General Sequential Episodic Memory Model

Arjun Karuvally 1  Terrence J. Sejnowski 2  Hava T. Siegelmann 1

Abstract
The state-of-the-art memory model is the General Associative Memory Model, a generalization of the classical Hopfield network. Like its ancestor, the general associative memory has a well-defined state-dependent energy surface, and its memories correlate with its fixed points. This is unlike human memories, which are commonly sequential rather than separated fixed points. In this paper, we introduce a class of General Sequential Episodic Memory Models (GSEMM) that, in the adiabatic limit, exhibit a dynamic energy surface, leading to a series of metastable states capable of encoding memory sequences. A multiple-timescale architecture enables the dynamic nature of the energy surface with newly introduced asymmetric synapses and signal propagation delays. We demonstrate its dense capacity under polynomial activation functions. GSEMM combines separate memories, short and long sequential episodic memories, under a unified theoretical framework, demonstrating how energy-based memory modeling can provide robust and scalable memory systems in static and dynamic memory cases.

Introduction
Episodic memory refers to the conscious recollection of facts or subjective past experiences and forms an essential component of long-term memory (Tulving, 2002; Duff et al., 2019; Renoult et al., 2019). The recollection process may have both singleton and sequence characteristics. Singleton retrieval is the associative recall of a single memory from a retrieval cue. This memory could be the description of a particular object of interest or important dates of events. Sequence retrieval leads to a recollection process that is not just a single memory but a trajectory of sequentially connected memories. Memories organized into these trajectories are called episodes. Memories may come together in episodes allowing us to link and retrieve sometimes distinct and representationally unrelated memories. The Sequence Episodic Memory (SEM) problem in Recurrent Neural Networks (RNNs) pertains to creating and manipulating these memories and their sequential relationships by encoding relevant information in some form in the synapses.

The energy paradigm plays a major role in singleton episodic memory modeling. The energy paradigm for memory was introduced by Hopfield (Hopfield, 1982; Amari, 2004), who defined energy as a quadratic function of the neural activity in symmetrically connected networks with binary neurons. A single memory is stored as a local minimum of the energy surface. The network states update such that it converges to one of the local minima and retrieves a stable activity state representing a single memory. The Hopfield network model has subsequently been generalized along two directions.

The first direction focuses on memory capacity. Capacity relates to the number of neurons required in the ensemble to store and retrieve a certain number of memories without corruption. The capacity of the original Hopfield model was 14% of the number of neurons, a small fraction of the number of neurons in the population (McEliece et al., 1987; Folli et al., 2016; Amit et al., 1985). A significant breakthrough in capacity came with the introduction of Dense Associative Memory (Krotov & Hopfield, 2016), which introduced a polynomial non-linearity to separate the contribution of each memory to the energy minimum. The non-linearity enabled the models to store more memories than the number of neurons (hence the term dense). Further studies extended these ideas to continuous state spaces, and exponential memory capacity (Demircigil et al., 2017). Currently, these models form the fundamental components of transformer architectures (Vaswani et al., 2017; Ramsauer et al., 2021) with high levels of performance on large-scale natural language processing tasks (Radford et al., 2018; Devlin et al., 2019) and computer vision (Carion et al., 2020) tasks. Recently, General Associative Memory Model (GAMM) (Krotov & Hopfield, 2021) unified these advances in associative mem-
Figure 1: System architecture and schematic retrieval process for the General Sequential Episodic Memory Model (GSEMM)

A The two-layer neural architecture with neural connectivity of GSEMM. The new delay-based synapses (Φ) we introduced (shown in red) are directed connections between neurons in the hidden layer of GAMM. Dotted curved lines in the figure indicate one-to-many connections.

B The schematic representation of how the network performs its computations. The new synapses create a delayed signal of the feature layer neurons, which is provided as input to the hidden layer.

