UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

CONTINUOUS ENERGY LANDSCAPE MODELS FOR PREDICTING COGNITIVE DECLINE IN BRAIN TUMOR PATIENTS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By Triet Tran
Norman, Oklahoma
2025

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/341819

CONTINUOUS ENERGY LANDSCAPE MODELS FOR PREDICTING COGNITIVE DECLINE IN BRAIN TUMOR PATIENTS

A DISSERTATION APPROVED FOR THE

GALLOGLY COLLEGE OF ENGINEERING

BY THE COMMITTEE CONSISTING OF

Dr. Sina Khanmohammadi, Chair

Dr. Chao Lan

Dr. Anindya Maiti

Dr. John Jiang

©Copyright by Triet Tran 2025
All Rights Reserved.

Acknowledgements

I would like to express my deepest gratitude to my advisor, Dr. Sina Khanmohammadi, for his constant support, guidance, and encouragement throughout my Ph.D. journey. His mentorship shaped not only the direction of this dissertation but also my growth as a researcher. I am sincerely grateful for his patience, trust, and the many opportunities he has provided for me to explore, learn, and succeed. I would also like to thank my dissertation committee for their valuable insights, constructive feedback, and generous support. Their expertise strengthened my work and helped me expand my knowledge in meaningful ways. I am grateful as well to the faculty and staff in the School of Data Science and Analytics for creating a supportive environment and for providing the resources that made this research possible. I would especially like to acknowledge my lab mates and collaborators, whose discussions, assistance, and encouragement made this work more enjoyable and enriching.

I am deeply indebted to my wife, Quynh, whose unwavering love, patience, and encouragement have been my greatest source of strength. She stood by me through every challenge and every long night of research, and this accomplishment would not have been possible without her. I am also profoundly grateful to my family for their continuous support, sacrifices, and belief in me from the very beginning of my academic journey.

Finally, I would like to thank my friends for their companionship and encouragement throughout these years. Their presence helped me maintain balance and motivation while completing this work.

To all of you, thank you.

ABSTRACT

Large-scale brain activity exhibits metastable dynamics, recurrently visiting a limited set of configurations whose transitions can be summarized by an energy landscape. Most neuroimaging studies estimate this landscape using pairwise maximum-entropy models applied to binarized signals, which discards amplitude information and limits the representation of higher-order interactions. This dissertation introduces two complementary advances to address these limitations. First, it extends the discrete framework to a high-order formulation that aggregates regions into ensembles using k-means clustering and represents interactions among ensembles through exact state enumeration at the group level. Second, it proposes a continuous energy landscape model that operates directly on real-valued time series data, estimating a normalized density that eliminates the need for binarization prior to energy calculation. This approach results in a more expressive model with richer parameterization.

The proposed methods were evaluated on synthetic data generated from Kuramoto oscillator networks and switching linear dynamical systems, assessing their ability to recover basin geometry, transition matrices, and latent state distributions. In addition, a real-world case study was considered using a resting-state functional magnetic resonance imaging (fMRI) dataset, where the proposed energy landscape model was applied to predict postoperative cognitive outcomes, such as working memory, executive function, and reaction time, in patients with brain tumors. Together, these findings underscore the effectiveness and translational potential of the proposed framework for modeling complex brain dynamics.

TABLE OF CONTENTS

1 Introduction and Literature Review **1**

1.1 Motivation and Problem Statement 1

1.2 Preliminaries 3

1.2.1 Intuition Behind Energy Landscape Models 3

1.2.2 Ising-based Energy Landscape Models 5

1.2.3 Limitations Of The Ising-based Energy Landscapes 6

1.3 Prior Work 9

1.3.1 Maximum Pseudo-Likelihood (MPL) 9

1.3.2 Minimum Probability Flow (MPF) 10

1.3.3 Noise-Contrastive Estimation (NCE) 12

1.3.4 Sampling-based Surrogates 13

1.3.5 Normalization Approximations 14

1.4 Overview of the Proposed Work 15

2 High-Order Discrete Energy Landscape (HODEL) **16**

2.1 High-Order Energy Landscape Framework 16

2.1.1 Identifying Functional Clusters 16

2.1.2 Extracting Brain States 17

2.1.3 Fitting Maximum Entropy Model 18

2.1.4 Constructing Energy Landscapes 20

2.2 Experimental Setup 20

2.2.1 Overview 20
2.2.2 Generator A: Switching Linear Dynamical System (SLDS) 21
2.2.3 Generator B: Kuramoto Oscillator Network With Template Locking . 23
2.2.4 Evaluation Metrics And Interpretation 26
2.3 Simulation Results: High-Order Vs. Low-Order
DEL 28
2.3.1 Kuramoto Oscillator Network 29
2.3.2 Switching Linear Dynamical System (SLDS) 29
2.3.3 Computational Complexity And Memory Requirements 30
2.4 Conclusion 31

3 High-Order Continuous Energy Landscape (HOCEL) 32
3.1 Continuous Energy Landscape Framework 32
3.1.1 Continuous Formulation of Energy Landscapes 32
3.1.2 Graph Parametrization of the Precision Matrix 35
3.1.3 Mixture-Based Energy Formulation for Multi-Basin Dynamics 39
3.2 Simulation Results: Continuous versus Discrete 41
3.2.1 Switching Linear Dynamical System 41
3.2.2 Kuramoto Oscillator Network 41
3.3 Conclusion 42

4 Applications in Predicting Cognitive Function in Brain Tumor Patients 44
4.1 Data Description 45
4.2 Prediction Framework 47
4.2.1 Feature Extraction 47
4.2.2 Model Training and Evaluation 48
4.3 Results 49
4.3.1 Working Memory 49

4.3.2 Executive Function 50
4.3.3 Reaction Time 52
4.4 Conclusion 53

5 Discussion **57**

5.1 Proposed Framework and Key Contributions 57
5.2 Practical Applications in Assessing Cognitive Function 59
5.3 Limitations and Potential Future Direction 60

LIST OF TABLES

4.1 Summary of Patient Cohort. 45
4.2 Summary of Cognitive Measures Used in the Study. 46

LIST OF FIGURES

1.1 Computational growth of the discrete MEM configuration space. The horizontal axis shows the number of brain regions N and the vertical axis shows the number of brain configurations on a logarithmic scale. The line is straight on the log scale, reflecting exponential growth. 7
1.2 Loss of amplitude information. The continuous signal (blue) contains distinct positive amplitudes and graded peaks; the binarized trace (orange) collapses all positive values to 1 and all negative values to 0. 9

1.3 Threshold sensitivity. Two plausible thresholds applied to the same continuous signal produce different binary sequences. The higher threshold reduces time labeled as 1, shortens runs of 1s, and moves flip locations toward local maxima, whereas the lower threshold increases duty cycle and merges nearby events. . 10

1.4 MPL schematic. Each region's binary series is regressed on the others to recover one row of W and the bias h_i, where aggregating across regions yields (W, h). Despite its computational efficiency, rare positive labels compress the logistic function's dynamic range and tend to push h_i toward negative values unless class-imbalance remedies are applied. Post hoc symmetrization, which averages W_{ij} and W_{ji}, stabilizes estimates but can attenuate localized effects when node-wise fits disagree. 11

1.5 The Minimum Probability Flow (MPF) objective vs. model parameters. The plot shows how the initial probability flow from data states to their one-bit-flip neighbors relates to the local MPF loss. The loss is minimized in the preferred parameter zone, where the local energy ordering around each observed configuration best matches the data: if the model is too flat, probability flows out of data states (underfitting); if it is too steep, the ordering among neighboring states becomes unstable (overfitting). This local, one-flip focus explains MPF's speed, but also its weakness when important transitions require changing multiple bits at once. 12

1.6 Noise-Contrastive Estimation (NCE) in state space. The plot shows the data distribution $p_{\text{data}}(x)$, the noise distribution $p_{\text{noise}}(x)$, and the learned log-odds $\log\Big(p_{\text{data}}(x)/p_{\text{noise}}(x)\Big)$ evaluated over the same state axis. Broad overlap between p_{data} and p_{noise} places the classifier in a preferred separation zone where the NCE objective is well-conditioned, yielding stable estimates of the energy $E(x)$ and its normalizing constant. When the overlap is too small, the classification problem becomes ill-conditioned, biasing both $E(x)$ and the normalizer. 14

2.1 **High-order discrete energy landscape framework. (A)** *Identifying Functional Clusters:* Resting-state fMRI signals are grouped into functionally similar clusters using k-means algorithm. Each cluster represents a group of brain regions with similar activity. **(B)** *Extracting Brain States:* The average signal from each cluster is converted into binary states (active = 1 and inactive = 0) based on a threshold (e.g. mean value), forming a sequence of brain states over time. **(C)** *Fitting Maximum Entropy Model:* The maximum entropy model is fitted to identify the probability distribution of brain states that maximizes entropy, while ensuring that the model matches the observed average activation of each cluster and the pairwise co-activation between clusters. **(D)** *Constructing Energy Landscapes:* The parameters of the resulting optimum probability distribution is utilized to calculate the energy value of each state. 17

2.2 Comparison of HODEL and LowDEL on Basin Recovery (BR), Transition Matrix Accuracy (TMA), and State Distribution Agreement (SDA) for the Kuramoto simulations. Paired Wilcoxon signed-rank p-values are shown in each panel. 30

2.3 Comparison of LowDEL and HODEL on Basin Recovery (BR), Transition Matrix Accuracy (TMA), and State Distribution Agreement (SDA) for the SLDS simulations. Paired Wilcoxon signed-rank p-values are shown in each panel. 30

3.1 **High-order continuous energy landscape framework. (A)** *Graph Construction:* Standardized time series $\mathbf{x}(t)$ and baseline $\boldsymbol{\mu}$ define the degree-normalized adjacency matrix $\tilde{\mathbf{A}}$. Each data group uses its own $\tilde{\mathbf{A}}$ built from $\mathbf{x}(t)$ with frozen hyperparameters. **(B)** *Embedding Generation:* A frozen trunk maps $\tilde{\mathbf{A}}$ to node embeddings $\mathbf{V}$. A two-layer GCN produces observation embeddings $\mathbf{H}$, and linear heads output latent factors $\mathbf{Z}$ and external field $\boldsymbol{h}$. **(C)** *Energy Landscape:* The precision matrix $\mathbf{S}$ combines latent factors and regularization. The single-well energy depends on deviation from baseline and external field, and its minimum represents the most probable configuration. 33

3.2 SLDS simulations comparing the discrete baseline and HOCEL on Basin Recovery, Transition Matrix Accuracy, and State Distribution Agreement across the simulation grid. Bars summarize paired results across repeats. Paired Wilcoxon p-values shown in-panel are 2.91×10^{-28}, 1.18×10^{-89}, and 5.65×10^{-90}. 41

3.3 Kuramoto simulations comparing the discrete baseline and HOCEL on Basin Recovery, Transition Matrix Accuracy, and State Distribution Agreement across the simulation grid. Bars summarize paired results across repeats. Paired Wilcoxon p-values shown in-panel are 0.0311, 1.11×10^{-13}, and 1.66×10^{-103}. 42

4.1 Working memory: accuracy across pipelines (mean ± SD across leave–one–out folds). HighCEL is highest and HighDEL close behind; their paired difference is not significant by McNemar's test. For both grouped pipelines, accuracy exceeds each network–subset baseline with Benjamini–Hochberg $FDR<0.05$ within the family of comparisons. 50

4.2 Working memory: F_1 across pipelines (mean ± SD across leave–one–out folds). HighCEL matches or slightly exceeds HighDEL; their paired difference is not significant by a subject–level permutation test. Both grouped pipelines exceed the network–subset baselines with adjusted FDR<0.05. 51

4.3 Working memory: ROC–AUC across pipelines (mean ± SD across leave–one–out folds). HighCEL attains the highest AUC, followed closely by HighDEL; both grouped pipelines outperform the network–subset baselines. HighCEL vs. HighDEL is not significant by DeLong's test. 52

4.4 Executive function: accuracy under leave–one–out validation (mean ± SD). HighCEL achieves the highest mean accuracy and significantly exceeds the network–subset baselines after Benjamini–Hochberg correction; the difference vs. HighDEL is not significant. 53

4.5 Executive function: F_1 under leave–one–out validation (mean ± SD). HighCEL attains the top mean F_1 and exceeds two of the three network–subset baselines after adjustment; the comparisons versus Salience and HighDEL are not significant. 54

4.6 Executive function: ROC–AUC under leave–one–out validation (mean ± SD). HighCEL yields the highest mean AUC, followed closely by HighDEL; both outperform the network–subset baselines. The HighCEL–HighDEL difference is not significant by DeLong's test. 55

4.7 Reaction time prediction: absolute error (ms) under leave–one–out cross–validation. Boxes summarize the distribution of per–patient errors. Paired Wilcoxon tests favor HighCEL over each comparator after Benjamini–Hochberg correction. Lower is better. 55

4.8 Reaction time prediction: per–subject R^2 contributions (leave–one–out). Boxes aggregate patient–level contributions to explained variance. Paired Wilcoxon tests favor HighCEL over each comparator after Benjamini–Hochberg correction. Higher is better. 56

Chapter 1

Introduction and Literature Review

1.1 Motivation and Problem Statement

Brains operate as high-dimensional dynamical systems whose activity evolves continuously over time. Across rest, task, and naturalistic conditions, neural populations exhibit coordinated fluctuations that recurrently occupy a limited set of configurations where the brain stays in each configuration for a certain amount of time before transitioning to others. This pattern of switching between stable states reflects brain activity related to thinking and behavior. It motivates models that describe both the set of states the brain uses and the rules for moving between them [1]. Whole-brain modeling further suggests that large-scale coordination arises from constrained trajectories on a low-dimensional manifold, making state occupancy and transition structure central explanatory targets rather than incidental artifacts of noise [2].

Capturing these properties requires moving beyond static summaries of connectivity and toward dynamic models that characterize how brain states emerge and change over time. Energy landscape analysis (ELA) provides such a framework by representing multivariate brain activity through an energy function, where local minima correspond to metastable brain states (brain states with relatively high stability) and the height of energy barriers

determines the ease of transitions between them. In the conventional formulation of energy landscape models, a pairwise maximum-entropy (Ising) model is fitted to time-series data, and a disconnectivity graph is derived to enumerate basins, state energies, and transition pathways. These quantities offer mechanistic summaries that can be compared across individuals, groups, or experimental conditions [3]. Recent studies have evaluated how consistently energy–landscape analysis (ELA) performs when applied to resting-state fMRI, examining properties such as test–retest reliability and the stability of estimated landscapes. These efforts provide essential validation, paving the way for research on individual differences and potential clinical applications [4].

Applications of ELA span multiple domains in which altered brain dynamics are hypothesized. In mood disorders, functional near-infrared spectroscopy studies report constrained dynamics with increased residence in low-energy basins and case–control separability, consistent with reduced flexibility of state exploration [5]. In neurodegeneration, resting-state fMRI analyses reveal altered occurrence and transition profiles among control and sensory-integration states that track cognitive performance, suggesting disease-specific remodeling of the state space [6]. In drug-resistant epilepsy, dietary therapy has been associated with remodeling of individualized state repertoires and transitions on the energy landscape, highlighting a potential network-level mechanism of treatment response [7]. Together, these studies illustrate how landscape-based descriptors can link pathophysiology, behavior, and intervention within a single, interpretable framework.

Despite this progress, several methodological gaps limit broader adoption of energy landscape models. First, pairwise maximum-entropy models capture second-order structure but can miss distributed multivariate dependencies that contribute to the overall system dynamics [8, 9, 10]. Second, classical Ising formulations often require binarization of time series, discarding amplitude fluctuations that reflect important aspects of brain physiology [11, 12, 13, 3]. Third, the discrete state space scales exponentially with the number of regions (2^N configurations), making energy landscape calculations intractable beyond small number

of regions [14, 15, 16]. Lastly, the energy landscape methods require calibration and robustness checks that remain computationally tractable while accounting for sampling variability, physiological and motion confounds, and reproducibility when moving from simulations to real cohorts [17, 18, 19]. Addressing these gaps is essential to make energy landscape models robust, reliable, and interpretable, enabling them to answer fundamental questions about neural dynamics, such as how brain states emerge, persist, and transition across conditions [20]. This dissertation addresses these gaps by developing energy–landscape models that capture higher-order neural interactions while preserving the continuous amplitude information in the dataset.

More specifically, this dissertation aims to answer the following three questions: First, do higher-order interactions at the group level provide better recovery of the underlying basin structure and transition organization? Second, can energy landscape values approximated using continuous signals improve the recovery of state-space properties? Finally, can such higher-order continuous energy landscape models provide valuable information about cognitive function and its decline in patients with brain tumors?

1.2 Preliminaries

1.2.1 Intuition Behind Energy Landscape Models

The energy landscape approach, inspired by statistical physics, represents high-dimensional brain activity using an energy function that characterizes which configurations occur and how they change over time. A configuration x is the multivariate pattern of brain activity across regions at a given time t. The energy function E is then formally defined as:

$$E : \mathcal{X} \to \mathbb{R}, \tag{1.1}$$

where $\mathcal{X}$ denotes the configuration space that represents the set of all possible activity patterns of the system. The lower energy values indicate more typical or stable configurations, and higher values indicate rarer or less stable ones. Brain activity over time can be conceptualized as a noisy trajectory on this surface, spending most time near low-energy valleys and occasionally crossing higher-energy barriers to reach other valleys [1, 2].

In this context, a state $\mathbf{s}$ is a locally preferred configuration, represented as a local minimum of the energy function E. Around each state is a basin of attraction $\mathcal{B}$, which includes nearby configurations that would move toward the minimum under relaxation dynamics. The depth and shape of this basin help explain how long the system stays in the state and how stable it is. In practice, similar patterns are often grouped together into larger, metastable macrostates to make the analysis easier to interpret.

The stability of a state is closely related to its likelihood of being visited. This relationship is captured by the Boltzmann–Gibbs distribution, which links energy and occupancy:

$$p(x) \propto \exp\{-E(x)\}, \tag{1.2}$$

where $p(x)$ is the probability of observing configuration x. Lower-energy configurations are therefore more probable, while higher-energy configurations occur less frequently.

Transitions between states are governed by the energy barriers that separate their basins. An energy barrier is the difference between the energy at the saddle point and the energy of the starting minimum. Higher barriers make transitions more difficult, whereas lower barriers allow greater flexibility in moving between states. To summarize this structure, two common representations are used: (i) *disconnectivity graphs*, which illustrate how basins merge as the energy threshold increases and encode the heights of barriers; and (ii) *state-transition graphs*, which simplify basins into nodes and annotate edges with empirical or model-based transition probabilities. Both provide interpretable ways to compare energy landscapes between subjects or conditions [3].

1.2.2 Ising-based Energy Landscape Models

The energy landscape framework typically involves three main steps [3]: (i) defining brain configurations, (ii) estimating their probability distribution, and (iii) summarizing properties of the energy landscape derived from that distribution. In the conventional Ising-based approach, we begin by defining binary brain configurations $\{\mathbf{q}^{(t)}\}_{t=1}^{T}$ obtained by thresholding the signal $\mathbf{y}$. Each binary configuration $\mathbf{q}^{(t)} \in \{0,1\}^N$ represents the activity pattern across N regions at time t. The configuration space enumerates all possible patterns $\{\mathbf{q}_k\}_{k=1}^{2^N}$, where k is the index of 2^N distinct configurations.

Next, the pairwise maximum-entropy model assigns probabilities $\{p(\mathbf{q}_k)\}$ by maximizing Shannon entropy subject to matching the empirical first and second-order moments:

$$\begin{aligned} \text{maximize: } & -\sum_{k=1}^{2^N} p(\mathbf{q}_k)\, \log p(\mathbf{q}_k) \\ \text{subject to: } & \sum_k p(\mathbf{q}_k)\, q_{ki} = \langle q_i \rangle_{\text{data}}, \\ & \sum_k p(\mathbf{q}_k)\, q_{ki} q_{kj} = \langle q_i q_j \rangle_{\text{data}}, \end{aligned} \tag{1.3}$$

where the first-order and second-order moments are calculated as:

$$\langle q_i \rangle_{\text{data}} = \frac{1}{T}\sum_{t=1}^{T} q_i^{(t)}, \qquad \langle q_i q_j \rangle_{\text{data}} = \frac{1}{T}\sum_{t=1}^{T} q_i^{(t)}\, q_j^{(t)}. \tag{1.4}$$

Here, $\langle q_i \rangle_{\text{data}}$ is the average activation of region i, and $\langle q_i q_j \rangle_{\text{data}}$ is the average co-activation of regions i and j across T time points. The solution of this optimization problem the Boltzmann form

$$p(\mathbf{q}_k) \;=\; \frac{\exp\Big[-E(\mathbf{q}_k)\Big]}{\mathcal{Z}}, \tag{1.5}$$

where $\mathcal{Z} = \sum_{k=1}^{2^N} \exp\Big[-E(\mathbf{q}_k)\Big]$ is the partition function that normalizes the probability

distribution and E is the energy function defined as:

$$E(\mathbf{q}_k) \;=\; -\sum_{i<j} W_{ij}\, q_{ki} q_{kj} \;-\; \sum_i h_i\, q_{ki}, \tag{1.6}$$

where W_{ij} represents the pairwise interaction strength between regions i and j, h_i is the bias term for region i, and $\mathbf{q}_k \in \{0,1\}^N$ denotes the k-th binary configuration of N regions. We could also obtain the equivalent matrix form as:

$$E(\mathbf{q}) \;=\; -\tfrac{1}{2}\,\mathbf{q}^\top \mathbf{W}\mathbf{q} \;-\; \mathbf{h}^\top \mathbf{q}, \qquad \mathbf{W} = \mathbf{W}^\top, \;\; W_{ii} = 0, \tag{1.7}$$

where the factor $\frac{1}{2}$ prevents double counting when the coupling matrix $\mathbf{W}$ is symmetric with zero diagonal. The parameters ($\mathbf{W}$ and $\mathbf{h}$) are estimated so that the model reproduces the empirical first and second-order moments in Eq. (1.4). Once the model parameters are estimated, we can compute properties of the resulting energy landscape, such as local minima, their basins of attraction, and disconnectivity graphs, to characterize the system's stability and transition structure.

