Optimal packing of attractor states in neural representations

John J. Vastola JOHN_VASTOLA@HMS.HARVARD.EDU
Department of Neurobiology, Harvard Medical School, Boston, MA, USA

Editors: Sophia Sanborn, Christian Shewmake, Simone Azeglio, Nina Miolane

Abstract

Animals' internal states reflect variables like their position in space, orientation, decisions, and motor actions—but how should these internal states be arranged? Internal states which frequently transition between one another should be close enough that transitions can happen quickly, but not so close that neural noise significantly impacts the stability of those states, and how reliably they can be encoded and decoded. In this paper, we study the problem of striking a balance between these two concerns, which we call an 'optimal packing' problem since it resembles mathematical problems like sphere packing. While this problem is generally extremely difficult, we show that symmetries in environmental transition statistics imply certain symmetries of the optimal neural representations, which allows us in some cases to exactly solve for the optimal state arrangement. We focus on two toy cases: uniform transition statistics, and cyclic transition statistics. Code is available at https://github.com/john-vastola/optimal-packing-neurreps23.

Keywords: optimization, Markov chain, neural representation, neural dynamics, sphere packing, symmetry

1. Introduction

Animals' internal states appear to reflect environmental variables, like their position in space and orientation relative to some reference (Barry and Burgess, 2014; Hulse and Jayaraman, 2020), as well as their interactions with the environment, like decisions (Gold and Shadlen, 2007) and motor actions (Cisek, 2005). As an animal acts in its environment, it must constantly update these internal states to reflect environmental changes and the results of internal computations; however, these updates cannot be *instantaneous*, since biophysical limitations force internal quantities to change in a somewhat continuous fashion. They are also not *error-free* due to noise in encoding, decoding, and neural dynamics (Faisal et al., 2008; van Vreeswijk and Sompolinsky, 1996).

How should internal states be arranged? On the one hand, an animal can act more quickly if the next relevant internal state is 'near' the current one, since it can be reached more quickly. This suggests that the structure of neural representations should reflect the structure of environmental transitions; this is consistent with what is known about circuits like the head direction system, whose latent geometry mirrors the circular nature of the variable it tracks (Ajabi et al., 2023), and theoretical ideas about smoothness as a constraint on neural codes (Stringer et al., 2019). On the other hand, the closer all internal states are to one another, the easier it is for neural noise to cause problems, either via noise-induced transitions (Burak and Fiete, 2012) or by increasing the likelihood of encoding and decoding errors. In principle, a 'good' arrangement of internal states strikes a balance between these two concerns: internal states must be packed closely enough that desired transitions can happen quickly, but not so closely that errors are likely.

Given that noise sets an effective length scale for separating internal states, this issue in some ways mathematically resembles an optimal packing problem. While often quite difficult, problems like optimal sphere packing (Zong, 2008) are made substantially easier to solve and understand by exploiting symmetry-related considerations. For example, Viazovska et al.'s solution of the sphere packing problem in dimensions 8 and 24 (Viazovska, 2017; Cohn et al., 2017) crucially uses symmetry properties of the E_8 and Leech lattices. In this paper, we attempt to formulate a toy version of the problem of constructing an 'optimal packing' of neural representations, and similarly turn to symmetry-related tools in order to say something meaningful about it. The particular symmetry-related claim we will motivate, and then use, is that an attractor-based neural representation of a Markov chain that exhibits a symmetry may also exhibit that symmetry.

2. Mathematical formulation of optimal packing problem

To formalize our optimal packing problem, we need five things: a model of environment state statistics, a model of internal state transition dynamics, an encoding model, a decoding model, and a cost function.

Environment dynamics. We will model the environment as a Markov chain on M states. In particular, we will assume that it can be characterized by a set of states $\mathcal{X} = \{1, ..., M\}$, a base probability of state occupancy $p_0(x)$ for all $x \in \mathcal{X}$, and a probability $p(y|x)$ of transitioning from any state $x \in \mathcal{X}$ to any state $y \in \mathcal{X}$ on some characteristic time scale. Since we are interested in finding representations that respect environmental transition structure *when a transition occurs*, we assume without loss of generality that $p(x|x) = 0$ for all $x \in \mathcal{X}$.

Internal state transition dynamics. Let $\mathcal{Z} = \mathbb{R}^D$ (for some $D \geq M$) denote the set of all possible internal states, and assume that each environment state $x \in \mathcal{X}$ is in one-to-one correspondence with an internal attractor state $\boldsymbol{z}_x \in \mathcal{Z}$. Assume also that the positive definite matrix $\boldsymbol{\Sigma}^{-1}$ can be used to compute the distance

$$D(\boldsymbol{z}_1, \boldsymbol{z}_2) := (\boldsymbol{z}_1 - \boldsymbol{z}_2)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{z}_1 - \boldsymbol{z}_2) \tag{1}$$

between any two internal states. The matrix $\boldsymbol{\Sigma}$ is intended to model how noisy different directions in $\mathcal{Z}$ are; different states are 'closer', in the sense of being easier to reach from one another, if the line connecting them corresponds to a particularly noisy direction.

Although it is possible to write down an extremely explicit model of internal state transition dynamics, we will consider a somewhat coarse description in order to keep our problem mathematically tractable. We will assume three things: first, that transitions are essentially between attractor basins, so that the relevant quantity is the discrete distribution $q(\boldsymbol{z}_y|\boldsymbol{z}_x)$; second, that there is a mechanism for destabilizing attractor states when a transition is desired, so that $q(\boldsymbol{z}_x|\boldsymbol{z}_x) = 0$ for all attractor states $\boldsymbol{z}_x$; and third, that transitions to good approximation only depend on the distances between states. The last assumption makes sense within a landscape picture of internal dynamics (involving M attractors of similar width and depth), and can be formally justified via appealing to, e.g., Kramers' theory

(Kramers, 1940; Hänggi et al., 1990). Explicitly, we will assume that

$$q(\boldsymbol{z}_y|\boldsymbol{z}_x) = (1-\delta_{xy})\frac{e^{-D(\boldsymbol{z}_y,\boldsymbol{z}_x)^2/2}}{Z(\boldsymbol{z}_x)}$$
$$Z(\boldsymbol{z}_x) = \sum_{a\neq x} e^{-D(\boldsymbol{z}_a,\boldsymbol{z}_x)^2/2} \ . \tag{2}$$

Encoding/decoding models. We will assume that the encoding of an environment state $x \in \mathcal{X}$ is noisy, and that when a mistake is made, x is more likely to be encoded as a state $\boldsymbol{z}_a$ *near* $\boldsymbol{z}_x$. Explicitly, we will assume

$$p_e(\boldsymbol{z}_a|x) = \frac{\delta_{ax}e^b + (1-\delta_{ax})e^{-D(\boldsymbol{z}_a,\boldsymbol{z}_x)^2/2}}{e^b + Z(\boldsymbol{z}_x)} \tag{3}$$

where increasing the bias $b \geq 0$ makes errors less likely. In the $b \to \infty$ limit, encoding is perfect. Assuming a uniform prior, we obtain a decoding model through Bayes' rule:

$$p_d(x|\boldsymbol{z}_a) = \frac{p_e(\boldsymbol{z}_a|x)}{\sum_x p_e(\boldsymbol{z}_a|x)} \ . \tag{4}$$

In the $b \to \infty$ limit, since $p(\boldsymbol{z}_a|x) = \delta_{ax}$, we also have perfect decoding.

