Multi-Pattern Real-Valued Spectral Associative Memories

R. G. Spencer

Texas A&M University
Electrical Engineering Department
College Station, Texas 77843-3128
ron.spencer@ieee.org

Abstract

A multi-pattern encoding and decoding scheme is presented that extends the family of spectral associative memories (SAMs) to include gray-level, or analog patterns. SAMs are frequency-domain formulations of associative memory that combine the extrinsic redundancy of neural networks with the in-phase, quadrature, and complex modulation schemes of communications. Considerable coding gain occurs at the level of modulation and these networks may be regarded as multi-channel, multi-carrier generalizations of amplitude modulation. Unlike multi-pattern bipolar SAMs, which are exclusively content-addressable, real-valued SAMs also have an address-addressable mode in which the recall of a particular memory may be forced. Band structures and anti-aliasing constraints are presented along with a probabilistic formulation in which virtual entanglement is a natural feature. Simulations are presented that demonstrate dual-memory recall for 6x6 gray-level patterns.

1 Introduction

Spectral associative memories (SAMs) were introduced in [1]-[3] for transmitting digital data patterns over noisy channels with built-in noise immunity. SAMs are frequency-domain formulations of associative memory that combine the extrinsic redundancy of neural networks with the in-phase, quadrature, and complex modulation schemes of telecommunications. Unlike conventional associative memories in which data patterns are stored as long-term, non-volatile attractors in a synaptic weight matrix, spectral memories are manifested as memory waves. Spectral attractors are not *stored* in a neural network or holographic plate - they are only *expanded* by the neural decoder. Data patterns, or codewords may be enfolded into an attractor wave by associative amplitude modulation (AAM) (spectral spread) and recalled with considerable noise immunity by recurrent associative amplitude demodulation (AAD) (spectral focus). AAM disperses the data pattern across the frequency domain, creating extrinsic redundancy and AAD re-focuses the pattern in the frequency domain for spatial re-expansion. Temporary basins of attraction are set up in the decoder that force the recall of one of the enfolded patterns and disappear when transmission of the memory wave ceases. The memory wave guides the state of the decoder into one of the basins of attraction like deBroglie's pilot wave in Bohm's wave/particle theory [4] and such a decoder is an uninstantiated, attractorless decoder with no net motive until activated by a coherent spectral attractor. Long-range connectivity is made virtually by spectral convolution and inner products are calculated in the frequency domain [5], allowing both coding and decoding networks to scale linearly with pattern dimension. The price is bandwidth, which scales quadratically with pattern dimension. Interestingly, memory formation and recall may be expressed in Dirac notation, the language of quantum mechanics [6] and a virtual analog of entanglement results from factorable memory waves.

In this paper, real-valued SAMs are presented in which *gray-level* patterns may be encoded by singular value composition and transmitted to a decoding network for recall. Unlike bipolar SAMs, the identity wave is subtracted at the decoding network and every memory pattern must have a unique weighting factor, or eigenvalue, l, in the spectral composition. As a result, analog associative memory patterns may be recalled by either *content* (content-addressability), in which the eigenvalue is allowed to adapt in the decoding network, or by *address* in which the eigenvalue is fixed (address-addressability). When l in the decoding network equals the eigenvalue of one of the encoded eigenvectors (patterns), the kernel of the corresponding linear transformation is biased in the direction of the corresponding eigenvector, all other attractor basins are enervated, and the corresponding pattern is recalled.

The conventional matrix-vector formulation of analog associative memory is given in section 2 followed by the wave formulation in section 3. Band structures and anti-aliasing constraints are given in section 4 and simulation results of a 37-dimensional network with two 6 x 6 gray-level patterns are given in section 5. A probabilistic formulation is given in section 6 followed by conclusions in section 7.

