Deep Quantum Geometry of Matrices

Xizhi Han (韩希之) and Sean A. Hartnoll

Department of Physics, Stanford University, Stanford, California 94305-4060, USA

We employ machine learning techniques to provide accurate variational wave functions for matrix quantum mechanics, with multiple bosonic and fermionic matrices. The variational quantum Monte Carlo method is implemented with deep generative flows to search for gauge-invariant low-energy states. The ground state (and also long-lived metastable states) of an $SU(N)$ matrix quantum mechanics with three bosonic matrices, and also its supersymmetric "mini-BMN" extension, are studied as a function of coupling and N. Known semiclassical fuzzy sphere states are recovered, and the collapse of these geometries in more strongly quantum regimes is probed using the variational wave function. We then describe a factorization of the quantum mechanical Hilbert space that corresponds to a spatial partition of the emergent geometry. Under this partition, the fuzzy sphere states show a boundary-law entanglement entropy in the large N limit.

Subject Areas: Computational Physics, String Theory

I. INTRODUCTION

A quantitative, first-principles understanding of the emergence of spacetime from nongeometric microscopic degrees of freedom remains among the key challenges in quantum gravity. Holographic duality has provided a firm foundation for attacking this problem; we now know that supersymmetric large N matrix theories can lead to emergent geometry [1,2]. What remains is the technical challenge of solving these strongly quantum mechanical systems and extracting the emergent spacetime dynamics from their quantum states. Recent years have seen significant progress in numerical studies of large N matrix quantum mechanics at nonzero temperature. Using Monte Carlo simulations, quantitatively correct features of emergent black hole geometries have been obtained; see, e.g., Refs. [3–5]. To grapple with questions such as the emergence of local spacetime physics, and its associated short-distance entanglement [6,7], new and inherently quantum mechanical tools are needed.

Variational wave functions can capture essential aspects of low-energy physics. However, the design of accurate many-body wave-function ansatze has typically required significant physical insight. For example, the power of tensor network states, such as matrix product states, hinges upon an understanding of entanglement in local systems [8,9]. In contrast, we are faced with models where there is an emergent locality that is not manifest in the microscopic interactions. This locality cannot be used *a priori*; it must be uncovered. Facing a similar challenge of extracting the most relevant variables in high-dimensional data, deep learning has demonstrated remarkable success [10–12], in tasks ranging from image classification [13] to game playing [14]. These successes, and others, have motivated tackling many-body physics problems with the machine learning toolbox [15]. For example, there has been much interest and progress in applications of restricted Boltzmann machines to characterize states of spin systems [16–19].

In this work, we solve for low-energy states of quantum mechanical Hamiltonians with both bosons and fermions, using generative flows (normalizing flows [20–22] and masked autoregressive flows [23–25], in particular) and the variational quantum Monte Carlo method. Compared with spin systems, the problem we are trying to solve contains continuous degrees of freedom and gauge symmetry, and there is no explicit spatial locality. Recent works have applied generative models to physics problems [26–28] and have aimed to understand holographic geometry, broadly conceived, with machine learning [29–31]. We use generative flows to characterize emergent geometry in large N multimatrix quantum mechanics. As we have noted above, such models form the microscopic basis of established holographic dualities.

We focus on quantum mechanical models with three bosonic large N matrices. These are among the simplest models with the core structure that is common to holographic theories. The bosonic part of the Hamiltonian takes the form

$$H_B = \mathrm{tr}\left(\frac{1}{2}\Pi^i\Pi^i - \frac{1}{4}[X^i, X^j][X^i, X^j] + \frac{1}{2}\nu^2 X^i X^i\right.$$
$$\left. + i\nu\epsilon^{ijk}X^iX^jX^k\right). \qquad (1.1)$$

Here, the X^i are N-by-N traceless Hermitian matrices, with $i = 1, 2, 3$. The Π^i are conjugate momenta, and ν is a mass deformation parameter. The potential energy in Eq. (1.1) is a total square: $V(X) = \frac{1}{4}\mathrm{tr}[(\nu\epsilon^{ijk}X^k + i[X^i, X^j])^2]$. The supersymmetric extension of this model [32], discussed below, can be thought of as a simplified version of the BMN matrix quantum mechanics [33]. We refer to the supersymmetric model as mini-BMN, following Ref. [34]. For the low-energy physics we are exploring, the large N planar diagram expansion in this model is controlled by the dimensionless coupling $\lambda \equiv N/\nu^3$. Here, λ can be understood as the usual dimensionful 't Hooft coupling of a large N quantum mechanics at an energy scale set by the mass term (cf. Ref. [35]).

The mass deformation in the Hamiltonian (1.1) inhibits the spatial spread of wave functions—which will be helpful for numerics—and leads to minima of the potential at

$$[X^i, X^j] = i\nu\epsilon^{ijk}X^k. \qquad (1.2)$$

In particular, one can have $X^i = \nu J^i$, with the J^i being, for example, the N-dimensional irreducible representation of the $\mathfrak{su}(2)$ algebra. This set of matrices defines a "fuzzy sphere" [36]. There are two important features of this solution. First, in the large N limit, the noncommutative algebra generated by the X^i approaches the commutative algebra of functions on a smooth two-dimensional sphere [37,38]. Second, the large ν limit is a semiclassical limit in which the classical fuzzy sphere solution accurately describes the quantum state. In this semiclassical limit, the low-energy excitations above the fuzzy sphere state are obtained from classical harmonic perturbations of the matrices about the fuzzy sphere [39]. See also Ref. [40] for an analogous study of the large-mass BMN theory. At large N and ν, these excitations describe fields propagating on an emergent spatial geometry.

By using the variational Monte Carlo method with generative flows, we obtain a fully quantum mechanical description of this emergent space. This result, in itself, is excessive given that the physics of the fuzzy sphere is accessible to semiclassical computations. Our variational wave functions will quantitatively reproduce the semiclassical results in the large ν limit, thereby providing a solid starting point for extending the variational method across the entire N and ν phase diagram. Exploring the parameter space, we find that the fuzzy sphere collapses upon moving into the small ν quantum regime. We consider two different "sectors" of the model, with different fermion

number R. The first will be purely bosonic states, with $R = 0$. The second will have an $R = N^2 - N$. In this latter sector, the fuzzy sphere state is supersymmetric at large positive ν, so we refer to it as the "supersymmetric sector." In the bosonic sector of the model, the fuzzy sphere is a metastable state, and it collapses in a first-order large N transition at $\nu \sim \nu_c \approx 4$. See Figs. 2 and 3 below. In the supersymmetric sector of the model, where the fuzzy sphere is stable, the collapse is found to be more gradual. See Figs. 6 and 7. In Fig. 8, we start to explore the small ν limit of the supersymmetric sector.

Beyond the energetics of the fuzzy sphere state, we define a factorization of the microscopic quantum mechanical Hilbert space that leads to a boundary-law entanglement entropy at large ν. See Eq. (5.14) below. This factorization both captures the emergent local dynamics of fields on the fuzzy sphere and also reveals a microscopic cutoff to this dynamics at a scale set by N. The nature of the emergent fields and their cutoffs can be usefully discussed in string theory realizations of the model. In string-theoretic constructions, fuzzy spheres arise from the polarization of D branes in background fields [41–44]. A matrix quantum mechanics theory such as Eq. (1.1) describes N "D0 branes"—see Ref. [32] and the discussion section below for a more precise characterization of the string theory embedding of mini-BMN theory—and the maximal fuzzy sphere corresponds to a configuration in which the D0 branes polarize into a single spherical D2 brane. There is no gravity associated with this emergent space; the emergent fields describe the low-energy worldvolume dynamics of the D2 brane. In this case, the emergent fields are a Maxwell field and a single scalar field corresponding to transverse fluctuations of the brane. In the final section of the paper, we discuss how richer, gravitating states may arise in the opposite small ν limit of the model.