C The typical retrieval process for GAMM with three stored memories (ξ₁, ξ₂, ξ₃). The energy surface is shown by the colormap with high energy denoted by red and low energy blue. The system (shown by the white ball) flows to the basin corresponding to the nearest attractor. Due to the energy surface’s static nature, GAMM stays in a low-energy memory. Consequently, the system cannot retrieve more than one memory (ξ₃ in the figure).

D Similar retrieval process for GSEMM. Unlike GAMM, the system changes the energy surface so that a new minimum is formed that connects to a sequentially related memory (ξ₃ → ξ₁ in the figure). The dynamic nature of the energy surface allows the system to adapt to the new minimum under the condition that the changes in the energy surface are adiabatic to changes in state. This feature of GSEMM enables it to retrieve more than one memory organized in sequence.
ory in a single theoretical framework. GAMM succeeded in explaining the capacity improvements through a simple energy function that characterized the long-term behavior of these models just like its predecessors. However, GAMM’s state-parameterized fixed energy surface restricts it to singleton memories.

The second research direction focuses on extending energy-based models to handle sequences. In contrast to memories in singleton episodic memory, sequence memories are meta-stable states in the dynamical evolution (Rabinovich et al., 2008; Durstewitz & Deco, 2008; Camera et al., 2019). Some of the early works to produce sequential meta-stable memory (Kleinfeld, 1986; Sompolinsky & Kanter, 1986) used a combination of symmetric interactions, asymmetric interactions, and delay signals to produce stable sequential activation of memory patterns. However, these models required additional mechanisms to selectively raise the energies of states, which added complications to the use of the energy paradigm and demonstrated the difficulty of reconciling the static nature of the energy surface with the dynamical nature of models required for sequential memory retrieval. One way to alleviate this difficulty is the introduction of stochasticity (Miller & Katz, 2010; Jones et al., 2007) with sufficient noise to push the system’s state beyond the basin of memory to another memory (Miller, 2016; Braun & Mattia, 2010). Models developed along these directions relaxed the symmetric constraints on the neural interactions of Hopfield Networks, resulting in a rich repertoire of dynamics (Asllani et al., 2018; Orhan & Pitkow, 2020). Theoretical proposals for meta-stable memory models used non-equilibrium landscapes where the energy function and a probability flux together determined the stability of memory states (Yan et al., 2013). In these models, stochasticity played a major role in determining the stability of meta-stable states. In contrast to the theoretical models, evidence from biology (Howard et al., 2014; Rolls & Mills, 2019; Umbach et al., 2020) has emphasized the importance of multiple timescales in SEM tasks. Empirical models (Kurikawa & Kaneko, 2021; Kurikawa, 2021) also use multiple timescales to generate SEM. However, current energy models do not take advantage of multiple timescales, and the SEM capacity is only about 12% of the number of neurons (Kurikawa & Kaneko, 2021).

In the GSEMM, we extend the static energy paradigm of the GAMM to the dynamic case using two timescales to define the dynamic behavior of the model. In the process, we discover mechanisms that significantly improve the memory capacity of sequence networks.

**General Sequential Episodic Memory Model (GSEMM)**

The General Sequential Episodic Memory Model unifies singleton and sequence memory retrieval. To facilitate this unification, we provide the mathematical description of GSEMM as a two layer system of interacting neurons organized according to the General Associative Memory Model (GAMM) (Krotov & Hopfield, 2021) with the addition of delay based intra-layer interactions between neurons in the hidden layer. The feature layer is mainly concerned with the input and output of the model. There are no synaptic connections between neurons in this layer. The hidden layer encodes abstract information about stored memories. Unlike the feature layer, the hidden layer neurons are connected using synapses that delay the signal from the feature layer. These intra-hidden layer connections enable interactions between memories. In the most general case, there is no restriction on the nature (in terms of symmetry) of the connections between neurons in this layer. In addition to these intra-layer connections, the neurons in the two layers are connected through symmetric synaptic interactions. The architecture for the model is shown in Figure 1B.