2 Matrix-Vector Formulation

2.1 Memory Formation and Recall

For an N-dimensional network, $P \le N$ orthonormal patterns may be encoded into a single memory matrix $\mathbf{M}$ by a weighted-sum of outer products:

$$\mathbf{M} = \sum_{p=0}^{P-1} l_p \mathbf{u}^{(p)} \mathbf{u}^{(p)\mathrm{T}} ; \quad \mathbf{u}^{(i)} \mathbf{u}^{(j)} = \ddot{a}_{i,j} , \tag{1}$$

where $\mathbf{u}^{(p)}$ is the p^{th} real-valued, orthonormal pattern vector to be encoded and l_p are real weighting factors, which must be different for each pattern. Real-valued coding is spectral decomposition in reverse:

$$\mathbf{M} = \mathbf{U} \Sigma \mathbf{U}^{\mathrm{T}} , \tag{2}$$

where $\mathbf{U}$ is a unitary matrix (orthogonal in this context) containing memory patterns $\mathbf{u}^{(p)}$ and $\mathbf{\Sigma}$ is a diagonal matrix containing the corresponding eigenvalues l_p. To recall one of the encoded memories by gradient descent, a recall potential may be defined as follows:

$$E \equiv \frac{1}{2} \left\| \mathbf{W} \mathbf{v} \right\|^2 ; \quad \mathbf{W} = l \, \mathbf{I} - \mathbf{M} \tag{3}$$

where $\mathbf{v}$ is the continuously refined state vector, $\mathbf{W}$ is the synaptic weight matrix, $\mathbf{I}$ is the identity matrix, and l is a continuously adapted variable which converges to the eigenvalue of one of the stored pattern vectors. It is straightforward to show that $\mathbf{v}$ and l may be adapted along the negative gradient of E by

$$\mathrm{d}\,\mathbf{v}/\mathrm{d}t = -c_v \mathbf{W} \mathbf{W} \mathbf{v} \tag{4a}$$

$$\mathrm{d}\,l/\mathrm{d}t = -c_l \mathbf{W} \mathbf{v} \mathbf{v} , \tag{4b}$$

where c_v and c_l are the corresponding adpatation rates.

In the special case when only one pattern is encoded i.e. when $\mathbf{M}$ has only one non-zero eigenvalue, then (4a) reduces to $-c_v \mathbf{W} \mathbf{v}$ since $\mathbf{W}\mathbf{W} = l \mathbf{W}$, where l may be absorbed into the adaptation rate and (4b) would be unnecessary.

2.2 Addressibility

To avoid spurious recall, every encoded memory must have a unique eigenvalue, which may serve as an address. Thus, depending on the relative adaptation rates, real-valued associative memories may be address-addressable or content-addressable. When $c_l = 0$, the eigenvalue drives the content and the network recalls the memory associated with l (address-addressibility). On the other hand, when $c_l > c_v$ the content in $\mathbf{v}$ drives the eigenvalue and the initial conditions determine the final state (content-addressibility).

3 Spectral Formulation

A wave formulation of real-valued associative memory replaces the synaptic weight matrix with an attractor wave and the neural state vector is replaced by a state wave. A key feature of spectral associative memory is that long-range connectivity may be achieved virtually in the frequency domain by spectral convolution, which allows linear scalability of the physical realization with pattern dimension, per encoded pattern. A continuously transmitting coding network is required for each pattern, and a composite memory wave is formed from which any one of the patterns may be recalled by recurrent associative amplitude demodulation.

3.1 Memory Formation

Memory formation may be divided into three parts: 1) pattern preparation, 2) synthesis, and 3) spectral convolution. See Fig. 1.

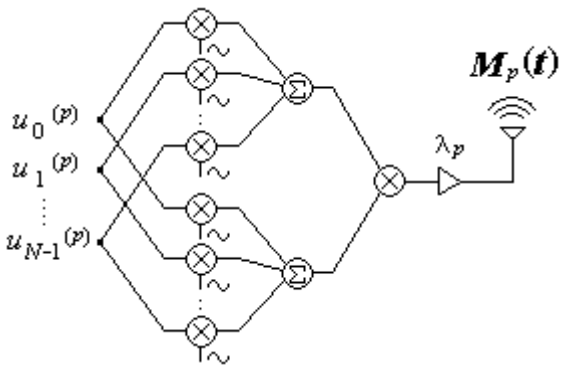

Fig. 1. Block diagram of a single associative amplitude modulator, or spectral coding network.

