are set to zero. Not only does this avoid the need for a multiplexed hologram which would introduce loss due to subdivision of the input signal, but the threshold response with variable offset is easier to realise optically than the variable slope of the Hopfield threshold (2). The resultant net consists of two layers of interconnections surrounding the central iterative loop. If the training data is chosen to be monopolar, then the first and last pattern association nets can be holograms implemented using intensity encoding.

Other equivalent forms are possible for the transformed MAXNET which offer further possibilities for implementation. If A_{mean} is normalised to unity on each iteration, the threshold offset can be <u>fixed</u> at $\theta = \varepsilon.M$. The amplifier can conveniently be combined with the offset threshold to give two elements: a fixed threshold with an offset of $\theta = \varepsilon.M$ and a slope of $(1+\varepsilon)$ and an automatic gain control for normalisation. The AGC must sum the amplitudes of all of the bits in the code, that is, it must integrate across the code (spatially for optics, e.g, using a lens to focus the pattern), and then adjust the gain to maintain a mean amplitude of unity. The gain will increase non-uniformly on successive iterations as more and more of the bits are set to zero by the threshold. A further equivalent form involves combining these elements into a single threshold with variable gain, which increases non-uniformly as the net iterates.

CONCLUSIONS

A matched filter model of the Hamming net has been developed and this has enabled a three layer pattern association net to be designed. The matched filter technique has also enabled a new two layer Hamming Associative Net to be designed, which is more suited to optical implementation. Novel equivalent forms for the MAXNET have also been derived.

REFERENCES

1. Hopfield, J., 1982, <u>Proc. Nat. Acad. Sci. USA</u>, <u>79</u>, 2554-2558.

2. Selviah, D.R., and Midwinter, J.E., 1989, "New Developments in Neural Computing" Conference Proceedings, eds. Taylor and Manion, IOP Publishing Ltd.

3. Selviah, D.R., Midwinter, J.E., River, A.W., and Lung, K.W., 1989, <u>IEE Proceedings Part F</u>, <u>136</u>(3), 143-148.

4. Selviah, D.R., and Midwinter, J.E., 1987, "Optical techniques for information processing", IEE Colloquium Digest No. 1987/105, 6/1-6/4.

5. Midwinter, J.E.M., and Selviah, D.R., 1989, <u>Neural Computing Architectures</u>, ed. I. Aleksander (North Oxford Academic Publishers Ltd) 258-278.