We now provide the mathematical description of GSEMM using the notations summarized in Table 1. We use standard linear algebra notations and indexed notations to denote states and synapses in our model. We use mainly indexed notation but switch to matrix and vector notation wherever convenient. Let $(V_f)_i$ be the current through the $i^{th}$ neuron of the feature layer, $\sigma_f(V_f)$ be the activation function for the feature layer, $(V_h)_j$ be the current through the $j^{th}$ neuron of the hidden layer, $\sigma_h$ be the activation function for the hidden layer, and $(V_d)_i$ be the delayed feature neuron signal from the $i^{th}$ feature neuron. The states $V_f, V_h, V_d$ evolve with characteristic timescales $T_f, T_h, T_d$ respectively. Let $\Xi_{ij}$ be the strength of the synaptic connection between the neuron $i$ in the feature layer to the neuron $j$ in the hidden layer, $\Phi_{kj}$ be the strength of the synaptic connection from the $k^{th}$ hidden neuron to the $j^{th}$ hidden neuron. Similar to how memories are loaded in GAMM, each column of the matrix $\Xi$ stores individual memories. We introduce two scalar parameters to control the strength of signals through the synapses. Let $\alpha_s, \alpha_c$ be the strength of signals through the synapses $\Xi$ and $\Phi$ respectively. The governing dynamics
are given by:

$$
\begin{align*}
\mathcal{T}_f \frac{d(V_f)_i}{dt} &= \sqrt{\alpha_s} \sum_{j=1}^{N_f} \Xi_{ij} (\sigma_h(V_h)_j)_{i} - (V_f)_i, \\
\mathcal{T}_h \frac{d(V_h)_i}{dt} &= \sqrt{\alpha_c} \sum_{i=1}^{N_h} \Xi_{ij} (\sigma_f(V_f)_j)_{i} + \\
\alpha_c \sum_{k=1}^{N_h} \sum_{i=1}^{N_f} \Phi_{kj} \Xi_{ik} (V_d)_i - (V_h)_j, \\
\mathcal{T}_d \frac{d(V_d)_i}{dt} &= (\sigma_f(V_f))_{i} - (V_d)_i.
\end{align*}
$$

(1)

The dynamic evolution equations are analogous to GAMM (Krotov & Hopfield, 2021) except with the addition of delayed in-layer synapses \( \Phi \), and two strength parameters \( \alpha_s \) and \( \alpha_c \). The timescale \( \mathcal{T}_d \) characterizes the timescale of delay and is assumed to be higher than the timescale of the feature and hidden layers. The delay signal is obtained by applying a continuous convolution operator (Kleinfield, 1986) of the feature layer signal.

\[
(V_d)_i = \frac{1}{\mathcal{T}_d} \int_0^\infty (\sigma_f(V_f(t-x)))_i \exp\left(-\frac{x}{\mathcal{T}_d}\right) dx. 
\]

(2)

We transformed the convolution operation to a dynamical state variable update \( \frac{dV_d}{dt} \) to simplify the theoretical analysis of the system.

Without intra-layer synapses in the hidden layer, the GSEMM has properties of associative memory. This means that for certain conditions on the set of functions \( \sigma_f \) and \( \sigma_h \), the long-term behavior of the state of the feature layer neurons converged to one of the stored memories. The necessary condition for convergence is that the dynamical trajectory of the system follows an energy function with minima near the stored memory states. The delay-based synapses we introduced enable the energy function to change with time, so the long-term behavior is not just a single memory but a sequence of related memories.

**Energy Dynamics**

The energy dynamics of the system is analyzed by considering the new delay variable \( V_d \) as a control parameter. We show that for a delay signal \( V_d \) that is changing sufficiently slowly compared to \( V_f \) and \( V_h \), the energy function evaluated at the instantaneous state \( V_d \) can still be used to characterize the dynamical nature of \( V_f \) and \( V_h \). The term sufficiently slowly means that \( V_f \) and \( V_h \) converge to their instantaneous attractor states before \( V_d \) changes the energy surface. To derive the energy function, we use two Lagrangian terms \( L_f \) and \( L_h \) for the feature and hidden neurons respectively (Krotov & Hopfield, 2021), defined as

\[
\sigma_f(V_f) = \mathcal{J}(L_f)^T, \quad \text{and} \quad \sigma_h(V_h) = \mathcal{J}(L_h)^T. 
\]