3.1.1 Pattern Preparation. A maximum of N patterns may be stored in an N-dimensional associative memory, but due to the orthogonality constraint only N-1 of them may be arbitrarily determined to some degree. P-1 channels must be reserved for orthonormalization, where $P < N$ is the number of patterns to be stored, and if pattern intensity is to be preserved then one channel must be allocated as a fixed reference. Thus, the total number of usable dimensions or channels per pattern is $n = N\text{-}P$.

3.1.2 Synthesis. Coding waveforms $s_1(t)$ and $s_2(t)$ for a *single pattern p* are defined as the sum of N waveforms, each carrying pattern information in the amplitude:

$$s_{1(2)}^{(p)}(t) = \sum_{i=0}^{N-1} u_i^{(p)} e^{j w_{S1(2)i} t} \tag{5}$$

where $w_{S1(2)i}$ are the i^{th} elements of the nonoverlapping carrier frequency vectors $\boldsymbol{\omega}_{\mathbf{S1(2)}}$, defined in the next section.

3.1.3 Spectral Convolution. The individual memory waves, $M_p(t)$ are formed by spectral convolution and added together to form a collective spectral composition:

$$M(t)=\sum_{p=0}^{N-1}M_p(t);\quad M_p(t)=l_p s_1^{(p)}(t)s_2^{(p)}(t)^*. \qquad (6)$$

Due to the complex coding in (5), the upper-sideband (USB) naturally cancels out in (6) and all available power goes into the lower-sideband (LSB) of the memory wave. Contained within $M(t)$ is extrinsic redundancy about the encoded data patterns, which facilitates noise immunity.

3.2 Memory Recall

Recall is achieved by recurrent associative amplitude demodulation in which spectral convolution generates inner product information in the frequency domain [5]. An *attractor wave* is continuously generated from the memory wave and spectrally convolved with the *state wave* of the decoding network. Contained within the result is a linear transformation in which the kernel adapts toward one of the encoded memories. Extrinsic information distributed over the entire attractor band is re-focused and transferred to the LSB of the gradient wave, which pilots the network into the strongest basin of attraction, influenced by the current eigenvalue address. White noise is filtered out over time as the memory wave provides the only coherent motive.

Recall may be divided into five parts: state wave synthesis, attractor wave generation, error wave synthesis, state update and eigenvalue update. See Fig. 2.

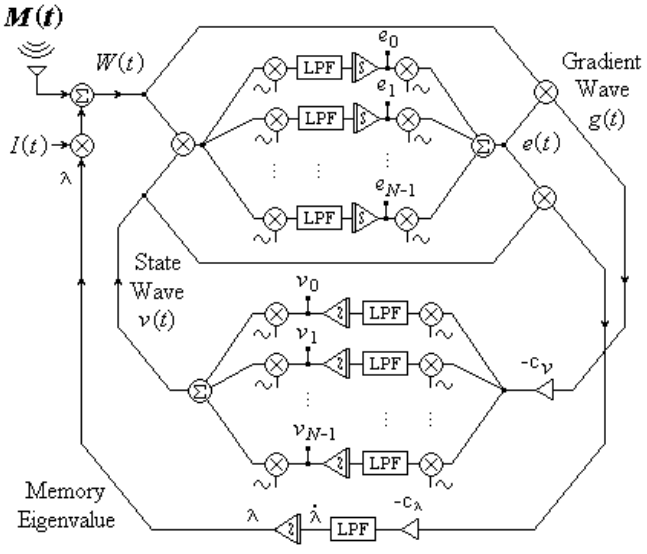

Fig. 2. Block diagram of an associative amplitude demodulator, or spectral decoding network.