Cost function. We want to penalize different possible *arrangements* of internal attractor states according to some objective function, so that optimizing that objective corresponds to identifying an optimal packing. 'Optimality' here means an arrangement which, as much as possible, produces internal dynamics (i.e., movement between attractors) whose transition statistics mirror the statistics of environmental transitions (Figure 1a). The interpretation of this is that, *in the absence of any external input*, the internal state is poised to change in the same way that the environment is likely to change.

Consider the way ring-attractor-like networks reckon with uncertainty as a concrete example of this feature: in the absence of external input, the bump representing heading direction diffuses (Kutschireiter et al., 2023), a purely internal state change that reflects the fact that moment-to-moment changes in heading direction will usually be small, and are equally likely to be clockwise or counterclockwise.

One way to formalize this desire mathematically is to ask that

$$p_{int}(y|x) := \sum_{a,b} p_d(y|\boldsymbol{z}_b)q(\boldsymbol{z}_b|\boldsymbol{z}_a)p_e(\boldsymbol{z}_a|x) \ . \tag{5}$$

on average matches $p(y|x)$, the function that determines the statistics of environmental transitions. (Equivalently: we can ask that the diagram in Figure 1a commutes.) More precisely, we want the Kullback-Leibler divergence between $p(y|x)$ and $p_{int}(y|x)$ to be small.

We also want to include a regularization term which enforces the fact that, all else being equal, we prefer configurations with low activity, i.e., configurations for which the norm

$$\|\boldsymbol{z}_x\|^2 := \boldsymbol{z}_x^T \boldsymbol{\Sigma}^{-1} \boldsymbol{z}_x \tag{6}$$

is small for all attractor states $\boldsymbol{z}_x$. (This can be viewed as a kind of firing rate penalty.) Hence, we will define

$$\begin{aligned} J[\{\boldsymbol{z}_x\}] &:= \mathbb{E}_{\boldsymbol{x}} \left\{ \; KL(p \| p_{int}) + \frac{\alpha}{2} \|\boldsymbol{z}_x\|^2 \; \right\} \\ &= \sum_{x,y} p_0(x) p(y|x) \left\{ \; \log p(y|x) - \log p_{int}(y|x) \right\} + \frac{\alpha}{2} \sum_x p_0(x) \|\boldsymbol{z}_x\|^2 \end{aligned} \tag{7}$$

as an objective over possible attractor state assignments $\{\boldsymbol{z}_x\}$. Heuristically, we can think of the objective as representing a contest between three competing interests: low firing rate, high encoding/decoding accuracy, and internal dynamics mirroring environmental transition structure (for example, in the sense depicted in Figure 1b). The firing rate penalty pushes all attractor states towards the origin; increasing encoding and decoding accuracy pushes all attractor states infinitely far apart; and having internal dynamics mirror environment dynamics incentivizes particular relationships between attractor states.

As usual, we can drop terms which do not depend on the $\boldsymbol{z}_x$, so we can redefine J as

$$J[\{\boldsymbol{z}_x\}] := -\sum_{x,y} p_0(x) p(y|x) \log p_{int}(y|x) + \frac{\alpha}{2} \sum_x p_0(x) \|\boldsymbol{z}_x\|^2 \; . \tag{8}$$

Our central concern in the following is: *under what conditions can we find optimal attractor state assignments* $\{\boldsymbol{z}_x\}$*?*

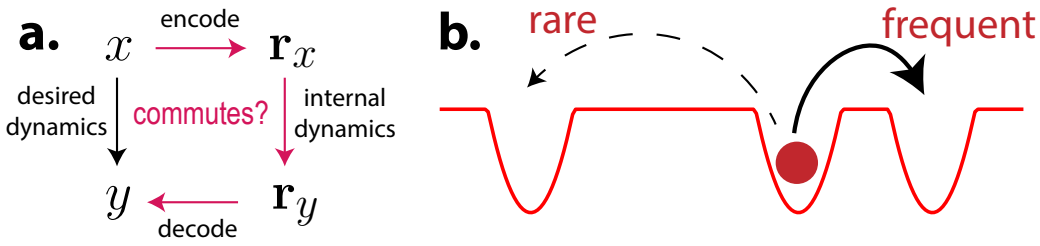

Figure 1: Schematic of optimal packing problem. **a.** We want environment transition statistics $p(y|x)$ to typically match the combination of encoding, internal dynamics, and decoding, or equivalently for this diagram to commute. **b.** Intuitively, the geometric structure of the attractor landscape should match the structure of the Markov chain; for example, states with frequent transitions ought to be closer together than states between which transitions are rare.

3. Initial observations: simplifications and symmetry

Eq. 8 defines a high-dimensional, nonlinear, and (as we will see) non-convex optimization problem which is in general difficult to solve. In this section, we will make several useful preliminary observations about it.

Simplifying the objective. The objective can be reparameterized in a way that makes various features of the problem clearer (see Appendix A). First, we can change variables from $\boldsymbol{z}$ to $\boldsymbol{r}$ with

$$D(\boldsymbol{z}_1, \boldsymbol{z}_2) = D(\boldsymbol{r}_1, \boldsymbol{r}_2) = \|\boldsymbol{r}_1 - \boldsymbol{r}_2\|_2^2 \qquad\qquad \boldsymbol{z}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{z} = \boldsymbol{r}^T \boldsymbol{r} \tag{9}$$

by using scaled eigenvectors of $\boldsymbol{\Sigma}$ as a basis for $\mathbb{R}^D$. By doing so, we can disentangle the impact of noise anisotropy from other aspects of the problem. Let $\{\boldsymbol{r}_x\}$ denote attractor assignments in the new coordinate system; in what follows, we will work exclusively with $\boldsymbol{r}_x$ instead of $\boldsymbol{z}_x$. Second, we can write J in terms of

$$\langle \boldsymbol{r} \rangle := \sum_x p_0(x) \boldsymbol{r}_x \qquad\qquad d_{xy} := \|\boldsymbol{r}_x - \boldsymbol{r}_y\|_2^2 = d_{yx} \ , \tag{10}$$

i.e., the average attractor state location (D scalars) and the pairwise distances between each attractor state ($M(M-1)/2$ scalars), each of which is independent of the others. There are typically less than DM degrees of freedom since the problem is both rotation- and reflection-invariant, and since the objective is defined in terms of averages.[1] In terms of these variables, we have

$$J[\{d_{xy}\}, \langle \boldsymbol{r} \rangle] = -\sum_{x,y} p_0(x) p(y|x) \log p_{int}(y|x) + \alpha \frac{\|\langle \boldsymbol{r} \rangle\|_2^2}{2} + \alpha \sum_{a \neq b} p_0(a) p_0(b) \frac{d_{ab}^2}{4} \ . \tag{11}$$

Furthermore, the optimal choice of $\langle \boldsymbol{r} \rangle$ is obvious, since it only appears in the quadratic regularization term: all optimal configurations have $\langle \boldsymbol{r} \rangle = \boldsymbol{0}$. This means that we only need to optimize the $M(M-1)/2$ pairwise distances d_{xy}.