1.2.3 Limitations Of The Ising-based Energy Landscapes

The Ising-based energy landscape model described in the previous section has two main limitations, which are discussed in detail here.

Exponential State Space And Computational Cost

The exact computation of partition function $\mathcal{Z}$ in equation 1.5 by direct summation requires evaluating 2^N terms. Model expectations, such as $\langle x_i \rangle_{\text{model}}$ and $\langle x_i x_j \rangle_{\text{model}}$, are also sums over $\mathcal{X}$ and therefore scale at least linearly with 2^N when computed exactly. The log-linear relation in Figure 1.1 shows that each additional region doubles the number of states, so a ten-region increase multiplies the configuration space by approximately one thousand. For example, $2^{20} \approx 10^6$, $2^{40} \approx 10^{12}$, and $2^{50} \approx 10^{15}$. Given that typical fMRI parcellations are on the

order of hundreds of regions, Ising-based energy landscape models become computationally infeasible for large datasets [21, 22].

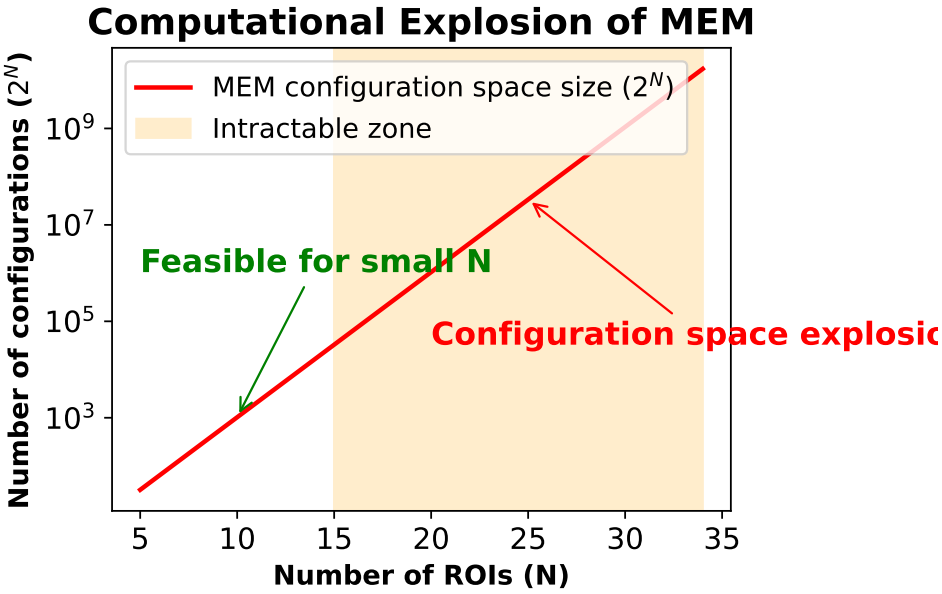


Figure 1.1: Computational growth of the discrete MEM configuration space. The horizontal axis shows the number of brain regions N and the vertical axis shows the number of brain configurations on a logarithmic scale. The line is straight on the log scale, reflecting exponential growth.

Information Loss Under Binarization

Let $\mathbf{y} = (y_1, \ldots, y_N)$ denote regional amplitudes, and let $\mathbf{q}$ be a binarized representation obtained via a deterministic coordinatewise threshold. For any external variable v, the mutual information satisfies:

$$\mathrm{MI}(\mathbf{y}; v) \;\geq\; \mathrm{MI}(\mathbf{q}; v), \tag{1.8}$$

by the data processing inequality, since $\mathbf{q}$ is a measurable function of $\mathbf{y}$ [23]. Thus, binarization cannot increase predictive information about v and typically results in information loss.

Binarization also attenuates second-order dependencies. Under a bivariate normal model for (y_i, y_j) with zero means and unit variances, and a zero threshold for binarization, the

correlation between binarized variables satisfies the arcsine law [24]:

$$\mathrm{Corr}(q_i, q_j) = \frac{2}{\pi} \arcsin(\rho), \tag{1.9}$$

where ρ is the original correlation between y_i and y_j. This transformation strictly reduces the magnitude of correlation for $|\rho| < 1$. For nonzero thresholds, attenuation persists, although the closed-form expression differs.

Figure 1.2 illustrates how binarization can lead to information loss. Specifically, long blue segments with varying amplitudes in the original signal are mapped to identical orange plateaus after binarization. For example, both a large peak and a small bump above the threshold become 1, while all sub-threshold variations collapse to 0. Near threshold crossings, small oscillations in the blue trace produce rapid $0 \leftrightarrow 1$ toggles in the orange trace, fragmenting continuous runs and introducing high-frequency label noise. These effects reduce mutual information and distort empirical moments: the mean of q_i reflects only the proportion of time above threshold (duty cycle), and the co-activation term $\langle q_i q_j \rangle$ depends on joint time above threshold rather than the strength of co-fluctuation. This explains the arcsine attenuation of correlations after binarization.

Figure 1.3 shows that changing the threshold alters event counts, event durations, and flip locations even when the underlying continuous signal remains unchanged. A higher threshold yields fewer 1s (lower duty cycle), more isolated spikes, and shorter run lengths, whereas a lower threshold merges adjacent peaks into longer 1-runs. These visible differences propagate directly to the empirical moments used by the MEM where $\langle q_i \rangle_{\mathrm{data}}$ decreases under a higher threshold, and $\langle q_i q_j \rangle_{\mathrm{data}}$ drops when near-threshold co-fluctuations no longer register as joint 1s. Consequently, changes in the thresholding rule alone can alter the estimated values of h_i and W_{ij} even when the underlying neural activity remains unchanged.

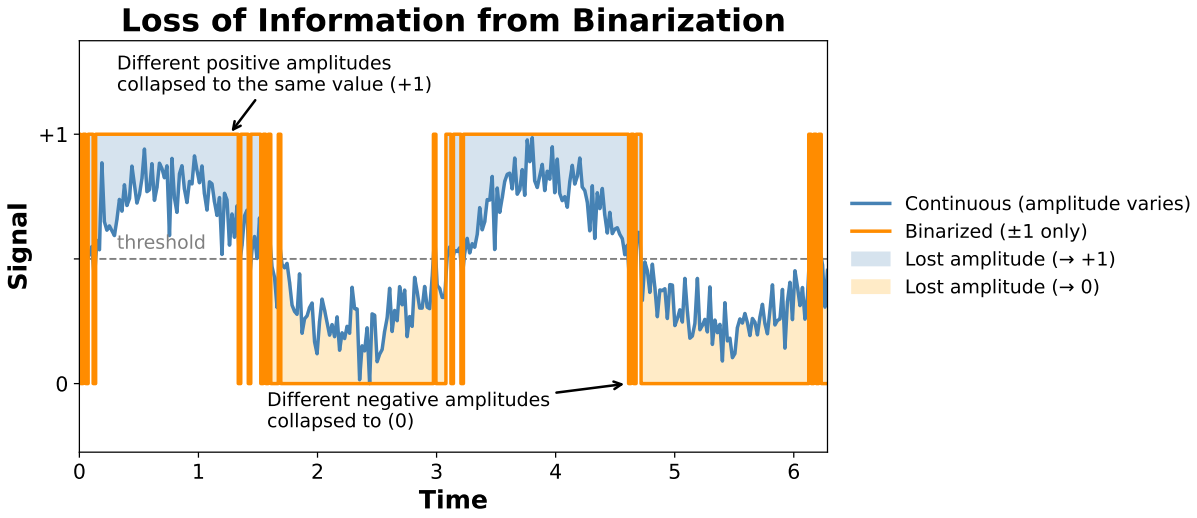


Figure 1.2: Loss of amplitude information. The continuous signal (blue) contains distinct positive amplitudes and graded peaks; the binarized trace (orange) collapses all positive values to 1 and all negative values to 0.

1.3 Prior Work

Several methods have been proposed in the literature to address the computational cost of energy landscape analysis. In this section, we provide an overview of the most recent approaches.

1.3.1 Maximum Pseudo-Likelihood (MPL)

MPL replaces joint normalization in the pairwise maximum-entropy model with a product of logistic conditionals [25, 26]. With $\mathbf{q} = (q_1, \ldots, q_N) \in \{0,1\}^N$ and

$$E(\mathbf{q}) \;=\; -\sum_{i<j} W_{ij}\, q_i q_j \;-\; \sum_i h_i\, q_i, \tag{1.10}$$

the conditional distribution is

$$p\Big(q_i = 1 \mid \mathbf{q}_{\setminus i}\Big) \;=\; \sigma\Big(h_i + \sum_{j\neq i} W_{ij}\, q_j\Big), \qquad \operatorname{logit}^{-1}(u) = \frac{1}{1+\exp(-u)}, \tag{1.11}$$

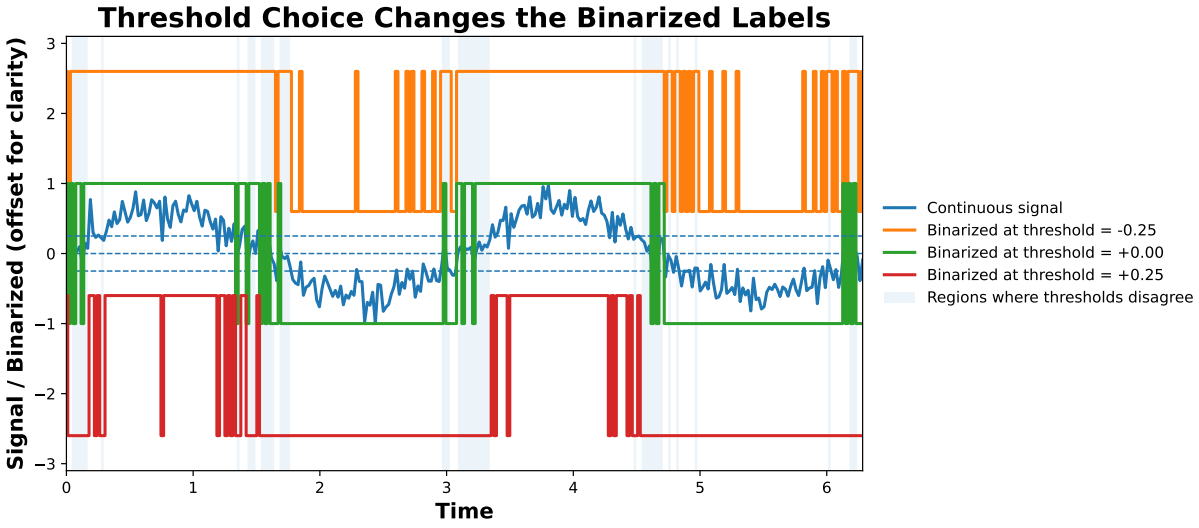


Figure 1.3: Threshold sensitivity. Two plausible thresholds applied to the same continuous signal produce different binary sequences. The higher threshold reduces time labeled as 1, shortens runs of 1s, and moves flip locations toward local maxima, whereas the lower threshold increases duty cycle and merges nearby events.

where, $\mathbf{q}_{\setminus i}$ denotes the configuration of all variables except q_i and u is the linear predictor

$$u = h_i + \sum_{j \neq i} W_{ij}\, q_j, \tag{1.12}$$

which aggregates the local bias h_i and the weighted contributions from neighboring variables.

MPL maximizes the sum of conditional log-likelihoods over i and time, which is equivalent to performing N node-wise logistic regressions of q_i on $\{q_j : j \neq i\}$, followed by symmetrization $W_{ij} \leftarrow (W_{ij} + W_{ji})/2$ and enforcing $W_{ii} = 0$. Nonetheless, MPL is an approximation, which ignores the global partition function and may lead to biased estimates when dependencies are strong. Figure 1.4 shows the schematic of MPL model.

1.3.2 Minimum Probability Flow (MPF)

The Minimum Probability Flow (MPF) fits $(\mathbf{W}, \mathbf{h})$ by minimizing the instantaneous probability outflow from observed configurations to their one-flip neighbors under a continuous-time Markov dynamics [27, 28]. Let $\mathcal{X}$ denote the configuration space and $\mathcal{S} \subset \mathcal{X}$ the set

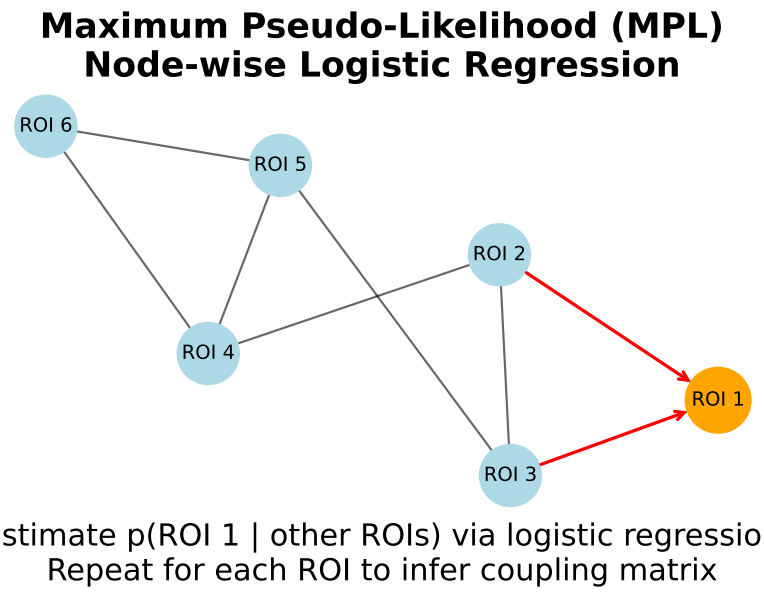


Figure 1.4: MPL schematic. Each region's binary series is regressed on the others to recover one row of W and the bias h_i, where aggregating across regions yields (W, h). Despite its computational efficiency, rare positive labels compress the logistic function's dynamic range and tend to push h_i toward negative values unless class-imbalance remedies are applied. Post hoc symmetrization, which averages W_{ij} and W_{ji}, stabilizes estimates but can attenuate localized effects when node-wise fits disagree.

of observed configurations, with $\mathcal{N}(\mathbf{q})$ representing the Hamming-1 neighborhood of $\mathbf{q}$. A standard objective is

$$\mathcal{L}_{\mathrm{MPF}}(\mathbf{W}, \mathbf{h}) \;=\; \sum_{\mathbf{q}\in\mathcal{S}} \sum_{\mathbf{q}'\in\mathcal{N}(\mathbf{q})} \exp\!\left\{\frac{1}{2}\Big(E(\mathbf{q}) - E(\mathbf{q}')\Big)\right\}, \tag{1.13}$$

where $E(\mathbf{q})$ is defined in (1.10). The gradients of $\mathcal{L}_{\mathrm{MPF}}$ depend only on local differences between neighboring configurations, which makes first-order optimization possible without sampling from the full model or computing the partition function $\mathcal{Z}$. However, MPF still requires checking all one-flip neighbors for each observed configuration, which becomes costly when the configuration space $\mathcal{X}$ or the number of regions N is large. In addition, because MPF minimizes instantaneous probability flow rather than the full likelihood, its estimates can be sensitive to how neighborhoods are defined and may perform poorly when the observed data cover only a small portion of $\mathcal{X}$. Figure 1.5 illustrates the schematic of the

MPF loss function landscape, highlighting its potential values across different parameter configurations.

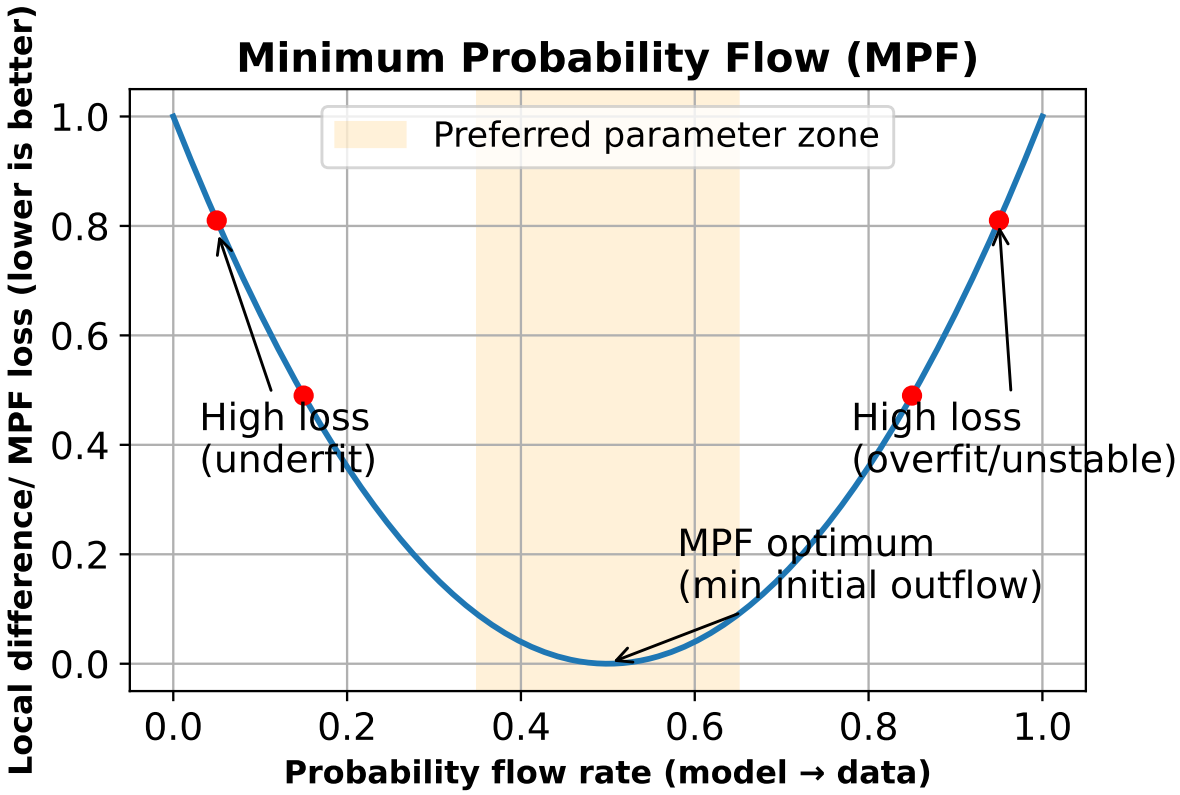


Figure 1.5: The Minimum Probability Flow (MPF) objective vs. model parameters. The plot shows how the initial probability flow from data states to their one-bit-flip neighbors relates to the local MPF loss. The loss is minimized in the preferred parameter zone, where the local energy ordering around each observed configuration best matches the data: if the model is too flat, probability flows out of data states (underfitting); if it is too steep, the ordering among neighboring states becomes unstable (overfitting). This local, one-flip focus explains MPF's speed, but also its weakness when important transitions require changing multiple bits at once.

1.3.3 Noise-Contrastive Estimation (NCE)

Noise-Contrastive Estimation (NCE) learns an unnormalized energy-based model by framing parameter estimation as a binary classification problem between observed data and samples from a known noise distribution $r(x)$ [29]. Let $\mathcal{X}$ denote the configuration space and $\mathcal{S} \subset \mathcal{X}$

the set of observed configurations. The unnormalized model is

$$\tilde{p}(x) = \exp\{-E(x)\}, \tag{1.14}$$

where $E(x)$ is the energy function defined in (1.10). NCE introduces a learned log-normalizer $\beta \in \mathbb{R}$ and optimizes

$$\begin{aligned} \max_{\mathbf{W},\mathbf{h},\beta} \sum_{x \in \mathcal{S}} \log \frac{1}{1 + \exp\Big(- (-E(x) + \beta - \log r(x))\Big)} + \\ \sum_{x \sim r} \log \frac{1}{1 + \exp\Big(- (E(x) - \beta + \log r(x))\Big)}, \end{aligned} \tag{1.15}$$

where $\exp(\cdot)$ denotes the exponential function, and the fraction represents the logistic function $\text{logit}^{-1}(u) = \frac{1}{1+\exp(-u)}$. The term $r(x)$ is the known noise distribution, and β acts as a learned log-normalizer that approximates $\log \mathcal{Z}$. This formulation allows NCE to avoid computing $\mathcal{Z}$ explicitly by reframing normalization as a classification problem. However, stable estimation requires sufficient overlap between the noise distribution $r(x)$ and the data distribution, as well as an adequate number of noise samples. Figure 1.6 illustrates the schematic of the NCE feature space and the overlap between data and noise distributions.