3.2.1 State Wave Synthesis. The state wave $v(t)$, is a weighted-sum of individual channel waves across the entire state band:

$$v(t)=\sum_{i=0}^{N-1}v_i e^{jw_{Vi}t} \qquad (7)$$

where w_{Vi} is the decoder synthesis frequency for the i^{th} virtual neuron. State vector **v**, which may initially contain random guesses or noisy or imcomplete content, converge to scaled versions of the encoded data patterns over the recall period and serve as the final system output.

3.2.2 Attractor Wave Generation. The attractor wave is formed by subtracting the memory wave from a scaled identity wave $I(t)$ containing energy at diagonal frequencies:

$$W(t)=lI(t)-M(t);\qquad I(t)=\sum_{i=0}^{N-1}e^{jw_{ILi}t} \qquad (8)$$

where w_{ILi} is the i^{th} element of the LSB identity frequency vector $\boldsymbol{\omega}_{\mathbf{IL}}$, defined in the next section. Subtracting $M(t)$ from $lI(t)$ generates a kernel wave space by adjusting the strength of virtual self-connectivity in the decoder. This step transforms the range space of the linear transformation $W(t)v(t)^*$ into a kernel space and should not be confused with the subtraction of an unscaled identity wave in the bipolar SAM which modifies the range space. Although l may be complex, it is restricted to real values in the present treatment.

3.2.3 Error Wave Synthesis. Error information is extracted from the attractor and state waves by direct conversion:

$$e_i=\text{LPF}\left[W(t)v(t)^* e^{-j\omega_{Ai}t}\right] \qquad (9)$$

where e_i is the i^{th} element of the spatial error vector **e**, w_{Ai} is the i^{th} element of analysis frequency vector $\boldsymbol{\omega}_{\mathbf{A}}$ and LPF refers to low-pass filtering [7] with at least 40dB of attenuation at the carrier separation frequency (defined in the next section). An error wave is then synthesized by

$$e(t)=\sum_{i=0}^{N-1}e_i e^{jw_{Ei}t}. \qquad (10)$$

where w_{Ei} is the error synthesis frequency for the i^{th} virtual neuron, which may be made equal to w_{Vi}. Together, these equations perform the function of a *comb filter* – a function commonly used to separate luminance and color information in TV signals.

3.2.4 State Update. Real-valued recall follows a continuous trajectory along the negative gradient of the recall potential, which is extracted by direct conversion:

$$\mathrm{d}\,\mathbf{v}/\mathrm{d}\,t = -c_v\,\mathrm{LPF}(g(t)e^{-j\boldsymbol{\omega}_A t})\,;\quad g(t) = W(t)e(t)^* \qquad (11)$$

where c_v is the eigenvector adaptation rate and $g(t)$ is the gradient wave formed by the spectral convolution.

3.2.5 Eigenvalue Update. The eigenvalue may be updated by low-pass filtering the spectral convolution of the error and state waves as follows:

$$\mathrm{d}\,l\,/\mathrm{d}\,t = -c_l\,\mathrm{LPF}(e(t)v(t)^*) \qquad (12)$$

since $e(t)$ and $v(t)$ are coherent and phase-locked, where c_l is the eigenvalue adaptation rate.

4 Band Structure and Anti-Aliasing Constraints

Not all frequencies facilitate associative coding and recall. Carrier frequencies must be separated by integer multiples of Δw, the *beat frequency*. Δw determines bandwidth, noise immunity, and speed of recall. The higher the value of Δw, the higher the noise immunity and allowable speed of recall, but at the expense of increased bandwidth.