6. Lippmann, R.P., 1987, IEEE ASSP Magazine, 4(2), 4-22.

7. White, H.J., 1988, SPIE, 963, Optical Computing 88, 570-575.

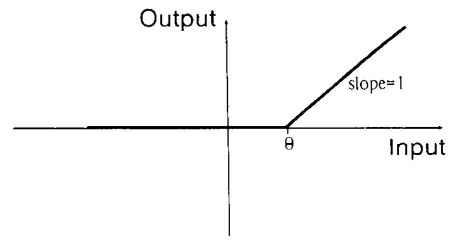

Fig. 1. Hamming Threshold Response

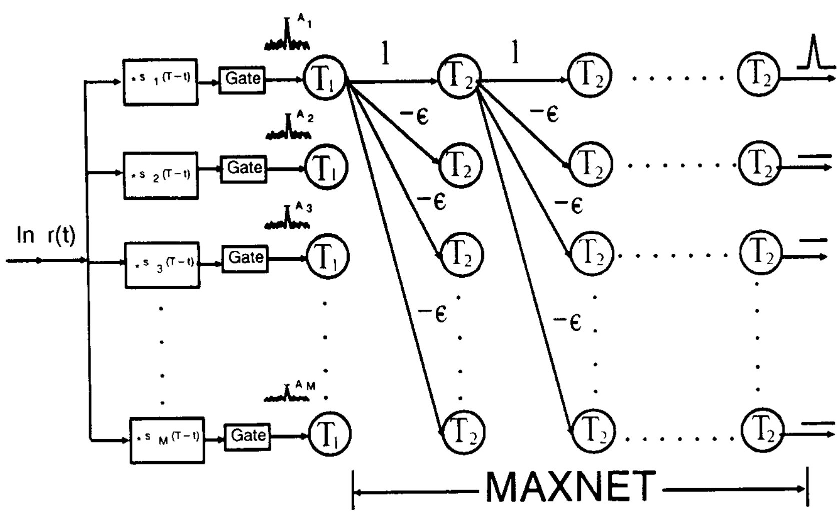

Fig. 2. Hamming Neural Net with the First Subnet Represented as a Matched Filter Bank and the Second Subnet as a Multilayer Net

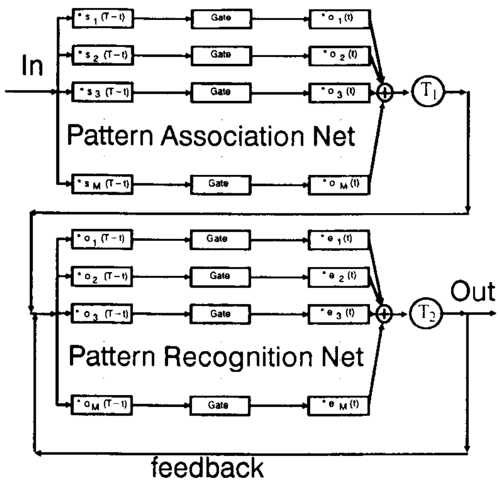

Fig. 3. Matched Filter Model of Hamming Neural Net

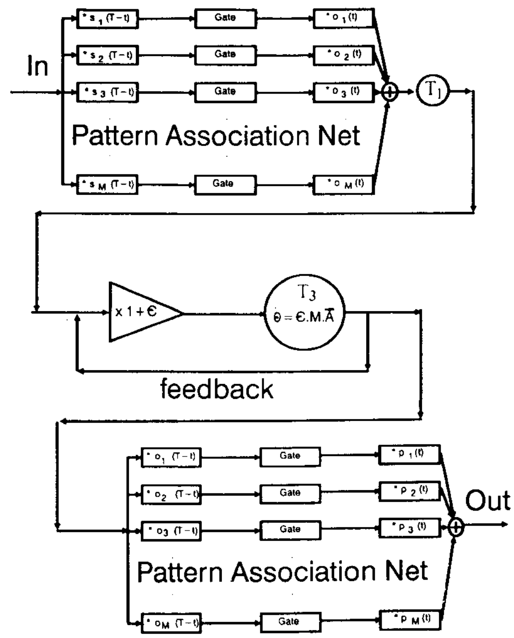

Fig. 5. Two Layer Hamming Net

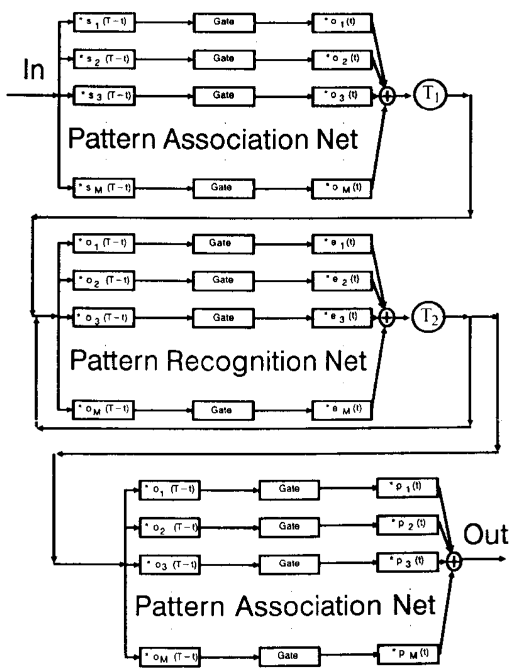

Fig. 4. Three Layer Hamming Net