Symmetry. Let $\pi : \{1, ..., M\} \to \{1, ..., M\}$ be a permutation of $\mathcal{X}$. Suppose that π is a symmetry of the Markov chain, i.e., that

$$p_0(\pi(x)) = p_0(x) \qquad\qquad p(\pi(y)|\pi(x)) = p(y|x) \ . \tag{12}$$

We will show that the objective function shares this symmetry. Consider the map that takes $d_{xy} \mapsto d_{\pi(x)\pi(y)}$. Because p_{int} only depends on pairwise distances, we have $p_{int}(y|x) \mapsto p_{int}(\pi(y)|\pi(x))$. The relevant part of the objective becomes

$$\begin{aligned} &-\sum_{x,y} p_0(x) p(y|x) \log p_{int}(\pi(y)|\pi(x)) + \alpha \sum_{a \neq b} p_0(a) p_0(b) \frac{d_{\pi(a)\pi(b)}^2}{4} \\ = \ &-\sum_{x,y} p_0(\pi(x)) p(\pi(y)|\pi(x)) \log p_{int}(\pi(y)|\pi(x)) + \alpha \sum_{a \neq b} p_0(\pi(a)) p_0(\pi(b)) \frac{d_{\pi(a)\pi(b)}^2}{4} \end{aligned} \tag{13}$$

where we have used the definition of the symmetry. But since we are summing over x, y, a, and b, it does not matter how we permute them; hence, the map we have introduced does not change the objective.

1. Note that $DM = (M-1)D + D \geq \frac{M(M-1)}{2} + D$. At worst, we have as many degrees of freedom as we started with, but we usually have fewer. This only works since we assumed $D \geq M$.

For convex optimization problems, one can show that the unique global minimum of the objective must share its symmetries. But our problem is probably not convex, so we must settle for something weaker: in the spirit of the Purkiss principle (Waterhouse, 1983), we can look for solutions that share the objective's symmetries. A more rigorous analysis of Eq. 11 may be able to show that Waterhouse's precise formulation of the Purkiss principle applies, although we do not pursue such an analysis here.

4. Results: optimal packing for uniform and cyclic topologies

In this section, we study the symmetric solutions of two classes of highly symmetric packing problems: the first assumes environment statistics are uniform (each state is equally likely to be next), and the second assumes statistics are cyclic (transitions occur on a ring).

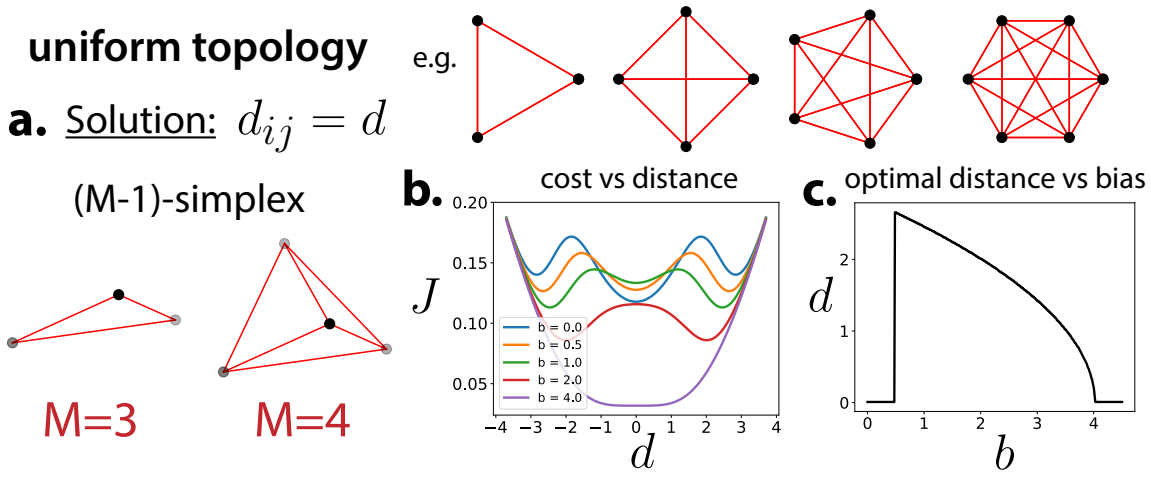

Figure 2: Solution of packing problem for a Markov chain with a uniform topology (see top right graphs). **a.** The optimal solution has the distance between all states equal, which means the geometry is that of a $(M-1)$-simplex. **b.** The objective function versus the distance d. Note the bifurcation as the bias decreases. **c.** The optimal distance as a function of the bias. It is zero for very small or large biases.

4.1. Optimal packing for uniform topology

If environmental transitions are completely uniform, then the corresponding Markov chain can be visualized as a complete graph (each vertex is connected to all others) on M vertices (Figure 2). Every possible permutation of $\mathcal{X}$ is a symmetry of this Markov chain, so our symmetric solution has $d_{ij} = d$ for all pairs. Hence, the geometry of this representation is completely determined: it is an $(M-1)$-simplex, an $(M-1)$-dimensional object (see Figure 2a for two examples).

After some algebra (see Appendix B), we find that the objective function can be written

$$J = 2\log\left[e^b + (M-1)e^{-d^2/2}\right] - \log\left[e^{2b} + 2(M-2)e^{b-d^2/2} + (M^2-3M+5)e^{-d^2}\right] + \alpha\frac{M-1}{M}\frac{d^2}{4}$$

up to unimportant additive constants. The cost J as a function of the distance d between attractor states is plotted in Figure 2b for different values of the bias b. We observe fairly interesting qualitative behavior: when the bias is small, encoding and decoding are highly noisy, and the optimal solution places all states at the origin; when the bias is large, performance is good even if all states are placed arbitrarily close together; finally, when the bias is moderate, a nontrivial solution exists. In the region with a nontrivial solution, the optimal distance decreases monotonically as the bias increases (Figure 2c).