4.1 Band Structures

The fine structure of one set of coding bands is illustrated in Fig. 3a and given by the following equations:

$$w_{S1_i} = w_{S1.L} + iN\Delta w\,,\qquad w_{S2_i} = w_{S2.L} + i\Delta w \qquad (13)$$

for i=0,1,…N-1. The diagonal vector is given by

$$\boldsymbol{\omega}_{\mathbf{IL}} = \boldsymbol{\omega}_{\mathbf{S1}} - \boldsymbol{\omega}_{\mathbf{S2}}\,. \qquad (14)$$

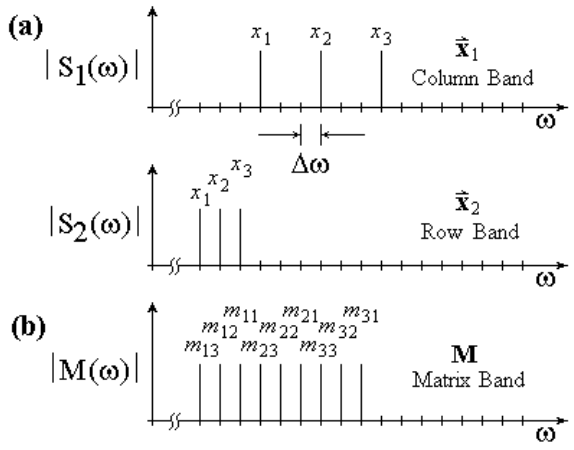

Fig. 3. Fine structure of 3-D AAM spectrums at the coding network: (a) coding bands and (b) transmitted memory wave.

The state and error bands may be placed one Δw higher than the attractor wave,

$$w_{E_i} = w_{V_i} = w_{V.H} - i\Delta w\,, \qquad (15)$$

where $w_{V.H} = B_{S.GAP}+B_{S1}+2B_{S2}+\Delta w$ is the lowest frequency of the state band and $B_{S.GAP}$ is the coding band gap between the end of the second band and the beginning of the first. $B_{S1} = N(N-1)\Delta w$ and $B_{S2} = (N-1)\Delta w$ are the widths of the respective coding bands. The analysis vector is given by

$$\boldsymbol{\omega}_{\mathbf{A}} = \boldsymbol{\omega}_{\mathbf{V}} - \boldsymbol{\omega}_{\mathbf{IL}}\,. \qquad (16)$$

The fine structure of these decoding bands, as well as the matrix-bands are shown in Fig. 4.

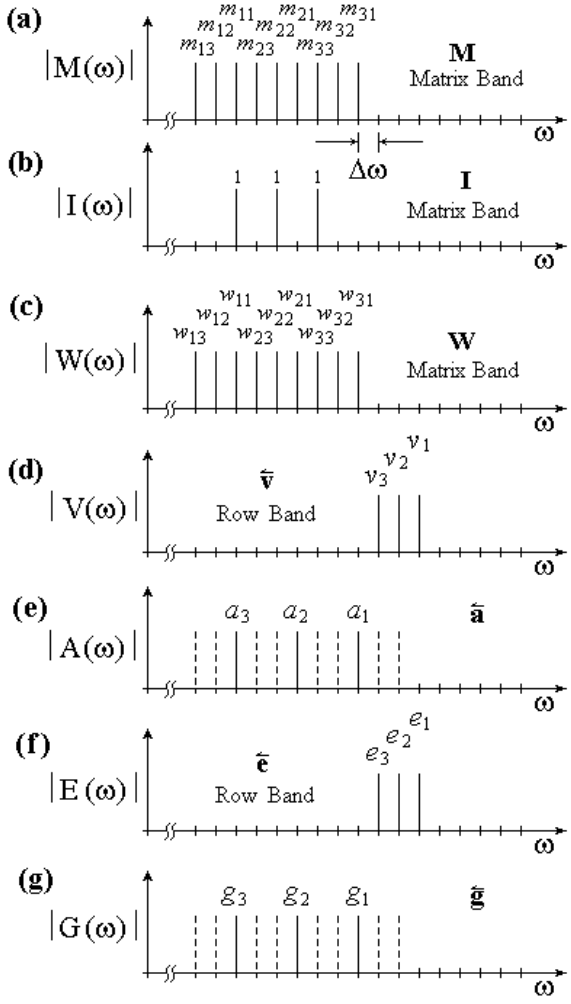

Fig. 4. Fine structure of 3-D AAD spectrums at the decoding network: (a) received memory wave, (b) identity wave, (c) attractor wave, (d) state wave, (e) activation wave, (f) error wave and (g) gradient wave.