1.3.4 Sampling-based Surrogates

Another group of methods is based on Contrastive Divergence (CD) and its persistent variant, which replace exact normalization with short-run Markov chain updates from data or from a persistent chain [30, 31, 32]. These methods approximate maximum likelihood but introduce bias that depends on mixing. While they reduce computational cost, they retain binarized inputs and exhibit sensitivity to thresholding.

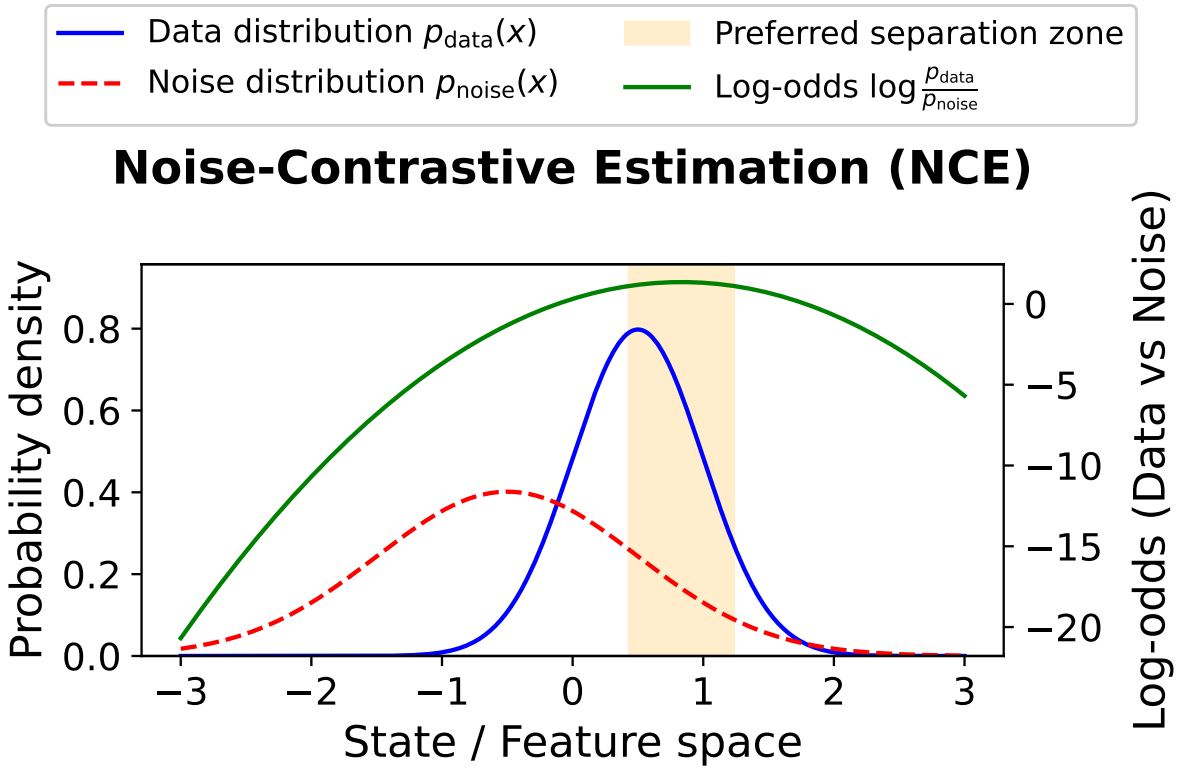


Figure 1.6: Noise-Contrastive Estimation (NCE) in state space. The plot shows the data distribution $p_{\text{data}}(x)$, the noise distribution $p_{\text{noise}}(x)$, and the learned log-odds $\log\big(p_{\text{data}}(x)/p_{\text{noise}}(x)\big)$ evaluated over the same state axis. Broad overlap between p_{data} and p_{noise} places the classifier in a preferred separation zone where the NCE objective is well-conditioned, yielding stable estimates of the energy $E(x)$ and its normalizing constant. When the overlap is too small, the classification problem becomes ill-conditioned, biasing both $E(x)$ and the normalizer.

1.3.5 Normalization Approximations

A different category of methods relies on mean-field and related expansions to approximate the log-partition function and moments using tractable surrogates, such as naive mean-field, TAP corrections, Bethe approximations, and loopy belief propagation [33, 34]. Adaptive Cluster Expansion (ACE) sums contributions from selected clusters and approaches maximum-likelihood fidelity when informative clusters remain small [35]. These approximations scale well in sparse or weakly coupled regimes but can degrade under strong dependencies and still operate on binarized data.

1.4 Overview of the Proposed Work

Most prior work has focused on reducing computational complexity by replacing intractable normalization with tractable surrogate objectives, such as pseudo-likelihood in neuroimaging applications and probability flow or contrastive estimation more broadly in the energy-based modeling literature, thereby making discrete energy landscapes feasible at realistic parcellation sizes. However, these approaches typically operate on binarized representations and estimate pairwise interactions during the modeling phase without explicitly addressing robustness or the information loss introduced by thresholding continuous signals. This limitation is particularly critical in neuroimaging and other high-dimensional domains, where binarization can obscure subtle but informative variations in connectivity patterns.

In this study, we propose a two-stage framework to mitigate these issues. First, we introduce a higher-order representation that groups signals into semantically coherent clusters prior to binarization, preserving mesoscale structure and reducing sensitivity to arbitrary thresholds. This clustering step captures shared dynamics across regions, enabling a more stable and interpretable discrete model. Second, we extend the framework to bypass binarization entirely by directly estimating the precision matrix from continuous signals using graph convolutional networks (GCNs). GCNs exploit the graph structure, including both spatial adjacency and functional similarity, allowing the model to learn localized dependencies while maintaining global consistency. By integrating these components, our approach bridges the gap between tractable energy-based modeling and the rich variability inherent in continuous data, offering improved reliability, robustness, and scalability.

Chapter 2

High-Order Discrete Energy Landscape (HODEL)

2.1 High-Order Energy Landscape Framework

In this section, we introduce the higher-order energy landscape model, in which signals from different regions of interest (ROIs) are first grouped into clusters before applying the maximum entropy framework. This clustering step enables the energy landscape model to capture higher-order interactions that may not be apparent from direct pairwise interactions. The overall higher-order energy landscape framework is illustrated in Figure 2.1, and each step of the framework is explained in detail below.

2.1.1 Identifying Functional Clusters

Let $\mathcal{G} = \{G_1, \ldots, G_K\}$ denote a partition of the N regions into K clusters, with $K \ll N$. Clusters can be obtained using k-means on standardized continuous regional signals $\mathbf{y}_i(t)$ or on features derived from connectivity. When available, network-guided constraints such as anatomical neighborhoods or structural connectivity can regularize clustering.

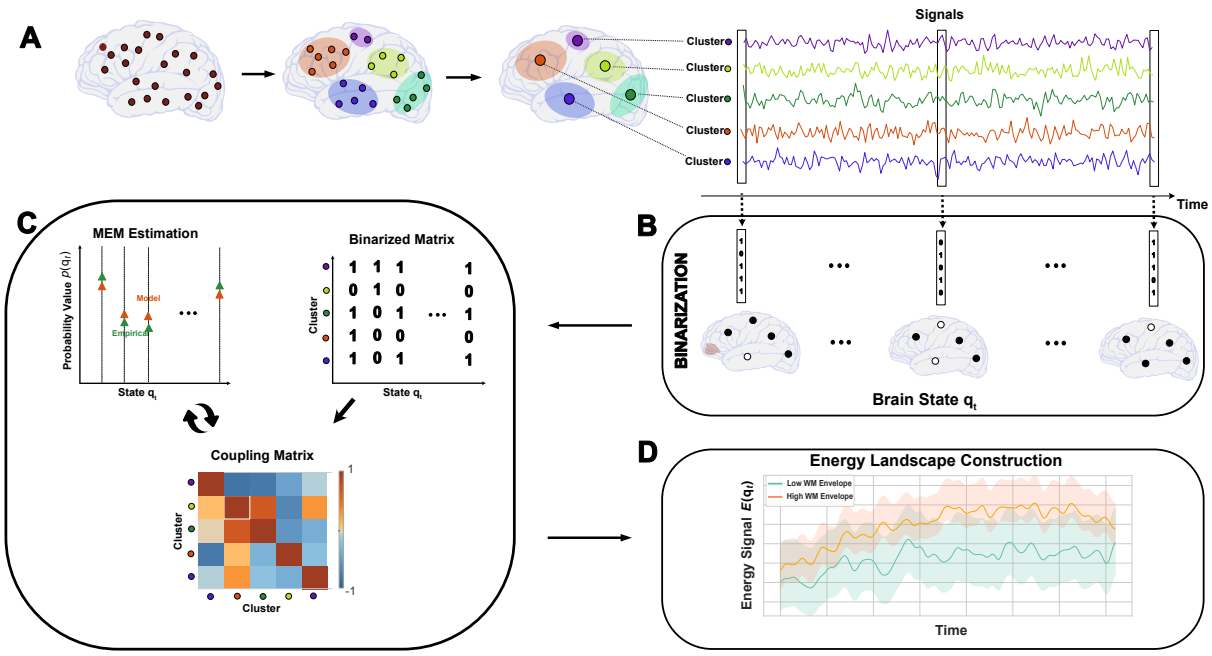


Figure 2.1: **High-order discrete energy landscape framework. (A)** *Identifying Functional Clusters:* Resting-state fMRI signals are grouped into functionally similar clusters using k-means algorithm. Each cluster represents a group of brain regions with similar activity. **(B)** *Extracting Brain States:* The average signal from each cluster is converted into binary states (active = 1 and inactive = 0) based on a threshold (e.g. mean value), forming a sequence of brain states over time. **(C)** *Fitting Maximum Entropy Model:* The maximum entropy model is fitted to identify the probability distribution of brain states that maximizes entropy, while ensuring that the model matches the observed average activation of each cluster and the pairwise co-activation between clusters. **(D)** *Constructing Energy Landscapes:* The parameters of the resulting optimum probability distribution is utilized to calculate the energy value of each state.

2.1.2 Extracting Brain States

Given the partition of regions $\mathcal{G} = \{G_1, \ldots, G_K\}$, continuous regional signals are first pooled within each group and then binarized at the group level:

$$\bar{\mathbf{y}}_{G_g}(t) = \frac{1}{|G_g|} \sum_{i \in G_g} \mathbf{y}_i(t). \tag{2.1}$$

$$q_g(t) = \mathbf{1}\Big\{ \bar{\mathbf{y}}_{G_g}(t) > \tau_g \Big\}, \tag{2.2}$$

$$\mathbf{q}(t) = \Big(q_1(t), \ldots, q_K(t)\Big). \tag{2.3}$$

Here, $\bar{\mathbf{y}}_{G_g}(t)$ denotes the average of the regional signals within group g at time t, $q_g(t) \in \{0, 1\}$ represents the binary state of group g, and $\mathbf{q}(t)$ is the binary brain configuration at time t. By default, the threshold τ_g is set to the temporal median of $\bar{\mathbf{y}}_{G_g}(t)$. If the signals $\mathbf{y}_i(t)$ are mean-centered, we also test alternative thresholds such as $\tau_g = 0$ or τ_g set to the average value of the pooled signal over time.

2.1.3 Fitting Maximum Entropy Model

The grouped Maximum Entropy model is defined over the configuration space $\mathcal{X} = \{0, 1\}^K$, with the pairwise Ising form:

$$p(\mathbf{q}) \;=\; \frac{\exp\{-E(\mathbf{q})\}}{\mathcal{Z}}, \qquad E(\mathbf{q}) \;=\; -\sum_{1 \le c < g \le K} W_{cg}\, q_c q_g \;-\; \sum_{g=1}^{K} h_g\, q_g, \tag{2.4}$$

Here, $p(\mathbf{q})$ denotes the probability of configuration $\mathbf{q} \in \mathcal{X}$, $E(\mathbf{q})$ is the energy function, $\mathcal{Z}$ is the partition function, $\mathbf{W} = \{W_{cg}\}$ represents the pairwise coupling matrix, and h_g are the local bias terms. For the energy function in (2.4), the negative log-likelihood of the observed configurations is:

$$\mathcal{L}(\mathbf{W}, \mathbf{h}) = T \log \mathcal{Z} + \sum_{t=1}^{T} E\Big(\mathbf{q}(t)\Big), \tag{2.5}$$

where the partition function is:

$$\mathcal{Z} \;=\; \sum_{\mathbf{q}} \exp\{-E(\mathbf{q})\}. \tag{2.6}$$

The model expectations are computed by summation:

$$\langle q_g \rangle_{\text{model}} = \sum_{\mathbf{q}} q_g \, \frac{\exp\{-E(\mathbf{q})\}}{\mathcal{Z}}, \qquad \langle q_c q_g \rangle_{\text{model}} = \sum_{\mathbf{q}} q_c q_g \, \frac{\exp\{-E(\mathbf{q})\}}{\mathcal{Z}}, \tag{2.7}$$

where c, g index groups within the partition $\mathcal{G}$. For numerical stability, $\log \mathcal{Z}$ is evaluated using the log–sum–exp transform. Let $E_{\min} = \min_{\mathbf{q}} E(\mathbf{q})$, then:

$$\log \mathcal{Z} = -E_{\min} + \log \sum_{\mathbf{q}} \exp\Big\{ - \big(E(\mathbf{q}) - E_{\min}\big) \Big\}. \tag{2.8}$$

Gradients of the negative log-likelihood are given by model–data moment differences:

$$\frac{\partial \mathcal{L}}{\partial h_g} = T \left(\langle q_g \rangle_{\text{model}} - \langle q_g \rangle_{\text{data}} \right), \qquad \frac{\partial \mathcal{L}}{\partial W_{cg}} = T \left(\langle q_c q_g \rangle_{\text{model}} - \langle q_c q_g \rangle_{\text{data}} \right) \quad (c < g). \tag{2.9}$$

To reduce the difference between model predictions and observed data (the residuals), we adjust the parameters using the following rule:

$$\begin{aligned} h_g &\leftarrow h_g + \alpha \Big(\langle q_g \rangle_{\text{data}} - \langle q_g \rangle_{\text{model}} \Big), \\ W_{cg} &\leftarrow W_{cg} + \alpha \Big(\langle q_c q_g \rangle_{\text{data}} - \langle q_c q_g \rangle_{\text{model}} \Big), \end{aligned} \tag{2.10}$$

where, α is the learning rate, and the term in parentheses represents the difference between the data statistics and the model statistics (the residual). After each update, we enforce $W_{cg} = W_{gc}$ and $W_{gg} = 0$ to maintain symmetry and prevent self-interactions in the coupling matrix $\mathbf{W}$. The learning process begins with an initial coupling matrix $W^{(\mathcal{G})}$ set near zero and bias terms $h^{(\mathcal{G})}$ initialized from empirical means to stabilize early iterations. It then proceeds iteratively and stops when the maximum absolute residual for first-order moments and the root-mean-square of second-order residuals fall below predefined tolerances. In addition to these criteria, we monitor the model's performance on held-out data to ensure good generalization. Furthermore, we verify that configurations occurring frequently in the dataset have lower energy values (and thus higher probability) under the model, while rare

configurations have higher energy. This alignment between energy and empirical frequency serves as a sanity check that the model ranks configurations consistently with the observed data.

2.1.4 Constructing Energy Landscapes

Once the optimal parameters $\mathbf{W}$ and $\mathbf{h}$ are fitted, we can compute the energy for each possible configuration using Equation (2.4), thereby obtaining the full energy landscape. These energy landscape features can then be used to characterize the structural properties of the system, such as identifying stable states, mapping transition pathways, and estimating the relative probabilities of different configurations.

2.2 Experimental Setup

2.2.1 Overview

We evaluated the proposed energy landscape models by assessing their ability to capture multiple attractor basins, their occupancy frequencies, and the switching behavior between basins. This evaluation was conducted using generated multivariate time series with a known multi-basin ground truth and a controlled transition matrix governing the dynamics. The time series data were generated using two complementary methods: (i) a Switching Linear Dynamical System (SLDS) to model piecewise linear Gaussian behavior with tunable dwell times and supervised transitions [36, 37], and (ii) a Kuramoto oscillator network to capture nonlinear phase-coupled behavior with coherent patterns after thresholding [38, 39, 40, 41, 42]. Together, these two methods encompass both linear–Gaussian and nonlinear phase-coupled dynamics, ensuring a fair comparison, as neither the lower-dimensional discrete energy model nor the higher-order model is biased toward a single type of underlying process.

For each experiment, we specified the parameters N, K, T, SNR, and $\mathbf{P}$ to generate simulated continuous signals $\mathbf{y}(t) \in \mathbb{R}^N$. Specifically, N denotes the number of regions, K

is the true number of attractor basins, T is the total number of time points, SNR controls the separation between basin centers relative to within-state variability, and $\mathbf{P}$ is the state transition matrix that governs switching dynamics among basins. This experimental design addresses three key questions: Do the configurations that the model predicts as most stable (lowest energy) actually occur most frequently in the data? Do the recovered macrostates accurately correspond to the known attractor basins? Do the reconfiguration summaries reflect the ground-truth switching dynamics imposed by $\mathbf{P}$?

2.2.2 Generator A: Switching Linear Dynamical System (SLDS)

We construct a first–order Markov chain $\gamma_{1:T} \in \{1, \ldots, K^*\}$ with transition matrix $\mathbf{P}$. The diagonal entries $P_{mm} \in [0.80, 0.95]$ control the expected dwell time in each state, where:

$$P_{mn} = \frac{1 - P_{mm}}{K^* - 1} \quad \text{for } n \neq m, \tag{2.11}$$

where $m, n \in \{1, \ldots, K^*\}$ index the discrete states of the Markov chain. The initial state is drawn from the stationary distribution $\boldsymbol{\varsigma}$ satisfying $\boldsymbol{\varsigma}^\top \mathbf{P} = \boldsymbol{\varsigma}^\top$.

Within this setting, a latent continuous state $\mathbf{a}_t \in \mathbb{R}^N$ evolves linearly according to state-specific parameters:

$$\mathbf{a}_{t+1} = \mathbf{L}^{(m)} \mathbf{a}_t + \mathbf{c}^{(m)} + \boldsymbol{\varepsilon}_t, \tag{2.12}$$

where $\boldsymbol{\varepsilon}_t$ is zero-mean Gaussian process noise with covariance $\boldsymbol{\Sigma}^{(m)}$. Here, $\mathbf{a}_t$ denotes the latent continuous state vector at time t, $\mathbf{L}^{(m)}$ is the linear dynamic matrix for Markov state m, and $\mathbf{c}^{(m)}$ is the latent dynamic offset for state m.

The observed continuous regional signal is:

$$\mathbf{y}_t = \mathbf{H} \mathbf{a}_t + \mathbf{b}^{(m)} + \boldsymbol{\nu}_t, \tag{2.13}$$

where $\boldsymbol{\nu}_t$ is zero-mean Gaussian observation noise with covariance $\boldsymbol{\Sigma}_{\text{obs}}$. Here, $\mathbf{y}_t$ denotes the

observed continuous signal, $\mathbf{H}$ is the observation matrix (set to I_N unless otherwise stated), and $\mathbf{b}^{(m)}$ is the observation offset for state m. The offsets $\mathbf{c}^{(m)}$ and $\mathbf{b}^{(m)}$ let each state have its own average level.

We construct each dynamic matrix $\mathbf{L}^{(m)}$ so that the system remains stable and each state behaves differently. To achieve this, we use an eigen-decomposition, where we choose random orthonormal eigenvectors (directions of variation) and eigenvalues that are all less than one. This guarantees that the dynamics do not diverge over time and that states exhibit diverse patterns of evolution. Let $\mathbf{M}^{(m)} \in \mathbb{R}^{N\times N}$ be a random matrix with i.i.d. standard normal entries. We compute its QR decomposition and denote by $\mathbf{U}^{(m)}$ the Q-factor. Next, draw eigenvalues $\lambda_1^{(m)}, \ldots, \lambda_N^{(m)} \overset{\text{iid}}{\sim} \text{Uniform}\Big([0, \gamma^{(m)}]\Big)$, where $\gamma^{(m)} \in [0.80, 0.98]$. Finally, we define the linear dynamics matrix

$$\mathbf{A}^{(m)} = \mathbf{U}^{(m)} \,\text{diag}\Big(\lambda_1^{(m)}, \ldots, \lambda_N^{(m)}\Big)\, \mathbf{U}^{(m)\top}. \tag{2.14}$$

Unless otherwise stated, we set the latent process noise covariance as $\mathbf{\Sigma}^{(m)} = \sigma_{\mathbf{a}}^2 \mathbf{I}_N$ and the observation noise covariance as $\mathbf{\Sigma}_{\text{obs}} = \sigma_{\mathbf{y}}^2 \mathbf{I}_N$. The pair $(\sigma_{\mathbf{a}}, \sigma_{\mathbf{y}})$ controls the within-state variability.

We impose K^* attractor basin centers $\{\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{K^*}\} \subset \mathbb{R}^N$ to control between-state separation. For reproducibility and balanced separation, we construct sign-patterns with fixed Hamming distance. Let $\mathbf{B} \in \{-1, +1\}^{N\times K^*}$ be the sign-pattern matrix, whose columns $\mathbf{B}_{:,s}$ represent basin prototypes such that each pair differs in approximately h entries (typically $h \approx N/2$). We define an amplitude $\delta > 0$ and set $\boldsymbol{\mu}_s = \delta\, \mathbf{B}_{:,s}$. We center the observation by $\mathbf{B}_{:,s} = \boldsymbol{\mu}_s$ and set the latent dynamic offset $\mathbf{c}^{(s)} = \mathbf{0}$ in Eq. (2.13). By setting $\mathbf{b}^{(s)} = \boldsymbol{\mu}_s$ and $\mathbf{c}^{(s)} = \mathbf{0}$, the observation model is centered at $\boldsymbol{\mu}_s$, so under weak noise the expected observation coincides with the basin center.