II. THE MINI-BMN MODEL

The mini-BMN Hamiltonian is [32]

$$H = H_B + \mathrm{tr}\left(\lambda^\dagger \sigma^k[X^k, \lambda] + \frac{3}{2}\nu\lambda^\dagger\lambda\right) - \frac{3}{2}\nu(N^2 - 1). \qquad (2.1)$$

The bosonic part H_B is given in Eq. (1.1). The σ^k are Pauli matrices. The λ are matrices of two-component SO(3) spinors. It can be useful to write the matrices in terms of the $\mathfrak{su}(N)$ generators T_A, with $A = 1, 2, \ldots, N^2 - 1$, which obey $[T_A, T_B] = if_{ABC}T_C$ and are Hermitian and orthonormal (with respect to the Killing form). In other words, $X^i = X_A^i T^A$ and $\lambda^\alpha = \lambda_A^\alpha T^A$. The ijk and ABC indices are freely raised and lowered. Lower $\alpha\beta$ indices are for spinors transforming in the 2 representation of SO(3), while upper indices are for $\bar{\mathbf{2}}$. We will not raise or lower spinor indices. The full Hamiltonian can then be written as

$$H = -\frac{1}{2}\frac{\partial^2}{(\partial X_A^i)^2} + \frac{1}{4}(f_{ABC}X_B^i X_C^j)^2 + \frac{1}{2}\nu^2(X_A^i)^2$$

$$-\frac{1}{2}\nu f_{ABC}\epsilon^{ijk}X_A^i X_B^j X_C^k + if_{ABC}\lambda_A^{\alpha\dagger}X_B^k \sigma_\alpha^{k\beta}\lambda_{C\beta}$$

$$+\frac{3}{2}\nu\lambda_A^{\alpha\dagger}\lambda_{A\alpha} - \frac{3}{2}\nu(N^2 - 1), \tag{2.2}$$

where $\lambda_A^{\alpha\dagger} \equiv (\lambda_{A\alpha})^\dagger$ and $\{\lambda_A^{\alpha\dagger},\lambda_{B\beta}\} = \delta_{AB}\delta_\beta^\alpha$ are complex fermion creation and annihilation operators. This Hamiltonian is seen to have four supercharges,

$$Q_\alpha = \left(-i\frac{\partial}{\partial X_A^i} + i\nu X_A^i - \frac{i}{2}f_{ABC}\epsilon_{ijk}X_B^j X_C^k\right)\sigma_\alpha^{i\beta}\lambda_{A\beta},$$

$$\bar{Q}^\alpha = (Q_\alpha)^\dagger, \tag{2.3}$$

which obey

$$\{Q_\alpha, \bar{Q}^\alpha\} = 4H. \tag{2.4}$$

States that are invariant under all supercharges therefore have vanishing energy.

Matrix quantum mechanics theories arising from microscopic string theory constructions are typically gauged. This means that physical states must be invariant under the $SU(N)$ symmetry. In particular, physical states are annihilated by the generators

$$G_A = -if_{ABC}\left(X_B^i\frac{\partial}{\partial X_C^i} + \lambda_B^{\alpha\dagger}\lambda_{C\alpha}\right). \tag{2.5}$$

A. Representation of the fermion wave function

The mini-BMN wave function can be represented as a function from bosonic matrix coordinates to fermionic states $\psi(X) = f(X)|M(X)\rangle$. Here, X denotes the three bosonic traceless Hermitian matrices. The function $f(X) \geq 0$ is the norm of the wave function at X, while $|M(X)\rangle$ is a normalized state of matrix fermions. A fermionic state with definite fermion number R is parametrized by a complex tensor $M_{A\alpha}^{ra}$ such that

$$|M\rangle \equiv \sum_{r=1}^D \prod_{a=1}^R \left(\sum_{\alpha=1}^2 \sum_{A=1}^{N^2-1} M_{A\alpha}^{ra}\lambda_A^{\alpha\dagger}\right)|0\rangle, \tag{2.6}$$

where $|0\rangle$ is the state with all fermionic modes unoccupied.

The definition (2.6) is parsed as follows: For any fixed r and a, $\eta^{ra\dagger} = \sum_{\alpha A} M_{A\alpha}^{ra}\lambda_A^{\alpha\dagger}$ is the creation operator for the matrix fermionic modes, where A runs over some orthonormal basis of the $\mathfrak{su}(N)$ Lie algebra and $\alpha = 1, 2$ for two fermionic matrices. Then, $\prod_a \eta^{ra\dagger}|0\rangle$ is a state of multiple free fermions created by $\eta^\dagger$. The final summation over r in Eq. (2.6) is a decomposition of a general fermionic state into a sum of free fermion states. Such a representation is

seen to be completely general (but not unique) if we have the number of free fermion states D sufficiently large.

For purely bosonic models, $|M(X)\rangle$ is simply the phase of the wave function.

B. Gauge invariance and gauge fixing

The generators (2.5) correspond to the following action of an element $U \in G = SU(N)$ on the wave function:

$$(U\psi)(X) = f(U^{-1}XU)|(UMU^{-1})(U^{-1}XU)\rangle. \tag{2.7}$$

In other words, the group acts by matrix conjugation. The wave function is required to be invariant under the group action; i.e., $U\psi = \psi$ for any $U \in G$.

Gauge invariance allows us to evaluate the wave function using a representative for each orbit of the gauge group. Let $\tilde{X}$ be the representative in the gauge orbit of X. Gauge invariance of the wave function implies that there must exist functions $\tilde{f}$ and $\tilde{M}$ such that

$$f(X) = \tilde{f}(\tilde{X}),$$

$$|M(X)\rangle = |U\tilde{M}(\tilde{X})U^{-1}\rangle, \quad \text{where } X = U\tilde{X}U^{-1}. \tag{2.8}$$

The functions $\tilde{f}$ and $\tilde{M}$ take gauge representatives as inputs or may be thought of as gauge-invariant functions. The wave function we use will be in the form (2.8). The functions $\tilde{f}$ and $\tilde{M}$ will be parametrized by neural networks, as we describe in Sec. III.

We proceed to describe the gauge fixing we use to select the representative for each orbit, as well as the measure factor associated with this choice. The $SU(N)$ gauge representative $\tilde{X}$ will be such that
(1) $X^i = U\tilde{X}^i U^{-1}$ for $i = 1, 2, 3$ and some unitary matrix U.
(2) $\tilde{X}^1$ is diagonal and $\tilde{X}_{11}^1 \leq \tilde{X}_{22}^1 \leq \ldots \leq \tilde{X}_{NN}^1$.
(3) $\tilde{X}_{i(i+1)}^2$ is purely imaginary, with the imaginary part positive for $i = 1, 2, \ldots, N - 1$.

The third condition is needed to fix the $U(1)^{N-1}$ residual gauge freedom after diagonalizing X^1. The representative $\tilde{X}$ is well defined except on a subspace of measure zero where the matrices are degenerate. Then, $\tilde{X}$ can be represented as a vector in $\mathbb{R}^{2(N^2-1)}$ with a positivity constraint on some components. The change of variables from X to $\tilde{X}$ leads to a measure factor given by the volume of the gauge orbit:

$$d^{3(N^2-1)}X = \Delta(\tilde{X})d^{2(N^2-1)}\tilde{X}, \tag{2.9}$$

with

$$\Delta(\tilde{X}) \propto \prod_{i\neq j=1}^N |\tilde{X}_{ii}^1 - \tilde{X}_{jj}^1| \prod_{i=1}^{N-1} |\tilde{X}_{i(i+1)}^2|. \tag{2.10}$$

Keeping track of this measure (apart from an overall prefactor) will be important for proper sampling in the

Monte Carlo algorithm. The derivation of (2.10) is shown in Sec. A of the Supplemental Material (SM) [45].

III. ARCHITECTURE DESIGN FOR MATRIX QUANTUM MECHANICS

In this work, we propose a variational Monte Carlo method with importance sampling to approximate the ground state of matrix quantum mechanics theories, leading to an upper bound on the ground-state energy. The importance sampling is implemented with generative flows. The basic workflow is sketched as follows:

(1) Start with a wave function ψ_θ with variational parameters θ. In our case, θ will characterize neural networks.

(2) Write the expectation value of the Hamiltonian to be minimized as

$$E_\theta = \langle \psi_\theta | H | \psi_\theta \rangle = \int dX |\psi_\theta(X)|^2 H_X[\psi_\theta]$$
$$= \mathbb{E}_{X \sim |\psi_\theta|^2}[H_X[\psi_\theta]]. \tag{3.1}$$

In the mini-BMN case, X denotes three traceless Hermitian matrices (indices omitted), and $H_X[\psi_\theta]$ is the energy density at X. Notationally, $\mathbb{E}_{X \sim p(X)}$ is the expectation value, with the random variable X drawn from the probability distribution $p(X)$.

(3) Generate random samples according to the wave-function probabilities $X \sim p_\theta(X) = |\psi_\theta(X)|^2$, and evaluate their energy densities $H_X[\psi_\theta]$. The variational energy (3.1) can then be estimated as the average of energy densities of the samples.