(3)

The new energy function (Appendix A.1) for GSEMM is derived as

\[
E = \left[ V_f^T \sigma_f(V_f) - L_f \right] + \left[ V_h^T \sigma_h(V_h) - L_h \right] \\
- \sqrt{\alpha_s} \left[ \sigma_f(V_f) \Xi \sigma_h(V_h) \right] - \alpha_c \left[ V_f^T \Xi \Phi \sigma_h(V_h) \right].
\]

(4)
At this point, it is instructive to note that without the additional synapses, \( \Phi \), and setting the strength parameter \( \alpha_s = 1 \), the system and the associated energy function reduce to GAMM energy with only singleton episodic memory.

In order to analyze how the dynamics of energy change with the introduction of delay based synapses, we take the time derivative of the GSEMM energy function along the dynamical trajectory of the system. We assume the conditions of positive semi-definite Hessians of the Lagrangian terms and bounded activation functions \( \sigma_f \) and \( \sigma_h \) (Krotov & Hopfield, 2021). It is to be noted that the entire state description of the system consists of three vectors \( \{V_f, V_d, V_h\} \). These states are grouped as a fast subsystem \( V_f \) and \( V_h \), and a slow subsystem \( V_d \). The analysis becomes easier when we consider the slow subsystem as a control variable of the fast subsystem. This allows the characterization of the state dynamics of the fast subsystem as instantaneous fixed point attractor dynamics modulated by input from the slow subsystem.

The dynamical evolution of the energy function after separating the slow and fast subsystems is given as (Appendix A.2),

\[
\frac{dE}{dt} = F(\frac{dV_f}{dt}, \frac{dV_h}{dt}) + G(\frac{dV_d}{dt}).
\]

\[
F(\frac{dV_f}{dt}, \frac{dV_h}{dt}) = -\left[ T_f \left( \frac{dV_f}{dt} \right)^\top \mathcal{H}(L_f) \frac{dV_f}{dt} + T_h \left( \frac{dV_h}{dt} \right)^\top \mathcal{H}(L_h) \frac{dV_h}{dt} \right].
\]

\[
G(\frac{dV_d}{dt}) = -\alpha_c \left[ \sigma_h(V_h)^\top \Phi^\top \Xi^\top \frac{dV_d}{dt} \right].
\]

\( F \) and \( G \) separate the contributions of the two timescales - the fast \( \{(T_f, T_h)\} \) and slow \( \{T_d\} \). It can be easily seen that among the two terms, only \( G \) is affected by the timescale of the delay signal. Just like in GAMM, under the assumption of positive semi-definite Hessian of the Lagrangian and bounded energy, we get,

\[
F(\frac{dV_f}{dt}, \frac{dV_h}{dt}) \leq 0.
\]

The inequality means that the fast subsystem can have two possible long-term behaviors when \( F \) eventually converges to zero. One behavior is convergence to a single stable state corresponding to minima of the energy function leading to fixed point attractor dynamics. The second possible behavior is when the system moves in an iso-energetic trajectory without convergence. In this paper, we focus only on the case of the fixed point attractor behavior of the system.

Like in GAMM, the fixed point attractor behavior of the system acts to stabilize the dynamics on the energy surface such that the energy is non-increasing and convergent, but unlike GAMM, delay based synapses lead to another term \( G \).

\[
G(\frac{dV_d}{dt}) = -\alpha_c \left[ \sigma_h(V_h)^\top \Phi^\top \Xi^\top \frac{dV_d}{dt} \right].
\]  

