Supplementary Math Note

Flexible modulation of sequence generation in the entorhinal-hippocampal system

Daniel C. McNamee*, Kimberly L. Stachenfeld, Matthew M. Botvinick, Samuel J. Gershman
*daniel.c.mcnamee@gmail.com

A Towards spectral modulation in biologically detailed EHC network architectures

A.1 Hexagonal spectral components based on non-negative generator factorization

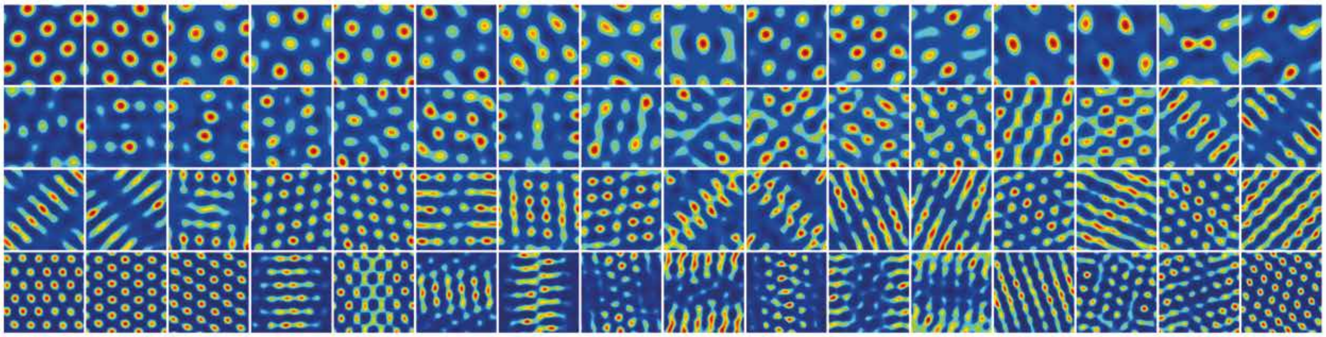

Figure N1. We present grid firing maps corresponding to all columns of the G matrix which optimizes Eqn. 20 for a random-walk generator O^{rw} in an open field environment. The ordering (induced from the G matrix) is approximately organized from large to small spatial scales. Note that hexagonal firing fields are evident across multiple spatial scales.

In the exposition in the main text, we associate generator eigenvectors with medial entorhinal grid cells. This association abstracts away several known features of grid cells in order to facilitate a parsimonious analysis of the generator-propagator framework. In order to pave the way for the implementation of spectral modulation in biologically plausible architectures, we show how features such as non-negative activation functions and hexagonally organized firing fields may be incorporated. Generator eigendecompositions correspond to the solution of the following constrained optimization problem:

$$\underset{G^\top G=I}{\text{argmin}} \, ||O - G\Lambda G^\top||_F^2 \tag{19}$$

where O is a generator and $||\cdot||_F$ is the Frobenius norm. In particular, if the matrix objective is minimized at zero and G is orthonormal (i.e., $G^\top G = I$) then $O = G\Lambda G^\top$ implying that $OG = G\Lambda$. Therefore the columns of G are generator eigenvectors and Λ is a diagonal matrix with eigenvalues being the only non-zero components. This is the scenario studied in the main text where $G^\top = W$ models the linear readout from MEC to HC.

We augment the objective Eqn. 19 with a non-negativity constraint leading to

$$\underset{G^\top G=I, G\geq 0}{\text{argmin}} \, ||O - G\Lambda G^\top||_F^2 \; . \tag{20}$$

In a recent theoretical study which unifies several models of grid coding via pattern formation theory, it was demonstrated analytically that the imposition of a non-negativity constraint on an analogous matrix decomposition necessarily leads to hexagonally organized grid fields[55]. We examined whether this principle also applies to the matrix decomposition of generators (Eqn. 20). In Fig. N1, we present the firing maps drawn for the grid code matrix G which optimizes Eqn. 20 for a random-walk generator O^{rw} based on the Generalized Hebbian algorithm and ReLU activation functions. Hexagonally organized firing fields were reliably observed at multiple spatial scales. In accordance with the analysis presented in Ref. [55], the emergence of such hexagonally organized fields are apparently due to the fact that the spatial covariance of generators are similar to those of difference-of-gaussian functions as commonly implemented for place cell activation maps in continuous attractor network models. This is because generators tend to have a center-surround spatial structure "inhibiting" locally and "exciting" at moderate ranges (note that this is "inverted" with respect to difference-of-gaussian place activation maps). Ref. [55] shows that the resulting spatial covariance, in conjunction with a non-negativity constraint, leads to hexagonally organized firing fields.

A.2 Distributed spectral modulation across a modular, non-orthogonal grid code

Another assumption implicit in our model that is not reflected in the biological reality of the entorhinal cortex is that the spectral components be orthogonal. Consider the generator similarity transformation $O = GDG^{-1}$ where G is not necessarily orthonormal and D is not diagonal. This implies that the spectral modulation mechanism can no longer be directly applied since the matrix exponential $S(D) = e^{-\tau^{-1}|D|^\alpha}$ becomes computationally costly. In this section, we elaborate on an extended model which implicitly "whitens" the incoming grid code thus facilitating spectral modulation in a distributed fashion. This extension is based on the modular architecture of grid cells in entorhinal cortex[30].

Diagonalizing $D = V\Lambda V^{-1}$ implies that the generator decomposition $O = GV\Lambda V^{-1}G^{-1}$ leads to the propagator $\rho_t = \rho_0 GVS(\Lambda)V^{-1}G^{-1}$. However, spectral modulation via $S(\Lambda)$ now applies to linear mixtures of grid cells based on V. Consider the calculation of entries in the matrix GVS:

$$[GVS]_{ik} = \sum_j G_{ij} S_{kk} V_{jk} \tag{21}$$

$$[GVS]_{ik'} = \sum_j G_{ij} S_{k'k'} V_{jk'} \tag{22}$$

We note that this calculation requires two distinct spectral modulations (by S_{kk} and $S_{k'k'}$) of the same grid component G_{ij}. In order to accommodate this computation in the pathway from MEC to HC, we construct a population $\tilde{G}$ of grid cells across grid modules as

$$\tilde{G}_{i(jk)} = G_{ik} \tag{23}$$

where i indexes spatial position, k indexes the grid module, and j indexes the cell within that module. In order to process a full-rank decomposition of D, each grid module k is composed of n_G copies of the same grid cell defined by k-th column of G though a low-rank approximation may be sufficient in general. The aforementioned matrix entries of GVS then become

$$[GVS]_{ik} = \sum_j \tilde{G}_{i(jk)} V_{jk} S_{kk} \tag{24}$$

$$[GVS]_{ik'} = \sum_j \tilde{G}_{i(jk')} V_{jk'} S_{k'k'} \quad . \tag{25}$$

Note that $\tilde{G}$ can be considered as a tensor of order 3 or a matrix (order-2 tensor) where the cell j and module k indices have been collapsed to a single index (jk). The receptive field ψ_j^k the jth cell in module k can be expressed in functional form as

$$\psi_j^k(x_i) = \tilde{G}_{i(jk)} \quad . \tag{26}$$

Indeed treating (jk) as a single cell-module index, we re-define the following cell-module spectral modulation matrix $\tilde{S}$ to have entries

$$\left[\, \tilde{S} \,\right]_{(jk)(jk)} = V_{jk} S_{kk} \tag{27}$$

$$\tag{28}$$

implying that

$$[GVS]_{ik} = \sum_j \left[\, \tilde{G}\,\tilde{S} \,\right]_{i(jk)} \quad . \tag{29}$$