(4) Update the parameters θ (via stochastic gradient descent) to minimize E_θ:

$$\theta_{t+1} = \theta_t - \alpha \nabla_{\theta_t} E_{\theta_t}, \tag{3.2}$$

where $t = 1, 2, \ldots$ denotes the steps of training and the parameter $\alpha > 0$ sets the learning rate. The gradient of energy is estimated from Monte Carlo samples:

$$\nabla_\theta E_\theta = \mathbb{E}_{X \sim p_\theta}[\nabla_\theta H_X[\psi_\theta]]$$
$$+ \mathbb{E}_{X \sim p_\theta}[\nabla_\theta(\ln p_\theta(X))(H_X[\psi_\theta] - E_\theta)]. \tag{3.3}$$

The method is applicable even if the probabilities are available only up to an unknown normalization factor.

(5) Repeat steps 3 and 4 until E_θ converges. Observables of physical interest are evaluated with respect to the optimal parameters after training.

In the following, we discuss details of parametrizing and sampling from gauge-invariant wave functions with fermions. Technicalities concerning the evaluation of $H_X[\psi_\theta]$ are spelled out in the SM, Sec. B [45]. More details concerning the training are given in the SM, Sec. D [45]. Benchmarks are presented at the end of this section.

A. Parametrizing and sampling the gauge-invariant wave function

We first describe how gauge invariance is incorporated into the variational Monte Carlo algorithm. As already discussed, an important step is to sample according to $X \sim |\psi(X)|^2$. From Eq. (2.8), for a gauge-invariant wave function, $|\psi(X)|^2 = |\tilde{f}(\tilde{X})|^2$. However, in sampling $\tilde{X}$, we must keep track of the measure factor $\Delta(\tilde{X})$ in Eq. (2.10). We do this as follows:

(1) Sample $\tilde{X}$ according to $p(\tilde{X}) = \Delta(\tilde{X})|\tilde{f}(\tilde{X})|^2$.
(2) Generate Haar random elements $U \in \mathrm{SU}(N)$.
(3) Output samples $X = U\tilde{X}U^{-1}$.

The correctness of this procedure is shown in the SM, Sec. A [45].

Conversely, at the evaluation stage, $\psi(X)$ can be computed in the following steps for gauge-invariant wave functions (2.8):

(1) Gauge fix $X = U\tilde{X}U^{-1}$ as discussed in the last section.
(2) Compute $\tilde{M}(\tilde{X})$ and $\tilde{f}(\tilde{X})$. Details of the structure of $\tilde{M}$ and $\tilde{f}$ will be discussed below.
(3) Return $\psi(X) = \tilde{f}(\tilde{X})|U\tilde{M}(\tilde{X})U^{-1}\rangle$ according to Eq. (2.8).

We now describe the implementation of $\tilde{M}$ and $\tilde{f}$ as neural networks. The basic building block, a multilayer fully connected (also called dense) neural network, is an elemental architecture capable of parametrizing complicated functions efficiently [12]. The neural network defines a function $F: x \mapsto y$ mapping an input vector x to an output vector y via a sequence of affine and nonlinear transformations:

$$F = A_\theta^m \circ \tanh \circ A_\theta^{m-1} \circ \tanh \circ \ldots \circ \tanh \circ A_\theta^1. \tag{3.4}$$

Here, $A_\theta^1(x) = M_\theta^1 x + b_\theta^1$ is an affine transformation, where the weights M_θ^1 and the biases b_θ^1 are trainable parameters. The hyperbolic tangent nonlinearity then acts elementwise on $A_\theta^1(x)$. We experimented with different activation functions; the final result is not sensitive to this choice. Similar mappings are applied m times, allowing M_θ^i and b_θ^i to be different for different layers i, to produce the output vector y. The mapping $F: x \mapsto y$ is nonlinear and capable of approximating any square integrable function if the number of layers and the dimensions of the affine transformations are sufficiently large [46].

The function $\tilde{M}(\tilde{X})$ is implemented as such a multilayer fully connected neural network, mapping from vectorized $\tilde{X}$ to $\tilde{M}$ in Eq. (2.6), i.e., $\mathbb{R}^{2(N^2-1)} \to \mathbb{R}^{DR2(N^2-1)}$. The implementation of $\tilde{f}(\tilde{X})$ is more interesting, as both evaluating $\tilde{f}(\tilde{X})$ and sampling from the distribution $p(\tilde{X}) = \Delta(\tilde{X})|\tilde{f}(\tilde{X})|^2$ are necessary for the Monte Carlo algorithm. Generative flows are powerful tools to efficiently

parametrize and sample from complicated probability distributions. The function $\tilde{f}(\tilde{X}) = \sqrt{p(\tilde{X})/\Delta(\tilde{X})}$, so we can focus on sampling and evaluating $p(\tilde{X})$, which will be implemented by generative flows.

Two generative flow architectures are implemented for comparison: a normalizing flow and a masked autoregressive flow. The normalizing flow starts with a product of simple univariate probability distributions $p(x) = p_1(x_1)\ldots p_M(x_M)$, where the p_i can be different. Values of x sampled from this distribution are passed through an invertible multilayer dense network as in Eq. (3.4). The probability distribution of the output y is then

$$q(y) = p(x)\left|\det\frac{Dy}{Dx}\right|^{-1} = p(F^{-1}(y))|\det DF|^{-1}. \quad (3.5)$$

The masked autoregressive flow generates samples progressively. It requires an ordering of the components of the input, say, $x_1, x_2, \ldots, x_M$. Each component is drawn from a parametrized distribution $p_i(x_i; F_i(x_1, \ldots, x_{i-1}))$, where the parameter depends only on previous components. Thus, x_1 is sampled independently, and for other components, the dependence F_i is given by Eq. (3.4). The overall probability is the product

$$q(x) = \prod_{i=1}^{M} p_i(x_i; F_i(x_1, \ldots, x_{i-1})). \quad (3.6)$$

When $p_i(x_i)$ are chosen as normal distributions, both flows are able to represent any multivariate normal distribution exactly. Features of the wave function (such as polynomial or exponential tails) can be probed by experimenting with different base distributions $p_i(x_i)$. Choices of the base distributions and performances of the two flows are assessed in the following benchmark subsection and also in the SM, Sec. D [45]. We use both types of flow in the numerical results of Sec. IV.

B. Benchmarking the architecture

In Ref. [34], the Schrödinger equation for the $N = 2$ mini-BMN model was solved numerically. Comparison with the results in that paper allows us to benchmark our architecture before moving to larger values of N. In Ref. [34], the Schrödinger equation is solved in sectors with a fixed fermion number

$$R = \sum_{A\alpha}\lambda_A^{\alpha\dagger}\lambda_{A\alpha}, \qquad [R, H] = 0, \quad (3.7)$$

and total SO(3) angular momentum $j = 0, 1/2$. We do not constrain j, but we do fix the number of fermions in the variational wave function.

The variational energies obtained from our machine learning architecture with $R = 0$ and $R = 2$ are shown as a function of ν in Fig. 1. We take negative ν to compare with the results given in Ref. [34], which uses an opposite sign convention. (There is a particle-hole symmetry of the
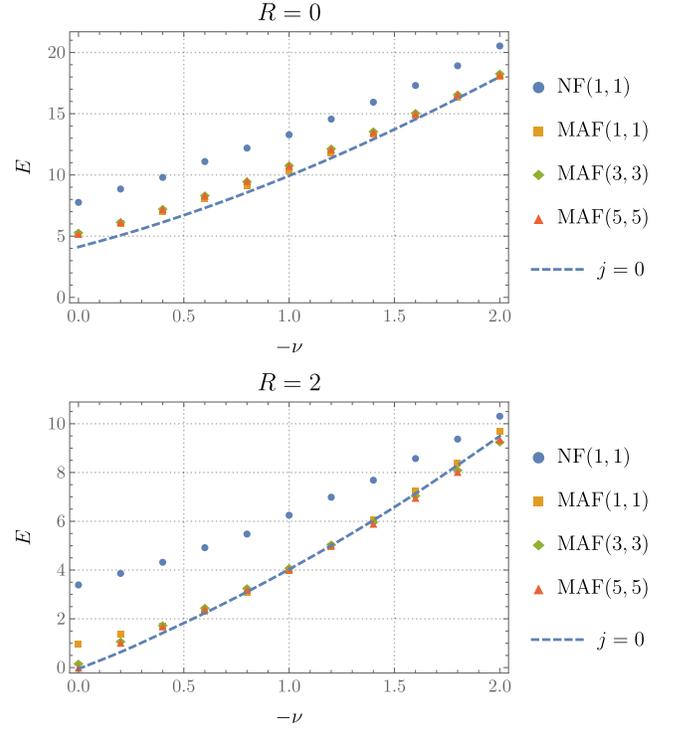

FIG. 1. Benchmarking the architecture: Variational ground-state energies for the mini-BMN model with $N = 2$ and fermion numbers $R = 0$ and $R = 2$ (shown as dots) compared to the exact ground-state energy in the $j = 0$ sector, obtained in Ref. [34] (shown as the dashed curve). Uncertainties are at or below the scale of the markers; in particular, the variational energies slightly below the dashed line are within the numerical error of the line. NF stands for normalizing flows and MAF for masked autoregressive flows. As described in the main text, the numbers in the brackets are, first, the number of layers in the neural networks and, second, the number of generalized normal distributions in each base distribution.