It may be difficult to specify the system’s behavior for any general choice of \( \Phi \), \( \sigma_h \), and \( V_d \). However, in the adiabatic limit of the slow subsystem (under the condition that \( T_d \gg 1 \) and \( T_d \gg T_f, T_h \)), the system can still exhibit a non-increasing energy function because \( \frac{dV_d}{dt} \rightarrow 0 \) in this limit and \( G \rightarrow 0 \). This condition is especially true when analyzing the dynamic properties of the fast subsystem \((\frac{dV_f}{dt} \neq 0 \) and \( \frac{dV_h}{dt} \neq 0 \)), which is the property that seems to be relevant in dynamic memory models. The delay signals thus have two functions. The slow-changing nature of the delay signal helps to stabilize the dynamics of the fast subsystem on the energy surface. The second function is that the delay signal changes the energy surface to create new minima and destroy old minima. In our numerical simulations, we consider high enough settings of \( T_d \) such that \( V_d \) changes sufficiently slowly for the energy function to characterize the dynamics but not so high as to prevent the system from exhibiting state transitions in a reasonable time. It is to be noted that although we consider a specific case for the delay signal, the adiabatic theory is general enough to apply to any adiabatic control signal to the previously introduced general associate memory model (GAMM).

**Dense GSEMM**

We next derive the activation functions which will make GSEMM dense and thus most useful both as human memory and for AI. Analogous to how practical models are derived from GAMM, we consider the diabatic limit of hidden neurons, \( T_h \rightarrow 0 \), such that

\[
(V_h)_j = \sqrt{\alpha_s} \sum_{i=1}^{N_f} \Xi_{ij} (\sigma_f(V_f))_i + \alpha_c \sum_{k=1}^{N_h} \sum_{i} \Phi_{kj} \Xi_{ik} (V_d)_i.
\]

Substituting this in the dynamical evolution of feature neurons we get,

\[
T_f \frac{d(V_f)_i}{dt} = \sqrt{\alpha_s} \sum_{j=1}^{N_h} \Xi_{ij} \sigma_h(\Xi^\top \sigma_f(V_f))_j + \alpha_c \Phi^\top \Xi^\top (V_d)_j - (V_f)_i.
\]
It can be seen from the dynamical evolution of the feature neurons that depending on the setting of \( \sigma_h \), the feature-hidden synapses may interact linearly with hidden-feature synapses and hidden-hidden synapses. In Dense GSEMM, we choose the feature layer activation function to be the tanh non-linearity \( \sigma_f(x) = \tanh(x) \), and the hidden layer activation to be the polynomial activation function of degree \( n \) \( \sigma_h(x) = x^n \) for \( n \geq 0 \). When \( n = 1 \), the parametric model has linear memory interactions, and \( n > 1 \) progressively increases the order of these interactions. These settings are analogous to polynomial interactions in associative memory models that improved their capacity significantly (Krotov & Hopfield, 2016).

According to the theoretical analysis above, the system follows the instantaneous minima of the energy function and flows from memory to memory via the slow updates to the energy function. We validate this for the case of \( n = 1 \) with simulation in Figure 2. We plot the energy function and the state of the system as time progresses for a simulated episode of the system. The momentary loss in stability of fixed points near memories that allow for state transition can be clearly observed in the figure. The addition of the adiabatic control signal provides a way to adaptively change the energy surface once the system converges so that dynamic properties can emerge from static energy models.

In Figure 3, we show that increasing the polynomial degree of the hidden layer activation increases the sequence memory capacity, similar to what is observed in associative memory models with polynomial activations. This indicates that the proposed polynomial non-linearity in GSEMM can effectively improve the capacity of sequence memories.

**Discussion**

The General Sequential Episodic Memory Model (GSEMM) introduced in this paper is an approach to encoding memories along with their sequential relationships. The key to this capability is the slow-fast timescale dynamics created by adding delay-based synapses in the hidden layer of the General Associative Memory Model. These delay-based
Figure 3: Memory capacity grows with the polynomial degree in the hidden layer. The plot shows the distribution of model capacities in 100 neuron network for different sets of random binary vectors along with the mean of the distributions (blue line) for 30 trials. The distributions can be compared against the baseline of 12 length sequence (black horizontal line). The baseline performance seems to be maximum number of memories that can be stored in a polynomial degree=1 Dense GSEMM. It can be seen that for polynomial degree=15, there are models (with the highest sequence length) capable of storing more than 100 memories making the SEM capacity dense in the number of neurons.