In particular, for (τ, α)-modulation, the cell-module spectral modulation matrix is

$$\tilde{S}_{(jk)(jk)} = V_{jk} e^{-\tau^{-1}|\Lambda_{kk}|^{\alpha}} \quad . \tag{30}$$

In complementary fashion, we define the readout tensor $\tilde{W}$ according to

$$\left[\, \tilde{W} \,\right]_{(jk)l} = \left[V^{-1}G^{-1}\right]_{kl} \quad . \tag{31}$$

The modular spectral modulation circuit in matrix format is then

$$\rho_t = \rho_0 \tilde{G}\,\tilde{S}\,\tilde{W} \quad . \tag{32}$$

A.3 Including generators in continuous attractor networks

Continuous attractor networks (CANs) have been used to model path integration[19,56,57], spatial localization using landmarks[18,20], and spatial navigation[55,58] in the entorhinal-hippocampal circuit. In this section, we discuss the relationship between CAN models and the generator model. In summary, continuous-time Markov processes (such as the generator model) form a subset of linear dynamical systems with specific constraints on the evolution matrix which ensure that the system state vector represents a valid probability distribution for all time. Therefore, the generator model can be conceptualized as a linearized "effective" model of the evolution of a neural population's firing rate vector encoding a spatial distribution. From this perspective, we outline the relevant connections to firing rate CANs.

In CANs, the firing rate vector s of a recurrent network containing a finite[1] population of N neurons evolves according to[2]

$$\begin{aligned}\dot{s} &= -s + \mathcal{F}\left[W^{\text{rec}} \cdot s + b\right] \\ \dot{s}_i(t) &= -s_i(t) + \mathcal{F}\left[\sum_{j=1}^{N} W_{ij}^{\text{rec}} s_j(t) + b_i\right]\end{aligned} \tag{33}$$

where the recurrent synaptic weight matrix W^{rec} serves as the evolution operator and $\mathcal{F}$ is a rectified linearity. We use a left matrix multiplication convention here in network equations which contrasts with the right matrix multiplication convention used for master equations. A range of combinations of $\mathcal{F}$ and W^{rec} (e.g., reflecting short-range excitation and long-range inhibition) can be specified which ensures that the system described by Eqn. 33 is an attractor network with a continuous set of steady-state activity bumps[19,20,59]. In models of path integration in entorhinal cortex, the firing rate vector s describes the activity of grid cells which are coupled to velocity-conjunctive cells v in the augmented dynamics equation

$$\dot{s} = -s + \mathcal{F}\left[W^{\text{rec}} \cdot s + gW^{\text{vel}} \cdot v\right] \tag{34}$$

where g_i are gains which control the period of the corresponding grid cell.

In a study of planning based on the successor representation[58], a network of grid cells s receives a representation of an initial position and a goal position in a coordinate system corresponding to low-dimensional projection of the random policy successor representation which we consider to be an SR grid code[28]. The network connectivity structure is constructed such that the latent SR representation smoothly evolves from the initial state SR grid code to the goal SR grid code over time. A separate population then continuously reads out the latent SR representation to a place code ρ resulting in a preplay-like mechanism. In our notation, this attractor network model is governed by the differential equation:

$$\dot{\rho} = -\rho + \mathcal{F}\left[(1-\epsilon)W^{\text{rec}}\rho + \kappa W^{\text{ff}} s^{\text{in}}\right] \tag{35}$$

where the weight matrices W^{rec} and W^{ff} correspond to recurrent and feedforward connections (from the SR grid population to place cells) respectively. The network is parametrized by the input strength κ, and the decay rate ϵ. The feedforward weight matrix W^{ff} projects an SR grid code s to the preferred location of each unit ρ_i in an analogous fashion to the W matrix in our formulation drawn from a generator eigendecomposition.

In the linear regime of $\mathcal{F}$, ρ evolves according to

$$\dot{\rho} \approx -\epsilon\rho + \kappa\rho^{in} \tag{36}$$

with the ratio between κ and ϵ controlling the speed at which the system evolves towards the input place code $\rho^{in} \approx W^{\text{ff}} s^{\text{in}}$. Incorporating a generator, we have

$$\dot{\rho} \approx -\epsilon\rho + \rho O + \kappa\rho^{in} \quad . \tag{37}$$

These place cell dynamics are generated by the attractor network:

$$\dot{\rho} = -\rho + \mathcal{F}\left[(1-\epsilon)W^{\text{hip}}\rho + W_{\tau,\alpha}^{\text{grid}}\rho + \kappa W^{\text{in}}\rho^{\text{in}}\right] \quad . \tag{38}$$

This architecture has two distinct recurrent pathways corresponding to a "small-loop" within hippocampus (with weight matrix W^{hip}) and a "big-loop" (implemented by weight matrix $W_{\tau,\alpha}^{\text{grid}}$) which filters hippocampal activity through the entorhinal cortex. Notably, a common feature of continuous attractor models of EHC is long-range inhibition in the recurrent weights W^{hip}. This stabilizes a "bump" of activity during self-localization. In our model, the same mechanism may serve to ensure that sampling propagates away from a rodent's location during diffusive exploration due to the strong inhibitory feedback from the simultaneous concurrent activation of multiple non-local place cells. The big-loop is composed of two layers $W_{\tau,\alpha}^{\text{grid}} = W^{\text{hip}\to\text{ec}} W^{\text{ec}\to\text{hip}}$ which are specified by the propagator eigendecomposition $e^{\tau^{-1}O} = GS_{\tau,\alpha}W$. The weights $W^{\text{hip}\to\text{ec}} \equiv GS_{\tau,\alpha}$ transform a place code to a spectrally modulated grid code and $W^{\text{ec}\to\text{hip}} \equiv W$ maps from a grid code to a predictive map place code. There is a single two-layer feedforward pathway integrating input information ρ^{in} filtered through the entorhinal cortex $W^{\text{in}} = W^{\text{in}\to\text{ec}} W^{\text{ec}\to\text{hip}}$.

[1] See[20] for the analogous definitions in terms of a "neural sheet" with an infinite number of neurons.

[2] Network models are usually endowed with a time constant parameter. We have set the intrinsic time constant of the network to 1 in order to avoid confusion with the propagator time constant τ.

A.4 Aligning the distribution of spectral components with the medial entorhinal distribution of grid scales

Although our approach based on decomposing a discrete generator results in a discretization of spatial scales in the resulting spectral components, this discretization of scales lacks a key feature that is well-established in the modularity of grid cell populations in the medial entorhinal cortex. Namely, that the spatial scales of grid firing maps are successively separated by a ratio of approximately $\sqrt{2}$ across distinct grid modules[30] (Fig. N2A). We sought to examine whether this scale-ratio principle posed an irreconcilable obstacle to our proposed spectral modulation mechanism. In a simulation-based experiment extending that presented in Fig. 2, we spectrally modulated random-walk sequence generation using two distinct propagators constructed from a *dense* population composed of all spectral components and a *scale-ratio* population composed of a randomly sampled subset with spatial scales adhering to the scale-ratio feature of MEC grid modules (Fig. N2B). For the dense population, sequence generation was shifted between diffusion and superdiffusion based on (τ, α)-spectral density formula (Eqn. 7). With respect to exploration efficiency, superdiffusive sequence generation ($\alpha = 0.5$) outperformed diffusive sequence generation ($\alpha = 1$) as expected (Fig. N2C). For the scale-ratio population G^{sr}, we optimized a spectral density by minimizing the KL-divergence between the induced scale-ratio propagator $P_{\alpha}^{\text{sr}} = G^{\text{sr}} S_{\alpha}^{\text{sr}} W^{\text{sr}}$ and the target (dense population) propagators $P_{\alpha=1}$ for diffusion and $P_{\alpha=0.5}$ for superdiffusion:

$$S_{\alpha}^{\text{sr}} = \underset{S}{\operatorname{argmin}} \, \text{KL} \left[G^{\text{sr}} S W^{\text{sr}} || P_{\alpha} \right] \quad . \tag{39}$$