We define SNR as the ratio between a target between–state separation κ and the typical

within–state standard deviation:

$$\mathrm{SNR} = \frac{\kappa}{\sqrt{\sigma^2_{\text{within}}}}. \tag{2.15}$$

For regime s, the stationary latent covariance $\mathbf{\Sigma}^{(s)}$ satisfies the discrete Lyapunov equation:

$$\mathbf{\Sigma}^{(s)} = \mathbf{A}^{(s)}\mathbf{\Sigma}^{(s)}\mathbf{A}^{(s)\top} + \mathbf{Q}^{(s)}. \tag{2.16}$$

The corresponding observation covariance is $\mathbf{H}\mathbf{\Sigma}^{(s)}\mathbf{H}^\top + \mathbf{R}$. The average within–state variance is

$$\sigma^2_{\text{within}} = \frac{1}{NK^*}\sum_{s=1}^{K^*} \mathrm{tr}\Big(\mathbf{H}\mathbf{\Sigma}^{(s)}\mathbf{H}^\top + \mathbf{R}\Big). \tag{2.17}$$

For sign–pattern centers with Hamming distance h, the Euclidean distance between two centers equals $2\delta\sqrt{h}$. Choosing

$$\delta = \frac{\mathrm{SNR}\,\sqrt{N\,\sigma^2_{\text{within}}}}{2\sqrt{h}} \tag{2.18}$$

sets the minimal pairwise separation relative to the typical within–state scale. We iterate once if needed to refine latent and observation scales so that the realized SNR matches the target within tolerance.

The SLDS simulator returns the standardized continuous time series $\mathbf{Y} \in \mathbb{R}^{T\times N}$ with rows $\mathbf{y}_t$, the binarized series $\mathcal{X}^{\text{bin}} \in \{0,1\}^{T\times N}$ with rows $\mathbf{q}^{\text{bin}}(t)$, the ground-truth latent state path $\mathbf{z}_{1:T}$, and the cluster centers $\{\boldsymbol{\mu}_k\}_{k=1}^K$.

2.2.3 Generator B: Kuramoto Oscillator Network With Template Locking

We simulate phases $\theta_i(t)$ for N oscillators on a network with mesoscale community structure. A switching label $\mathbf{z}_t \sim \mathbf{P}$ selects which phase template is active at time t. Dynamics follow

a stochastic Kuramoto system with network coupling and weak template locking:

$$\frac{d\theta_i}{dt} \;=\; \omega_i \;+\; \sum_{j=1}^{N} \mathbf{W}_{ij}\,\sin\!\Big(\theta_j - \theta_i\Big) \;+\; \nu\,\sin\!\Big(\phi_i^{(\mathbf{z}_t)} - \theta_i\Big) \;+\; \xi_i(t), \tag{2.19}$$

where ω_i is the intrinsic frequency of oscillator i, $\mathbf{W} = \{\mathbf{W}_{ij}\}$ is the pairwise coupling matrix, $\nu > 0$ is a weak gain toward one of K phase templates $\{\phi^{(1)}, \ldots, \phi^{(K)}\}$, and $\xi_i(t)$ is a small stochastic drive.

We generate an undirected adjacency matrix $\mathbf{A} \in \{0,1\}^{N\times N}$ with community structure using a stochastic block model [43], with within-community probability p_{in} and between-community probability $p_{\text{out}} < p_{\text{in}}$. Degrees are $d_i = \sum_j \mathbf{A}_{ij}$. The pairwise coupling matrix $\mathbf{W}$ is degree-normalized:

$$\mathbf{W}_{ij} = \chi\,\frac{\mathbf{A}_{ij}}{\max(1, d_i)} \quad (i \neq j), \qquad \mathbf{W}_{ii} = 0, \tag{2.20}$$

where $\chi > 0$ controls overall coupling strength. Intrinsic frequencies are drawn from a narrow band:

$$\omega_i \sim \text{Unif}\Big(\omega_0 - \gamma_\omega,\; \omega_0 + \gamma_\omega\Big), \tag{2.21}$$

where ω_0 is the central frequency and γ_ω is the bandwidth (half-width). Noise is white with variance density σ^2:

$$\xi_i(t) = \sqrt{2\sigma^2}\,\eta_i(t), \tag{2.22}$$

where $\eta_i(t)$ is the standard white noise.

The stochastic differential equation 2.19 is integrated with Euler–Maruyama [44], wrapping phases to $[0, 2\pi)$ after each update:

$$\theta_i(t+\Delta t) = \theta_i(t) + \Delta t\left[\omega_i + \sum_{j=1}^{N} \mathbf{W}_{ij}\,\sin\!\Big(\theta_j(t) - \theta_i(t)\Big) + \chi_\phi\,\sin\!\Big(\phi_i^{(z_t)} - \theta_i(t)\Big)\right] + \sqrt{2\sigma^2\Delta t}\,\varepsilon_i(t), \tag{2.23}$$

where $\varepsilon_i(t) \sim \mathcal{N}(0,1)$. The label process is drawn as $z_1 \sim \boldsymbol{\varsigma}$ and $z_{t+1} \sim \mathbf{P}(z_t, \cdot)$ for $t = 1, \ldots, T-1$, using the same $\mathbf{P}$ as in the SLDS generator. A short burn-in is discarded to remove transients.

After simulating the phase dynamics $\theta_i(t)$ using the Euler-Maruyama scheme, we map oscillator phases to regional observables as:

$$\mathbf{y}_i(t) = \sin\Big(\theta_i(t)\Big), \tag{2.24}$$

where $\mathbf{y}_i(t)$ denotes the continuous signal for region i. Each region's signal is then standardized to zero mean and unit variance over time:

$$\mathbf{y}_i^{\text{std}}(t) = \frac{\mathbf{y}_i(t) - \bar{\mathbf{y}}_i}{\sigma(\mathbf{y}_i)}, \tag{2.25}$$

where $\bar{\mathbf{y}}_i$ is the temporal mean and $\sigma(\mathbf{y}_i)$ is the standard deviation.

Finally, region-wise binarization follows the procedure in Section 2.1:

$$\mathbf{q}_i(t) = \mathbf{1}\Big\{\, \mathbf{y}_i^{\text{std}}(t) > \tau_i \,\Big\}, \qquad \tau_i \in \{\text{mean}, \text{median}\}, \tag{2.26}$$

where τ_i is the binarization threshold for region i. The resulting binary configuration $\mathbf{q}(t)$ belongs to the configuration space $\mathcal{X}$ defined earlier.

We prevent trivial global synchrony by choosing the between-state separation parameter κ and the oscillator frequency bandwidth γ_ω so that the Kuramoto order parameter

$$\ell(t)\, e^{\mathrm{i}\bar{\theta}(t)} = \frac{1}{N} \sum_{i=1}^{N} e^{\mathrm{i}\theta_i(t)} \tag{2.27}$$

remains below a preset threshold outside short locking episodes. Here, $\ell(t)$ denotes the magnitude of phase synchrony and $\bar{\theta}(t)$ the mean phase. Dwell time and separation are tuned by (ν, σ), where increasing ν raises template adherence and occupancy purity, while increasing σ

raises within-state variability. Target separation at the observable level is matched by a short pilot run that estimates within-state variance of $\mathbf{y}_t$ and adjusts ν via a one-dimensional line search so that the ratio of inter-template distance to within-state standard deviation meets the specified SNR tolerance.

The Kuramoto simulator returns the standardized continuous time series $\mathbf{Y} \in \mathbb{R}^{T\times N}$ with rows $\mathbf{y}_t$, the binarized series $\mathcal{X}^{\text{bin}} \in \{0,1\}^{T\times N}$ with rows $\mathbf{q}^{\text{bin}}(t)$, and the ground-truth latent state path $\mathbf{z}_{1:T}$.

2.2.4 Evaluation Metrics And Interpretation

In order to evaluate our energy landscape models using simulated data, we identify local minima (recovered basins $\hat{\mathcal{B}}$) via greedy single-bit descent. Starting from an observed binary brain configuration $\mathbf{q} \in \{0,1\}^N$, we flip one coordinate at a time only if the energy value $E(\mathbf{q})$ strictly decreases, and stop when no single flip lowers energy. The terminal configuration obtained after this process is considered a local minimum and represents one of the minimum-energy states in the set $\mathbf{s}$.

To quantify similarity between configurations, we use the Hamming distance h. For two configurations $\mathbf{q}$ and $\mathbf{q}'$, it is defined as:

$$h(\mathbf{q},\mathbf{q}') = \sum_{i=1}^{N} \mathbb{1}\{\mathbf{q}_i \neq \mathbf{q}'_i\}. \tag{2.28}$$

If more than K^* distinct minima are found, we keep the K^* most frequent as representatives and merge each remaining minimum to the nearest representative by Hamming distance, breaking ties by larger representative frequency and then by lower representative energy E. After merging, each observed time point t is assigned to the representative basin corresponding to the local minimum reached by greedy descent. This produces a sequence of discrete state labels $\hat{\mathbf{z}}_{1:T}$, where $\hat{\mathbf{z}}_t$ indicates the basin associated with time point t. Recovered macro basins are matched to ground truth with the Hungarian algorithm [45] using

basin signatures. For SLDS, the true signature is the center μ_k of the corresponding state, whereas for the Kuramoto model it is the observable template obtained by averaging the continuous signal $\mathbf{y}(t)$ within each state. The signature of a recovered basin $\hat{\mathcal{B}}$ is computed as the mean of $\mathbf{y}(t)$ over its assigned time points. A match is accepted when the Euclidean distance between signatures is below the basin matching tolerance value ζ. Once matching is complete, we calculate the following three metrics to evaluate performance.

Basin Recovery (BR)

The Basin Recovery (BR) score is defined as the fraction of true basins $\mathcal{B}_k$ that have a matched recovered basin $\hat{\mathcal{B}}_{\tilde{k}}$ within a tolerance ζ. This tolerance is set to the 10th percentile of the pairwise Euclidean distances between true basin centers for the given condition. The BR metric ranges from 0 to 1, where 1 indicates perfect recovery. Formally, we define:

$$\mathrm{BR} = \frac{1}{K^*} \sum_{k=1}^{K^*} \mathbb{1}\Big(\Big\| \bar{\mathbf{y}}^{(\hat{\mathcal{B}}_{\tilde{k}})} - \mu_k \Big\|_2 \leq \zeta \Big), \tag{2.29}$$

where $\mathbb{1}(\cdot)$ is the indicator function, $\bar{\mathbf{y}}^{(\hat{\mathcal{B}}_{\tilde{k}})}$ is the average signal of the recovered basin $\hat{\mathcal{B}}_{\tilde{k}}$ matched to true basin $\mathcal{B}_k$, μ_k is the true basin center, and $\|\cdot\|_2$ denotes the Euclidean norm.

Transition Matrix Accuracy (TMA)

The second metric evaluates how well the inferred transition matrix matches the true one. Let $\widehat{\mathbf{P}}$ be the row-stochastic maximum-likelihood estimate computed from the recovered latent discrete state sequence $\mathbf{z}_{1:T}$, with entries

$$\widehat{\mathbf{P}}_{mn} = \frac{\varrho_{mn}}{\sum_{n'=1}^{K^*} \varrho_{mn'}}, \qquad \varrho_{mn} = \Big| \{ t : \mathbf{z}_t = m,\ \mathbf{z}_{t+1} = n \} \Big|, \tag{2.30}$$

where ϱ_{mn} denotes the count of transitions from state m to state n. Let $\mathbf{P}$ denote the ground-truth transition matrix. We define the Transition Matrix Agreement (TMA) score

as

$$\text{TMA} = 1 - \frac{1}{2K^*} \left\| \widehat{\mathbf{P}} - \mathbf{P} \right\|_1, \qquad \left\| \widehat{\mathbf{P}} - \mathbf{P} \right\|_1 = \sum_{m=1}^{K^*} \sum_{n=1}^{K^*} \left| \widehat{\mathbf{P}}_{mn} - \mathbf{P}_{mn} \right|. \tag{2.31}$$

This metric lies in $[0, 1]$ because each row of a transition matrix is stochastic (sums to 1), so the maximum possible difference between two rows is two. Consequently, the total difference across K^* rows is at most $2K^*$, and dividing by $2K^*$ ensures the normalized score falls between zero and one.

State Distribution Agreement (SDA)

The final metric evaluates how well the empirical probability vector over states matches the true stationary distribution. Let $\widehat{\boldsymbol{\varsigma}}$ denote the empirical probability vector computed from the recovered state sequence $\mathbf{z}_{1:T}$, and let $\boldsymbol{\varsigma}_\star$ denote the stationary probability vector derived from the ground-truth transition matrix $\mathbf{P}$. We define:

$$\text{SDA} = 1 - \text{TV}(\widehat{\boldsymbol{\varsigma}}, \boldsymbol{\varsigma}_\star), \qquad \text{TV}(p, q) = \tfrac{1}{2} \sum_{k=1}^{K^*} |p_k - q_k|, \tag{2.32}$$

where $\text{TV}(p, q)$ is the total variation distance between two probability vectors p and q. The SDA score lies in $[0, 1]$, with 1 indicating perfect agreement and 0 indicating maximum discrepancy.

2.3 Simulation Results: High-Order Vs. Low-Order DEL

For the experimental results, we generate data using the Switching Linear Dynamical System (SLDS) and the Kuramoto oscillator network with template locking, as described in the previous sections. We vary the total number of regions $N \in \{6, 7, \ldots, 14\}$, the true number of attractor basins $K^* \in \{3, 4, 5\}$, the total number of time points $T \in \{500, 1000\}$, and three signal-to-noise ratio levels SNR $\in \{1.0, 1.5, 2.0\}$ corresponding to low, medium, and

high noise conditions. Each condition is repeated 50 times with independent random seeds.

For each simulation condition, we summarize performance using the mean of the 50 repeated simulation units and report accompanying 95% bootstrap confidence intervals. To compute these intervals, we resample the 50 simulation units with replacement 10,000 times. When comparing two methods, we keep the pairing intact by using the same resampled index set for both pipelines within each bootstrap replicate. Confidence intervals are then taken as the percentile range of these paired differences. To ensure a fair comparison across methods, all simulation units share the same underlying settings, including $(N, K^*, T, \mathrm{SNR}, \mathrm{seed})$. After constructing the bootstrap distribution, we assess the significance of the performance difference by testing whether the median paired difference equals zero, using a two-sided Wilcoxon signed-rank test [46].

2.3.1 Kuramoto Oscillator Network

Figure 2.2 shows that the High-Order Discrete Energy Landscape (HODEL) model outperforms the Low-Order Discrete Energy Landscape (LowDEL) model across all evaluation metrics. The paired differences are statistically significant for each metric: BR ($p = 1.92 \times 10^{-10}$), TMA ($p = 3.4 \times 10^{-10}$), and SDA ($p = 6.82 \times 10^{-5}$). Across the parameter grid, TMA ranges from approximately 0.91 to 0.99, and SDA from 0.86 to 1.00. In every case, HODEL shifts the distribution upward relative to LowDEL, indicating consistently improved performance.

2.3.2 Switching Linear Dynamical System (SLDS)

Figure 2.3 shows that HODEL achieves consistent and statistically significant improvements over LowDEL across all evaluation metrics in the SLDS simulations as well. The paired differences are significant for BR ($p = 3.47 \times 10^{-9}$), TMA ($p = 1.43 \times 10^{-7}$), and SDA ($p = 0.0116$). Across the parameter grid, BR ranges from approximately 0.70 to 1.00, TMA from 0.88 to 1.00, and SDA from 0.95 to 1.00, indicating accurate basin recovery and transition structure under both models. In all cases, HODEL achieves higher scores than

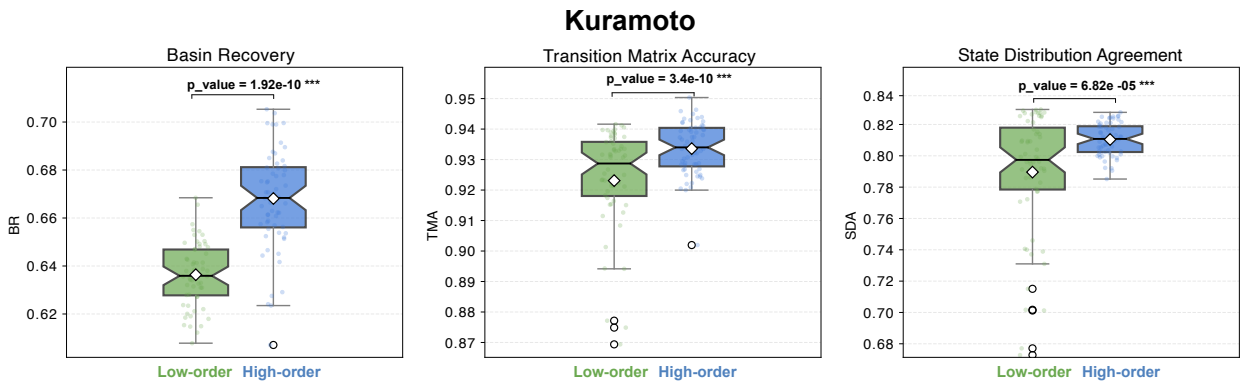


Figure 2.2: Comparison of HODEL and LowDEL on Basin Recovery (BR), Transition Matrix Accuracy (TMA), and State Distribution Agreement (SDA) for the Kuramoto simulations. Paired Wilcoxon signed-rank p-values are shown in each panel.

LowDEL.

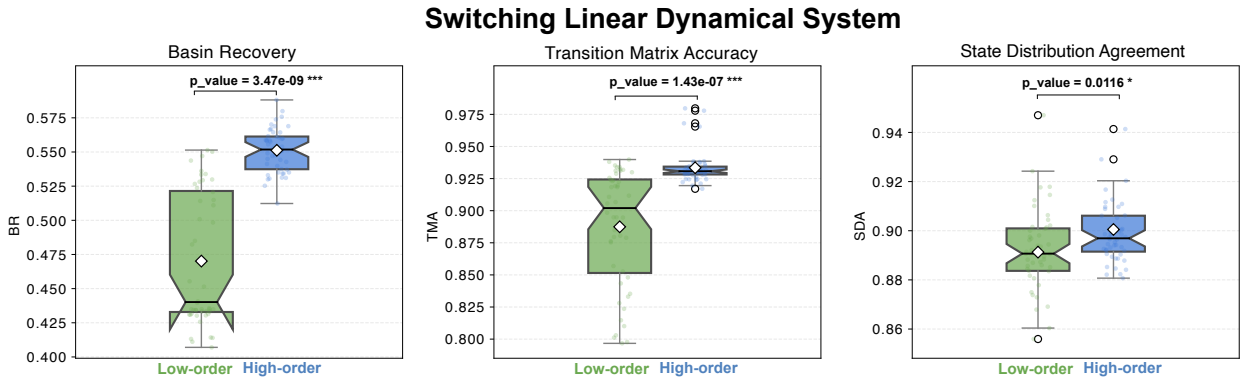


Figure 2.3: Comparison of LowDEL and HODEL on Basin Recovery (BR), Transition Matrix Accuracy (TMA), and State Distribution Agreement (SDA) for the SLDS simulations. Paired Wilcoxon signed-rank p-values are shown in each panel.

2.3.3 Computational Complexity And Memory Requirements

Computational cost arises from fitting the maximum entropy model and evaluating the energy landscape. For N regions, the model has $N(N-1)/2+N$ parameters (or $K(K-1)/2+K$ for the grouped model with K ensembles). Given $(\mathbf{W}, \mathbf{h})$, computing the energy of a configuration x costs $\mathcal{O}(\mathrm{nnz}(\mathbf{W}))$ in the sparse case or $\mathcal{O}(N^2)$ in the dense case, and evaluating T samples scales as $\mathcal{O}\big(T\,\mathrm{nnz}(\mathbf{W})\big)$. Any step requiring exact scoring or enumeration in the grouped space $\{0,1\}^K$ is exponential in K, with total cost $\mathcal{O}\big(2^K\,\mathrm{nnz}(\mathbf{W}_{\text{group}})\big)$ and memory

$\Omega(2^K)$. Grouping therefore reduces exponential dependence from 2^N to 2^K and quadratic dependence from N^2 to K^2, making exhaustive landscape mapping tractable when $K \ll N$.

2.4 Conclusion

This chapter introduces the high-order energy landscape model by exact summation in the reduced state space $\{0,1\}^K$, which is feasible when K is small. Grouping reduces the parameter scale from $\mathcal{O}(N^2)$ to $\mathcal{O}(K^2)$ and replaces the exponential dependence on 2^N with 2^K. As a result, exact normalization, exact model moments, and full landscape enumeration, often intractable in the original parcel space, become practical in the grouped space. Because HODEL and LowDEL use the same discrete Maximum Entropy model and the same mode-extraction procedure, the performance gains come from how the data are represented. Grouping regions into ensembles boosts the effective signal-to-noise ratio, reduces unstable near-threshold flips, and produces cleaner minima in Hamming space. These advantages translate into more accurate basin recovery, more precise transition structure, and better occupancy alignment in both the Kuramoto and SLDS simulations.