4.2. Optimally packing four attractor states

The simplest possible cyclic Markov chain that is not uniform has $M = 4$ states. The relevant graph is a square, and the relevant symmetry group is D_4, the dihedral group of order 4. Symmetry constraints (in particular, rotation symmetry) tell us that $d_{12} = d_{23} = d_{34} = d_{41} = d$, and that $d_{13} = d_{24} = L$; the precise values of d and L must be determined by optimizing the objective.

Naively, we might expect that the answer should be a square in neural activity space; however, this is not necessarily true, even after our correction for noise anisotropy. First off, non-square arrangements of four points exist with equal edge lengths and diagonals. For example, a square folded along one diagonal, with its angles slightly distorted, satisfies the distance constraints (Figure 3a).

After some algebra (see Appendix C), we find that the objective function can be written

$$J = \log Z + 2\log(e^b + Z) + \frac{d^2}{2} - \log\left[e^{2b} + 4e^b e^{-L^2/2} + 3e^{-L^2} + 4e^{-d^2}\right] + \frac{\alpha}{4}\left(\frac{d^2}{2} + \frac{L^2}{4}\right),$$

which we plot in Figure 3b for different values of the bias b. We see a similar phase transition as the one we saw in the uniform case: for small bias, the optimal solution places all states at the origin because it is too noisy; for a moderate bias, there is a nontrivial solution. Unlike before, in the case of a large bias, not all states are placed at the origin: the meaning of $d = 0$ and $L \neq 0$ is essentially that states are 'glued' together so that the quadrilateral becomes a line.

Perhaps unexpectedly, the optimal d and L do not produce a square representation in general, since $L/d \neq \sqrt{2}$ except at a special bias value (Figure 3c). The actual arrangement produced (effectively three-dimensional, since only four points are involved) is depicted for a few bias values in Figure 3d. Note that if the bias is *just* large enough for a nontrivial solution to exist, the arrangement is approximately square.

4.3. Optimal packing for cyclic topology

If environmental transitions are cyclic, then the corresponding Markov chain can be visualized as a cycle graph C_M. By rotational symmetry, we have as many distances to determine as there are distinct vertex-vertex distances in a regular polygon (and this number differs depending on whether M is even or odd). The relevant symmetry group is D_M, the dihedral

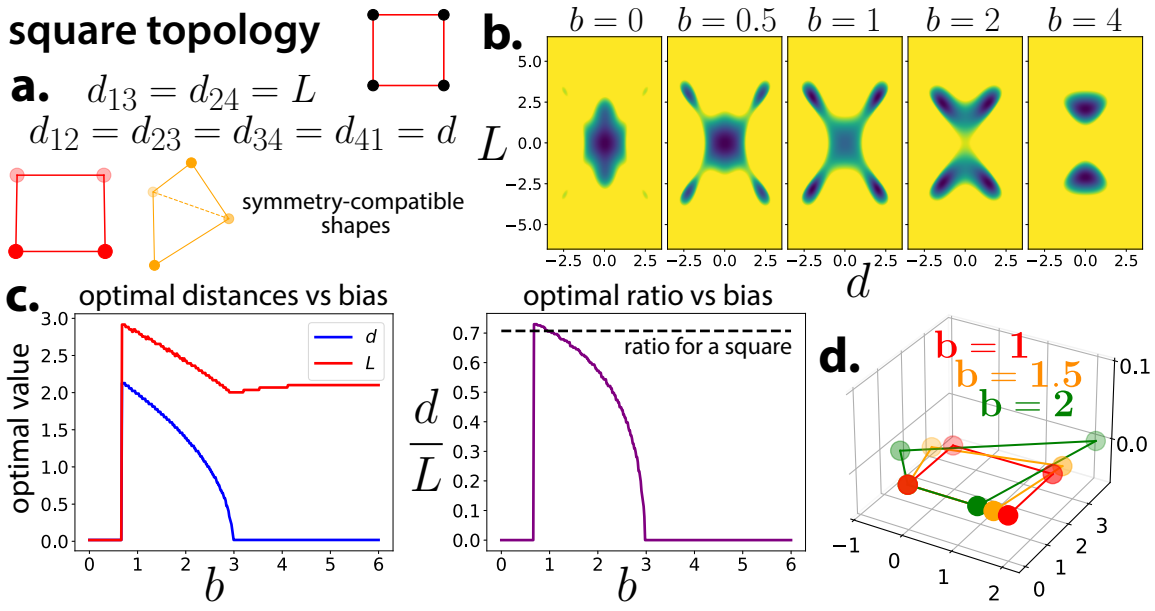

Figure 3: Solution of packing problem for a Markov chain with a square topology. **a.** The optimal solution has two undetermined distances d and L; depending on their ratio, non-square solutions are possible. **b.** The objective function versus d and L for different bias values. Note the phase transitions for small and large biases. **c.** The optimal d and L values generally do not have $L/d = \sqrt{2}$ when a nontrivial solution exists. **d.** The optimal arrangement is generally not square.

group of order M. In principle, this problem could be analyzed using the same approach we used to analyze the $M = 4$ case; however, this is extremely tedious. An interesting special case that may be tractable is where the bias b is large, in which case p_{int} is nearly (see Appendix D)

$$p_{int} \approx \frac{e^{-d_1^2/2}}{Z} \left\{ 1 + 2e^{-b} \left[2 \sum_{k=1}^{M/2-1} e^{-d_k^2/2 - d_{k+1}^2/2 + d_1^2/2} - Z \right] \right\} \qquad Z = e^{-d_N^2/2} + \sum_{k=1}^{M/2-1} 2e^{-d_k^2/2} \,,$$

where d_1, d_2, and so on are the various state-state distances. Our intuition from the $M = 4$ case should suggest the following: the optimal configuration should have all state-state distances (approximately) equal to those of a regular polygon near the minimal bias sufficient to support a nontrivial configuration. We conjecture that a result like this formally holds when M and α are both large.

5. Discussion

We have attempted to formulate the problem of packing attractor states in a neural representation so that internal transition statistics match environmental transition statistics as much as possible. Here, we will make comments mainly about two things: other potentially tractable Markov chain topologies, and possible generalizations of our formulation.

Other classes that may be tractable include straightforward generalizations of the cyclic topology (e.g., spherical or toroidal) and a translation-invariant lattice topology. It may be somewhat surprising that the optimization problem proved somewhat difficult *even* in the case of a cyclic topology. This suggests that exact solutions may be hard to obtain, even for models with a high degree of symmetry, unless an approximation (e.g., high bias, large M) or special trick is used. Simulation may be a more effective route towards understanding the behavior of this problem. For example: is the global minimum generally unique? It was in the cases we examined, but this does not imply much about the general case.