The spectral density parameters in the diagonal matrix S were initialized according to the (τ, α)-spectral density formula. We hypothesized that including a spectral optimization step may endow the scale-ratio population with the ability to adaptively mitigate for the absence of spectral components which do not adhere to the scale-ratio modularity principle. In all cases, τ was chosen to equalize the modal propagation across diffusive and superdiffusive propagators. As for the dense population, superdiffusive sequences significantly outperformed diffusive sequences in terms of exploration efficiency (Fig. N2D). Furthermore, the sequences generated by the scale-ratio population reflected the same qualitative structure as those generated the dense population as predicted by our theory. Specifically, diffusive sequences (Fig. N2I, J) random-walked based on localized, approximately Gaussian, propagators, and superdiffusive sequences (Fig. N2K, L) interleaved local steps with large jumps. These simulations indicate that the scale-ratio modularity of MEC can be incorporated in our model. However we emphasize that our theory does not necessarily predict this feature of the population grid code. In order to address this, we suggest that efficient coding models employed to explain the scale-ratio principle in terms of hierarchical spatial encoding may be fruitfully extended to the case of propagator encoding[60,61].

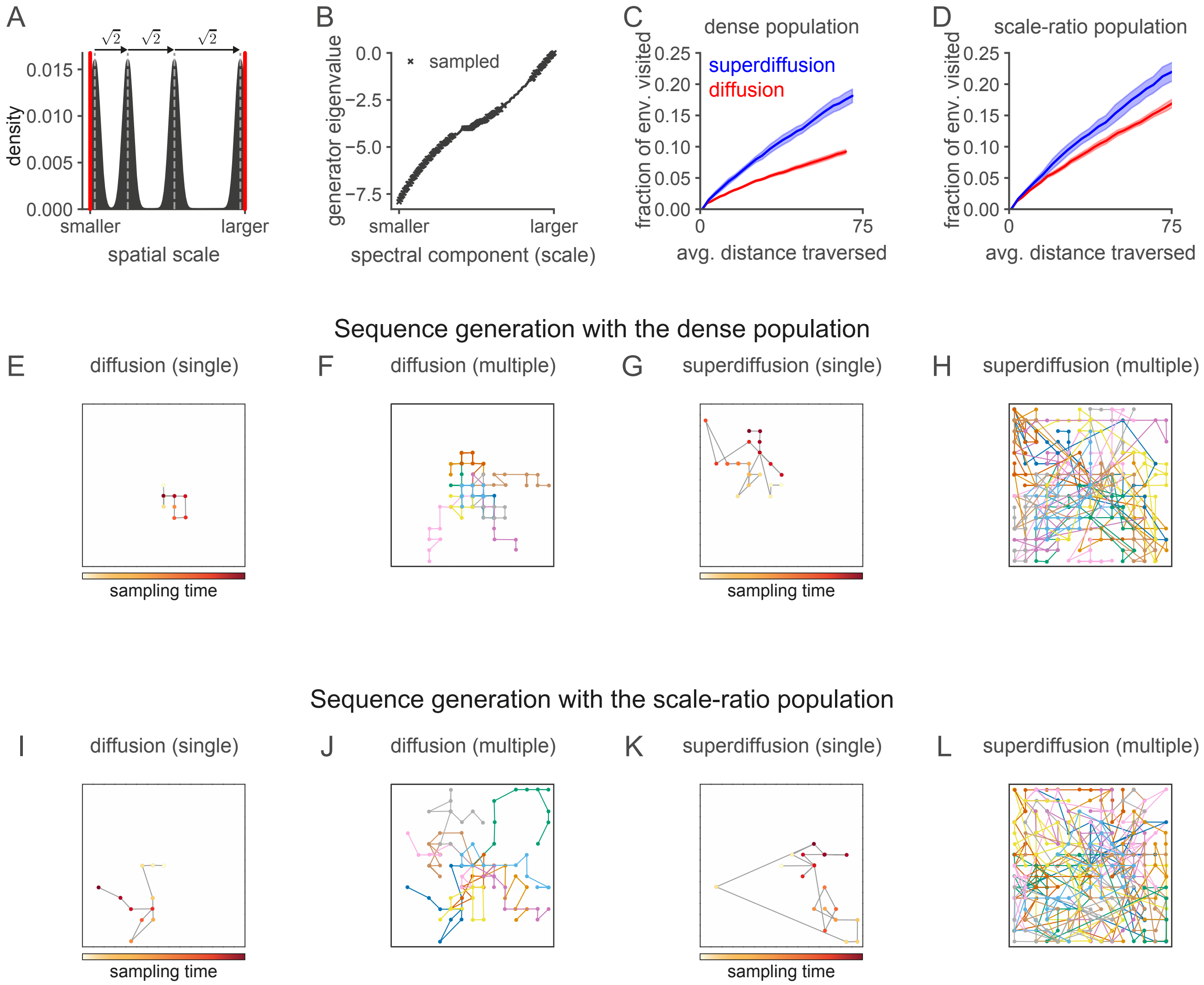

Figure N2. **A.** Sampling density of spectral components consistent with the scale-ratio modularity principle. **B.** Generator eigenvalues of sampled spectral components. **C.** Exploration efficiency curves for dense population of spectral components (i.e., including all generator eigenvectors). Curves and error bands reflect the mean and standard error respectively for $n = 10$ simulated trajectories. **D.** Exploration efficiency curves for scale-ratio population of spectral components. **E-H.** Diffusive and superdiffusive sequences generated by the dense propagator. **I-L.** Diffusive and superdiffusive sequences generated by the scale-ratio propagator.

B Theory of generators and propagators

B.1 Generator constraints

A generator matrix O must satisfy the following constraints[23]

$$\begin{aligned} O_{ij} &\geq 0 \ , \ \ x_i \neq x_j & (40) \\ O_{ii} &\leq 0 & (41) \\ \sum_{j=1}^{|\mathcal{X}|} O_{ij} &= 0 & (42) \end{aligned}$$

for all states $x_i, x_j \in \mathcal{X}$. These three contraints have intuitive interpretations. The first constraint (Eqn. 40) implies an attractive "force" between states, the second constraint (Eqn. 41) enforces a repulsive "force" from any state, and the third constraint (Eqn. 42) ensures that the total state probability is conserved through time. That is, it is impossible for a particle stochastically evolving in this state-space to "leave" the state-space.

B.2 Generator interpretation

We have modeled sequence sampling using a master equation (Eqn. 1, main text, omitting τ for simplicity):

$$\dot{\rho} = \rho O \tag{43}$$

where the notation $\dot{\rho}$ indicates the time derivative of ρ. This equation describes the continuous-time evolution of the state probability vector ρ and has the solution

$$\rho_t = \rho_0 e^{tO} \ . \tag{44}$$

We consider its Taylor expansion in t around the zero timepoint

$$\begin{aligned} \rho_t &\approx \rho_0 + \left[\frac{d\rho_t}{dt}\right]_{t=0} (t-0) + \frac{1}{2}\left[\frac{d^2\rho_t}{dt^2}\right]_{t=0} (t-0)^2 + \cdots & (45) \\ &= \rho_0 + \rho_0 t O + \frac{1}{2}\rho_0 t^2 O^2 + \cdots \ . & (46) \end{aligned}$$

Ignoring second-order terms above, we see that O plays the role of the time Jacobian in a linearization of the Markov process dynamics. Put another way, the generator matrix O defines the state transition dynamics at very short timescales

$$O_{ij} = \lim_{\Delta t \to 0} \frac{\mathrm{P}(X_{t+\Delta t} = x_j | X_t = x_i)}{\Delta t} \ \ , \ \ i \neq j \ . \tag{47}$$

Generators in our continuous-time formulation are analogous to stochastic matrices in discrete-time Markov chains and transition matrices in reinforcement learning models.