Hamiltonian (2.2) via $\nu \to -\nu$, $\lambda \to \lambda^\dagger$, $\lambda^\dagger \to \lambda$, and $X \to -X$). The masked autoregressive flow yields better (lower) variational energies. These energies are seen to be close to the $j = 0$ results obtained in Ref. [34]. The variational results seem to be asymptotically accurate as $|\nu| \to \infty$, while remaining a reasonably good approximation at small ν. Small ν is an intrinsically more difficult regime, as the potential develops flat directions (visualized in Ref. [34]), and hence the wave function is more complicated, possibly with long tails. In the supersymmetric $R = 2$ sector, where quantum mechanical effects at small ν are expected to be strongest, further significant improvement at the smallest values of ν is seen with deeper autoregressive networks and more flexible base distributions, as we describe shortly. Analogous improvements in these regimes will also be seen at larger N in Sec. IV C and the SM, Sec. D [45].

In Fig. 1, the base distributions $p_i(x_i)$, introduced in the previous subsection, are chosen to be a mixture of s generalized normal distributions:
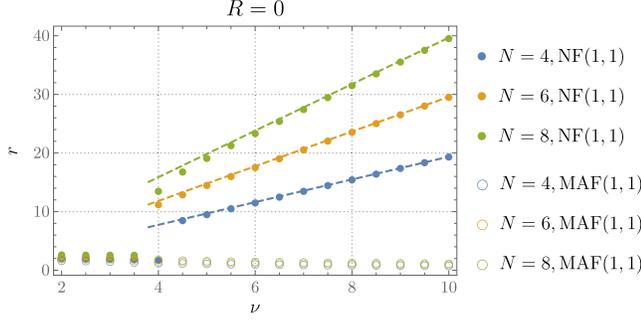
FIG. 2. Expectation value of the radius in the zero fermion sector of the mini-BMN model, for different N and ν. The dashed lines are the semiclassical values (4.4). Solid dots are initialized near the fuzzy sphere configuration, and the open markers are initialized near zero. We have used normalizing and autoregressive flows, respectively, as these produce more accurate variational wave functions in the two different regimes.

$$p_i(x_i) = \sum_{r=1}^{s} k_r^i \frac{\beta_r^i}{2\alpha_r^i \Gamma(1/\beta_r^i)} e^{-(|x_i - \mu_r^i|/\alpha_r^i)^{\beta_r^i}}, \quad \sum_{r=1}^{s} k_r^i = 1. \quad (3.8)$$

Here, the k_r^i are positive weights for each generalized normal distribution in the mixture. In Eq. (3.8), the k_r^i, α_r^i, β_r^i, and μ_r^i are learnable (i.e., variational) parameters. For autoregressive flows, these parameters further depend on x_j, with $1 \le j < i$, according to Eq. (3.4).

Because of the gauge fixing conditions 2 and 3 in Sec. II B, some components x_i are constrained to be positive. In the normalization flow, this case is implemented by an additional map $x_i \mapsto \exp(x_i)$. For the autoregressive flows, we have a more refined control over the base distributions; in this case, for components x_i that must be positive, we draw from Gamma distributions instead:

$$p_i(x_i > 0) = \sum_{r=1}^{s} k_r^i \frac{(\beta_r^i)^{\alpha_r^i}}{\Gamma(\alpha_r^i)} (x_i)^{\alpha_r^i - 1} e^{-\beta_r^i x_i}, \quad \sum_{r=1}^{s} k_r^i = 1, \quad (3.9)$$

where again the k_r^i, α_r^i, and β_r^i depend on x_j, with $1 \le j < i$, according to Eq. (3.4).

In Fig. 1, we have shown mixtures with $s = 1, 3, 5$ distributions. The number of layers in Eq. (3.4) has been increased with s to search for potential improvements in the space of variational wave functions. As noted, the only improvement within the autoregressive flows in going beyond one layer and one generalized normal distribution is seen at the smallest values of ν with $R = 2$. On the other hand, the gap between the variational energies of the two types of flows in Fig. 1 suggests that the wave function is complicated in this regime, so the more sophisticated MAF architecture shows an advantage. The recursive nature of the MAF flows means that they are already "deep" with only a single layer. The complexity of the small ν wave function should be contrasted with the fuzzy sphere phase at large positive ν as discussed in Sec. IV and shown in,
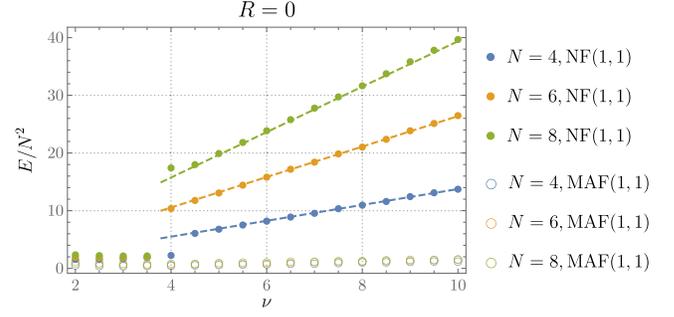

FIG. 3. Variational energies in the zero fermion sector of the mini-BMN model, for different N and ν. The dashed lines are semiclassical values: $E = -\frac{3}{2}\nu(N^2 - 1) + \Delta E|_{\mathrm{bos}}$, with $\Delta E|_{\mathrm{bos}}$ given in Eq. (4.8). As in Fig. 2, solid dots are initialized near the fuzzy sphere configuration, and the open markers are initialized near zero.

e.g., Figs. 2 and 3. The wave function in this semiclassical regime is almost Gaussian, and indeed, the NF(1, 1) and MAF(1, 1) flows give similar energies when initialized near fuzzy sphere configurations. The NF architecture, in fact, gives slightly lower energies in this regime, so we have used normalizing flows in Figs. 2 and 3 for the fuzzy sphere.

The numerics above and below are performed with $D = 4$ in Eq. (2.6), so the fermionic wave function $|M(X)\rangle$ is a sum of four free fermion states for each value of the bosonic coordinates X. In the SM, Sec. D [45], we see that increasing D above 1 lowers the variational energy at small ν, indicating that the fermionic states are not Hartree-Fock in this regime.

IV. EMERGENCE OF GEOMETRY

A. Numerical results, bosonic sector

The architecture described above gives a variational wave function for low-energy states of the mini-BMN model. With the wave function in hand, we can evaluate observables. We start with the purely bosonic sector of the model (i.e., $R = 0$). Then, we add fermions. An important difference between the bosonic and supersymmetric cases is that the semiclassical fuzzy sphere state is metastable in the bosonic theory but stable in the supersymmetric theory.

Figure 2 shows the expectation value of the radius

$$r = \sqrt{\frac{1}{N} \mathrm{tr}(X_1^2 + X_2^2 + X_3^2)}, \quad (4.1)$$

for runs initialized close to a fuzzy sphere configuration (solid circles) and close to zero (open circles). For large ν, a fuzzy sphere state with large radius is found, in addition to a "collapsed" state without significant spatial extent. Below $\nu_c \approx 4$, the fuzzy sphere state ceases to exist. The nature of the transition at ν_c can be understood from the variational energy of the states, plotted in Fig. 3. The bosonic semiclassical fuzzy sphere state is seen to be metastable at large
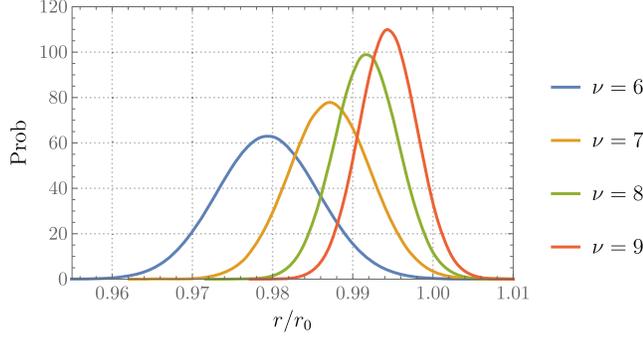
FIG. 4. Probability distribution, from the variational wave function, for the radius in the fuzzy sphere phase for $N = 8$ and different ν. The horizontal axis is rescaled by the semi-classical value of the radius r_0, given in Eq. (4.4) below. The width of the distribution in units of the classical radius becomes smaller as ν is increased.