Acknowledgements
This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Agreement No. HR00112190041. We thank Peter Delmas-trio for valuable insights leading to the development of the model. We thank Edward Rietman, Andre Pietrzykowski, Joshua Russel, Adam Kohan, and the reviewers for valuable feedback to improve the manuscript.

References


A. Energy

The most important aspect of the model we discussed is the energy function. We use the function to show the behavior of the system in the adiabatic case and compute instantaneous attractors.

A.1. Energy Function for GSEMM

Here, we will derive the Energy function of GSEMM starting from a previously derived energy function used for associative memory. Assume a signal $\mathcal{I}_h$ applied to the neurons in the hidden layer.

$$E = \left[V_f^T \sigma_f(V_f) - L_f\right] + \left[(V_h - \mathcal{I}_H)^T \sigma_h(V_h) - L_h\right] - \left[\sqrt{\alpha_s} \sigma_f(V_f) \Xi \sigma_h(V_h)\right]$$

(11)

In our case, the input signal comes from the delay signal activity $V_d$ and is given as $\mathcal{I}_h = \alpha_c \Phi^\top \Xi^\top V_d$ from our governing dynamics. Substituting this in the energy equation

$$E = \left[V_f^T \sigma_f(V_f) - L_f\right] + \left[(V_h - \alpha_c \Phi^\top \Xi^\top V_d)^T \sigma_h(V_h) - L_h\right] - \left[\sqrt{\alpha_s} \sigma_f(V_f) \Xi \sigma_h(V_h)\right]$$

(12)

Expanding this equation, we get

$$E = \left[V_f^T \sigma_f(V_f) - L_f\right] + \left[V_h^T \sigma_h(V_h) - L_h\right] - \left[\sqrt{\alpha_s} \sigma_f(V_f) \Xi \sigma_h(V_h)\right] - \alpha_c \left[V_d^T \Xi \Phi \sigma_h(V_h)\right]$$

(13)

A.2. Energy Function Dynamics

To find how the energy function behaves along the dynamical trajectory of the system. Taking the derivative of the energy function with respect time

$$\frac{dE}{dt} = \left[V_f^T J(\sigma_f) \frac{dV_f}{dt} + \sigma_f(V_f)^T \frac{dV_f}{dt} - \frac{dL_f}{dt}\right] + \left[V_h^T J(\sigma_h) \frac{dV_h}{dt}\right] + \left[\sqrt{\alpha_s} \sigma_f(V_f) \Xi J(\sigma_h) \frac{dV_h}{dt}\right]$$

(14)

$$+ \left[\sqrt{\alpha_s} \sigma_h(V_h) \Xi^T J(\sigma_f) \frac{dV_f}{dt}\right] - \alpha_c \left[V_d^T \Xi \Phi J(\sigma_h) \frac{dV_h}{dt} + \sigma_h(V_h)^T \Xi^T \Phi^T \frac{dV_d}{dt}\right]$$

The derivatives of the lagrangian terms can be converted as $\frac{dL_f}{dt} = \sigma_f(V_f)^T \frac{dV_f}{dt}$ and $\frac{dL_h}{dt} = f(V_h)^T \frac{dV_h}{dt}$. Substituting these.

$$\frac{dE}{dt} = \left[V_f^T J(\sigma_f) \frac{dV_f}{dt} + V_f^T J(\sigma_h) \frac{dV_h}{dt}\right] - \left[\sqrt{\alpha_s} \sigma_f(V_f) \Xi J \frac{dV_h}{dt}\right]$$

$$+ \left[\sqrt{\alpha_s} \sigma_h(V_h) \Xi^T J(\sigma_f) \frac{dV_f}{dt}\right] - \alpha_c \left[V_d^T \Xi \Phi J(\sigma_h) \frac{dV_h}{dt} + \sigma_h(V_h)^T \Xi^T \Phi^T \frac{dV_d}{dt}\right]$$

(15)
We simulated using this procedure, multiple instantiations of Dense GSEMM with polynomial degrees ranging from 0.