B.3 Efficient propagation to non-local timepoints

The master solution $\rho_t = \rho_0 e^{tO}$ requires the computation of a matrix exponential. Therefore, in general, a direct calculation of ρ_t is infeasible since it requires an infinite sum over matrix powers:

$$e^{tO} = \sum_{n=0}^{\infty} \frac{t^n}{n!} O^n \ . \tag{48}$$

However, the solution (Eqn. 2) can be computed efficiently using the eigendecomposition of the infinitesimal generator O. Consider the diagonalization $O = G\Lambda W$ where Λ is a diagonal matrix containing the corresponding eigenvalues of O and $W = G^{-1}$, then

$$\begin{aligned} e^{tO} &= \sum_{n=0}^{\infty} \frac{t^n (G\Lambda W)^n}{n!} \\ &= I + tG\Lambda W + 2^{-1}t^2 G\Lambda(WG)\Lambda W + \cdots \\ &= I + tG\Lambda W + 2^{-1}t^2 G\Lambda^2 W + \cdots \\ &= G\left(I + t\Lambda + 2^{-1}t^2\Lambda^2 + \cdots\right) W \\ &= G\left(\sum_{n=0}^{\infty} \frac{t^n \Lambda^n}{n!}\right) W \\ &= Ge^{t\Lambda}W \ . & (49) \end{aligned}$$

The key observation is that the matrix exponential of a diagonal matrix Λ is relatively trivial since multiplying a diagonal matrix by itself simply squares the individual diagonal values.

B.4 Generator eigenspectrum

Each eigenvalue $\lambda_k = [\Lambda]_{kk}$ defines the rate of the exponential decay in the corresponding eigenvector over time (note that $\lambda_k \leq 0$ for any generator matrix O). In the limit $t \to \infty$, the contributions of all eigenvectors decay to zero since $e^{t\lambda} \to 0$ for $\lambda < 0$ except the the zeroth eigenvector corresponding to the zero eigenvalue (for any t, $e^{t\lambda} = 1$ if $\lambda = 0$). Therefore, the unique eigenvector with an eigenvalue of zero encodes the stationary density of the process in the limit as $t \to \infty$. In particular, it is the (elementwise) square-root of the stationary distribution.

B.5 The random walk generator

We describe the continuous-time random walk process on an undirected, weighted graph $\mathcal{X}$. The random walk generator O^{rw} is constructed from the weight matrix X of the graph $\mathcal{X}$. The weight matrix components X_{ij} take a strictly positive value whenever there is an edge connecting two states x_i and x_j, and is zero otherwise. With respect to the stochastic dynamics of a random walk on $\mathcal{X}$, the value X_{ij} reflects how likely it is that a step occurs between these two states. Specifically, if $X_{ij} > X_{ik}$, then the process is more likely to step to state x_j than to x_k from state x_i. Given a weight matrix X, the random walk generator O^{rw} is defined as

$$O^{\text{rw}}_{ij} = \begin{cases} X_{ij} & i \neq j \\ -\sum_{j' \neq i} X_{ij'} & i = j \end{cases} \tag{50}$$

or $O^{\text{rw}} = X - D$ in matrix notation where D is the diagonal *degree matrix* of the graph $\mathcal{X}$ with entries $D_{ii} = \sum_{j'} X_{ij'}$ equal to the "degree" of x_i. From a graph-theoretical point of view, $O^{\text{rw}} \equiv -L_{\mathcal{X}}$ where $L_{\mathcal{X}}$ is the unnormalized graph Laplacian[62]. Note that this implies that $\sum_j O^{\text{rw}}_{ij} = 0$ for all $x_j \in \mathcal{X}$ consistent with the definition of a generator[23].

B.6 To the spatial continuum limit in one dimension

Although we develop our model in the context of discrete state-spaces, it recovers an analogous continuous state-space model in the continuum limit. We summarize the connection between graph random walks (Section. B.5) and diffusions in continuous spaces. Localizing the master equation at a particular state $x_i \in \mathcal{X}$, we have

$$\dot{\rho}_i = \sum_{j=1}^{|\mathcal{X}|} [\rho_j O_{ji} - \rho_i O_{ij}] \tag{51}$$

This derivation implies that the time derivative of the x_i state probability is given by the "current" of particles $\rho_j O_{ji}$ jumping into x_i minus the "current" of particles $\rho_i O_{ij}$ jumping away from x_i. We analyze this local form of the master equation in the special case where the state-space $\mathcal{X}$ is an infinite one-dimensional lattice embedded in the real line at spatial intervals Δx. This puts strong constraints on the form of the O-matrix – it can only be nonzero for neighbouring states. Assuming that $O_{ij} = \kappa$ is constant for neighbouring states only (i.e., O_{ij} is only nonzero for $j = i - 1$ or $j = i + 1$), the local form of the master equation is

$$\begin{aligned} \dot{\rho}_i &= \sum_{j=1}^{|\mathcal{X}|} [\rho_j O_{ji} - \rho_i O_{ij}] \\ &= \underbrace{(\kappa\rho_{i+1} - \kappa\rho_i)}_{\text{forward difference}} + \underbrace{(\kappa\rho_{i-1} - \kappa\rho_i)}_{\text{backward difference}} \\ &= \kappa [\rho_{i+1} + \rho_{i-1} - 2\rho_i] \\ &= \kappa\Delta x^2 \left[\frac{\rho_{i+1} + \rho_{i-1} - 2\rho_i}{\Delta x^2} \right] \end{aligned} \tag{52}$$

Now consider the limit $\Delta x^2 \to 0$ such that κ is scaled in order to keep $K = \kappa\Delta x^2$ constant. That is, as $\Delta x^2 \to 0$, $\kappa \to \infty$ in order to reflect the same transition rate K integrated over shorter and shorter spatial scales. The other component in the above expression is the central finite difference scheme for the Laplacian operator ∇^2 (the backward and forward differences have been highlighted), therefore we arrive at the diffusion equation in the limit $\Delta x \to 0$:

$$\dot{\rho} = K\nabla^2\rho \tag{53}$$

where ρ is now a function $\rho(x, t)$ of both space and time. This derivation highlights how the random walk generator O on a lattice generalizes the diffusion operator $K\nabla^2$ (in one dimension).