ν, as the collapsed state has lower energy. For $\nu < \nu_c$, the fuzzy sphere is no longer even metastable. We gain a semiclassical understanding of this transition in Sec. IV B.

Figures 2 and 3 show that the radius and energy of the fuzzy sphere state are accurately described by semi-classical formulas (derived in the following section) for all $\nu > \nu_c$. In particular, E/N^3 and r/N are rapidly converging towards their large N values. Figure 4 further shows that the probability distribution for the radius r becomes strongly peaked about its semiclassical expectation value at large ν.

Analogous behavior to that shown in Figs. 2 and 3 has previously been seen in classical Monte Carlo simulations of a thermal analogue of our quantum transition [47–49]. These papers study the thermal partition function of models similar to Eq. (1.1) in the classical limit, i.e., without the Π^2 kinetic energy term. The fuzzy geometry emerges in a first-order phase transition as a low-temperature phase in these models. We see that, in our quantum mechanical context, the geometric phase is associated with the presence of a specific boundary-law entanglement.

B. Semiclassical analysis of the fuzzy sphere

The results above describe the emergence of a (meta-stable) geometric fuzzy sphere state at $\nu > \nu_c$. In this section, we recall that in the $\nu \to \infty$ limit, the fluctuations of the geometry are classical fields. For finite $\nu > \nu_c$, the background geometry is well defined at large N, but fluctuations will be described by an interacting (noncommutative) quantum field theory.

In the large ν limit, the wave function can be described semiclassically [39,40]. We now briefly review this limit, with details given in the SM, Sec. C [45]. These results provide a further useful check on the numerics and will guide our discussion of entanglement in Sec. V.

The minima of the classical potential occur at

$$[X^i, X^j] = i\nu\epsilon^{ijk}X^k. \qquad (4.2)$$

These are supersymmetric solutions of the classical theory, annihilated by the supercharges (2.3) in the classical limit; therefore, they have vanishing energy. The solution of Eq. (4.2) is

$$X^i = \nu J^i, \qquad (4.3)$$

where the J^i are representations of the $\mathfrak{su}(2)$ algebra, $[J^i, J^j] = i\epsilon^{ijk}J^k$. Here, we are interested in maximal, N-dimensional irreducible representations. (Reducible representations can also be studied, corresponding to multiple polarized D branes.)

The $\mathfrak{su}(2)$ Casimir operator suggests a notion of "radius" given by

$$r^2 = \frac{1}{N}\sum_{i=1}^{3}\text{tr}(X^i)^2 = \frac{\nu^2(N^2 - 1)}{4}. \qquad (4.4)$$

Indeed, the algebra generated by the X^i matrices tends towards the algebra of functions on a sphere as $N \to \infty$ [37,38]. At finite N, a basis for this space of matrices is provided by the matrix spherical harmonics $\hat{Y}_{jm}$. These harmonics obey

$$\sum_{i=1}^{3}[J^i, [J^i, \hat{Y}_{jm}]] = j(j+1)\hat{Y}_{jm}, \quad [J^3, \hat{Y}_{jm}] = m\hat{Y}_{jm}. \qquad (4.5)$$

We construct the $\hat{Y}_{jm}$ explicitly in the SM, Sec. C [45]. The j index is restricted to $0 \le j \le j_{\max} = N - 1$. The space of matrices therefore defines a regularized or "fuzzy" sphere [36].

Matrix spherical harmonics are useful for parametrizing fluctuations about the classical state (4.3). Writing

$$X^i = \nu J^i + \sum_{jm}y^i_{jm}\hat{Y}_{jm}, \qquad (4.6)$$

the classical equations of motion can be perturbed about the fuzzy sphere background to give linear equations for the parameters y^i_{jm}. The solutions of these equations define the classical normal modes. We find the normal modes in the SM, Sec. C [45], proceeding as in Refs. [39,40]. The normal mode frequencies are found to be $\nu\omega$, with

$$\omega^2 = 0 \quad \text{multiplicity } N^2 - 1,$$
$$\omega^2 = j^2 \quad \text{multiplicity } 2(j - 1) + 1,$$
$$\omega^2 = (j + 1)^2 \quad \text{multiplicity } 2(j + 1) + 1. \qquad (4.7)$$

Recall that $1 \le j \le j_{\max} = N - 1$. The three different sets of frequencies in Eq. (4.7) correspond to the group theoretic

$\mathfrak{su}(2)$ decomposition $j \otimes 1 = (j-1) \oplus j \oplus (j+1)$. Here, j is the "orbital" angular momentum, and the 1 is due to the vector nature of the X^i. We will give a field theoretic interpretation of these modes shortly. The modes give the following semiclassical contribution to the energy of the fuzzy sphere state:

$$\Delta E|_{\text{bos}} = \frac{|\nu|}{2} \sum |\omega| = \frac{4N^3 + 5N - 9}{6}|\nu|. \quad (4.8)$$

This energy is shown in Fig. 3. The scaling as N^3 arises because there are N^2 oscillators, with maximal frequency of order N. This semiclassical contribution will be canceled out in the supersymmetric sector studied in Sec. IV C below.

The normal modes (4.7) can be understood by mapping the matrix quantum mechanics Hamiltonian onto a non-commutative gauge theory. The analogous mapping for the classical model has been discussed in Ref. [50]. We carry out this map in the SM, Sec. C [45]. The original Hamiltonian (1.1) becomes the following noncommutative U(1) gauge theory on a unit spatial S^2 [setting the sphere radius to 1 in the field theory description will connect easily to the quantized modes in Eq. (4.7)]:

$$H = \nu \int d\Omega \left(\frac{1}{2}(\pi^i)^2 + \frac{1}{4}(f^{ij})^2 \right) + \text{const.} \quad (4.9)$$

The noncommutative star product $\star$ is defined in the SM [45] and

$$f^{ij} \equiv i(L^i a^j - L^j a^i) + \epsilon^{ijk} a^k + i\sqrt{\frac{4\pi}{N\nu^3}}[a^i, a^j]_\star, \quad (4.10)$$

where the derivatives generate rotations on the sphere $L^i = -i\epsilon_{ijk}x^j\partial_k$ and $[f, g]_\star \equiv f\star g - g\star f$. In Eqs. (4.9) and (4.10), the vector potential a^i can be decomposed into two components tangential to the sphere, which become the two-dimensional gauge field, and a component transverse to the sphere, which becomes a scalar field. This decomposition is described in the SM, Sec. C [45]. The normal modes (4.7) are coupled fluctuations of the gauge field and the transverse scalar field. The zero modes in Eq. (4.7) are pure gauge modes, given in Eq. (4.11) below. In Eq. (4.10), the effective coupling controlling quantum field theoretic interactions is seen to be $1/(N\nu)^{3/2}$. The extra $1/N$ arises because the commutator $[a^i, a^j]_\star$ vanishes as $N \to \infty$; see the SM [45]. Corrections to the Gaussian fuzzy sphere state are therefore controlled by a different coupling than that of the 't Hooft expansion (recall $\lambda = N/\nu^3$).

The SU(N) gauge symmetry generators (2.5) are realized in an interesting way in the noncommutative field theory description. We see in the SM [45] that upon mapping to noncommutative fields, the gauge transformations become
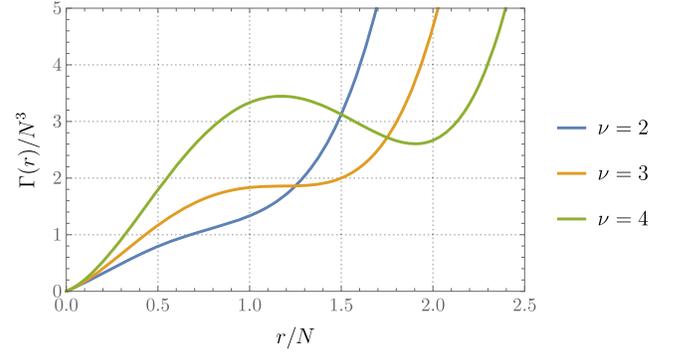

FIG. 5. One-loop effective potential $\Gamma(r)$ for the radius of the bosonic ($R = 0$) fuzzy sphere as $N \to \infty$. The fuzzy sphere is only metastable when $\nu > \nu_{\text{c},N=\infty}^{\text{1-loop}} \approx 3.03$; see the SM [45].

$$\delta a^i = -iL^i y - \sqrt{\frac{4\pi}{N\nu^3}}(\boldsymbol{n} \times \nabla y \cdot \nabla)a^i. \quad (4.11)$$

Here, $\boldsymbol{n}$ is the normal vector and $y(\theta, \phi)$ a local field on the sphere. The first term in Eq. (4.11) is the usual U(1) transformation. The second term describes a coordinate transformation with infinitesimal displacement $\boldsymbol{n} \times \nabla y$. Indeed, it is known that noncommutative gauge theories mix internal and spacetime symmetries, which, in this case, are area-preserving diffeomorphisms of the sphere [51,52]. The emergent U(1) noncommutative gauge theory thereby realizes the large N limit of the microscopic SU(N) gauge symmetry, as area-preserving diffeomorphisms [37,38].