At least superficially, our packing problem somewhat resembles the problem of finding Euclidean graph embeddings (Cai et al., 2018). (There are important qualitative differences in the functional form of the objective, however.) It may be possible to adapt some results from that setting to provide insight here.

A variety of generalizations are possible, both to make the problem more mathematically interesting and to make it more relevant to neuroscience. If we return to our definition of $p_{int}(y|x)$ (Eq. 5), a few become obvious: we can use more realistic encoding/decoding models, like Poisson spiking models or probabilistic population codes (Ma et al., 2006; Vastola et al., 2023); we can use a more realistic dynamics model, like a recurrent neural network; and we can define distances or dynamics on a non-Euclidean space. It is unclear which more complex choices would still yield a somewhat tractable mathematical problem.

An interesting phenomenon potentially related to the problem we consider here is the fact that neural representations tend to drift (Driscoll et al., 2022; Masset et al., 2022). If representational drift happens on a longer time scale than representation optimization, our formulation suggests that drift may be a consequence of changing environmental transition statistics (or a changing internal model of environmental statistics). This is somewhat compatible with other ideas about possible advantages of representational drift, e.g., for multi-task learning.

References

Zaki Ajabi, Alexandra T. Keinath, Xue-Xin Wei, and Mark P. Brandon. Population dynamics of head-direction neurons during drift and reorientation. *Nature*, 615(7954): 892–899, Mar 2023. ISSN 1476-4687. doi: 10.1038/s41586-023-05813-2. URL https://doi.org/10.1038/s41586-023-05813-2.

C. Barry and N. Burgess. Neural mechanisms of self-location. *Current Biology*, 24(8): R330–R339, 2014. ISSN 0960-9822. doi: https://doi.org/10.1016/j.cub.2014.02.049. URL https://www.sciencedirect.com/science/article/pii/S0960982214002176.

Yoram Burak and Ila R. Fiete. Fundamental limits on persistent activity in networks of noisy neurons. *Proceedings of the National Academy of Sciences*, 109(43):17645–17650, 2012. doi: 10.1073/pnas.1117386109. URL https://www.pnas.org/doi/abs/10.1073/pnas.1117386109.

HongYun Cai, Vincent W. Zheng, and Kevin Chen-Chuan Chang. A comprehensive survey of graph embedding: Problems, techniques, and applications. *IEEE Transactions on Knowledge and Data Engineering*, 30(9):1616–1637, 2018. doi: 10.1109/TKDE.2018.2807452.

Paul Cisek. Neural representations of motor plans, desired trajectories, and controlled objects. *Cognitive Processing*, 6(1):15–24, Mar 2005. ISSN 1612-4790. doi: 10.1007/s10339-004-0046-7. URL https://doi.org/10.1007/s10339-004-0046-7.

Henry Cohn, Abhinav Kumar, Stephen Miller, Danylo Radchenko, and Maryna Viazovska. The sphere packing problem in dimension 24. *Annals of Mathematics*, 185(3):1017 – 1033, 2017. doi: 10.4007/annals.2017.185.3.8. URL https://doi.org/10.4007/annals.2017.185.3.8.

Laura N. Driscoll, Lea Duncker, and Christopher D. Harvey. Representational drift: Emerging theories for continual learning and experimental future directions. *Current Opinion in Neurobiology*, 76:102609, 2022. ISSN 0959-4388. doi: https://doi.org/10.1016/j.conb.2022.102609. URL https://www.sciencedirect.com/science/article/pii/S0959438822001039.

A. Aldo Faisal, Luc P. J. Selen, and Daniel M. Wolpert. Noise in the nervous system. *Nature Reviews Neuroscience*, 9(4):292–303, Apr 2008. ISSN 1471-0048. doi: 10.1038/nrn2258. URL https://doi.org/10.1038/nrn2258.

Joshua I. Gold and Michael N. Shadlen. The neural basis of decision making. *Annual Review of Neuroscience*, 30(1):535–574, 2007. doi: 10.1146/annurev.neuro.29.051605.113038. URL https://doi.org/10.1146/annurev.neuro.29.051605.113038. PMID: 17600525.

Peter Hänggi, Peter Talkner, and Michal Borkovec. Reaction-rate theory: fifty years after Kramers. *Rev. Mod. Phys.*, 62:251–341, Apr 1990. doi: 10.1103/RevModPhys.62.251. URL https://link.aps.org/doi/10.1103/RevModPhys.62.251.

Brad K. Hulse and Vivek Jayaraman. Mechanisms underlying the neural computation of head direction. *Annual Review of Neuroscience*, 43(1):31–54, 2020. doi: 10.1146/annurev-neuro-072116-031516. URL https://doi.org/10.1146/annurev-neuro-072116-031516. PMID: 31874068.

H.A. Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7(4):284–304, 1940. ISSN 0031-8914. doi: https://doi.org/10.1016/S0031-8914(40)90098-2. URL https://www.sciencedirect.com/science/article/pii/S0031891440900982.

Anna Kutschireiter, Melanie A. Basnak, Rachel I. Wilson, and Jan Drugowitsch. Bayesian inference in ring attractor networks. *Proceedings of the National Academy of Sciences*, 120(9):e2210622120, 2023. doi: 10.1073/pnas.2210622120. URL https://www.pnas.org/doi/abs/10.1073/pnas.2210622120.

Wei Ji Ma, Jeffrey M. Beck, Peter E. Latham, and Alexandre Pouget. Bayesian inference with probabilistic population codes. *Nature Neuroscience*, 9(11):1432–1438, Nov 2006. ISSN 1546-1726. doi: 10.1038/nn1790. URL https://doi.org/10.1038/nn1790.

Paul Masset, Shanshan Qin, and Jacob A. Zavatone-Veth. Drifting neuronal representations: Bug or feature? *Biological Cybernetics*, 116(3):253–266, Jun 2022. ISSN 1432-0770. doi: 10.1007/s00422-021-00916-3. URL https://doi.org/10.1007/s00422-021-00916-3.

Carsen Stringer, Marius Pachitariu, Nicholas Steinmetz, Matteo Carandini, and Kenneth D. Harris. High-dimensional geometry of population responses in visual cortex. *Nature*, 571(7765):361–365, Jul 2019. ISSN 1476-4687. doi: 10.1038/s41586-019-1346-5. URL https://doi.org/10.1038/s41586-019-1346-5.

C. van Vreeswijk and H. Sompolinsky. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274(5293):1724–1726, 1996. doi: 10.1126/science.274.5293.1724. URL https://www.science.org/doi/abs/10.1126/science.274.5293.1724.