B.7 Plane wave solutions in the continuum limit and mathematical conventions

The (continuous) diffusion equation (Eqn. 53) can be solved via separation of variables leading to a family of solutions in the Fourier eigenfunction basis composed of plane waves e^{ikx} for real k. In analogy with the proposed EHC mechanism, matrix multiplication by G (Fig. 2A) corresponds to the Fourier transformation of the initial condition. Then, spectral modulation corresponds to multiplying each of the resulting spectral components (i.e., weighted plane waves parameterized by wavenumbers) by $e^{-\tau^{-1}k^2 t}$ and the linear readout via W is the inverse transformation which superimposes the temporally decayed plane waves to form the propagator at another timepoint. In the continuum limit, the wavenumber can be any positive real number $k \in \mathbb{R}_+$, however in discretizing the underlying space on a lattice (Eqn. 52), the wavenumbers are discretized. Generator eigenvalues $\{\lambda_k\}_{k=1,\ldots,|\mathcal{X}|}$ are analogous to discretized wavenumbers. Specifically, $\lambda \equiv -k^2$ in the continuum limit for some k. This convention implies that $\sqrt{|\lambda_k|} \equiv k$. Note that this convention induces a reversal of spatial scale ordering. In the continuum limit, the wavenumber k characterizes the frequency of the plane wave, so higher k corresponds to higher frequencies and smaller spatial scales. Similarly, the absolute eigenvalues $\{|\lambda_k|\}_{k=1,\ldots,|\mathcal{X}|}$ share this ordering such that higher $|\lambda_k|$ corresponds to smaller spatial scales. However, the generator eigenvalues themselves are ordered in reverse since they are confined to be negative real numbers. When plotting, we order spectral components, density values, and generator eigenvalues, from smaller to larger spatial scale consistent with generator eigenvalue order. This ordering is inverted with respect to that of wavenumbers in the continuum limit and generator eigenvalue magnitudes since $|\lambda_k| = -\lambda_k$. We use this unusual convention in order to align the mathematical derivations with the linear algebraic analysis as closely as possible and to simplify the (τ, α) parametrization.

B.8 Consistency of propagator sampling

We model sequence generation as sampling from the ρ at successive discrete time points. Note that the propagator can be initialized at any time t and applied recursively

$$\rho_{t+\Delta t_1+\Delta t_2} = \rho_t G e^{(\Delta t_1+\Delta t_2)\Lambda} W = \rho_t G e^{\Delta t_1 \Lambda} e^{\Delta t_2 \Lambda} W = \rho_t G e^{\Delta t_1 \Lambda} W G e^{\Delta t_2 \Lambda} W = \rho_{t+\Delta t_1} G e^{\Delta t_2 \Lambda} W \ . \tag{54}$$

This equation implies that propagation dynamics are congruent across time. That is, propagating from time t to time $t + \Delta t_1 + \Delta t_2$ results in the same state distribution as first propagating by time interval Δt_1 and then by time interval Δt_2. Therefore, sampling a state $x_t \sim \rho_0 G e^{t\Lambda} W$ at time t is equivalent to sampling from from a time-step propagation $x_t \sim \rho_{t-1} G e^{\Lambda} W$. This implies that generating sequences of states via recursive sampling will result in state sequences that accurately reflect the probabilistic dynamics defined by the generator O. Notably, this consistency is not guaranteed in the turbulent regime ($\alpha > 1$, see Section 2.6). Technically, this is due to the fact that turbulent propagators are analogous to unstable Lévy distributions in the continuum limit[24].

B.9 Embedding a transition matrix in a generator

We summarize a simple relationship between transition matrices governing discrete-time Markov chains and generator matrices controlling continuous-time Markov chains. Given a generator O, we can re-express it in terms of a vector $\eta = -\operatorname{diag} O$ and a transition matrix T:

$$T_{ij} = \begin{cases} \frac{O_{ij}}{\eta_i} & i \neq j \\ 0 & i = j \end{cases} \tag{55}$$

$$\eta_i = \sum_{j \neq i} O_{ij} \tag{56}$$

$$O = \eta(T - I) \ . \tag{57}$$

By construction, T is a transition matrix which defines the *embedded discrete-time Markov chain* generated by O[23]. The parameters η are known as the *jump rates*. This equivalence can be applied in the reverse direction. In simulations where the generator O_T is derived from a transition matrix T based on Eqn. 57, we consistently set the jump rate to $\eta = 15$ for all states as a default.

B.10 Localizing states in time

The jump time J_i for a state x_i is the time interval from when a random process arrives at state x_i to when it leaves to another state. Jump rates η_i are so-called because they parametrize the exponential distribution of *jump times* $J_i \sim \text{Exp}(\eta_i)$[23]. In particular, $\mathbb{E}[J_i] = \eta_i^{-1}$. Jump rates determine a temporal localization of state occupation which can be explicitly observed in the cumulative density function $P(J_i \leq \Delta t) = 1 - e^{-\eta_i \Delta t}$ where Δt is the time that the process remains at state x_i after arriving there. Note that the probability of remaining at a state decays exponentially over time according to the jump rate η_i.

Given that J_i scales as $v_i J_i \sim \mathrm{Exp}(v_i^{-1}\eta_i)$ under positive $v_i > 0$, decreasing the exponential parameter leads to an increase in expected jump time (and vice versa). Notably, this slows (or speeds) down the generative process at state x_i, increasing (or decreasing) its temporal imprint while leaving the spatial relationship, encoded by the off-diagonal elements of the generator O, relatively unchanged. This analysis highlights a key distinction between transition matrices and generators. In contrast to (discrete-time) transition matrices, generator matrices contain an "extra" parameter for each state which controls each states temporal representation.

C Further technical details

C.1 Motivating α-modulation based on the theory of stable distributions in the continuous domain

The stability parameter α is judiciously introduced into the parametrized power spectrum (Eqn. 7, main text) based on the theory of stable distributions in the continuous domain which characterize anomalous diffusion[24]. In probability theory, a distribution p is stable if positive linear sums of random variables which are distributed according to this distribution are also distributed according to p. This property is stated mathematically as $aX + bY \sim p$ for all $X \sim p, Y \sim p$ and $a, b \in \mathbb{R}_+$. Stable distributions form a family of probability distributions; many of which can only be expressed through the Fourier transform i.e., through various parametrizations of the characteristic function φ of the distribution. One parameter in particular, usually denoted α, interpolates between Gaussian distributions and a more general class of stable distributions known as symmetric Lévy α-stable distributions according to the following characteristic function:

$$\varphi_{\alpha,\mu,c}(k) = e^{ik\mu-|ck|^\alpha} \tag{58}$$

$$p_{\alpha,\mu,c}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \varphi_{\alpha,\mu,c}(k) e^{-ikx}\, dk \tag{59}$$

where μ is the mean of the distribution and c is a scale parameter. Comparing this characteristic function $\varphi_{\alpha,\mu,c}(k)$ of symmetric Lévy α-stable distributions and the parametrized power spectrum (Eqn. 7, main text), we see the following parameter equivalences

$$\begin{aligned} \mu &\rightarrow 0 \\ c^\alpha &\rightarrow \tau^{-1} \\ k &\rightarrow \lambda_k \ . \end{aligned} \tag{60}$$

The μ parameter is zero in the random walk propagator since it is centered at the initial position. The scale parameter c is analogous to the parameter τ in the propagator where the tempo is paramerized in order to control the scale independent of α. Finally, the frequency k in the Fourier transform is equivalent to the generator eigenvalues with the integral transform being performed by a continuous integral in the continuum limit and by matrix multiplication by W in the discrete state-space formulation pursued through the main text.

We use the term *stability* for the α parameter since it is technically relevant based on the theory of α-stable distributions and descriptively appropriate with respect to the changes manifested in generated sequences as α varies. Technically, the stable property underpins the recursive consistency of propagation. That is, it implies that taking multiple steps in a sequence is equivalent to taking one large step. Violating this property undermines the consistency between space and time as observed in the turbulent regime (Fig. 8) which correspond to an unstable distribution in the continuum limit. Descriptively, diffusive sequences are the "most stable" since they exhibit a constant equivalence between space and time in expectation whereas superdiffusive sequences contain a degree of instability in this intuitive sense and turbulent sequences are apparently highly unstable. We emphasize that the technical and intuitive senses of the term "stability" as used here are not completely aligned. Superdiffusions are stable in the technical sense but, as we suggest, intuitively reflect a type of instability from a spatial processing perspective. This latter, descriptive, conceptualization of instability is referred to as "anomalous" in the physics literature (e.g., superdiffusions are a form of "anomalous diffusion").