The fluctuation modes about the fuzzy sphere background allow a one-loop quantum effective potential for the radius to be computed in the SM, Sec. C [45]. The potential at $N \to \infty$ is shown in Fig. 5. At large ν, the effective potential shows a metastable minimum at $r \sim N\nu/2$. For $\nu < \nu_{\text{c},N=\infty}^{\text{1-loop}}$, this minimum ceases to exist. The large N, one-loop analysis therefore qualitatively reproduces the behavior seen in Figs. 2 and 3. The quantitative disagreement is mainly due to finite N corrections. The transition is only sharp as $N \to \infty$.

C. Numerical results, supersymmetric sector

We now consider states with fermion number $R = N^2 - N$. The fuzzy sphere background is now supersymmetric at large positive ν [32]. The contribution of the fermions to the ground-state energy is seen in the SM, Sec. C [45], to cancel the bosonic contribution (4.8) at one loop:

$$-\frac{3}{2}\nu(N^2 - 1) + \Delta E|_{\text{fer}} + \Delta E|_{\text{bos}} = 0. \quad (4.12)$$

In Fig. 6, the variational upper bound on the energy of the fuzzy sphere state remains close to zero for all values of ν. Figure 7 shows the radius as a function of ν. Probing the smallest values of ν requires a more powerful
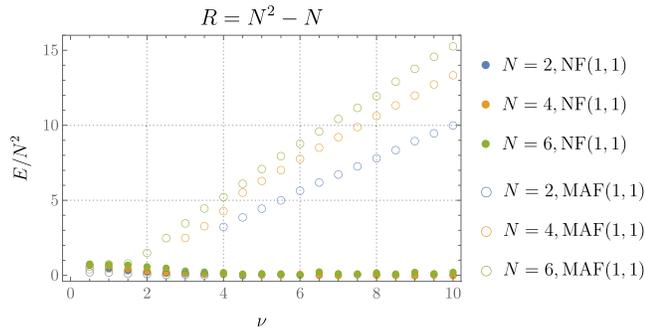
FIG. 6. Variational energies in the SUSY sector of the mini-BMN model, for different N and ν. Solid dots are initialized near the fuzzy sphere configuration, and the open markers are initialized near zero. We use normalizing and autoregressive flows, respectively, as these produce more accurate variational wave functions in the two different regimes.

wave-function ansatz than those of Figs. 6 and 7. We will consider that regime shortly.

In contrast to the states with zero fermion number in Fig. 3, here the fuzzy sphere is seen to be the stable ground state at large ν. However, the fuzzy sphere appears to merge with the collapsed state below a value of ν that decreases with N, which is physically plausible: While the classical fuzzy sphere radius $r^2 \sim \nu^2 N^2$ decreases at small ν, quantum fluctuations of the collapsed state are expected to grow in space as $\nu \to 0$. This is because the flat directions in the classical potential of the $\nu = 0$ theory, given by commuting matrices, are not lifted in the presence of supersymmetry [53]. Eventually, the fuzzy sphere should be subsumed into these quantum fluctuations. This smoother large N evolution towards small ν (relative to the bosonic sector) is mirrored in the thermal behavior of classical supersymmetric models [54,55].

Indeed, exploring the small ν region with more precision, we observe a physically expected feature. In Fig. 8, we see that as ν decreases towards zero, the radius not only ceases to follow the semiclassical decreasing behavior but turns
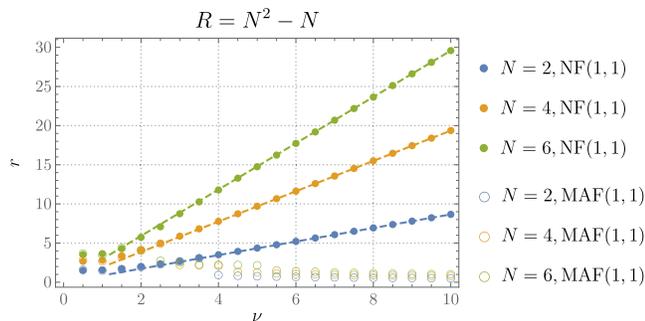

FIG. 7. Expectation value of radius in the SUSY sector of the mini-BMN model, for different N and ν. Solid dots are initialized near the fuzzy sphere configuration, and the open markers are initialized near zero. The dashed lines are the semiclassical values (4.4).
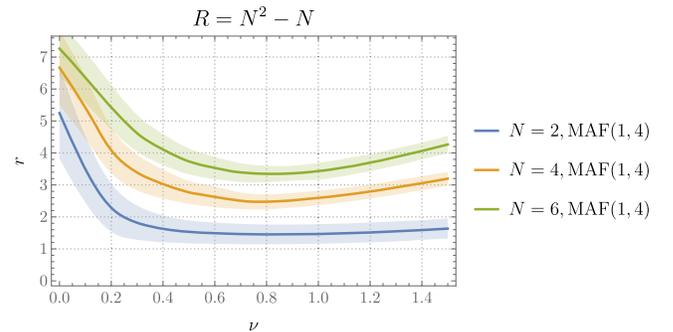

FIG. 8. Distribution of radius for different N and small ν. Bands show the standard deviation of the quantum mechanical distribution of $r = \sqrt{(1/N) \sum \text{tr} X_i^2}$, not to be confused with numerical uncertainty of the average. Recall that the numbers in the brackets are, first, the number of layers in the neural networks and, second, the number of generalized normal distributions in each base distribution.

around and starts to increase. The variance in the distribution of the radius is also seen to increase towards small ν, revealing the quantum mechanical nature of this regime. These behaviors (nonmonotonicity of radius and increasing variance) are expected—and proven for $N = 2$—because the flat directions of the classical potential at $\nu = 0$ show that the extent of the wave function is set by purely quantum mechanical effects in this limit.

The small ν regime here is furthermore an opportunity to test the versatility of our variational ansatz away from semiclassical regimes. In the SM, Sec. D [45], we see that for small ν, MAFs achieve much lower energies than NFs. Increasing the number of distributions in the mixture and the number D of free fermions states in Eq. (2.6) further lowers the energy. These facts mirror the behavior we found in our $N = 2$ benchmarking in Sec. III B at small ν, increasing our confidence in the ability of the network to capture this regime for large N also. The error in a variational ansatz is, as always, not controlled, and therefore, further exploration of this regime is warranted before very strong conclusions can be drawn. We plan to revisit this regime in future work, to search for the possible presence of emergent "throat" geometries, as we discuss in Sec. VI below.

V. ENTANGLEMENT ON THE FUZZY SPHERE

In this section, we show that the large ν fuzzy sphere state discussed above contains boundary-law entanglement. To compute the entanglement, one must first define a factorization of the Hilbert space. For our emergent space at finite N and ν, the geometry is both fuzzy and fluctuating, and hence lacks a canonical spatial partition. The fuzziness of the sphere is captured by a toy model of a free field on a sphere with an angular-momentum cutoff. Recall from Sec. IV that the noncommutative nature of the fuzzy sphere amounts to an angular-momentum cutoff $j_{\text{max}} = N - 1$.

We will start, then, by defining a partition of the space of functions with such a cutoff.

A. Free field with an angular-momentum cutoff

Consider a free massive complex scalar field $\varphi(\theta, \phi)$ on a unit two-sphere with the following Hamiltonian:

$$H = \int_{S^2} d\Omega[|\pi|^2 + |\nabla\varphi|^2 + \mu^2|\varphi|^2]. \qquad (5.1)$$

Here, π is the field conjugate to φ. We impose a cutoff $j \leq j_{\max}$ on the angular momentum, rending the quantum mechanical problem well defined. The fields can therefore be decomposed into a sum of spherical harmonic modes:

$$\varphi(\theta, \phi) = \sum_{0 \leq j \leq j_{\max}}^{|m| \leq j} a_{jm} Y_{jm}(\theta, \phi). \qquad (5.2)$$

The wave functional of the quantum field $\varphi(\theta, \phi)$ is then a mapping from coefficients a_{jm} to complex amplitudes. The ground-state wave functional of the Hamiltonian (5.1) is

$$\psi(a_{jm}) \propto e^{-\sum_{jm} \sqrt{j(j+1)+\mu^2}|a_{jm}|^2}. \qquad (5.3)$$

To calculate entanglement for quantum states, a factorization of the Hilbert space $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$ is prescribed. To motivate the construction of such a factorization in the fuzzy sphere case, we now review the general framework of defining entanglement in (factorizable) quantum field theories.