The mechanism is similar across both simulation regimes. In the Kuramoto setting, coordinated phase dynamics operate mainly at the ensemble level, and the grouped variables capture these mesoscale patterns more effectively than parcel-wise binarization. In the SLDS setting, grouping reduces within-state variability and increases the separation between regime means, leading to clearer minima and more accurate macro-transition estimates even without strong higher-order interactions. Together, these results show that modeling at the ensemble level produces a more stable and interpretable energy landscape, one that is easier to enumerate, more reflective of the underlying dynamics, and consistent with the modular organization of brain networks. This motivates the continuous formulation in the next chapter, which preserves higher order dependencies while avoiding the information loss and artifacts introduced by binarizing the data.

Chapter 3

High-Order Continuous Energy Landscape (HOCEL)

3.1 Continuous Energy Landscape Framework

In this section, we introduce the continuous high-order energy landscape model, in which the complete signal spectrum is used instead of binarizing it. We provide complete derivations of the resulting likelihood, state and prove properties that guarantee normalizability and identifiability, and present the structured precision parameterization based on a graph convolutional network, together with the training objective and gradients. The overall framework is shown in Figure 2.1.

3.1.1 Continuous Formulation of Energy Landscapes

The discrete Ising landscape in Chapter 2 assigns energy to a binary configuration $\mathbf{q} \in \{0,1\}^N$ by

$$E(\mathbf{q}) = -\tfrac{1}{2}\,\mathbf{q}^\top \mathbf{W}\,\mathbf{q} - \mathbf{h}^\top \mathbf{q}, \tag{3.1}$$

where $\mathbf{W}$ is a symmetric pairwise coupling matrix with $W_{ii} = 0$. To obtain a normalizable continuous energy on $\mathbb{R}^N$ that preserves the quadratic interaction and linear field while

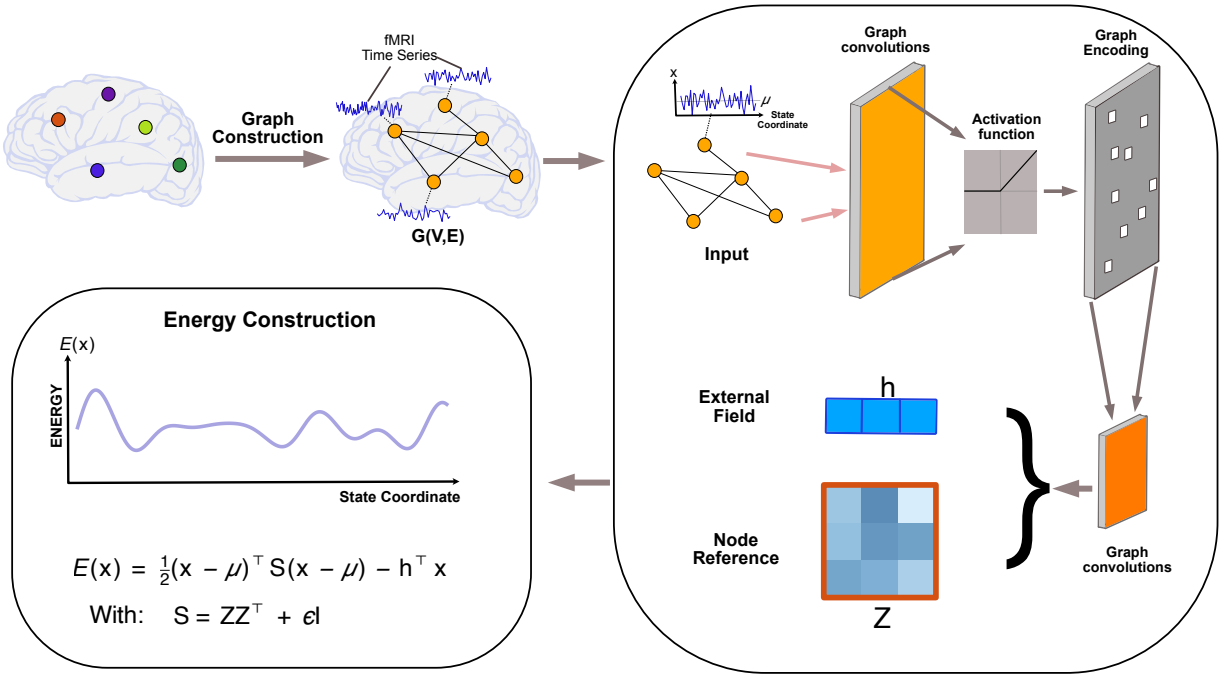


Figure 3.1: **High-order continuous energy landscape framework.** **(A)** *Graph Construction:* Standardized time series $\mathbf{x}(t)$ and baseline $\boldsymbol{\mu}$ define the degree-normalized adjacency matrix $\tilde{\mathbf{A}}$. Each data group uses its own $\tilde{\mathbf{A}}$ built from $\mathbf{x}(t)$ with frozen hyperparameters. **(B)** *Embedding Generation:* A frozen trunk maps $\tilde{\mathbf{A}}$ to node embeddings $\mathbf{V}$. A two-layer GCN produces observation embeddings $\mathbf{H}$, and linear heads output latent factors $\mathbf{Z}$ and external field $\boldsymbol{h}$. **(C)** *Energy Landscape:* The precision matrix $\mathbf{S}$ combines latent factors and regularization. The single-well energy depends on deviation from baseline and external field, and its minimum represents the most probable configuration.

eliminating binarization, we introduce a continuous configuration vector:

$$\mathbf{x} = \boldsymbol{\mu} + \tilde{\mathbf{y}},$$

where $\mathbf{x} \in \mathbb{R}^N$ denotes the continuous configuration (a relaxation of the discrete state $\mathbf{q}$), $\boldsymbol{\mu}$ is the baseline mean (subject-level average), and $\tilde{\mathbf{y}}$ represents deviations from this baseline.

We extend (3.1) to $\mathbb{R}^N$ via a second-order Taylor expansion around $\boldsymbol{\mu}$:

$$E(\mathbf{x}) \approx E_0 + \mathbf{g}^\top(\mathbf{x} - \boldsymbol{\mu}) + \tfrac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \mathcal{H}(\mathbf{x} - \boldsymbol{\mu}), \tag{3.2}$$

where

$$\mathbf{g} = \nabla E(\mathbf{x})\Big|_{\mathbf{x}=\boldsymbol{\mu}} = -\mathbf{W}\boldsymbol{\mu} - \mathbf{h}, \qquad \mathcal{H} = -\mathbf{W}.$$

The Hessian $\mathcal{H} = -\mathbf{W}$ may not be positive-definite, so the energy could diverge to $-\infty$. To ensure integrability, we replace $\mathcal{H}$ with a positive-definite precision matrix:

$$\mathbf{S} = \alpha\mathbf{I} - \mathbf{W}, \qquad \alpha > \lambda_{\max}(\mathbf{W}), \tag{3.3}$$

where $\mathbf{S}$ is the precision matrix, $\mathbf{I}$ is the identity matrix, and $\lambda_{\max}(\mathbf{W})$ denotes the largest eigenvalue of $\mathbf{W}$. This construction preserves interaction directions while guaranteeing $\mathbf{S} \succ 0$. Absorbing constants and linear terms gives:

$$E(\mathbf{x}) = \tfrac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \mathbf{S}(\mathbf{x} - \boldsymbol{\mu}) - \mathbf{h}^\top \mathbf{x}. \tag{3.4}$$

Completing the square:

$$E(\mathbf{x}) = \tfrac{1}{2}(\mathbf{x} - \boldsymbol{\mu} - \mathbf{S}^{-1}\mathbf{h})^\top \mathbf{S}(\mathbf{x} - \boldsymbol{\mu} - \mathbf{S}^{-1}\mathbf{h}) - \tfrac{1}{2}\mathbf{h}^\top \mathbf{S}^{-1}\mathbf{h} - \mathbf{h}^\top \boldsymbol{\mu}. \tag{3.5}$$

Thus, $p(\mathbf{x}) \propto e^{-E(\mathbf{x})}$ is a multivariate Gaussian with mean $\boldsymbol{\mu} + \mathbf{S}^{-1}\mathbf{h}$ and covariance $\mathbf{S}^{-1}$. The normalization constant involves $\det(\mathbf{S})$, ensuring the density is well-defined.

Because $\mathbf{S} \succ 0$, the energy $E(\mathbf{x})$ is strictly convex with a unique minimizer:

$$\mathbf{x}_* = \boldsymbol{\mu} + \mathbf{S}^{-1}\mathbf{h}, \qquad E_{\min} = -\mathbf{h}^\top \boldsymbol{\mu} - \tfrac{1}{2}\mathbf{h}^\top \mathbf{S}^{-1}\mathbf{h}.$$

Completing the square gives:

$$E(\mathbf{x}) = \tfrac{1}{2}(\mathbf{x} - \mathbf{x}_*)^\top \mathbf{S}(\mathbf{x} - \mathbf{x}_*) + E_{\min},$$

so $E(\mathbf{x}) \geq E_{\min} > -\infty$. Consequently, $p(\mathbf{x})$ is a well-defined multivariate Gaussian on $\mathbb{R}^N$.

3.1.2 Graph Parametrization of the Precision Matrix

Optimizing an unconstrained precision matrix $\mathbf{S} \in \mathbb{R}^{K \times K}$ directly requires estimating $K(K+1)/2$ parameters, which can lead to overfitting when the number of time points T is large. To mitigate this, we represent $\mathbf{S}$ using a low-rank structure, learned via a graph convolutional network (GCN) defined on an ensemble graph. In order to construct this graph, we first compute the Pearson correlation matrix

$$\mathbf{C} \in \mathbb{R}^{K \times K}$$

from variance-normalized local fields $u(t)$. To enforce comparable sparsity across subjects, we fix an edge density $\psi \in (0, 1)$ and threshold the off-diagonal absolute correlations at the $(1 - \psi)$-quantile:

$$\tau = \text{quantile}\Big(\{|C_{ij}| : i < j\},\, 1 - \psi\Big), \qquad A_{ij} = \mathbb{1}\{|C_{ij}| \geq \tau\}.$$

This yields a binary adjacency matrix

$$\mathbf{A} \in \{0, 1\}^{K \times K}$$

with edge density ψ. Next, define the weighted adjacency matrix

$$\mathbf{G} \in \mathbb{R}^{K \times K}, \qquad G_{ij} = |C_{ij}|\, A_{ij}.$$

Compute the diagonal degree matrix of $(\mathbf{I} + \mathbf{G})$:

$$D_{ii} = \sum_j (\mathbf{I} + \mathbf{G})_{ij}.$$

Finally, construct the degree-normalized adjacency matrix:

$$\tilde{\mathbf{A}} = \mathbf{D}^{-1/2}(\mathbf{I} + \mathbf{G})\mathbf{D}^{-1/2}.$$

We now define the graph convolutional network (GCN) used to learn node embeddings $\mathbf{V}$ and the precision matrix $\mathbf{S}$. Let $\boldsymbol{\Xi} \in \mathbb{R}^{N \times f_0}$ denote the node feature matrix, where N is the total number of regions and f_0 is the input feature dimension. When no exogenous features are available, we set $\boldsymbol{\Xi} = \mathbf{I}$, the $N \times N$ identity matrix. Two graph convolution layers produce embeddings:

$$\mathbf{V}_1 = \varphi\Big(\tilde{\mathbf{A}}\,\boldsymbol{\Xi}\,\boldsymbol{\Theta}_0 + \mathbf{1}\,\boldsymbol{\upsilon}_0^\top\Big), \qquad \mathbf{V}_2 = \varphi\Big(\tilde{\mathbf{A}}\,\mathbf{V}_1\,\boldsymbol{\Theta}_1 + \mathbf{1}\,\boldsymbol{\upsilon}_1^\top\Big) \equiv \mathbf{V} \in \mathbb{R}^{N \times \ell}, \tag{3.6}$$

where $\tilde{\mathbf{A}} \in \mathbb{R}^{N \times N}$ is the degree-normalized adjacency matrix, $\boldsymbol{\Theta}_0 \in \mathbb{R}^{f_0 \times \ell}$ and $\boldsymbol{\Theta}_1 \in \mathbb{R}^{\ell \times \ell}$ are layer-specific weight matrices, $\boldsymbol{\upsilon}_0, \boldsymbol{\upsilon}_1 \in \mathbb{R}^\ell$ are bias vectors, $\varphi(\cdot)$ denotes a pointwise nonlinearity (ReLU unless stated otherwise), and $\mathbf{1} \in \mathbb{R}^N$ is the all-ones vector used for bias broadcasting.

Then, a linear head maps node embeddings to q latent factors, producing the latent feature matrix $\mathbf{Z}$:

$$\mathbf{Z} = \mathbf{V}\,\boldsymbol{\Theta}_Z + \mathbf{1}\,\boldsymbol{\upsilon}_Z^\top \;\in\; \mathbb{R}^{K \times q}, \tag{3.7}$$

where $\mathbf{V} \in \mathbb{R}^{K \times \ell}$ is the node embedding matrix, $\boldsymbol{\Theta}_Z \in \mathbb{R}^{\ell \times q}$ is the weight matrix, and $\boldsymbol{\upsilon}_Z \in \mathbb{R}^q$ is the bias vector.

The precision matrix is then defined as:

$$\mathbf{S} = \mathbf{Z}\mathbf{Z}^\top + \kappa\,\mathbf{I}, \qquad \kappa > 0, \tag{3.8}$$

where $\mathbf{S}$ denotes the precision matrix, $\mathbf{Z} \in \mathbb{R}^{K \times q}$ is the latent representation matrix, κ is the between-state separation parameter, and $\mathbf{I}$ is the identity matrix. This formulation guarantees that $\mathbf{S} \succ 0$ (positive definite) for any $\mathbf{Z}$ and any $\kappa > 0$.

A separate linear head maps the node embeddings $\mathbf{V}$ to the external field $\mathbf{h} \in \mathbb{R}^K$:

$$\mathbf{h} = \mathbf{V}\,\boldsymbol{\upsilon}_h + \tau_h\,\mathbf{1}, \tag{3.9}$$

where $\mathbf{V}$ is the node embedding matrix, $\boldsymbol{\upsilon}_h \in \mathbb{R}^\ell$ is the weight vector associated with the field mapping, and $\tau_h \in \mathbb{R}$ is a scalar offset. Here, $\mathbf{1}$ denotes the all-ones vector.

We include all trainable weights and biases from the GCN and the linear heads into the parameter vector $\boldsymbol{\theta}$. Although the precision matrix $\mathbf{S}$ and the external field $\mathbf{h}$ do not explicitly include $\boldsymbol{\theta}$ in their definitions (e.g., (3.8)), they depend on $\boldsymbol{\theta}$ through the latent representation matrix $\mathbf{Z}(\boldsymbol{\theta})$ and the mapping heads. Hence, the overall loss function becomes:

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}) \;=\; \tfrac{1}{2}\sum_{t=1}^{T}\Big(u(t)-\boldsymbol{\mu}-\mathbf{S}(\boldsymbol{\theta})^{-1}\mathbf{h}(\boldsymbol{\theta})\Big)^{\top}\mathbf{S}(\boldsymbol{\theta})\Big(u(t)-\boldsymbol{\mu}-\mathbf{S}(\boldsymbol{\theta})^{-1}\mathbf{h}(\boldsymbol{\theta})\Big) \\ -\;\tfrac{T}{2}\,\log|\mathbf{S}(\boldsymbol{\theta})| \;+\; \lambda\,\|\mathbf{S}(\boldsymbol{\theta})\|_F^2.\end{aligned} \tag{3.10}$$

This loss combines three components: the first term measures the Mahalanobis distance between observed local fields $\mathbf{u}(t)$ and their model-predicted mean under the precision matrix $\mathbf{S}$; the second term, the log-determinant, corresponds to the normalization constant in a Gaussian likelihood and encourages well-conditioned precision matrices; and the third term, the Frobenius norm penalty $\|\mathbf{S}\|_F^2$, stabilizes training and prevents overfitting by controlling the magnitude of $\mathbf{S}$.

Once the loss function $\mathcal{L}$ is defined, we derive the gradients required for optimization. The gradients with respect to the external field $\mathbf{h}$ and the precision matrix $\mathbf{S}$ are:

$$\frac{\partial\mathcal{L}}{\partial\mathbf{h}} \;=\; -\sum_{t=1}^{T}\mathbf{e}(t), \qquad \frac{\partial\mathcal{L}}{\partial\mathbf{S}} \;=\; \tfrac{1}{2}\Big(\widehat{\boldsymbol{\Sigma}}_{\mathbf{e}} + \mathbf{r}\,(\mathbf{S}^{-1}\mathbf{h})^{\top} + (\mathbf{S}^{-1}\mathbf{h})\,\mathbf{r}^{\top}\Big) - \tfrac{T}{2}\,\mathbf{S}^{-1} + 2\,\vartheta\,\mathbf{S}, \tag{3.11}$$

where

$$\mathbf{e}(t) \;=\; \mathbf{u}(t) - \boldsymbol{\mu} - \mathbf{S}^{-1}\mathbf{h}, \qquad \widehat{\boldsymbol{\Sigma}}_{\mathbf{e}} \;=\; \frac{1}{T}\sum_{t=1}^{T}\Big(\mathbf{e}(t) - \bar{\mathbf{e}}\Big)\Big(\mathbf{e}(t) - \bar{\mathbf{e}}\Big)^{\top}, \qquad \mathbf{r} \;=\; \sum_{t=1}^{T}\Big(\mathbf{u}(t) - \boldsymbol{\mu}\Big). \tag{3.12}$$

The $\bar{\mathbf{e}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{e}(t)$ denotes the average residual. The first term in $\partial\mathcal{L}/\partial\mathbf{S}$ captures the covariance structure and its interaction with the external field; the second term arises from the log-determinant component; and the last term corresponds to Frobenius norm regularization.

To compute gradients with respect to the parameters $\boldsymbol{\theta}$, we use the chain rule because both $\mathbf{h}$ and $\mathbf{S}$ depend on $\boldsymbol{\theta}$. Specifically,

$$\mathbf{S}(\boldsymbol{\theta}) = \mathbf{Z}(\boldsymbol{\theta})\,\mathbf{Z}(\boldsymbol{\theta})^{\top} + \vartheta\,\mathbf{I}, \tag{3.13}$$

where $\mathbf{Z}$ is the latent representation matrix and ϑ is the regularization strength.

Let

$$\boldsymbol{\Gamma} \;=\; \frac{\partial\mathcal{L}}{\partial\mathbf{S}}, \tag{3.14}$$

be the gradient of the loss $\mathcal{L}$ with respect to the precision matrix $\mathbf{S}$. Since

$$d\mathbf{S} \;=\; d\mathbf{Z}\,\mathbf{Z}^{\top} \;+\; \mathbf{Z}\,d\mathbf{Z}^{\top}, \tag{3.15}$$

the gradient with respect to $\mathbf{Z}$ becomes

$$\frac{\partial\mathcal{L}}{\partial\mathbf{Z}} \;=\; 2\,\mathrm{sym}(\boldsymbol{\Gamma})\,\mathbf{Z}, \qquad \mathrm{sym}(\boldsymbol{\Gamma}) \;=\; \tfrac{1}{2}\Big(\boldsymbol{\Gamma} + \boldsymbol{\Gamma}^{\top}\Big). \tag{3.16}$$

We use the symmetric part

$$\mathrm{sym}(\boldsymbol{\Gamma}) = \tfrac{1}{2}\Big(\boldsymbol{\Gamma} + \boldsymbol{\Gamma}^{\top}\Big)$$

during optimization, where $\mathrm{sym}(\boldsymbol{\Gamma})$ enforces symmetry on the $\mathbf{S}$ path. The explicit depen-

dence of $\mathbf{e}(t)$ on $\mathbf{S}^{-1}\mathbf{h}$ contributes through

$$\frac{\partial \mathcal{L}}{\partial \mathbf{S}} \quad \text{and} \quad \frac{\partial \mathcal{L}}{\partial \mathbf{h}}$$

as in (3.11). Gradients with respect to $\mathbf{W_Z}$ and $\mathbf{b_Z}$ then follow from

$$\frac{\partial \mathbf{Z}}{\partial \mathbf{W_Z}} = \mathbf{H}, \qquad \frac{\partial \mathbf{Z}}{\partial \mathbf{b_Z}} = \mathbf{1},$$

and the remaining parameters backpropagate through the graph convolutions.