We used a fixed point finding algorithm to find the fixed points of the energy surface in Figure 1. The algorithm uses an iterative process to find the fixed points of the energy surface evaluated from a certain point in the state space. Starting from the neuron state on the energy landscape, the state is updated to follow the direction of the energy gradient till no more updates are possible, indicating convergence to a fixed point on the energy surface. This fixed point is also one meta-stable point in the network dynamics since it loses stability over time.

B. Simulation Procedures - Figure 1

We used the fourth order Runge-Kutta numerical procedure with step size 0.01 for numerical simulations. The output for the Dense GSEMM is the state of the feature neurons. The similarity between the output and memory is evaluated using the overlap of the feature neuron state with each memory in the system, which is defined as $m_i(V_f) = (1/N_f) \sum_i^N_j (\xi^{j(i)}_i) \sigma_f(V_f)$, where $\xi^{j(i)}_i$ is the $i^{th}$ memory in the system. Each memory in the model is a random binary vector such that $\Pr(\xi^{j(i)}_i = +1) = \Pr(\xi^{j(i)}_i = -1) = 1/2$. These memories are organized as 2 separate cyclical episodes: $\xi_1 \rightarrow \xi_2 \rightarrow \xi_3 \rightarrow \xi_1$ and $\xi_4 \rightarrow \xi_5 \rightarrow \xi_6 \rightarrow \xi_7 \rightarrow \xi_4$ with their sequential relationships stored as an adjacency matrix in $\Phi$. Two key factors were considered when we used these two episodes for evaluation. One factor is to demonstrate the ability of the model to extract only the related stored episode, even in the presence of other episodes. The second factor is that successfully generating the stored episode requires long-term non-equilibrium behavior with no stable states in the dynamics. The input to the system is the memory $\xi_5$ which retrieves the 4 length cycle.

B.1. Parameter Settings

We simulated Dense GSEMM with $N_f = 100, \alpha_s = 0.05, \alpha_c = 0.007, T_f = 1.0, T_d = 20.0$. The power of the polynomial non-linearity is 1 for the energy surface shown in the figure. The code for the simulations are available in the repository: https://github.com/arjun23496/gsemm.

B.2. Fixed point analysis

We used a fixed point finding algorithm to find the fixed points of the energy surface in Figure 1. The algorithm uses an iterative process to find the fixed points of the energy surface evaluated at a certain point in the state space. Starting from the neuron state on the energy landscape, the state is updated to follow the direction of the energy gradient till no more updates are possible, indicating convergence to a fixed point on the energy surface. This fixed point is also one meta-stable point in the network dynamics since it loses stability over time.

C. Simulation Procedures - Figure 3

We used the fourth order Runge-Kutta numerical procedure with step size 0.01 for numerical simulations. Each memory in the model is a random binary vector such that $\Pr(\xi^{j(i)}_i = +1) = \Pr(\xi^{j(i)}_i = -1) = 1/2$. These memories are organized in a single cyclical episode with $K$ memories such that: $\xi_1 \rightarrow \xi_2 \rightarrow \xi_3 \rightarrow \xi_1$ with their sequential relationships stored as an adjacency matrix in $\Phi$. $\Phi$ is thus a circulant matrix with a single cycle.

We simulated using this procedure, multiple instantiations of Dense GSEMM with polynomial degrees ranging from 1 to 15 with 30 different seeds for each polynomial degree. The parameters $N_f = 100, T_f = 1.0, T_d = 20.0, \alpha_s = 1.0$ are fixed. The parameter $\alpha_c \in \{0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$ is treated as hyperparameter and optimized to find the maximum number of memories that can be stored for each [seed, polynomial] pair.
We define that the model successfully retrieves the $K$ size memory cycle if the output of the model can successfully retrieve the memories in the cycle in the correct order at least 3 times with no errors.