In quantum mechanics, a quantum state is a function from the configuration space Q to complex numbers, and the Hilbert space of all quantum states is commonly that of square integrable functions $\mathcal{H} = L^2(Q)$. In quantum field theories, the space Q is furthermore a linear space of functions on some geometric manifold M, and thus an orthogonal decomposition $Q = Q_1 \oplus Q_2$ induces a factorization of $\mathcal{H} = L^2(Q_1) \otimes L^2(Q_2)$, which can be exploited to define entanglement.

To define entanglement, it then suffices to find an orthogonal decomposition of the space of fields on the fuzzy sphere. Without an angular-momentum cutoff, i.e., with $j_{\max} \to \infty$, there is a natural choice for any region A on the sphere, which sets Q_1 to be all functions supported on A, and Q_2 all functions supported on $\bar{A}$, the complement of A. Any function f on M can be uniquely written as a sum of $f_1 \in Q_1$ and $f_2 \in Q_2$, where $f_1 = f\chi_A$ and $f_2 = f(1 - \chi_A)$. Here, χ_A is the function on the sphere that is 1 on A and 0 otherwise. Note that the map of multiplication by χ_A, $f \mapsto f\chi_A$, acts as the projection $Q_1 \oplus Q_2 \to Q_1$. Conversely, given any orthogonal projection operator $P: Q \to Q$, we can decompose $Q = \text{im}P \oplus \ker P$.

When the cutoff $j_{\max}$ is finite, multiplication by χ_A will generally take the function out of the subspace of functions

with $j \leq j_{\max}$. However, we can still do our best to approximate the projector P_A^∞ of multiplication by χ_A, as defined in the previous paragraph, with a projector $P_A^{j_{\max}}$ that lives in the subspace with $j \leq j_{\max}$. Formally, let $Q^{j_{\max}}$ be the space of functions on the sphere spanned by $Y_{jm}(\theta, \phi)$, with $j \leq j_{\max}$. Define the orthogonal projector $P_A^{j_{\max}}: Q^{j_{\max}} \to Q^{j_{\max}}$ to minimize the distance $\|P_A^{j_{\max}} - P_A^\infty\|$. The projector $P_A^{j_{\max}}$ annihilates all functions in the orthogonal complement of $Q^{j_{\max}}$, when viewed as an operator acting on Q^∞. It is convenient to choose $\|\cdot\|$ to be the Frobenius norm, and in the SM, Sec. E [45], an explicit formula for $P_A^{j_{\max}}$ is obtained.

The projector $P_A^{j_{\max}}$ then defines a factorization of the Hilbert space $L^2(Q^{j_{\max}}) = L^2(\text{im}P_A^{j_{\max}}) \otimes L^2(\ker P_A^{j_{\max}})$ for any region A, and entanglement can be evaluated in the usual way. In particular, the second Rényi entropy of a pure state $|\psi\rangle$ on a region A is

$$\begin{aligned}
S_2(\rho_A) &= -\ln \int dx_A dx_{\bar{A}} dx'_A dx'_{\bar{A}} \psi(x_A + x_{\bar{A}})\psi^*(x'_A + x_{\bar{A}}) \\
&\quad \times \psi(x'_A + x'_{\bar{A}})\psi^*(x_A + x'_{\bar{A}}) \\
&= -\ln \int dx dx' \psi(x)\psi^*(Px' + (I - P)x)\psi(x') \\
&\quad \times \psi^*(Px + (I - P)x'), \qquad (5.4)
\end{aligned}$$

where $x_A = Px$ and $x_{\bar{A}} = (I - P)x$ are integrated over $\text{im}P$ and $\ker P$, for $P = P_A^{j_{\max}}$, and x_A and $x_{\bar{A}}$ can be more compactly combined into a field x with $j \leq j_{\max}$. Note that the various x's in Eq. (5.4) denote functions on the sphere.

The projector $P_A^{j_{\max}}$ is found to have two important geometric features:

(1) The trace of the projector, which counts the number of modes in a region, is proportional to the size of the region. Specifically, at large $j_{\max}$, $\text{tr}P_A^{j_{\max}} \propto j_{\max}^2 |A|$ as is seen numerically in Fig. 9 and understood analytically in the SM, Sec. E [45].
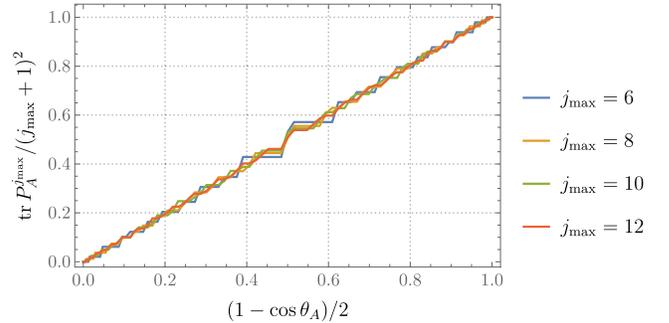

FIG. 9. Trace of the projector versus fractional area of the region (a spherical cap with polar angle θ_A), with different angular-momentum cutoffs $j_{\max}$. A linear proportionality is observed at large $j_{\max}$. The discreteness in the plot arises because the finite $j_{\max}$ space of functions cannot resolve all angles.
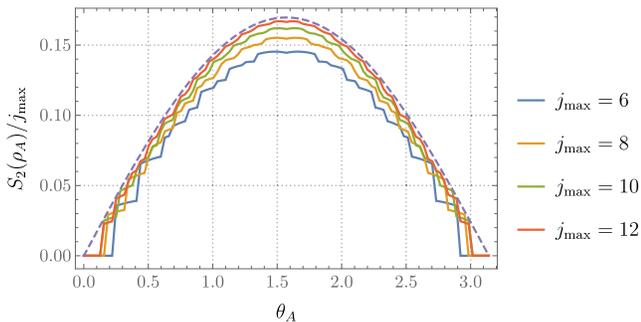
FIG. 10. The second Rényi entropy for a complex scalar free field (with mass $\mu = 1$) versus the polar angle θ_A of a spherical cap. The entropy with different cutoffs j_{max} is shown. At large j_{max}, the curve approaches the boundary law $0.03 \times 2\pi \sin\theta_A$, shown as a dashed line. Discreteness in the plot is again due to the finite j_{max} space of functions.

(2) The second Rényi entropy defined by the projector follows a boundary law. At large j_{max}, with the mass fixed to $\mu = 1$, the entropy $S_2 \approx 0.03 j_{max} |\partial A|$ as is seen numerically in Fig. 10 and understood analytically in the SM, Sec. E [45].

This boundary entanglement law in Fig. 10 is of course precisely the expected entanglement in the ground state of a local quantum field [6,7]. As the cutoff j_{max} is removed, the entanglement grows unboundedly.

The partition we have just defined can now be adapted to the fluctuations about the large ν fuzzy sphere state in the matrix quantum mechanics model. We do this in the following subsection. Intuitively, we would like to replace the $j(j+1) + \mu^2$ spectrum of the free field in the wave function (5.3) with the matrix mechanics modes (4.7). Recall that the matrix modes are cut off at angular momentum $j_{max} = N - 1$.

B. Fuzzy sphere in the mini-BMN model

Now, we address two additional subtleties that arise when adapting the free field ideas above to the mini-BMN fuzzy sphere. First, the mini-BMN theory is an $SU(N)$ gauge theory. It is known that entanglement in gauge theories may depend upon the choice of gauge-invariant algebras associated with spatial regions [56]. Different prescriptions correspond to different boundary or gauge conditions [57]. However, for a fuzzy geometry, the boundaries of regions and gauge edge modes are not sharply defined. To introduce the fewest additional degrees of freedom, we choose to factorize the physical Hilbert space, instead of an extended one [58,59], to evaluate entanglement in the mini-BMN model. This method is similar to the "balanced center" procedure in Ref. [56], where edge modes are absent [60].

Second, the emergent fields include fluctuations of the geometry itself. The factorization that we have discussed in the previous subsection is tailored to a region on the sphere

and does not need to approximate a spatial region in other geometries. The partition is even less meaningful in nongeometric regions of the Hilbert space. The variational wave function we have constructed can be used to compute entanglement for any given factorization of the Hilbert space, but it is unclear if preferred factorizations exist away from geometric limits. In this work, we focus on the entanglement in the $\nu \to \infty$ limit where the fields are infinitesimal and hence do not backreact on the spherical geometry. In this limit, the factorization is precisely—up to issues of gauge invariance—that of the free-field case discussed in the previous subsection.