3.1.3 Mixture-Based Energy Formulation for Multi-Basin Dynamics

Building on the previous formulation, where the precision matrix $\mathbf{S}$ and external field $\mathbf{h}$ were parameterized through the latent representation $\mathbf{Z}(\boldsymbol{\theta})$, we now extend the model to capture multiple recurring basins in the configuration space $\mathcal{X}$. This extension introduces a mixture-based energy formulation that generalizes the single-basin structure to a multimodal landscape. To capture multiple recurring basins in $\mathbb{R}^K$, the energy is defined as the negative log of a finite Gaussian mixture. Let R denote the number of mixture components. Each component $r \in \{1, \ldots, R\}$ has precision matrix $\mathbf{S}_r \succ 0$ and external field $\mathbf{h}_r$, with a shared baseline mean $\boldsymbol{\mu}$. The model-implied mean of component r is:

$$\boldsymbol{\mu}_r = \boldsymbol{\mu} + \mathbf{S}_r^{-1}\mathbf{h}_r. \tag{3.17}$$

We denote the Gaussian density in $\mathbb{R}^K$ with mean vector $\boldsymbol{\mu}$ and precision matrix $\mathbf{S}$ by $\varkappa(\mathbf{x}; \boldsymbol{\mu}, \mathbf{S})$, where $\mathbf{x}$ represents the brain configuration. The mixture weights $\pi_r > 0$ satisfy

$\sum_{r=1}^{R} \pi_r = 1$. The Gaussian mixture-based energy is then defined as:

$$E(\mathbf{x}) = -\log\left(\sum_{r=1}^{R} \pi_r \, \mathcal{N}\big(\mathbf{x}; \boldsymbol{\mu}_r, \mathbf{S}_r\big)\right). \tag{3.18}$$

and the normalized density is defined as:

$$p(\mathbf{x}) = \sum_{r=1}^{R} \pi_r \, \mathcal{N}\big(\mathbf{x}; \boldsymbol{\mu}_r, \mathbf{S}_r\big). \tag{3.19}$$

This construction is integrable over $\mathbb{R}^K$, generally nonconvex, and admits multiple wells centered near $\boldsymbol{\mu}_r$, with anisotropy governed by $\mathbf{S}_r$.

The negative log-likelihood with a Frobenius penalty is then:

$$\mathcal{L}(\boldsymbol{\Theta}) = -\sum_{t=1}^{T} \log\left(\sum_{r=1}^{R} \pi_r \, \mathcal{N}\big(\mathbf{x}(t); \boldsymbol{\mu}_r, \mathbf{S}_r\big)\right) + \vartheta \sum_{r=1}^{R} \|\mathbf{S}_r\|_F^2. \tag{3.20}$$

Define the residual for component r at time t as $\boldsymbol{r}_r(t) = \mathbf{x}(t) - \boldsymbol{\mu}_r - \mathbf{S}_r^{-1}\boldsymbol{h}_r$, and let $\mathcal{R}_r = \sum_{t=1}^{T} \digamma_{t,r}$ denote the total responsibility for component r. The gradients of the mixture-based loss take the same structural form as the single-basin case in Eq. (3.11), but are weighted by responsibilities $\digamma_{t,r}$:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \boldsymbol{h}_r} &= -\sum_{t=1}^{T} \digamma_{t,r} \, \boldsymbol{r}_r(t), \\
\frac{\partial \mathcal{L}}{\partial \mathbf{S}_r} &= \tfrac{1}{2} \sum_{t=1}^{T} \digamma_{t,r}\Big(\boldsymbol{r}_r(t)\boldsymbol{r}_r(t)^\top + \boldsymbol{r}_r(t)(\mathbf{S}_r^{-1}\boldsymbol{h}_r)^\top + (\mathbf{S}_r^{-1}\boldsymbol{h}_r)\boldsymbol{r}_r(t)^\top\Big) - \tfrac{\mathcal{R}_r}{2}\,\mathbf{S}_r^{-1} + 2\vartheta\,\mathbf{S}_r.
\end{aligned} \tag{3.21}$$

The weight gradient is:

$$\frac{\partial \mathcal{L}}{\partial \pi_r} = -\sum_{t=1}^{T} \frac{\digamma_{t,r}}{\pi_r},$$

with the simplex constraint enforced through a softmax parameterization. In practice, the heads output $\mathbf{Z}_r$ and $\boldsymbol{h}_r$, and we set $\mathbf{S}_r = \mathbf{Z}_r\mathbf{Z}_r^\top + \epsilon\mathbf{I}$. The determinant lemma and Woodbury identity provide stable and efficient evaluation of $\log|\mathbf{S}_r|$ and $\mathbf{S}_r^{-1}v$ per component, making the mixture objective fully differentiable and tractable.

3.2 Simulation Results: Continuous versus Discrete

We compare HOCEL to the discrete Ising MEM baseline using the same generators, grids, and paired testing protocol defined in Section 2.2. Metrics are Basin Recovery, Transition Matrix Accuracy, and State Distribution Agreement. Between-method tests use the paired Wilcoxon signed-rank procedure with Benjamini–Hochberg control across factorized tests.

3.2.1 Switching Linear Dynamical System

Across all SLDS conditions and seeds, HOCEL shows higher performance on all three metrics. Paired Wilcoxon p-values annotated in the figure are 2.91×10^{-28} for Basin Recovery, 1.18×10^{-89} for Transition Matrix Accuracy, and 5.65×10^{-90} for State Distribution Agreement. The paired distributions in Figure 3.2 show consistent gains across the grid.

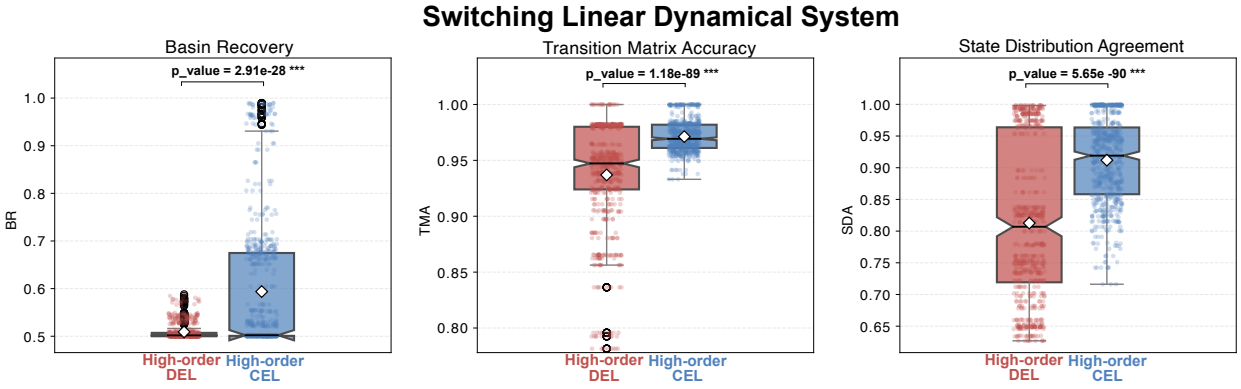


Figure 3.2: SLDS simulations comparing the discrete baseline and HOCEL on Basin Recovery, Transition Matrix Accuracy, and State Distribution Agreement across the simulation grid. Bars summarize paired results across repeats. Paired Wilcoxon p-values shown in-panel are 2.91×10^{-28}, 1.18×10^{-89}, and 5.65×10^{-90}.

3.2.2 Kuramoto Oscillator Network

On the nonlinear Kuramoto generator, paired Wilcoxon p-values in Figure 3.3 are 0.0311 for Basin Recovery, 1.11×10^{-13} for Transition Matrix Accuracy, and 1.66×10^{-103} for State Distribution Agreement. Differences are therefore significant on all three metrics in the

continuous high-order setting. The panels show the direction and magnitude of the paired effects across the grid.

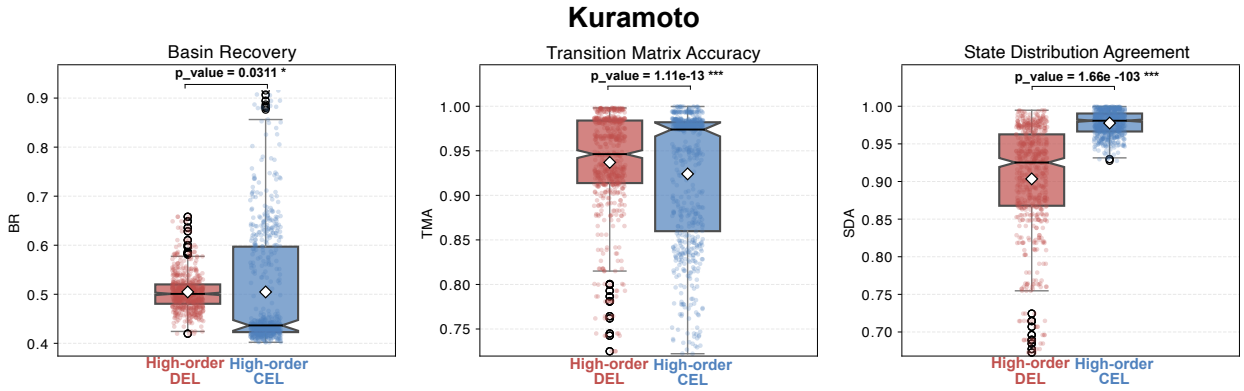


Figure 3.3: Kuramoto simulations comparing the discrete baseline and HOCEL on Basin Recovery, Transition Matrix Accuracy, and State Distribution Agreement across the simulation grid. Bars summarize paired results across repeats. Paired Wilcoxon p-values shown in-panel are 0.0311, 1.11×10^{-13}, and 1.66×10^{-103}.

3.3 Conclusion

This chapter introduces the high-order continuous energy landscape (HOCEL) model, which extends the discrete approach by operating directly on amplitudes $\mathbf{x} \in \mathbb{R}^K$. Using mixture wells, the continuous formulation improves recovery of basin geometry and occupancy and often matches or exceeds transition fidelity compared to the discrete baseline. It avoids thresholding, preserves graded cofluctuations, and replaces the intractable partition function $\mathcal{Z}$ with a closed-form Gaussian likelihood, producing well-calibrated energies $E(\mathbf{x})$. The precision matrix is defined as $\mathbf{S} = \mathbf{Z}\mathbf{Z}^\top + \epsilon\mathbf{I}$, where $\mathbf{Z}$ is the latent representation and ϵ ensures positive definiteness. The Gaussian mixture components introduce multiple wells without sacrificing analytic gradients $\nabla E(\mathbf{x})$. Together, these design choices stabilize training and align with best practices for transparent modeling and reliable time-varying analyses.

The mechanism reflects the cost of dichotomization: thresholding suppresses amplitude variance and compresses covariances, whereas HOCEL retains amplitude structure through quadratic terms and external fields. Mixture components define basin centers and curvature,

linking geometry to kinetics and supporting interpretable metrics such as occupancy and transition probabilities. This representation aligns with network neuroscience and clinical practice, where amplitude-preserving features capture subject-level variability lost under binarization. Together, these results show that continuous modeling produces a more stable and interpretable energy landscape than discrete approaches.

Chapter 4

Applications in Predicting Cognitive Function in Brain Tumor Patients

The previous chapters developed energy–landscape models that summarize whole–brain dynamics through a scalar function $E(\cdot)$ whose local minima behave like putative brain states and whose geometry constrains reconfiguration. In discrete form, a pairwise Maximum Entropy (Ising) model captures recurring binary configurations; in continuous form, a quadratic energy over real–valued signals yields Gaussian basins with explicit centers and curvatures. In both cases, high–dimensional time series are compressed into a small set of interpretable descriptors: which states are expressed, how often they occur, and how readily the system moves between them [3].

In this chapter, we will utilize the developed energy landscape models to predict postoperative cognitive decline in brain tumor patients using resting-state fMRI data. The main hypothesis is that the energy landscape of brain states provides critical insights into the stability and flexibility of neural dynamics. Specifically, we propose that alterations in the energy landscape, such as reduced basin depth or increased transition probabilities, reflect impaired network integration and segregation, which are associated with cognitive decline following surgery. By modeling these changes, we aim to identify predictive biomarkers that

can inform personalized treatment strategies and improve postoperative outcomes.

4.1 Data Description

We used a publicly available dataset of adults with intracranial tumors who underwent multimodal MRI and neuropsychological testing at baseline and approximately six months after surgical resection [47]. After quality control, data from a total of 20 patients were included in the analysis. Table 4.1 summarizes demographics, laterality, and tumor volume.

Table 4.1: Summary of Patient Cohort.

Characteristic	Value
Age (years)	57.8 ± 12.6 [39–77]
Sex	14 female (70%), 6 male (30%)
Tumor types present	Glioma (35%), Meningioma (65%)
Tumor laterality	Left 9 (45%); Right 9 (45%); Bilateral 2 (10%); Midline 0 (0%)
Tumor size (cm^3)	median 12.88 [IQR: 2.12–24.03]

All imaging was acquired on a 3T system [47]. Structural T1–weighted scans were obtained at 1 mm isotropic resolution. Resting–state fMRI used gradient–echo EPI with repetition time approximately 2100–2400 ms, echo time approximately 27 ms, and voxel size $3 \times 3 \times 3\,\mathrm{mm}^3$, with a few hundred volumes per run.

Preprocessing followed the original pipeline in [47]: slice–timing correction, motion correction, brain extraction, coregistration to T1 anatomy, and spatial normalization to MNI space. High–pass temporal filtering at approximately 0.01 Hz mitigated slow drifts. ROI time series were extracted using the FreeSurfer Desikan–Killiany cortical atlas (fsaverage), excluding voxels overlapping the tumor mask before ROI averaging. Unless otherwise noted, LowDEL uses ROIs from the Default Mode (DMN), Salience (SN), and Limbic networks, whereas HighDEL/HighCEL use all cortical ROIs grouped into K ensembles.

Neuropsychological testing was performed at baseline and repeated approximately six months after surgical resection [47], with the six-month scores serving as the prediction

targets in our framework. The three domains of measured cognitive function include working memory, executive function, and processing speed. Working memory is indexed by the Spatial Span (SSP) score, which records the maximum reproduced span of visuospatial locations and typically ranges from 2 to 9; higher scores indicate better performance. Executive function is assessed by a composite derived from the Stockings of Cambridge (SOC) task, which measures planning and problem-solving ability; higher composite values indicate more efficient planning. Processing speed is indexed by reaction time (RT) in milliseconds; lower values indicate faster responses. For modeling, six-month SSP and SOC scores are analyzed as balanced binary outcomes. When clinical or cohort-conventional thresholds are available, those thresholds are adopted. Otherwise, to avoid data leakage and preserve class balance, the threshold is set as the median computed on training subjects within each outer split and then applied to the held-out subject. Reaction time (RT) remains continuous and is z-scored for model fitting with results reported both in standardized units and in milliseconds.

Table 4.2: Summary of Cognitive Measures Used in the Study.

Measure	Construct	Raw unit / typical range	Use in this work
Spatial Span (SSP)	Working memory (visuospatial span)	Integer span, typically 2–9	Six-month score; binarized by the training-split median within each outer fold (balanced)
Stockings of Cambridge (SOC composite)	Executive planning/problem solving	Bounded composite (approx. 2–11)	Six-month score; binarized by the training-split median within each outer fold (balanced)
Reaction Time (RT)	Processing speed	Milliseconds (ms)	Six-month score; continuous; standardized for modeling and reported in ms

4.2 Prediction Framework

4.2.1 Feature Extraction

Once the pre-operative fMRI data are preprocessed, the regional time series serve as inputs to the energy landscape modeling framework. Different energy landscape pipelines produce patient-level feature vectors that are subsequently used in the predictive machine learning pipeline. Specifically, we utilize Low-Order Discrete Energy Landscape (LowDEL) using canonical brain networks as the baseline, High-Order Discrete Energy Landscapes (HODEL), and High-Order Continuous Energy Landscapes (HOCEL).

LowDEL implements a compact, hypothesis–driven discrete landscape on a small subset of ROIs in Default Mode, Salience, and Limbic. Regional time series are variance–normalized and binarized per region; a pairwise maximum–entropy model is fitted; local minima are identified; and time points are assigned to basins by iterative energy descent. Patient–level features include the number of recovered minima, occupancy proportions of the top basins, the mean and variance of energies among occupied states, a flexibility index defined as one minus the average self–transition probability in the empirical transition matrix, and the average row entropy of that matrix. When enumeration at this reduced dimensionality is feasible, calibration summaries are also recorded.

HighDEL applies the same discrete landscape after data–driven grouping as in Chapter 2. All available regions are clustered into K ensembles, with K much smaller than the number of ROIs. Within each ensemble, activity is pooled, binarized, and modeled with the same pairwise form. Basin identification, time–point labeling, and the feature menu match LowDEL to enable like–for–like comparison. Sensitivity to grouping level is assessed by one–up and one–down variants around the prespecified K.

HighCEL applies the continuous model of Chapter 3 using the identical grouping partition as HighDEL. Within each ensemble, activity remains real–valued and is modeled by a unimodal quadratic energy whose precision is parameterized via a shared graph–convolutional

trunk with a linear head for the external field. Patient–level features parallel the discrete one for comparability and include counts of active ensembles, occupancy proxies from the distribution of normalized energies (e.g., ensemble-wise means and inter-quartile ranges of Energy values), a flexibility index computed from low-energy dwell fractions, and a transition entropy computed from empirical transitions between low-energy quantile regions. Using the identical grouping isolates the effect of continuous versus discrete modeling while keeping the real-data CEL unimodal as in the paper.

4.2.2 Model Training and Evaluation

We utilized a random forest and ridge regression as our predictive models. For classification tasks (SSP and SOC), the random forest models were tuned over the number of trees $\{200, 500, 1000\}$, maximum depth $\{\text{None}, 2, 4, 6, 8\}$, and minimum samples per leaf $\{1, 2, 4\}$ using inner leave-one-subject-out (LOSO) cross-validation. Across outer folds, the classifier produced one out-of-fold probability per subject and outcome; stacking these subject-level probabilities with the corresponding ground-truth labels yielded the vectors used to compute ROC–AUC. For regression (RT), Ridge regression was tuned over $\{10^{-3}, 10^{-2}, \ldots, 10^{3}\}$ on a logarithmic grid. Between–pipeline comparisons use paired tests on subject–level contributions from the outer loop, with metric-appropriate choices: DeLong's test for correlated AUC, McNemar's test for accuracy on discordant pairs, a paired permutation test for F_1, and a two–sided Wilcoxon signed–rank test on per-subject absolute errors for RT; R^2 is reported descriptively. Where families of contrasts are reported within an outcome, the false discovery rate is controlled by the Benjamini–Hochberg procedure at $FDR = 0.05$.

To test whether gains reflect meaningful mesoscale organization rather than group–count alone, a random–network control is included. For each patient and each outer split, ROIs are assigned uniformly at random into K groups with the same sizes as the HighDEL/HighCEL partition. The control is run 100 times per split with independent seeds; features are computed for each randomization and carried through the same learning and evaluation pipeline.

Performance distributions across randomizations characterize an empirical null for grouping with the same K and group sizes but no data–driven structure, mirroring the null used in Chapter 2. All preprocessing and model selection occur strictly within the training split of each outer fold. Thresholds and grouping (if applicable), graph operators, and GCN trunk parameters are learned using training subjects only and then frozen before scoring the held-out subject. Feature scaling is fit on training subjects and applied to the held-out subject without refitting. Hyperparameters are chosen by an inner cross-validation loop on training data.

4.3 Results

4.3.1 Working Memory

Figure 4.1 summarizes accuracy across pipelines. HighCEL attains the highest accuracy, with HighDEL close behind, and all three network–subset baselines (Default Mode, Salience, Limbic) substantially lower. The HighCEL–HighDEL difference is not significant by McNemar's test at this sample size. For both grouped pipelines, accuracy improvements over each network–subset baseline remain significant after Benjamini–Hochberg control within the family of comparisons. The point estimate for HighCEL is higher than HighDEL, indicating a consistent trend favoring HighCEL on accuracy.

The F_1 results in Figure 4.2 show the same ordering. HighCEL equals or slightly exceeds HighDEL, and their paired difference is not significant by a subject-level permutation test. For both grouped pipelines, F_1 exceeds each network–subset baseline with adjusted $FDR<0.05$ within the family of comparisons, indicating improvements in both sensitivity and positive predictive value.

Figure 4.3 reports ROC–AUC. HighCEL has the highest average AUC, followed closely by HighDEL; all three baselines are lower. DeLong's paired test does not find a significant AUC difference between HighCEL and HighDEL. For both grouped pipelines, AUC exceeds each

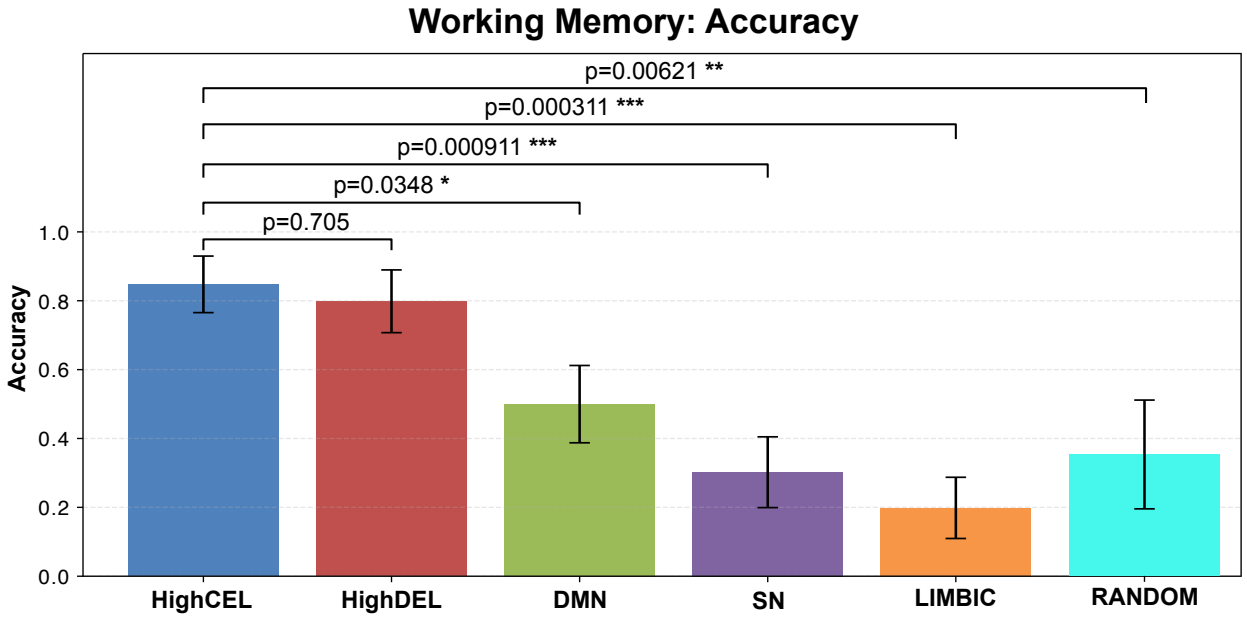


Figure 4.1: Working memory: accuracy across pipelines (mean ± SD across leave–one–out folds). HighCEL is highest and HighDEL close behind; their paired difference is not significant by McNemar's test. For both grouped pipelines, accuracy exceeds each network–subset baseline with Benjamini–Hochberg FDR<0.05 within the family of comparisons.

baseline with Benjamini–Hochberg FDR<0.05, showing that both grouped approaches separate low versus high working–memory outcomes more effectively than the network–subset models.