4.2 Anti-Aliasing Constraints

In complex (C) and quadrature (I/Q) formulations of spectral associative memory, upper-sidebands of mixing operations cancel out and all energy is focused into lower-sidebands. As a result, there are no anti-aliasing constraints to be satisfied other than the well-known Nyquist rate for sampled-data implementations. On the other hand, for in-phase (I) formulations in which the complex exponentials are replaced by cosines, upper-sidebands do not cancel out and aliasing must be avoided. Thus, the anti-aliasing constraints provided in this section give the worst case. Thus, they also work for C and I/Q formulations.

Two anti-aliasing constraints must be satisfied for in-phase coding. First, lower bounds must be placed on the lowest coding frequency to prevent sideband aliasing by the USB of the attractor wave. Second, the width of the coding band gap must be wide enough to allow for alias-free mixing in the decoder. The coding band gap and lowest coding frequency may be expressed as $B_{S.GAP} = \text{ceil}(N/2)\Delta w$ and $w_{S2.L} = (N^2+1)\Delta w$, respectively, which allow the calculation of all other frequency vectors.

Consider a 7-D example. The second coding band would start at $w_{S2.L} = 50\Delta w$ and the band gap would be $4\Delta w$. Coding vectors would be $\boldsymbol{\omega}_{\mathbf{S1}}$=[60 67 74 81 88 95 102]$^{\mathrm{T}}\Delta w$, $\boldsymbol{\omega}_{\mathbf{S2}}$=[50 51 52 53 54 55 56]$^{\mathrm{T}}\Delta w$, and $\boldsymbol{\omega}_{\mathbf{IL}}$=[10 16 22 28 34 40 46]$^{\mathrm{T}}\Delta w$. The highest state frequency would be $w_{V.H} = B_{S.GAP}+B_{S1}+2B_{S1}+\Delta w = 59\Delta w$, leading to $\boldsymbol{\omega}_{\mathbf{V}} = \boldsymbol{\omega}_{\mathbf{E}}$= [59 58 57 56 55 54 53]$^{\mathrm{T}}\Delta w$ and $\boldsymbol{\omega}_{\mathbf{A}}$=[49 42 35 28 21 14 7]$^{\mathrm{T}}\Delta w$.

5 Simulations

5.1 Dual-Pattern Example

Simulations were performed for 6 x 6 = 36-dimensional gray-level images, with an additional channel for preserving pattern intensity. Two gray-level patterns were continuously encoded and transmitted in parallel, shown in Fig. 5. The results of two simulations demonstrating different modes of addressability are shown in Fig. 6 and Fig. 7. In Fig. 6, the decoding network was provided with noisy initial conditions biased toward pattern *A* and the network converged to pattern *A*, demonstrating content-addressability. In Fig. 7, the network was provided with noisy initial conditions biased toward pattern *A*, but fixed eigenvalue address of pattern *B*, and the network converged to pattern *B*, demonstrating address-addressability.

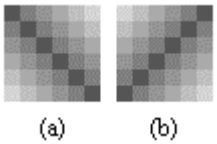

Fig. 5. Encoded 6 x 6 gray-level memory patterns *A* and *B*.

First order filters were employed and the bilinear transform [7] was used to map to discrete time. The beat frequency was Δf = 1kHz, the computational step time was 50 ns, and snapshots of the state vector were taken 40ms apart. Adaptation rates c_v and c_l were adjusted empirically until the transient response was slightly underdamped.

