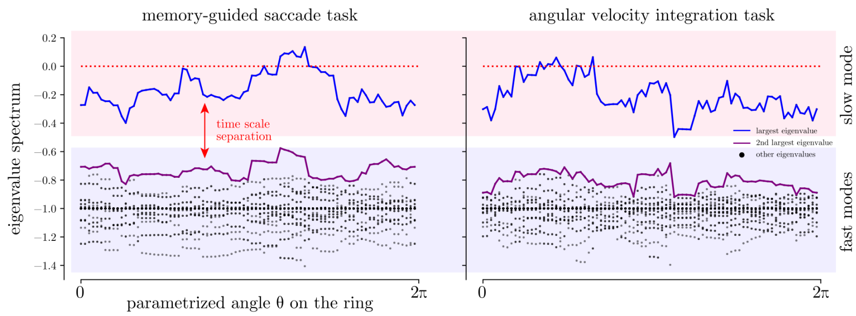

Figure R1: **Empirical demonstration of the slow-fast time scale separation in task-trained RNNs.** The eigenvalue spectrum of the linearization along the invariant manifold shows a gap between the first two largest eigenvalues. **(A)** Network trained on the memory-guided saccade task (Fig.4B in main text). **(B)** Network trained on the angular velocity integration task (Fig.4C in main text).

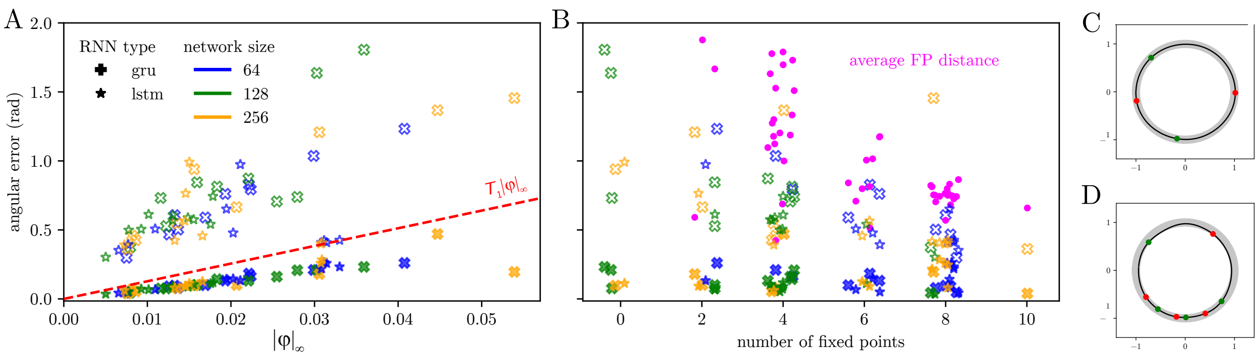

Figure R2: **GRU and LSTM results.** The different measures for memory capacity reflect the generalization properties implied by the topology of the found solution. **(A)** The average accumulated angular error vs. the uniform norm on the vector field shown. Angular error at T_1 = trial length (filled markers) and $\lim T_1 \to \infty$ (hollow markers). Points are jittered to aid legibility. **(B)** The number of fixed points vs. average accumulated angular error, with the average distance between neighboring fixed points (magenta). **(C,D)** Invariant manifold (black) of a trained LSTM (C) and GRU (D) with stable fixed points (green) and saddle nodes (red).

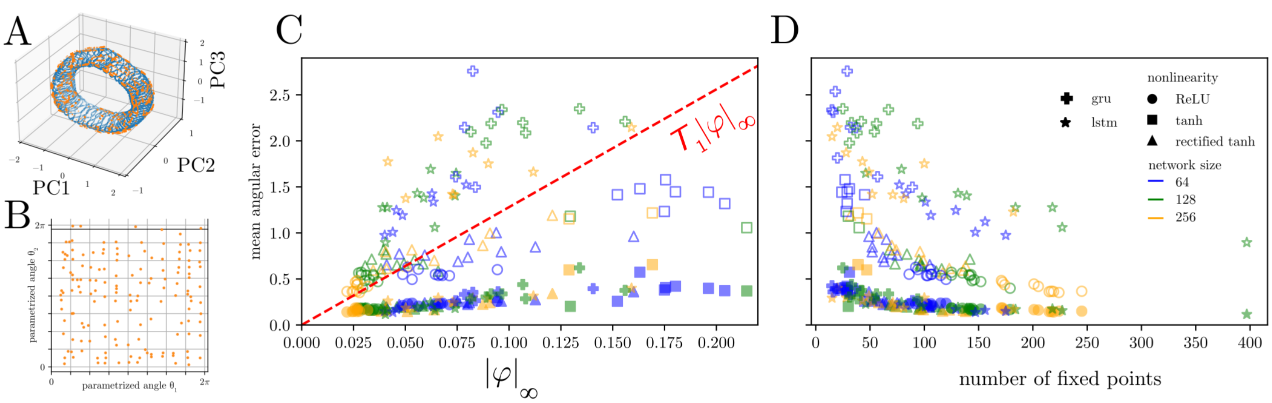

Figure R3: **2D continuous attractor in RNNs.** Networks trained on a double angular velocity integration task. **(A)** Initializations (blue) and fixed points (orange) of an example network. **(B)** Fixed points on a 2D parametrization of the torus for the example network. **(C)** The sum of the total mean angular error (sum of the two seperate angular errors over the two rings) is bounded by the uniform norm of the vector field. **(D)** Generalization for longer memory depends on the number of fixed points in the network.