C.2 Relationship between propagators and successor representations

The successor representation (SR) M[28,39] is defined as

$$M = \sum_{t=0}^{\infty} \gamma^t T^t = (I - \gamma T)^{-1} \tag{61}$$

where T is a transition matrix, I the identity matrix, and $0 < \gamma < 1$ is a discount factor. An entry T_{ij} quantifies the probability of an agent transitioning from state x_i to x_j in a single (discrete) timestep. In a Markov decision process, such a transition matrix T may be induced from an agent's policy π in combination with the environment dynamics P via

$$T_{ij} = \sum_{a \in \mathcal{A}_i} \pi(a|x_i) P(x_j|x_i, a) \tag{62}$$

where $\mathcal{A}_i$ is the set of actions available to the agent at state x_i. Under an eigendecomposition of the transition matrix $T = UDU^{-1}$ where D is diagonal, we have

$$M = \sum_{k=0}^{\infty} \gamma^k T^k = \sum_{k=0}^{\infty} \gamma^k U D^k U^{-1} = U \left(\sum_{k=0}^{\infty} \gamma^k D^k \right) U^{-1} = U(I - \gamma D)^{-1} U^{-1} = U \left(I - \gamma^{-1} D^{-1} \right) U^{-1} \ . \tag{63}$$

Therefore the SR M and the transition matrix T share eigenvectors[28]. We assume that there exists a generator matrix O such that

$$T = e^O \tag{64}$$

and therefore T is a propagator evaluated over one iteration $t = 1$. Then, we can express the SR M in terms of the generator eigendecomposition $O = G\Lambda W$ as

$$M = \sum_{t=0}^{\infty} \gamma^t e^{tO} = \sum_{t=0}^{\infty} \gamma^t G e^{t\Lambda} W = G \left(\sum_{t=0}^{\infty} \gamma^t e^{t\Lambda} \right) W = G \left(I - \gamma e^{\Lambda} \right)^{-1} W = G \left(I - \gamma^{-1} e^{-\Lambda} \right) W \tag{65}$$

essentially by substituting G for U and $e^{t\Lambda}$ for D in Eqn. 63. Therefore the generator O, the transition matrix (or time-step propagator) T, and the SR M can all be represented in the same basis set. If an eigenvalue λ_k is associated with the eigenvector ϕ_k of the generator O, then the transition matrix $T = e^O$ and SR M have eigenvalues e^{λ_k} and $1 - \gamma^{-1} e^{-\lambda_k}$ respectively associated to the same eigenvector ϕ_k. In particular, the power spectrum s_{SR} associated with the successor representation is

$$s_{\text{SR}}(\lambda) = 1 - \gamma^{-1} e^{-\lambda} \ . \tag{66}$$

C.3 State transition velocities: Lévy flights versus Lévy walks

In the physics literature, Lévy distributed jumps with infinite traversal velocities are known as *Lévy flights* and thus are manifestly unphysical for objects with nonzero mass (such as foraging rodents)[63]. In contrast, if state traversals occur with finite velocities, and no further jumps are possible during the cumulative time taken to traverse between states, then the resulting process is known as a Lévy walk[64]. Sequence generation, which may occur at a sampling rate of over 200Hz in the EHC and contain arbitrarily large state transitions, may be considered to approximate a Lévy flight when considered on a behavioral timescale measured in seconds to minutes.

In the continuum limit, for an unbounded state-space, Lévy flights pose analytic difficulties since the expected jump length variance diverges to infinity due to the nonzero probability of arbitrarily long jumps[65]. For example, the mean squared displacement measure $\langle r^2(t) \rangle = \infty$. Several methods have been proposed to circumvent this problem which may be collectively referred to as pseudo mean squared displacement measures[65]. These methods are based on either (1) rescaling the time-space relationship to avoid divergence asymptotically, or (2) smoothly eliminating the possibility of extremely long jumps by adaptively bounding the state-space[66]. The discrete spatial domains considered here are all bounded and so, effectively, the latter technique is imposed automatically[67].

C.4 Generator-based sampling as a Markov chain Monte Carlo method

In Markov chain Monte Carlo estimation, samples are first generated by a Markov chain and are then used to construct a Monte Carlo estimator. The challenging aspect of the first step is to design the Markov chain such that it will (eventually) emit a set of samples which are representative of a target distribution p. If this is the case, then the associated Monte Carlo estimate will be unbiased. There are many proposals for how this can be accomplished with the Metropolis-Hastings (MH) strategy being the archetype[17,68]. Each iteration of a Metropolis-Hastings sampler is composed of two steps. First, a sample x' is drawn from a proposal distribution q centered on the previously sampled value x_t. Second, the sample x' is noisily accepted at a rate determined by the acceptance ratio $\beta \propto p(x') / p(x_t)$. If accepted, the next iteration is initialized at $x_{t+1} = x'$, otherwise the proposed sample is rejected and $x_{t+1} = x_t$. Intuitively, the acceptance ratio is a noisy approximation of the curvature of the target density p at x_t. The accept/reject step encourages the chain to remain at x_t if the gradient declines steeply (with a small probability of exploration) otherwise the chain moves away from x_t in order to find regions of high probability density. In this way, samples are drawn in accordance with the density of p. With respect to the generator framework, we conceptualize Metropolis-Hastings strategy as being analogous to sequence generation on a lattice based whereby the propagator corresponds to the proposal distribution q. For example, diffusive sequence generation based on a random walk generator O^{rw} (Section B.5) corresponds to the proposal distribution of the random-walk Metropolis MCMC algorithm in the continuum limit[68].

Neurally, random walk propagation for MCMC estimation would require an EHC-extrinsic mechanism to accept or reject samples. However, since generators can flexibly encode non-Euclidean structure, an arbitrary density p can be embedded within a generator in a way that renders the accept/reject step unnecessary, as is the case in so-called "rejection-free" MCMC samplers. Specifically, an arbitrary density p can be encoded in a generator such that diffusive sequence generation equilibrates to p. The necessary generator is the negative Laplacian matrix $O[p] = -L[p]$ which is defined over a lattice embedded in the ambient space supporting p and approximates the curvature of the density Δp on the embedded lattice[62,69]. The matrix $L[p]$ is known as the Laplacian matrix since it converges to the Laplacian operator Δ in the continuum limit on a Euclidean space or, more generally, the Laplace-Beltrami operator on a manifold. Eigenvectors of $L[p]$ form what is known as a Laplacian eigenmap which has previously been encountered in non-linear dimensionality reduction[69]. Whereas previously Laplacian eigenmaps were fitted to sampled data in order to recover an interpretable representation of p, here we are taking a generative perspective. A propagator constructed from the negative Laplacian eigenmap $-L[p]$ (including all spectral components) can be used to sample from p. Sampling from this propagator as a proposal mechanism results in sequences $\mathbf{x}$ which will be distributed according to p asymptotically. From the MH-MCMC perspective, this means that samples need not be rejected in order to converge to the target density p. Geometrically, implementing MH-MCMC with an agnostic proposal distribution requires an accept/reject step based on a sampled approximation to the curvature of p. In contrast, in generator-based sampling, the curvature of p can be embedded in the spectral components of the $O[p]$-propagator.