Random–network control. Across random partitions per outer split, accuracy and F_1 for the random control cluster near or below the network–subset baselines and remain well below HighDEL and HighCEL. The distribution of AUC across randomizations does not overlap the HighCEL point estimate and only rarely overlaps HighDEL. These observations indicate that the grouped gains derive from data–driven structure rather than from the group count K alone.

4.3.2 Executive Function

HighCEL attains the highest mean accuracy across patients (Figure 4.4). Relative to the network–subset baselines, its advantage is clear after Benjamini–Hochberg correction within the

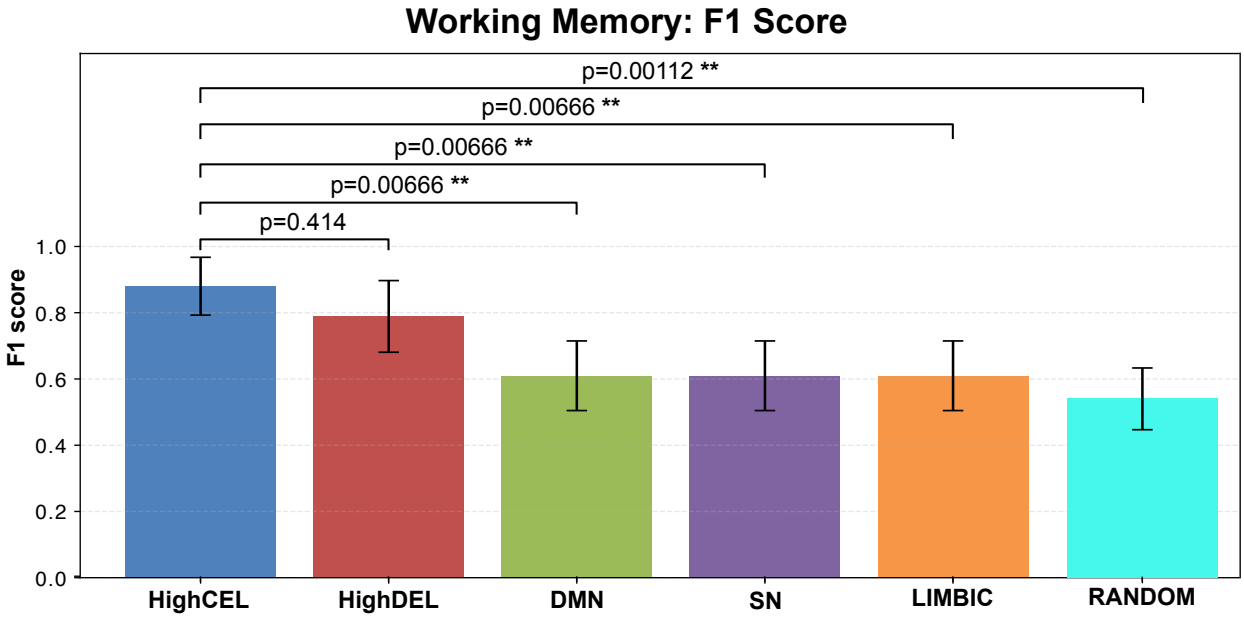


Figure 4.2: Working memory: F_1 across pipelines (mean ± SD across leave–one–out folds). HighCEL matches or slightly exceeds HighDEL; their paired difference is not significant by a subject–level permutation test. Both grouped pipelines exceed the network–subset baselines with adjusted $FDR<0.05$.

family of comparisons. The comparison between HighCEL and HighDEL is not significant at this sample size, although the HighCEL mean remains higher.

F_1 follows the same trend (Figure 4.5). HighCEL yields the highest mean F_1 and materially exceeds two of the three network–subset baselines after adjustment; the differences relative to Salience and to HighDEL are not significant.

ROC–AUC results (Figure 4.6) are consistent with accuracy and F_1. HighCEL achieves the highest mean AUC, followed closely by HighDEL; both grouped pipelines sit above the network–subset baselines. Although the HighCEL–HighDEL gap is modest and not statistically significant here, the ordering is stable across metrics.

Random–network control. As with working memory, the random–grouping control remains well below the grouped pipelines and near the lower end of the network–subset baselines across accuracy, F_1, and AUC. This supports the interpretation that the advantage of HighDEL and HighCEL reflects informative mesoscale grouping rather than an arbitrary

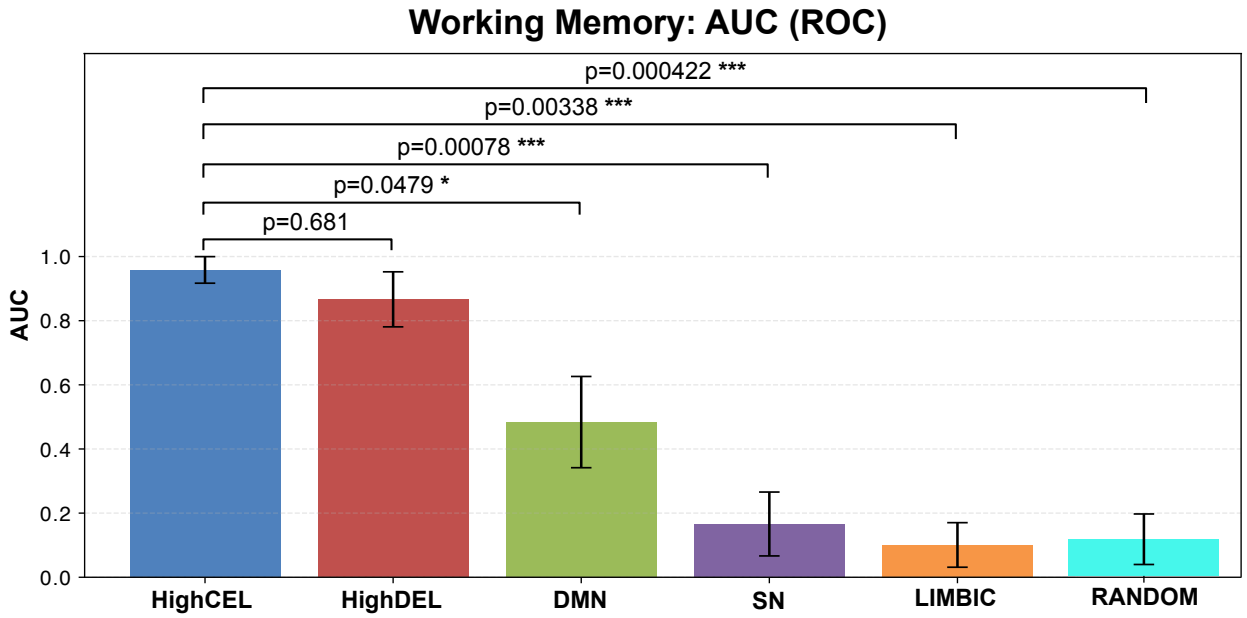


Figure 4.3: Working memory: ROC–AUC across pipelines (mean ± SD across leave–one–out folds). HighCEL attains the highest AUC, followed closely by HighDEL; both grouped pipelines outperform the network–subset baselines. HighCEL vs. HighDEL is not significant by DeLong's test.

aggregation.

4.3.3 Reaction Time

Figure 4.7 shows leave–one–out absolute error in milliseconds. HighCEL attains the lowest dispersion and the lowest median error among all pipelines. Paired Wilcoxon tests on per–patient absolute errors favor HighCEL over each comparator after Benjamini–Hochberg correction within the family of contrasts.

Figure 4.8 reports per–subject R^2 contributions under the same cross–validation. HighCEL achieves the highest median contribution and a clearly positive distribution across patients. Paired Wilcoxon tests again favor HighCEL over every other pipeline after correction. HighDEL improves upon the network–subset baselines but falls short of HighCEL on both central tendency and tail behavior.

Random–network control. Absolute errors from the random–grouping control are consis-

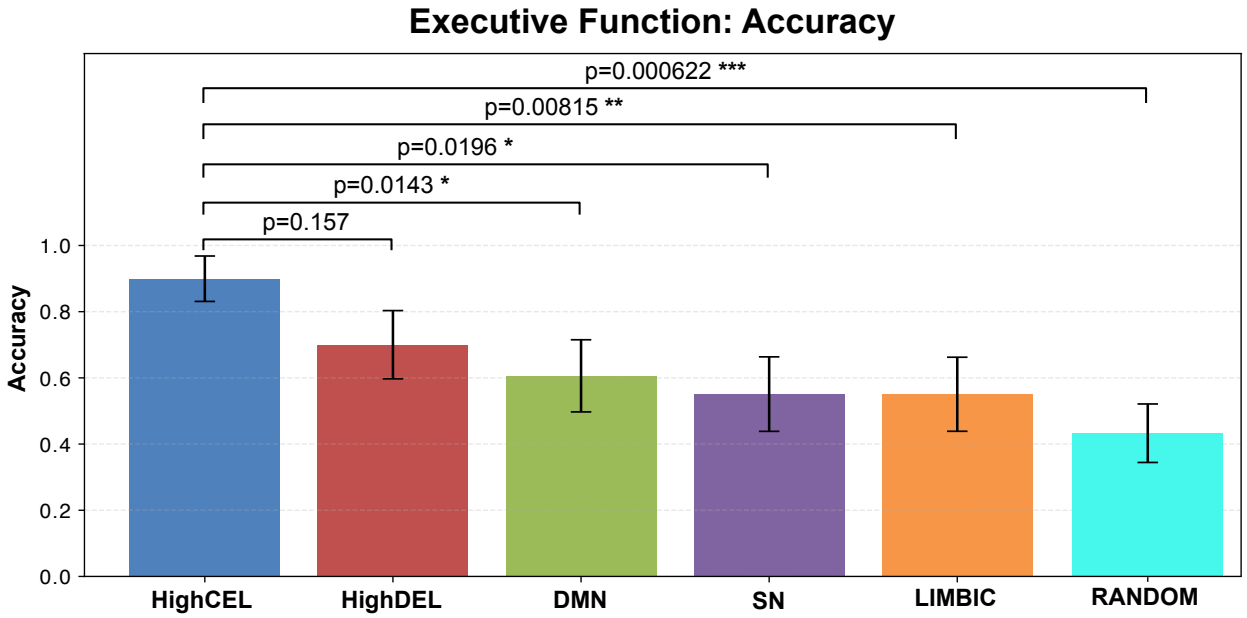


Figure 4.4: Executive function: accuracy under leave–one–out validation (mean ± SD). HighCEL achieves the highest mean accuracy and significantly exceeds the network–subset baselines after Benjamini–Hochberg correction; the difference vs. HighDEL is not significant.

tently larger than those of HighDEL and HighCEL and comparable to, or worse than, the network–subset baselines. The corresponding R^2 contributions are near zero or negative, further indicating that arbitrary groupings do not recover predictive structure.

4.4 Conclusion

Our results confirm the hypothesis that large-scale brain dynamics, as captured by energy landscape models, provide critical insights into the neural stability and flexibility underlying cognitive function in brain tumor patients. Across working memory, executive function, and processing speed, the proposed energy landscape models consistently outperformed small network–subset baselines, with the continuous, amplitude-preserving formulation achieving the highest average performance. These findings suggest that clinically relevant variation is embedded not only in the recurrence of discrete patterns but also in the strength of regional signal fluctuations, their coherence within distributed ensembles, and the trajectories among

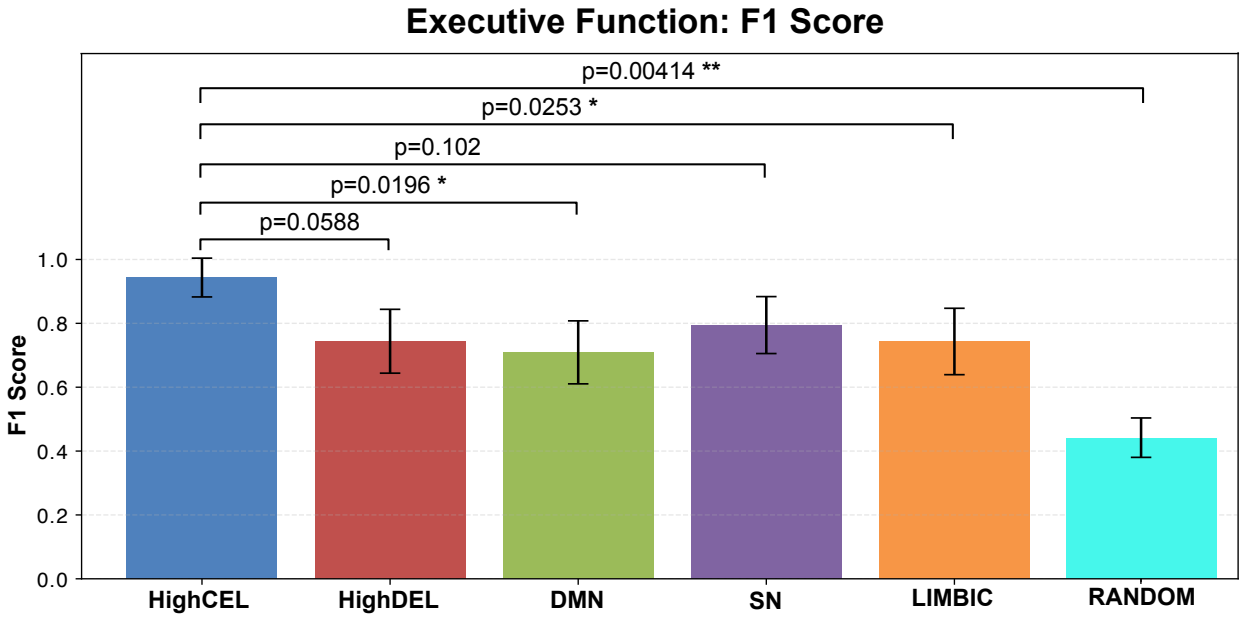


Figure 4.5: Executive function: F_1 under leave–one–out validation (mean ± SD). HighCEL attains the top mean F_1 and exceeds two of the three network–subset baselines after adjustment; the comparisons versus Salience and HighDEL are not significant.

a compact set of metastable configurations.

A recurring practical theme is the importance of mesoscale grouping. Tumors and their treatment impact distributed systems rather than isolated regions, and grouping aggregates ROI time series into ensembles that better reflect where pathology and compensatory processes occur. In our analyses, grouping enhances the predictive power of discrete landscapes and stabilizes patient-level features, while combining grouping with a continuous energy formulation yields the strongest overall performance. Two patients may share similar binary coactivation patterns yet differ in the amplitude of those patterns, in spread around basin centers, and in curvature governing how quickly trajectories return to those centers after perturbation. The continuous landscape captures these degrees of freedom through basin centers and positive–definite precisions, avoiding information loss from binarization and leveraging graded modulations that precede state switches.

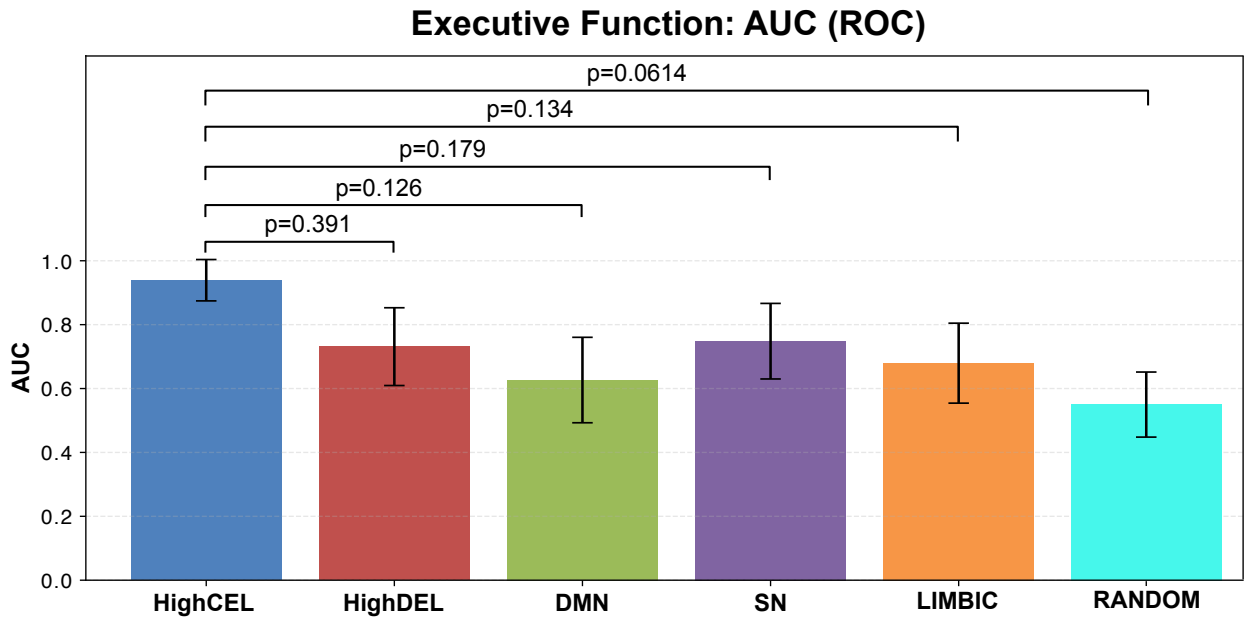


Figure 4.6: Executive function: ROC–AUC under leave–one–out validation (mean ± SD). HighCEL yields the highest mean AUC, followed closely by HighDEL; both outperform the network–subset baselines. The HighCEL–HighDEL difference is not significant by DeLong's test.

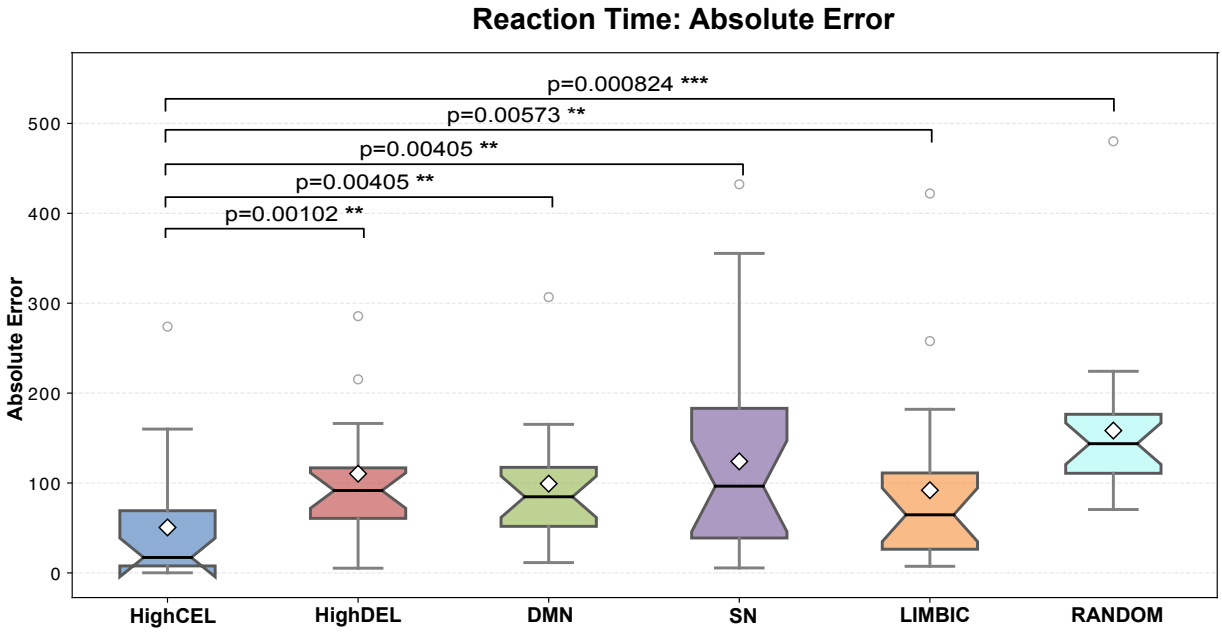


Figure 4.7: Reaction time prediction: absolute error (ms) under leave–one–out cross–validation. Boxes summarize the distribution of per–patient errors. Paired Wilcoxon tests favor HighCEL over each comparator after Benjamini–Hochberg correction. Lower is better.