John J. Vastola, Zach Cohen, and Jan Drugowitsch. Is the information geometry of probabilistic population codes learnable? In Sophia Sanborn, Christian Shewmake, Simone Azeglio, Arianna Di Bernardo, and Nina Miolane, editors, *Proceedings of the 1st NeurIPS Workshop on Symmetry and Geometry in Neural Representations*, volume 197 of *Proceedings of Machine Learning Research*, pages 258–277. PMLR, 03 Dec 2023. URL https://proceedings.mlr.press/v197/vastola23a.html.

Maryna S. Viazovska. The sphere packing problem in dimension 8. *Annals of Mathematics*, 185(3):991–1015, 2017. ISSN 0003486X. URL http://www.jstor.org/stable/26395747.

William C. Waterhouse. Do symmetric problems have symmetric solutions? *The American Mathematical Monthly*, 90(6):378–387, 1983. ISSN 00029890, 19300972. URL http://www.jstor.org/stable/2975573.

Chuanming Zong. *Sphere packings*. Springer Science & Business Media, 2008.

Appendix A. Simplifying the objective

In this appendix, we will simplify the objective. Assuming a diagonalizable covariance matrix, we have

$$\boldsymbol{\Sigma}^{-1} = \boldsymbol{Q}^T \boldsymbol{\Lambda}^{-1} \boldsymbol{Q} \tag{14}$$

and can make a change of variables to

$$\boldsymbol{r} = \boldsymbol{\Lambda}^{-1/2} \boldsymbol{Q} \boldsymbol{z} \ . \tag{15}$$

In terms of this variable, the objective reads

$$J[\{\boldsymbol{r}_x\}] = -\sum_{x,y} p_0(x) p(y|x) \log p_{int}(y|x) + \alpha \sum_x p_0(x) \frac{\|\boldsymbol{r}_x\|_2^2}{2} \tag{16}$$

where p_{int} is now exclusively a function of the pairwise distances between the $\boldsymbol{r}_a$.

We can reparameterize this objective in terms of a center $\langle \boldsymbol{r} \rangle$ and pairwise distances d_{ij} between the $\boldsymbol{r}_a$. This is obvious for the Kullback-Leibler divergence term, so we only need to write the regularization term in terms of them. Note,

$$\|\boldsymbol{r}_x\|_2^2 = \|\boldsymbol{r}_x - \langle \boldsymbol{r} \rangle + \langle \boldsymbol{r} \rangle\|_2^2 = \|\boldsymbol{r}_x - \langle \boldsymbol{r} \rangle\|_2^2 + \|\langle \boldsymbol{r} \rangle\|_2^2 + 2(\boldsymbol{r}_x - \langle \boldsymbol{r} \rangle) \cdot \langle \boldsymbol{r} \rangle \ . \tag{17}$$

Next,

$$\sum_x p_0(x) \frac{\|\boldsymbol{r}_x\|_2^2}{2} = \sum_x p_0(x) \ \frac{1}{2} \left[\ \|\boldsymbol{r}_x - \langle \boldsymbol{r} \rangle\|_2^2 + \|\langle \boldsymbol{r} \rangle\|_2^2 \ \right] \ . \tag{18}$$

After some algebra, we can show

$$\|\boldsymbol{r}_x - \langle \boldsymbol{r} \rangle\|_2^2 = \sum_{a,b} p_0(a) p_0(b) \frac{1}{2} \left[\ \|\boldsymbol{r}_x - \boldsymbol{r}_a\|_2^2 + \|\boldsymbol{r}_x - \boldsymbol{r}_b\|_2^2 - \|\boldsymbol{r}_a - \boldsymbol{r}_b\|_2^2 \ \right] \ . \tag{19}$$

Finally,

$$\begin{aligned} \sum_x p_0(x) \frac{\|\boldsymbol{r}_x\|_2^2}{2} &= \sum_x p_0(x) \ \left[\ \sum_{a,b} p_0(a) p_0(b) \left(\frac{d_{xa}^2}{4} + \frac{d_{xb}^2}{4} - \frac{d_{ab}^2}{4} \right) + \frac{\|\langle \boldsymbol{r} \rangle\|_2^2}{2} \ \right] \\ &= \frac{\|\langle \boldsymbol{r} \rangle\|_2^2}{2} + \sum_{a,b} p_0(a) p_0(b) \frac{d_{ab}^2}{4} \ . \end{aligned} \tag{20}$$

Using this result, we get the reparameterized objective that appears in the main text.

Appendix B. Details of uniform calculation

Consider a Markov chain for which $p_0(x) = 1/M$ for all x, and $p(y|x) = \frac{1-\delta_{xy}}{M-1}$. This Markov chain is maximally symmetric, so our symmetric solution should have $d_{ij} = d$ for all $i, j \in \mathcal{X}$. This means

$$\begin{aligned} q(\boldsymbol{r}_y|\boldsymbol{r}_x) &= \frac{1-\delta_{xy}}{M-1} \\ p_e(\boldsymbol{r}_a|x) &= \frac{\delta_{ax}e^b + (1-\delta_{ax})e^{-d^2/2}}{e^b + (M-1)e^{-d^2/2}} \\ p_d(x|\boldsymbol{r}_a) &= p_e(\boldsymbol{r}_a|x) \ . \end{aligned} \tag{21}$$

Note that $Z(x) = Z$ is state-independent, and has value

$$Z = (M-1)e^{-d^2/2} \ . \tag{22}$$

Multiplying out the encoding and decoding models, we have

$$\begin{aligned} p_{int}(y|x) &= \frac{1}{M-1}\sum_{a,b} p_e(\boldsymbol{r}_b|y)(1-\delta_{ab})p_e(\boldsymbol{r}_a|x) \\ &= \frac{1}{(M-1)(e^b+Z)^2}\sum_{a\neq b}\left[\delta_{ax}e^b + (1-\delta_{ax})e^{-d^2/2}\right]\left[\delta_{by}e^b + (1-\delta_{by})e^{-d^2/2}\right] \\ &= \frac{1}{(M-1)(e^b+Z)^2}\left\{e^{2b} + 2(M-2)e^{b-d^2/2} + \left[M(M-1) - 2(M-2) + 1\right]e^{-d^2}\right\} \\ &= \frac{1}{(M-1)(e^b+Z)^2}\left\{e^{2b} + 2(M-2)e^{b-d^2/2} + (M^2-3M+5)e^{-d^2}\right\} \ . \end{aligned}$$

Taking a logarithm,

$$-\log p_{int} = \log(M-1) + 2\log(e^b+Z) - \log\left[e^{2b} + 2(M-2)e^{b-d^2/2} + (M^2-3M+5)e^{-d^2}\right] \ .$$

The objective becomes

$$\begin{aligned} J &= -\frac{1}{M(M-1)}\sum_{x\neq y}\log p_{int} + \alpha\frac{M(M-1)}{M^2}\frac{d^2}{4} \\ &= -\log p_{int} + \alpha\frac{M(M-1)}{M^2}\frac{d^2}{4} \\ &= \log(M-1) + 2\log(e^b+Z) - \log\left[e^{2b} + 2(M-2)e^{b-d^2/2} + (M^2-3M+5)e^{-d^2}\right] + \alpha\frac{M(M-1)}{M^2}\frac{d^2}{4} \ . \end{aligned}$$