C.4.1 Standard techniques to minimize sample autocorrelations in Markov chain Monte Carlo methods

The fundamental challenge in MCMC estimation is that it may take a long time for a high quality of set of samples to be generated. The quality of samples are primarily determined by two factors[17,68]. First and foremost, the samples must be representative of the target distribution so that the Monte Carlo estimator is unbiased. Second, the more uncorrelated the samples are, the more accurate the Monte Carlo estimator for fewer samples. Fundamentally, both objectives are undermined by the same issue, namely generative autocorrelations in the Markov chain. With respect to the former, generative autocorrelations underpin an *initialization bias* implying that the chain has not "forgotten" where it started and therefore it has not converged. With respect to the second objective, autocorrelations imply that, even when the chain has converged, successive samples are strongly related and therefore each sample does not provide much extra information independent of the previous. Note that generative autocorrelations are similarily problematic both for MCMC algorithms (such as Metropolis-Hastings) and for generator-based sampling either in the diffusive or superdiffusive regime (Fig. 5F).

In order to extract a set of independent samples which are representative of the target distribution from an autocorrelated sequence, heuristics such as "burn-ins" and "thinning" are applied in order to decorrelate the generated sequences with respect to the initialization and at equilibrium respectively[68]. The former technique refers to removing an arbitrary number of initial samples under the assumption that they are not drawn from the stationary distribution. Thinning the sequence refers to the strategy of only accepting samples generated at a period greater than 1 in order to avoid correlations between successive samples after convergence. Ideally, a sample sequence may be thinned according to the estimated autocorrelation time Δt_{ac}. These heuristics are problematic in that they reduce the number of samples available for estimation and thus exacerbate the fundamental problem of Markov chain Monte Carlo methods, namely that they can take a relatively long time to become effectively accurate. For example, thinning a sequence at a modest period of ten eliminates 90% of the samples and thus the sampling chain must be run ten times as long in order to achieve the same estimator variance as that based on independent samples. This motivates our proposal that the EHC may seek an alternative mechanism to minimize autocorrelated sequence generation via spectral modulation.

C.5 Spectral minimization of generative autocorrelations

Assuming a set of samples are independently drawn from $x^{(i)} \sim p(x)$, estimation uncertainty can be quantified by the sample variance $\mathbb{V}_{\mathrm{ind}}[\hat{r}]$ of the estimator $\hat{r}$

$$\mathbb{V}_{\mathrm{ind}}[\hat{r}] = N^{-1}\mathbb{V}_p[r] \quad . \tag{67}$$

where $\mathbb{V}_p$ refers to the variance of the distribution p. Therefore Monte Carlo estimators become more accurate as the number of samples N is increased, converging in the limit as $N \to \infty$. Unfortunately, directly generating independent samples from a stationary distribution is not possible in general. In order address this problem, a widely employed approach, known as Markov chain Monte Carlo, is to draw samples from random processes which converge on the target distribution in the limit of infinite time[68]. Estimation based on sequences generated by the generator model falls into this category.

There are two major limitations of sampling algorithms based on Markovian dynamics such as the generator model or Markov Chain Monte Carlo estimators[17]. First, assuming that the sequence is not initialized at the (unknown) stationary distribution, then any samples generated prior to convergence will not be reflective of the target stationary distribution and thus suffer from an *initialization bias*. Second, *equilibrium autocorrelations* imply that, even after the chain has converged, successive samples from the chain are correlated and thus are not independent (see Section C.4.1 for a more detailed discussion). Fundamentally, both of these issues relate to the autocorrelation function of the generative sequence model. If the generated sequences are strongly autocorrelated then they will take a long time to "forget" the initialization and then will still be correlated at equilibrium. Statistically, these autocorrelations imply that the sample variance $\mathbb{V}_{\text{dyn}}[\hat{r}]$ of a Markov chain Monte Carlo estimator is proportional to the *integrated autocorrelation time* Δt_{ac}[17]

$$\mathbb{V}_{\text{dyn}}[\hat{r}] \quad = \quad \Delta t_{\text{ac}} N^{-1} \mathbb{V}_p[r] = \Delta t_{\text{ac}} \mathbb{V}_{\text{ind}}[\hat{r}] \tag{68}$$

where

$$\Delta t_{\text{ac}} = \sum_{\Delta t=0}^{\infty} C_X(0, \Delta t) \tag{69}$$

and $C_X(0, \Delta t)$ is the autocorrelation function[3] of the random state variable X at lag Δt. Note that the Markov chain Monte Carlo estimator achieves the same sample variance (Eqn. 67) of independent samples when $\Delta t_{\text{ac}} = 1$ which is the minimum possible value of Δt_{ac} (since $C_X(0,0) = 1$).

Standard practice in Markov chain Monte Carlo estimation is to implement a range of heuristics to minimize the autocorrelations within the set of samples used for estimation (see Section C.4.1 for a brief description). Fundamentally, these heuristics are typically based on discarding many autocorrelated samples. In contrast, we have suggested that a better solution for a biological agent endowed with grid-like representations is to directly minimize generative autocorrelations via spectral modulation. In order to study the autocorrelation function for sequence generation in discrete state-spaces, we consider a binary one-hot random vector X such that, on each time step t, $X_i(t) = 1$ if and only if state $x_i \in \mathcal{X}$ is sampled on time step t. Otherwise the value of $X_j, j \neq i$, is zero. We refer to this random process as an *occupator*. The autocorrelation function $C_X(0, \Delta t)$ of a random process X at lag Δt from an arbitrary initial time step 0 is the expectation

$$C_X(0, \Delta t) = \langle X(0) X(\Delta t) \rangle \ . \tag{70}$$

This expression captures the repetitiveness, the degree of localization, or the periodicity in the values taken by the process X as a function of time lag Δt. The occupator autocorrelation is then

$$\begin{aligned} C_X(0, \Delta t) &= \langle X(0) \cdot X(\Delta t) \rangle & (71) \\ &= \left\langle \sum_{i=1}^{|\mathcal{X}|} X_i(0) X_i(\Delta t) \right\rangle & (72) \\ &= \sum_{i=1}^{|\mathcal{X}|} \langle X_i(0) X_i(\Delta t) \rangle \ . & (73) \end{aligned}$$

Since $X_i(0) X_i(\Delta t) = 1$ if and only if state x_i is sampled at time 0 and at lag Δt,

$$\begin{aligned} \langle X_i(0) X_i(\Delta t) \rangle &= p(X_i(0) = 1, X_i(\Delta t) = 1) \\ &= p(x = x_i, t = 0)\, p(x = x_i, t = \Delta t | x = x_i, t = 0) \ . & (74) \end{aligned}$$

We show that C_X can be expressed analytically in terms of the power spectrum $s(\lambda)$. Define the diagonal matrix S with diagonal entries drawn from the power spectrum $s(\lambda)$. Then

$$\begin{aligned} C_X(0, \Delta t) &= \sum_{i=1}^{|\mathcal{X}|} \langle X_i(0) X_i(\Delta t) \rangle & (75) \\ &= \sum_{i=1}^{|\mathcal{X}|} p(x = x_i, t = 0)\, p(x = x_i, t = \Delta t | x = x_i, t = 0) & (76) \\ &= \sum_{i=1}^{|\mathcal{X}|} \rho_0(x_i) \left[G S^{\Delta t} W \right]_{ii} & (77) \end{aligned}$$

[3] Note that in the main text (e.g. Eqn. 9), we use the compressed notation $C_X(t) \equiv C_X(0, t)$ implicitly setting a time point of 0 if not otherwise stated.