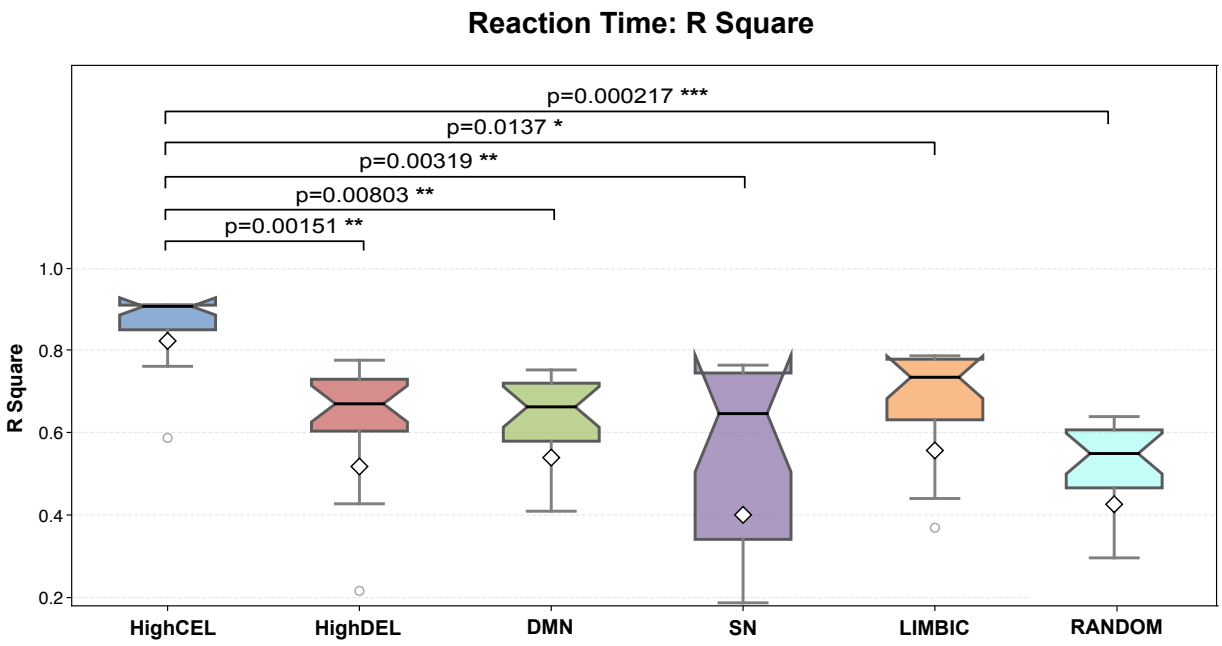


Figure 4.8: Reaction time prediction: per–subject R^2 contributions (leave–one–out). Boxes aggregate patient–level contributions to explained variance. Paired Wilcoxon tests favor HighCEL over each comparator after Benjamini–Hochberg correction. Higher is better.

Chapter 5

Discussion

5.1 Proposed Framework and Key Contributions

This dissertation focuses on energy landscape models and introduces two important extensions in this domain. The first extension represents activity at a mesoscale by aggregating parcels into $K \ll N$ ensembles, allowing recurrent coordination to emerge as joint modulation of ensembles. The second extension retains amplitude information and estimates a normalizable energy on $\mathbb{R}^K$ whenever possible, so that the energy geometry reflects not only which ensembles coactivate but also how strongly they fluctuate and how variability is shaped within a regime. The grouped discrete model (HODEL) implements the first approach using a pairwise maximum-entropy form over $\{0,1\}^K$, while the continuous model (HOCEL) implements the second by defining a quadratic or mixture energy directly on ensemble amplitudes with positive-definite precision.

The results from controlled simulations and resting-state fMRI data from brain tumor patients support the effectiveness of the proposed energy landscape models. Grouping improved recovery of basin identity, occupancy, and empirical transition structure across both a switching linear dynamical system and a phase-coupled oscillator regime. This improvement occurred under matched T, noise levels, and input dimensionality, indicating that it

resulted from a representational change rather than just a simpler estimation task. Furthermore, preserving amplitudes with HOCEL matched or exceeded HODEL in recovering basin geometry and reconfiguration tendencies, and delivered the strongest clinical prediction on average. Because both models share the same grouping and evaluation pipeline, the comparison isolates the effect of modeling amplitudes directly from the effect of coarse-graining.

A practical implication of the proposed energy landscape models is improved computational efficiency. In discrete landscapes, all quantities that require coverage of the state space scale with 2^K rather than 2^N, making exhaustive scoring and disconnectivity mapping feasible when $K \ll N$. In continuous landscapes, there is no combinatorial state explosion. With a low-rank plus ridge precision matrix $\mathbf{S} = \mathbf{Z}\mathbf{Z}^\top + \epsilon\mathbf{I}$, the log-determinant and linear solves required by the likelihood reduce to operations on low-rank matrices, yielding stable training and principled model selection. Such computational efficiency helps move energy landscape analysis from a small number of ROIs to large networks, providing a scalable framework for high-dimensional neural datasets.

Another advantage of the proposed higher-order energy landscape models is their robustness to noise. Energy landscape analyses inherit sensitivity to preprocessing and sample size that is common in dynamic functional connectivity. Motion, physiology, and acquisition artifacts can imprint structure on time series and therefore on estimated landscapes [48, 49, 50]. Utilizing higher-order interactions reduces the reliance on pairwise correlations, making the inferred landscapes less susceptible to spurious dependencies introduced by noise or confounds. By capturing structured dependencies beyond second-order statistics, these models provide a more faithful representation of underlying neural dynamics, even in the presence of imperfect data. This robustness enables more reliable inference across heterogeneous datasets and supports generalization to large-scale networks.

5.2 Practical Applications in Assessing Cognitive Function

Our findings demonstrate that energy landscape features are strongly related to cognitive performance across multiple domains. More specifically, utilizing pre-surgical functional MRI data, we show that energy landscape models can accurately predict postoperative cognitive function in brain tumor patients. This aligns with our prior results indicating that patients with poorer working memory and executive function exhibit landscapes characterized by deep energy basins and infrequent transitions, suggesting a rigid and energetically demanding architecture [51]. Such rigidity implies that shifting between cognitive states requires greater neural effort, likely reflecting disrupted structural connectivity and reduced capacity for functional reconfiguration, both critical for maintaining working memory and cognitive flexibility [52, 53, 54].

Furthermore, these patients demonstrated larger energy gaps between states, which may slow the initiation of neural transitions and manifest behaviorally as prolonged reaction times. In contrast, individuals with better cognitive outcomes showed shallower basins and more frequent transitions, indicative of a dynamic and adaptable network organization that supports efficient state switching and rapid processing. Taken together, these results suggest that energy landscapes provide a mechanistic link between network stability, flexibility, and cognitive performance, offering a novel framework for understanding how brain dynamics influence cognition and behavior.

These insights have important implications for neuro-oncology and cognitive rehabilitation. By quantifying the energetic cost of state transitions, energy landscape models could serve as biomarkers for predicting postoperative cognitive outcomes and guiding personalized interventions. For example, patients exhibiting rigid landscapes with high transition costs may benefit from targeted strategies aimed at enhancing network flexibility, such as cognitive training or neuromodulation techniques. Moreover, the ability to map energy landscapes at

scale opens opportunities for integrating these models into preoperative planning, enabling clinicians to anticipate cognitive risks and optimize surgical approaches. Ultimately, this framework bridges the gap between network-level dynamics and patient-centered outcomes, paving the way for precision medicine in neuro-oncology and beyond.

5.3 Limitations and Potential Future Direction

Current grouping strategies in our framework rely primarily on functional similarity derived from within-subject correlations at a fixed density, which may not fully capture structural constraints. This limitation can lead to unstable curvature estimates and reduced interpretability, particularly when functional connectivity is noisy or sparse. Future work could incorporate anatomical priors into grouping, such as Laplacian smoothness penalties or tract-based ordering, to improve stability and biological plausibility. Multi-scale graphs that combine structural and functional information represent a promising direction for achieving more robust basin geometry and transition estimates, especially in heterogeneous patient populations.

In the continuous energy formulation, HOCEL assumes Gaussian residuals within basins, which limits flexibility in capturing heavy tails, skewness, and nonstationarity. This assumption may mask subtle dynamics in pathological states and reduce the model's ability to represent complex variability. Future work could relax these constraints by incorporating non-Gaussian energy models using approaches such as score matching or generative gradient-based methods to better represent heterogeneous distributions and nonlinear coupling observed in neural systems.

The proposed energy models in this dissertation provide point estimates without formal uncertainty quantification, limiting confidence in patient-level predictions and hindering clinical translation. While bootstrap intervals offer partial solutions, they lack distribution-free guarantees and do not fully capture model uncertainty. Future work should explore Bayesian

variants of HOCEL with conjugate priors to produce reliable confidence intervals for basin centers and curvatures. Conformal prediction offers an alternative route to calibrated uncertainty for downstream predictive models without strong distributional assumptions. These approaches should be complemented by diagnostic checks, such as posterior predictive comparisons and energy-frequency alignment, to ensure reliability.

Lastly, the proposed framework was evaluated exclusively on resting-state fMRI without incorporating modalities that offer higher temporal resolution, such as electroencephalography (EEG). This limitation reduces the ability to capture rapid neural dynamics and cross-modal interactions that are critical for understanding cognition. Future work should focus on joint modeling of EEG-fMRI landscapes using shared structural priors to enhance temporal resolution and mechanistic insight. Achieving multimodal integration will require harmonized preprocessing pipelines and scalable algorithms to ensure tractability across diverse data types while maintaining interpretability.

References

[1] Braden AW Brinkman, Han Yan, Arianna Maffei, Il Memming Park, Alfredo Fontanini, Jin Wang, and Giancarlo La Camera. Metastable dynamics of neural circuits and networks. *Applied Physics Reviews*, 9(1), 2022.

[2] Jakub Vohryzek, Joana Cabral, Peter Vuust, Gustavo Deco, and Morten L Kringelbach. Understanding brain states across spacetime informed by whole-brain modelling. *Philosophical Transactions of the Royal Society A*, 380(2227):20210247, 2022.

[3] Naoki Masuda, Saiful Islam, Si Thu Aung, and Takamitsu Watanabe. Energy landscape analysis based on the ising model: Tutorial review. *PLOS Complex Systems*, 2(5):e0000039, 2025.

[4] Pitambar Khanra, Johan Nakuci, Sarah Muldoon, Takamitsu Watanabe, and Naoki Masuda. Reliability of energy landscape analysis of resting-state functional mri data. *European Journal of Neuroscience*, 60(3):4265–4290, 2024.

[5] Yushan Wu, Shi Qiao, Jitao Zhong, Lu Zhang, Juan Wang, Bin Hu, and Hong Peng. Fnirs-based energy landscape analysis to signify brain activity dynamics of individuals with depression. *CNS Neuroscience & Therapeutics*, 30(11):e70139, 2024.

[6] Le Xing, Zhitao Guo, and Zhiying Long. Energy landscape analysis of brain network dynamics in alzheimer's disease. *Frontiers in Aging Neuroscience*, 16:1375091, 2024.

[7] Hong Li, Xiaotong Shao, Fang He, Chenming He, Yuyu Yang, Yi Ge, Ruotong Chen,

Zijian Wang, Yuting Gong, Xipeng Long, et al. Neural dynamics remodeling induced by ketogenic diet therapy in drug-resistant epilepsy: A functional magnetic resonance imaging study. *Epilepsia*, 2025.

[8] Elad Schneidman, Michael J Berry, Ronen Segev, and William Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(7087):1007–1012, 2006.

[9] Andrea Santoro, Federico Battiston, Maxime Lucas, Giovanni Petri, and Enrico Amico. Higher-order connectomics of human brain function reveals local topological signatures of task decoding, individual identification, and behavior. *Nature Communications*, 15(1):10244, 2024.

[10] H Chau Nguyen, Riccardo Zecchina, and Johannes Berg. Inverse statistical problems: from the inverse ising problem to data science. *Advances in Physics*, 66(3):197–261, 2017.

[11] Nikos K Logothetis, Jon Pauls, Mark Augath, Torsten Trinath, and Axel Oeltermann. Neurophysiological investigation of the basis of the fmri signal. *nature*, 412(6843):150–157, 2001.

[12] Hong Yang, Xiang-Yu Long, Yihong Yang, Hao Yan, Chao-Zhe Zhu, Xiang-Ping Zhou, Yu-Feng Zang, and Qi-Yong Gong. Amplitude of low frequency fluctuation within visual areas revealed by resting-state functional mri. *Neuroimage*, 36(1):144–152, 2007.

[13] Qi-Hong Zou, Chao-Zhe Zhu, Yihong Yang, Xi-Nian Zuo, Xiang-Yu Long, Qing-Jiu Cao, Yu-Feng Wang, and Yu-Feng Zang. An improved approach to detection of amplitude of low-frequency fluctuation (alff) for resting-state fmri: fractional alff. *Journal of neuroscience methods*, 172(1):137–141, 2008.

[14] Takamitsu Watanabe, Satoshi Hirose, Hiroyuki Wada, Yoshio Imai, Toru Machida,

Ichiro Shirouzu, Seiki Konishi, Yasushi Miyashita, and Naoki Masuda. Energy landscapes of resting-state brain networks. *Frontiers in neuroinformatics*, 8:12, 2014.

[15] Takahiro Ezaki, Takamitsu Watanabe, Masayuki Ohzeki, and Naoki Masuda. Energy landscape analysis of neuroimaging data. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 375(2096):20160287, 2017.

[16] Erik Aurell and Magnus Ekeberg. Inverse ising inference using all the data. *Physical review letters*, 108(9):090201, 2012.

[17] Rikkert Hindriks, Mohit H Adhikari, Yusuke Murayama, Marco Ganzetti, Dante Mantini, Nikos K Logothetis, and Gustavo Deco. Can sliding-window correlations reveal dynamic functional connectivity in resting-state fmri? *Neuroimage*, 127:242–256, 2016.

[18] Timothy O Laumann, Abraham Z Snyder, and Caterina Gratton. Challenges in the measurement and interpretation of dynamic functional connectivity. *Imaging Neuroscience*, 2:1–19, 2024.

[19] Scott Marek, Brenden Tervo-Clemmens, Finnegan J Calabro, David F Montez, Benjamin P Kay, Alexander S Hatoum, Meghan Rose Donohue, William Foran, Ryland L Miller, Timothy J Hendrickson, et al. Reproducible brain-wide association studies require thousands of individuals. *Nature*, 603(7902):654–660, 2022.

[20] Luca Pasquini, Kyung K Peck, Mehrnaz Jenabi, and Andrei Holodny. Functional mri in neuro-oncology: state of the art and future directions. *Radiology*, 308(3):e222028, 2023.

[21] Francisco Barahona. On the computational complexity of ising spin glass models. *Journal of Physics A: Mathematical and General*, 15(10):3241, 1982.

[22] Mark Jerrum and Alistair Sinclair. Polynomial-time approximation algorithms for the ising model. *SIAM Journal on computing*, 22(5):1087–1116, 1993.

[23] Thomas M Cover. *Elements of information theory*. John Wiley & Sons, 1999.

[24] John H Van Vleck and David Middleton. The spectrum of clipped noise. *Proceedings of the IEEE*, 54(1):2–19, 2005.

[25] Julian Besag. Statistical analysis of non-lattice data. *Journal of the Royal Statistical Society Series D: The Statistician*, 24(3):179–195, 1975.

[26] Pradeep Ravikumar, Martin J Wainwright, and John D Lafferty. High-dimensional ising model selection using l 1-regularized logistic regression. 2010.

[27] Jascha Sohl-Dicksteinad, Peter Battaglinobd, and Michael R DeWeesebcd. Minimum probability flow learning. *arXiv preprint arXiv:0906.4779*, 2009.

[28] Jascha Sohl-Dickstein, Peter B Battaglino, and Michael R DeWeese. New method for parameter estimation in probabilistic models: minimum probability flow. *Physical review letters*, 107(22):220601, 2011.

[29] Michael Gutmann and Aapo Hyvärinen. Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 297–304. JMLR Workshop and Conference Proceedings, 2010.

[30] Geoffrey E Hinton. Training products of experts by minimizing contrastive divergence. *Neural computation*, 14(8):1771–1800, 2002.

[31] Tijmen Tieleman. Training restricted boltzmann machines using approximations to the likelihood gradient. In *Proceedings of the 25th international conference on Machine learning*, pages 1064–1071, 2008.

[32] Miguel A Carreira-Perpinan and Geoffrey Hinton. On contrastive divergence learning. In *International workshop on artificial intelligence and statistics*, pages 33–40. PMLR, 2005.

[33] Timm Plefka. Convergence condition of the tap equation for the infinite-ranged ising spin glass model. *Journal of Physics A: Mathematical and general*, 15(6):1971, 1982.

[34] Jonathan S Yedidia, William T Freeman, and Yair Weiss. Constructing free-energy approximations and generalized belief propagation algorithms. *IEEE Transactions on information theory*, 51(7):2282–2312, 2005.

[35] Simona Cocco and Rémi Monasson. Adaptive cluster expansion for inferring boltzmann machines with noisy data. *Physical review letters*, 106(9):090601, 2011.

[36] Zoubin Ghahramani and Geoffrey E Hinton. Variational learning for switching state-space models. *Neural computation*, 12(4):831–864, 2000.

[37] Emily Fox, Erik Sudderth, Michael Jordan, and Alan Willsky. Nonparametric bayesian learning of switching linear dynamical systems. *Advances in neural information processing systems*, 21, 2008.

[38] Yoshiki Kuramoto. Self-entrainment of a population of coupled non-linear oscillators. In *International symposium on mathematical problems in theoretical physics: January 23–29, 1975, kyoto university, kyoto/Japan*, pages 420–422. Springer, 2005.

[39] Steven H Strogatz. From kuramoto to crawford: exploring the onset of synchronization in populations of coupled oscillators. *Physica D: Nonlinear Phenomena*, 143(1-4):1–20, 2000.

[40] Juan A Acebrón, Luis L Bonilla, Conrad J Pérez Vicente, Félix Ritort, and Renato Spigler. The kuramoto model: A simple paradigm for synchronization phenomena. *Reviews of modern physics*, 77(1):137–185, 2005.

[41] Michael Breakspear. Dynamic models of large-scale brain activity. *Nature neuroscience*, 20(3):340–352, 2017.

[42] Joana Cabral, Etienne Hugues, Olaf Sporns, and Gustavo Deco. Role of local network oscillations in resting-state functional connectivity. *Neuroimage*, 57(1):130–139, 2011.

[43] Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.

[44] Peter E Kloeden and RA Pearson. The numerical solution of stochastic differential equations. *The ANZIAM Journal*, 20(1):8–12, 1977.

[45] Harold W Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955.

[46] Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics bulletin*, 1(6):80–83, 1945.

[47] Hannelore Aerts, Nigel Colenbier, Hannes Almgren, Thijs Dhollander, Javier Rasero Daparte, Kenzo Clauw, Amogh Johri, Jil Meier, Jessica Palmer, Michael Schirner, et al. Pre-and post-surgery brain tumor multimodal magnetic resonance imaging data optimized for large scale computational modelling. *Sci Data*, 9(1):676, 2022.

[48] R Matthew Hutchison, Thilo Womelsdorf, Elena A Allen, Peter A Bandettini, Vince D Calhoun, Maurizio Corbetta, Stefania Della Penna, Jeff H Duyn, Gary H Glover, Javier Gonzalez-Castillo, et al. Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage*, 80:360–378, 2013.

[49] Rastko Ciric, Daniel H Wolf, Jonathan D Power, David R Roalf, Graham L Baum, Kosha Ruparel, Russell T Shinohara, Mark A Elliott, Simon B Eickhoff, Christos Davatzikos, et al. Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. *Neuroimage*, 154:174–187, 2017.

[50] Jonathan D Power, Kelly A Barnes, Abraham Z Snyder, Bradley L Schlaggar, and Steven E Petersen. Spurious but systematic correlations in functional connectivity mri networks arise from subject motion. *Neuroimage*, 59(3):2142–2154, 2012.

[51] Triet M Tran and Sina Khanmohammadi. Neural energy landscapes predict working memory decline after brain tumor resection. *arXiv preprint arXiv:2507.23057*, 2025.

[52] Hikaru Takeuchi, Atsushi Sekiguchi, Yasuyuki Taki, Satoru Yokoyama, Yukihito Yomogida, Nozomi Komuro, Tohru Yamanouchi, Shozo Suzuki, and Ryuta Kawashima. Training of working memory impacts structural connectivity. *Journal of Neuroscience*, 30(9):3297–3303, 2010.

[53] Jessica R Cohen, Courtney L Gallen, Emily G Jacobs, Taraz G Lee, and Mark D'Esposito. Quantifying the reconfiguration of intrinsic networks during working memory. *PloS one*, 9(9):e106636, 2014.

[54] Urs Braun, Axel Schäfer, Henrik Walter, Susanne Erk, Nina Romanczuk-Seiferth, Leila Haddad, Janina I Schweiger, Oliver Grimm, Andreas Heinz, Heike Tost, et al. Dynamic reconfiguration of frontal brain networks during executive cognition in humans. *Proceedings of the National Academy of Sciences*, 112(37):11678–11683, 2015.