Up to unimportant additive constants,

$$J = 2\log\left[e^b + (M-1)e^{-d^2/2}\right] - \log\left[e^{2b} + 2(M-2)e^{b-d^2/2} + (M^2-3M+5)e^{-d^2}\right] + \alpha\frac{M-1}{M}\frac{d^2}{4} \ .$$

Appendix C. Details of four state calculation

Consider a Markov chain for which $p_0(x) = 1/M$ for all x, and $p(y|x) = (\delta_{y,x+1} + \delta_{y,x-1})/2$ (where addition is done modulo M, although note that our state labels begin at 1 rather than 0). Here, we consider the case $M = 4$. The relevant symmetry group is D_4, the dihedral group of order 4. Symmetry constraints tell us that $d_{12} = d_{23} = d_{34} = d_{41} = d$, and that $d_{13} = d_{24} = L$.

Relevant quantities include

$$\begin{aligned} Z &= 2e^{-d^2/2} + e^{-L^2/2} \\ q(\boldsymbol{r}_b|\boldsymbol{r}_a) &= \frac{(\delta_{b,a+1} + \delta_{b,a-1})e^{-d^2/2} + \delta_{b,a+2}e^{-L^2/2}}{Z} \\ p_e(\boldsymbol{r}_a|x) &= \frac{\delta_{ax}e^b + Zq(\boldsymbol{r}_a|\boldsymbol{r}_x)}{e^b + Z} \\ p_d(y|\boldsymbol{r}_b) &= p_e(\boldsymbol{r}_b|y) \ . \end{aligned} \tag{23}$$

Note also that all functions are symmetric, e.g., $q(\boldsymbol{r}_b|\boldsymbol{r}_a) = q(\boldsymbol{r}_a|\boldsymbol{r}_b)$. We can compute p_{int}:

$$\begin{aligned} p_{int} &= \sum_{a,b} p_e(\boldsymbol{r}_b|y)q(\boldsymbol{r}_b|\boldsymbol{r}_a)p_e(\boldsymbol{r}_a|x) \\ &= \frac{1}{(e^b + Z)^2} \sum_{a,b} \left[\delta_{ax}e^b + Zq(\boldsymbol{r}_a|\boldsymbol{r}_x)\right] \left[\delta_{by}e^b + Zq(\boldsymbol{r}_b|\boldsymbol{r}_y)\right] q(\boldsymbol{r}_b|\boldsymbol{r}_a) \\ &= \frac{1}{(e^b + Z)^2} \left\{ e^{2b}q(\boldsymbol{r}_y|\boldsymbol{r}_x) + 2Ze^b \sum_{a \neq x,y} q(\boldsymbol{r}_y|\boldsymbol{r}_a)q(\boldsymbol{r}_a|\boldsymbol{r}_x) + Z^2 \sum_{a,b} q(\boldsymbol{r}_y|\boldsymbol{r}_b)q(\boldsymbol{r}_b|\boldsymbol{r}_a)q(\boldsymbol{r}_a|\boldsymbol{r}_x) \right\} \ . \end{aligned}$$

The above expression involves two- and three-step probabilities. These can be computed exactly in this case, although it is tedious. They technically only need to be computed (by symmetry) for one of the adjacent transitions (e.g., $1 \to 2$), since those are the only transitions that contribute to the objective function.

Let 1 label the starting state, 2 and 3 be adjacent vertices, and 4 be the farther vertex. The two-step probabilities are (9 relevant paths, quantified using Z):

$$\begin{aligned} q^{(2)}(1 \to 1) &= q_{12}q_{21} + q_{13}q_{31} + q_{14}q_{41} = \frac{2e^{-d^2} + e^{-L^2}}{Z^2} \\ q^{(2)}(1 \to 2) &= q_{13}q_{32} + q_{14}q_{42} = \frac{2e^{-d^2/2 - L^2/2}}{Z^2} \\ q^{(2)}(1 \to 4) &= q_{12}q_{24} + q_{13}q_{34} = \frac{2e^{-d^2}}{Z^2} \ . \end{aligned} \tag{24}$$

The relevant three-step probability is a sum of the probabilities of several paths:

$$\begin{aligned} \frac{2e^{-\frac{3}{2}d^2} + e^{-d^2/2 - L^2}}{Z^3} &= q_{12}\left[q_{21}q_{12} + q_{23}q_{32} + q_{24}q_{42}\right] \\ \frac{2e^{-\frac{3}{2}d^2}}{Z^3} &= q_{13}\left[q_{34}q_{42} + q_{31}q_{12}\right] \\ \frac{2e^{-d^2/2 - L^2}}{Z^3} &= q_{14}\left[q_{43}q_{32} + q_{41}q_{12}\right] \ . \end{aligned} \tag{25}$$

Overall, the relevant three-step probability is

$$q^{(3)}(1 \to 2) = \frac{4e^{-\frac{3}{2}d^2} + 3e^{-d^2/2 - L^2}}{Z^3} \ . \tag{26}$$

Finally, we can write that the relevant part of p_{int} is

$$\begin{aligned} p_{int} &= \frac{1}{(e^b + Z)^2 Z} \left\{ e^{2b} e^{-d^2/2} + 2e^b (2e^{-d^2/2 - L^2/2}) + (4e^{-\frac{3}{2}d^2} + 3e^{-d^2/2 - L^2}) \right\} \\ &= \frac{e^{-d^2/2}}{(e^b + Z)^2 Z} \left\{ e^{2b} + 4e^b e^{-L^2/2} + (4e^{-d^2} + 3e^{-L^2}) \right\} \ . \end{aligned} \tag{27}$$

Taking a logarithm,

$$-\log p_{int} = \log Z + 2\log(e^b + Z) + \frac{d^2}{2} - \log \left[e^{2b} + 4e^b e^{-L^2/2} + 3e^{-L^2} + 4e^{-d^2} \right] \ . \tag{28}$$

The objective becomes

$$\begin{aligned} J &= \log Z + 2\log(e^b + Z) + \frac{d^2}{2} - \log \left[e^{2b} + 4e^b e^{-L^2/2} + 3e^{-L^2} + 4e^{-d^2} \right] + \frac{\alpha}{4} \frac{2}{M^2} (4d^2 + 2L^2) \\ &= \log Z + 2\log(e^b + Z) + \frac{d^2}{2} - \log \left[e^{2b} + 4e^b e^{-L^2/2} + 3e^{-L^2} + 4e^{-d^2} \right] + \frac{\alpha}{4} \left(\frac{d^2}{2} + \frac{L^2}{4} \right) \end{aligned}$$

where the regularization term comes from counting the number of pairings of each kind.