Define the tensor Γ of spectral components with entries

$$\Gamma_{kij} = \phi_k(x_i)\psi_k(x_j) \tag{78}$$

recalling that $\phi_k = G_{\cdot k}$ and $\psi_k = W_{k\cdot}$ are the column vectors of G and row vectors of W respectively. Then, the autocorrelation function is

$$C_X(0, \Delta t) \quad = \quad \sum_{k=1}^{|\mathcal{X}|} \sum_{i=1}^{|\mathcal{X}|} \rho_0(x_i)\Gamma_{kii} s_k^{\Delta t} \tag{79}$$

where $\rho_0(x_i)$ is an initial distribution of states. Given a spectrum s, this expression measures the probability that the sequence generation process will return to a state for each lag Δt. Linear constraints are necessary to ensure that a spectrally-modulated propagator defines a valid transition matrix[23]. Specifically, for all $i, j = 1 \ldots |\mathcal{X}|$, we require that

$$\begin{aligned} \sum_{k=1}^{|\mathcal{X}|} \Gamma_{kij} s_k &\geq 0 \\ \sum_{j,k=1}^{|\mathcal{X}|} \Gamma_{kij} s_k &= 1 \ . \end{aligned} \tag{80}$$

The first constraint ensures that all elements in the propagation matrix are positive and the second constraint ensures that each row sums to one thereby preserving probability density under time evolution. Furthermore, spectrum bounds are necessary to ensure that the stationary state distribution (encoded by the first eigenvector ϕ_1) in the diffusive regime is preserved:

$$s_1 \quad = \quad 1 \tag{81}$$

$$|s_i| \quad \leq \quad 1 \ , \quad i > 1 \ . \tag{82}$$

In summary, the optimal spectrum s_{mac} defining the sequence generation process with minimal integrated autocorrelation time Δt_{ac} is defined as

$$s_{\text{mac}} \quad = \quad \underset{s}{\operatorname{argmin}}\, \Delta t_{\text{ac}}(s) \tag{83}$$

$$\begin{aligned} &= \underset{s}{\operatorname{argmin}} \sum_{\Delta t} \sum_{k=1}^{|\mathcal{X}|} \sum_{i=1}^{|\mathcal{X}|} \left[\rho_0(x_i)\Gamma_{kii} s_k^{\Delta t}\right] \\ &\text{s.t.} \begin{cases} \sum_{k=1}^{|\mathcal{X}|} \Gamma_{kij} s_k & \geq 0 \\ \sum_{j,k=1}^{|\mathcal{X}|} \Gamma_{kij} s_k & = 1 \\ s_1 & = 1 \\ |s_i| & \leq 1 \ , \quad i > 1 \ . \end{cases} \end{aligned} \tag{84}$$

In practice, we optimized for s_{mac} using standard minimization routines in SciPy[70]. Note that the integrated autocorrelation time objective is dominated by the time-step dynamics $\Delta t = 1$ (since $C_X(1) \geq C_X(\Delta t > 1)$) which is linear in the spectrum. This characterization of autocorrelation depends only on the time lag Δt and is computed with respect to the prior state distribution ρ_0 at a distinguished initial time point $t = 0$. This prior may be taken to be a one-hot vector encoding physical position in an environment or the stationary state distribution ρ_∞. In the former case, minimizing Δt_{ac} corresponds to eliminating the initialization bias, whereas in the latter scenario, this procedure resembles equilibrium autocorrelation minimization.

C.6 List of major symbols

symbol	meaning
$\mathcal{X}$	state-space
x	position or state, $x \in \mathcal{X}$
	may be indexed by location $(x_i, x_j, \ldots)$ or time $(x_0, x_1, \ldots, x_t)$
$\mathbf{x}$	position or state sequence vector
O	infinitesimal generator
τ	tempo parameter controlling spatiotemporal scale
α	stability parameter controlling non-locality
P	propagator
ρ_t	position/state density at time t
$\mathbf{1}_x$	position/state density with x one-hot
G	grid cell matrix and generator matrix factor
S	spectral modulation matrix
$s(\lambda)$	spectral modulation as a function of generator eigenvalue
s_k	spectral modulation value associated with spectral component k
W	linear readout from MEC to HC and generator matrix factor
ϕ_k	k-th generator eigenvector $\phi_k = G_{\cdot k}$
P	place cell matrix
Λ	generator eigenvalue matrix
λ	generator eigenvalue taking values $\lambda \in (-\infty, 0]$
	intuitively associated with spatial scale
λ_k	generator eigenvalue associated with eigenvector number k
	note that lower eigenvectors have larger spatial scales
ψ_k	row vector of W, $\psi_k = W_{k\cdot}$
Γ	spectral tensor
C_X	autocorrelation function for random variable X
Δt	time displacement
Δt_{ac}	integrated autocorrelation time
s_{mac}	minimal autocorrelation spectrum
X	weight/adjacency matrix of a graph
L_X	unnormalized graph Laplacian associated with X
∇^2	Laplacian operator ($\nabla^2 = \nabla \cdot \nabla$ where ∇ is the gradient operator)
T	transition matrix
M	successor representation
γ	discount factor
η	jump rate
J	jump time
$d(t)$	spatial distance (e.g., Euclidean) as a function of time
K	diffusion constant

References

55. Sorscher, B., Mel, G., Ganguli, S. & Ocko, S. A unified theory for the origin of grid cells through the lens of pattern formation. *Advances in Neural Information Processing Systems* (2019).

56. Samsonovich, A. & McNaughton, B. L. Path integration and cognitive mapping in a continuous attractor neural network model. *Journal of Neuroscience* **17,** 5900–5920 (1997).

57. Zilli, E. Models of grid cell spatial firing published 2005-2011. *Frontiers in Neural Circuits* **6,** 16 (2012).

58. Corneil, D & Gerstner, W. Attractor Network Dynamics Enable Preplay and Rapid Path Planning in Maze–like Environments. *Advances in Neural Information Processing Systems 28,* 1675–1683 (2015).

59. McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I. & Moser, M.-B. Path integration and the neural basis of the 'cognitive map'. *Nature reviews. Neuroscience* **7,** 663–678 (2006).

60. Stemmler, M., Mathis, A. & Herz, A. Connecting Multiple Spatial Scales to Decode the Population Activity of Grid Cells. *Science Advances* **in press,** 1–12 (2015).

61. Wei, X. X., Prentice, J. & Balasubramanian, V. A principle of economy predicts the functional architecture of grid cells. *eLife* **4,** 1–29 (2015).

62. Chung, F. R. *Spectral graph theory* (American Mathematical Soc., 1997).

63. Shlesinger, M. F., Zaslavsky, G. M. & Frisch, U. Lévy flights and related topics in physics. *Lecture notes in physics* **450,** 52 (1995).

64. Zaburdaev, V., Denisov, S. & Klafter, J. Lévy walks. *Reviews of Modern Physics* **87** (2015).

65. Metzler, R. & Klafter, J. The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Physics Reports* **339,** 1–77 (2000).

66. Jespersen, S., Metzler, R. & Fogedby, H. C. Lévy flights in external force fields: Langevin and fractional Fokker-Planck equations and their solutions. *Physical Review E* **59,** 2736 (1999).

67. Dybiec, B., Gudowska-Nowak, E., Barkai, E. & Dubkov, A. A. Lévy flights versus Lévy walks in bounded domains. *Physical Review E* **95** (2017).

68. Brooks, S., Gelman, A., Jones, G. & Meng, X.-L. *Handbook of markov chain monte carlo* (CRC press, 2011).

69. Belkin, M. & Niyogi, P. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation* **15,** 1373–1396 (2003).

70. Virtanen, P *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* **17,** 261–272 (2020).