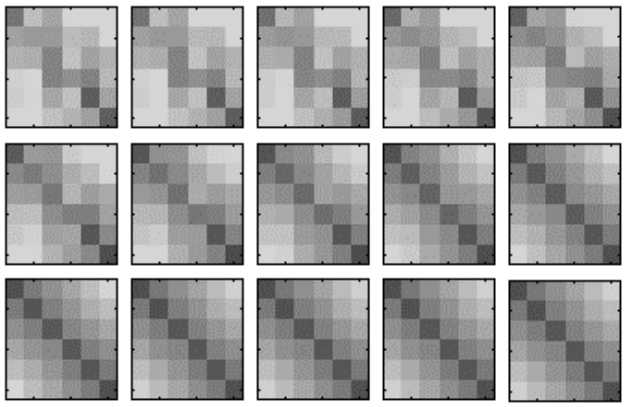

Fig. 6. Transient response of 37-D network showing content-addressibility. Provided with noisy initial conditions biased in favor of pattern *A*, the network converged to pattern *A*.

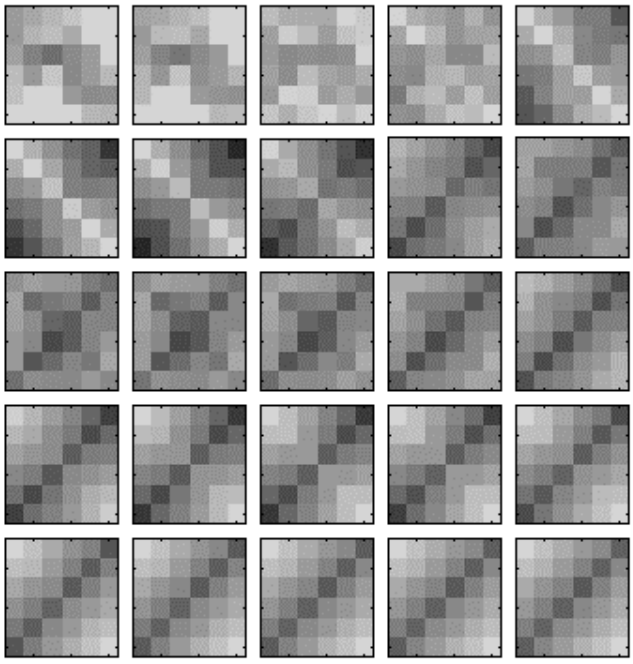

Fig. 7. Transient response of a 37-D network demonstrating address-addressibility. Provided with noisy initial conditions biased in favor of pattern *A*, but fixed eigenvalue address of pattern *B*, the network converged to pattern *B*.

6 Probabilistic Formulation

From an empirical perspective one may describe the encoding process as a biasing of recall in favor of one or more memory patterns for random initial conditions. This

forms a biased probability space, where the basis states are the encoded patterns. Using Dirac notation, in which "kets" are used to represent states, the biased probability state $|\Psi\rangle$ may be expressed as a weighted sum of encoded states:

$$|\Psi\rangle = \sum_{i=1}^{P} c_p |\Psi_p\rangle; \quad c_p = r_p e^{jj_p} \tag{17}$$

where c_p are complex weighting factors in which the phase j_p has been assumed to be zero until now, but may be non-zero due to the wave formulation.[1] The probability p_j of recalling a particular pattern j is calculated from the complex coefficients as follows:

$$p_j = |c_j| \left(\sum_{p=1}^{P} |c_p| \right)^{-1}. \tag{18}$$

Thus, due to the orthonormal coding and conservation of probability, multiplying the decoder possibility state by a real scalar r, does not change the probability of recall, assuming stability is maintained in the decoder:

$$r|\Psi\rangle \Leftrightarrow |\Psi\rangle \tag{19}$$

Interestingly, the encoding of patterns leads to a biased recall probability state in which the observed value of one channel is not independent of the other channels and cannot be factored. For example, coding a single 3-D pattern in which the pattern is $\mathbf{x} = [x_1\ x_2\ x_3]^T$ leads to the state:

$$|\Psi\rangle = \frac{1}{2}|x_1\rangle|x_2\rangle|x_3\rangle + \frac{1}{2}|-x_1\rangle|-x_2\rangle|-x_3\rangle \tag{20}$$

because the negative of an eigenvector is still an eigenvector and the two patterns are equally likely. More importantly, this expression cannot be factored into a product of other ket vectors and implies a kind of "virtual entanglement"; however, the memory wave which facilitates this state in the *a priori* formulation *is* factorable, and is actually formed by a tensor product as given in (6). This result is carried over from the conventional matrix-vector formulation; there is nothing unique about the spectral formulation other than the fact that computational entanglement is achieved in the frequency domain.

7 Conclusions

Multi-pattern, real-valued spectral associative memories were presented that allow the coding and recall of multiple analog data patterns over noisy channels. A complex-exponential formulation was presented in which upper-sidebands naturally cancel out and all available energy is spread over the lower-sideband of the memory waves. Due to the spectral representation of attractors, connectivity is made virtually in the frequency domain and real-valued SAMs scale linearly like bipolar SAMs. A notion of "entanglement" is implied by the *a posteriori* formulation - a degree of interconnectedness to which coding gain and noise immunity may be attributed.

Real-valued SAMs may be thought of as multi-channel coder/decoders (CODECs) and modulator/demodulators (MODEMs) due to the built-in extrinsic redundancy provided at the level of modulation. For single patterns, the decoding network may be simplified by eliminating eigenvalue adaptation, error analysis, and error synthesis, and the activation wave may serve as the gradient wave. Because spectral attractors disappear when transmission of the memory waves ceases, a stream of single patterns may be transmitted over a noisy channel with increased SNR. For multiple patterns, multiple "video" streams may be transmitted simultaneously over the same noisy channel, and multiple decoding networks tuned to the same decoding frequencies may recall different video streams based on address or content.

[1] Unlike the conventional formulation, like-patterns can destructively interfere if the corresponding memory waves are out of phase.

Acknowledgements

I would like to acknowledge Jack Sarfatti, whose physical theories of consciousness triggered many of the key ideas in this work. I also thank Dan Ventura and Mitja Perus for meaningful discussions on the role of entanglement in associative memory.

References

[1] R. Spencer, "Bipolar spectral associative memories," *IEEE Trans. on Neural Networks*, Vol. 12, No. 3, May 2001.

[2] R. Spencer, "Nonlinear spectral associative memories: Neural encoders and decoders for digital communications," In: *Intelligent Engineering Systems Through Artificial Neural Networks*, C.Dagli, A. Buczac, J.Ghosh, M.Embrechts, O.Ersoy, and S.Kercel (Eds), ASME Press, Vol 10, pp 971-976, 2000.

[3] R. Spencer, "Nonlinear heteroassociative spectral memories: Virtually partitioned neural encoders and decoders for digital communications," *Proceedings of the 2000 ICSC Symposia on Intelligent Systems & Applications (ISA 2000)*, F. Naghdy, F. Kurfess, H. Ogata, E. Szczerbicki, H. Bothe, and H. Tlanfield (eds.), Wollongong, Australia, ICSC Academic Press, Dec, 2000.

[4] D. Bohm and B. J. Hiley, *The Undivided Universe*, Routledge, Mar. 1995.

[5] A. Mondragon, R. Carvajal, J. Pineda de Gyvez, and E. Sanchez-Sinencio, "Frequency domain intrachip communication schemes for CNN," *5th IEEE Int'l Workshop on Cellular Neural Networks & Their Applications*, London, England, Apr., 1998.

[6] P.A.M. Dirac, *The Principles of Quantum Mechanics*, Fourth Ed., Oxford University Press Inc., New York, Feb. 1982.

[7] R. Schaumann and M.E. Van Valkenburg, *Design of Analog Filters*, Oxford University Press, New York, 2001.