Appendix D. Details of general cyclic topology calculation

As in the previous appendix, consider a Markov chain for which $p_0(x) = 1/M$ for all x, and $p(y|x) = (\delta_{y,x+1} + \delta_{y,x-1})/2$ (where addition is done modulo M). For simplicity, assume that M is even, i.e., that $M = 2N$ for some integer $N \geq 1$. Symmetry constraints tell us that distances should only depend on two vertices' relative positions along the 'ring'. For example,

$$d_{12} = d_{23} = \cdots = d_{x,x+1} \tag{29}$$

for any $x \in \mathcal{X}$. For M even, there are $M/2$ unique distances that we must optimize (i.e., one hop away, two hops away, and so on); two vertices can only be at most $M/2$ edges apart. We will label these distances as d_1, d_2, ..., $d_{M/2}$.

Relevant quantities include

$$\begin{aligned} Z &= e^{-d_N^2/2} + \sum_{k=1}^{N-1} 2e^{-d_k^2/2} \\ q(\boldsymbol{r}_b|\boldsymbol{r}_a) &= \frac{\delta_{b,a+N} e^{-d_N^2/2} + \sum_{k=1}^{N-1} (\delta_{b,a+k} + \delta_{b,a-k}) e^{-d_k^2/2}}{Z} \\ p_e(\boldsymbol{r}_a|x) &= \frac{\delta_{ax} e^b + Z q(\boldsymbol{r}_a|\boldsymbol{r}_x)}{e^b + Z} \\ p_d(y|\boldsymbol{r}_b) &= p_e(\boldsymbol{r}_b|y) \ . \end{aligned} \tag{30}$$

As in the previous appendix, we can write p_{int} in terms of two- and three-step transition probabilities:

$$p_{int}(y|x) = \frac{e^{2b} q(\boldsymbol{r}_y|\boldsymbol{r}_x) + 2Ze^b \sum_{a \neq x,y} q(\boldsymbol{r}_y|\boldsymbol{r}_a) q(\boldsymbol{r}_a|\boldsymbol{r}_x) + Z^2 \sum_{a,b} q(\boldsymbol{r}_y|\boldsymbol{r}_b) q(\boldsymbol{r}_b|\boldsymbol{r}_a) q(\boldsymbol{r}_a|\boldsymbol{r}_x)}{(e^b + Z)^2} \ .$$

The only difference is that these probabilities are now slightly more annoying to compute. Fortunately, only a single transition—the nearest neighbor transition, from any x to $x+1$ (or equivalently, to $x-1$)—contributes to the objective, which by symmetry is equal to

$$\begin{aligned} J &= -\log p_{int}(2|1) + \frac{2\alpha}{4M^2} \sum_{k=1}^{M-1} (M-k) d_k^2 \\ &= -\log p_{int}(2|1) + \frac{2\alpha}{4M} \left[\frac{1}{2} d_N^2 + \sum_{k=1}^{N-1} d_k^2 \right] \end{aligned} \tag{31}$$

where $x = 1$ has been chosen arbitrarily, and where the details of the regularization term come from counting the pairwise distances of each kind.

The second term in $p_{int} := p_{int}(2|1)$ involves $M-2$ nonzero terms, and evaluates to

$$2Ze^b \sum_{a \neq x,y} q(\boldsymbol{r}_y|\boldsymbol{r}_a) q(\boldsymbol{r}_a|\boldsymbol{r}_x) = \frac{4e^b}{Z} \sum_{k=1}^{N-1} e^{-d_k^2/2 - d_{k+1}^2/2} \ . \tag{32}$$

As mentioned in the main text, we will simplify this calculation by considering the $b \to \infty$ limit. In this limit, the three-step transition probability term can be ignored since it is of order e^{-2b}. We have

$$\begin{aligned} p_{int} &= \frac{q(\boldsymbol{r}_y|\boldsymbol{r}_x) + 2Ze^{-b}\sum_{a\neq x,y} q(\boldsymbol{r}_y|\boldsymbol{r}_a)q(\boldsymbol{r}_a|\boldsymbol{r}_x) + Z^2e^{-2b}\sum_{a,b} q(\boldsymbol{r}_y|\boldsymbol{r}_b)q(\boldsymbol{r}_b|\boldsymbol{r}_a)q(\boldsymbol{r}_a|\boldsymbol{r}_x)}{(1+Ze^{-b})^2} \\ &\approx \left[q(\boldsymbol{r}_y|\boldsymbol{r}_x) + 2Ze^{-b}\sum_{a\neq x,y} q(\boldsymbol{r}_y|\boldsymbol{r}_a)q(\boldsymbol{r}_a|\boldsymbol{r}_x)\right]\left[1-2Ze^{-b}\right] \\ &\approx q(\boldsymbol{r}_y|\boldsymbol{r}_x) + 2Ze^{-b}\left\{\sum_{a\neq x,y} q(\boldsymbol{r}_y|\boldsymbol{r}_a)q(\boldsymbol{r}_a|\boldsymbol{r}_x) - q(\boldsymbol{r}_y|\boldsymbol{r}_x)\right] \end{aligned}$$

to first order in e^{-b}. Explicitly, since only the $y = x+1$ term matters,

$$\begin{aligned} p_{int} &\approx \frac{e^{-d_1^2/2} + 2e^{-b}\left[2\sum_{k=1}^{N-1} e^{-d_k^2/2 - d_{k+1}^2/2} - Ze^{-d_1^2/2}\right]}{Z} \\ &= \frac{e^{-d_1^2/2}}{Z}\left\{1 + 2e^{-b}\left[2\sum_{k=1}^{N-1} e^{-d_k^2/2 - d_{k+1}^2/2 + d_1^2/2} - Z\right]\right\} . \end{aligned}$$

We can now write

$$\begin{aligned} J &= \frac{d_1^2}{2} + \log Z - \log\left\{1 + 2e^{-b}\left[2\sum_{k=1}^{N-1} e^{-d_k^2/2 - d_{k+1}^2/2 + d_1^2/2} - Z\right]\right\} + \frac{2\alpha}{4M}\left[\frac{1}{2}d_N^2 + \sum_{k=1}^{N-1} d_k^2\right] \\ &\approx \frac{d_1^2}{2} + \log\left[e^{-d_N^2/2} + \sum_{k=1}^{N-1} 2e^{-d_k^2/2}\right] - 2e^{-b}\left[2\sum_{k=1}^{N-1} e^{-d_k^2/2 - d_{k+1}^2/2 + d_1^2/2} - Z\right] + \frac{2\alpha}{4M}\left[\frac{1}{2}d_N^2 + \sum_{k=1}^{N-1} d_k^2\right] . \end{aligned}$$