The matrices corresponding to the infinitesimal fields on the fuzzy sphere are, cf. Eq. (4.6),

$$A^i = X^i - \nu J^i, \tag{5.5}$$

which should be thought of as living in the tangent space at $X^i = \nu J^i$. At large ν, the wave function is strongly supported on the classical configuration, and hence in this limit, the infinitesimal description is accurate. Gauge transformations then act as

$$A^i \to A^i + i\epsilon[Y, \nu J^i] + \dots, \tag{5.6}$$

where ϵ is infinitesimal and Y is an arbitrary Hermitian matrix. The $\epsilon[Y, A^i]$ term is omitted in Eq. (5.6) as it is of higher order. Gauge invariance of the state is manifested as

$$\psi(\nu J^i + A^i) = \psi(\nu J^i + A^i + i\epsilon[Y, \nu J^i]). \tag{5.7}$$

Physical states are wave functions on gauge orbits $[A^i]$, the set of infinitesimal matrices differing from A^i by a gauge transformation (5.6). Similarly to the discussion of free fields above, a partition of the space of gauge orbits is specified by a projector P. We now explain how this projector is constructed. Given a projector P' acting on infinitesimal matrices A^i, a projector acting on gauge orbits can be defined as

$$P([A^i]) = [P'(A^i)]. \tag{5.8}$$

However, for P to be well defined, P' must preserve gauge directions:

$$P'(A^i + i\epsilon[Y, \nu J^i]) = P'(A^i) + i\epsilon[Y', \nu J^i], \tag{5.9}$$

for any A^i, Y and some Y' dependent on Y. Let V be the subspace of gauge directions:

$$V = \{i[Y, J^i] : Y \text{ is Hermitian}\}. \tag{5.10}$$

Then, Eq. (5.9) is equivalent to the requirement that $P'(V) \subset V$. The strategy for finding the projector P is to solve for the projector P' that minimizes $\|P' - \chi_A\|$ subject

to the constraint that Eq. (5.9) is satisfied. Then, P is defined via P' as in Eq. (5.8).

The problem of minimizing $\|P' - \chi_A\|$ for orthogonal projectors P' such that $P'(V) \subset V$ is exactly solvable as follows. The condition that $P'(V) \subset V$ is equivalent to imposing that $P' = P_V \oplus P_{V_\perp}$, where P_V is some projector in the subspace V and $P_{V_\perp}$ in its orthogonal complement $V_\perp$. Here, $\|P' - \chi_A\|$ is minimized if and only if $\|P_V - \chi_A|_V\|$ and $\|P_{V_\perp} - \chi_A|_{V_\perp}\|$ are both minimized. Via the correspondence between matrix spherical harmonics $\hat{Y}_{jm}$ and spherical harmonic functions $Y_{jm}(\theta, \phi)$—in the SM, Sec. C [45]—both of these minimizations become the same problem as in the free field case, with a detailed solution in the SM, Sec. E [45].

The second Rényi entropy, in terms of gauge orbits, is evaluated similarly to Eq. (5.4):

$$
\begin{aligned}
S_2(\rho_A) = -\ln \int &d[A]d[A']\Delta([A])\Delta([A']) \\
&\times \psi_{\text{inv}}([A])\psi_{\text{inv}}^*(P[A'] + (I-P)[A]) \\
&\times \psi_{\text{inv}}([A'])\psi_{\text{inv}}^*(P[A] + (I-P)[A']),
\end{aligned}
\tag{5.11}
$$

where Δ are measure factors for gauge orbits and $\psi_{\text{inv}}([A]) = \psi(\nu J + A)$. Recall that ψ is gauge invariant according to Eq. (5.7). The formula (5.11) as displayed does not involve any gauge choice. However, there are some gauges where evaluating Eq. (5.11) is particularly convenient. The gauge we choose for this purpose, which is different from that in Sec. II B, is that $A \in V_\perp$; i.e., the fields are perpendicular to gauge directions. In this gauge, measure factors are trivial, and the projector is simply $P_{V_\perp}$, which minimizes $\|P_{V_\perp} - \chi_A|_{V_\perp}\|$:

$$
\begin{aligned}
S_2(\rho_A) = -\ln \int_{V_\perp} &dAdA'\psi_\perp(A)\psi_\perp^*(P_{V_\perp}A' + (I-P_{V_\perp})A) \\
&\times \psi_\perp(A')\psi_\perp^*(P_{V_\perp}A + (I-P_{V_\perp})A'),
\end{aligned}
\tag{5.12}
$$

where $\psi_\perp(A)$ is defined as $\psi(\nu J + A)$ for $A \in V_\perp$ [61].

The bosonic fuzzy sphere wave function can be written in the $\nu \to \infty$ limit as follows. As in Eq. (4.6), the perturbations can be decomposed as $A^i = \sum_a \delta x_a \sum_{jm} y^i_{jma} \hat{Y}_{jm}$, where the y^i_{jma} diagonalize the potential energy at quadratic order in A so that $V = (\nu^2/2)\sum_a \omega_a^2(\delta x_a)^2 + \cdots$ (see SM, Sec. C [45]). The wave function is then, analogously to Eq. (5.3),

$$
\psi_\perp(A) \propto e^{-\frac{|\nu|}{2}\sum_a |\omega_a|(\delta x_a)^2}.
\tag{5.13}
$$

The frequencies are given by Eq. (4.7), excluding the pure gauge zero modes. Using this wave function, the Rényi entropy (5.12) can be computed exactly and is shown as a solid line in Fig. 11. As $N \to \infty$, these curves approach a boundary law

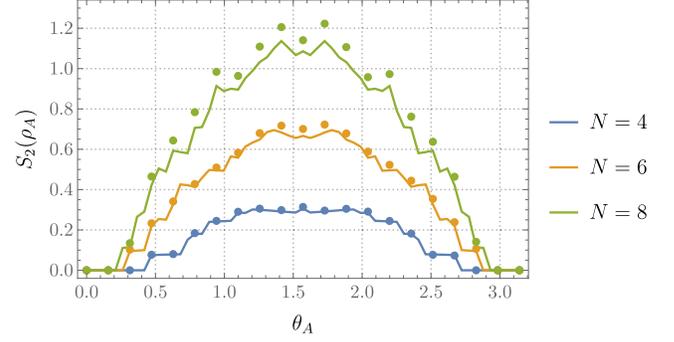FIG. 11. The second Rényi entropy for a spherical cap on the matrix theory fuzzy sphere versus the polar angle θ_A of the cap. Solid curves are exact values at $\nu = \infty$, and dots are numerical values from variational wave functions at $\nu = 10$ for different N. The wave functions are NF(1, 1) in the zero fermion sector as shown in Figs. 2 and 3.

$$
S_2(\rho_A) \approx 0.03N|\partial A|.
\tag{5.14}
$$

Here, $|\partial A| = 2\pi \sin \theta_A$ is again the circumference of the spherical cap A [in units where the sphere has radius one, consistent with the field theoretic description in Eq. (4.9)]. The result (5.14) is the same as that of the toy model in Fig. 10, with j_{max} now set by the microscopic matrix dynamics to be $N - 1$. (A simpler instance of entanglement revealing the inherent graininess of a spacetime built from matrices is two-dimensional string theory [62,63]). This regulated boundary-law entanglement underpins the emergent locality on the fuzzy sphere at large N and ν. Recall from the discussion around Eq. (4.9) that there are only two emergent fields on the sphere: a Maxwell field and a scalar field. The perpendicular gauge choice we have made translates into the Coulomb gauge for the emergent Maxwell field, cf. the discussion around Eq. (4.11) above. The factor of N in Eq. (5.14) is due to the microscopic cutoff at a scale $L_{\text{fuzz}} \sim L_{\text{sph}}/N$.

Previous work on the entanglement of a free field on a fuzzy sphere involved similar wave functions but a different factorization of the Hilbert space, which was inspired instead by coherent states [64–67]. Those results did not always produce boundary-law entanglement. Here, we see that the UV/IR mixing in noncommutative field theories does not preclude a partition of the large N and large ν Hilbert space with a boundary-law entanglement.

We can also evaluate the entropy (5.12) using the large ν variational wave functions, without assuming the asymptotic form (5.13). The results are shown as dots in Fig. 11. However, we stress that only the $\nu \to \infty$ limit has a clear physical meaning, where fluctuations are infinitesimal. The variational results are close to the exact values in Fig. 11, showing that the neural network ansatz captures the entanglement structure of these matrix wave functions.

The results in this section are for the bosonic fuzzy sphere. The projection we introduced in order to partition

the space of matrices can be extended in a similar, but more involved, way to factorize the fermionic Hilbert space.

VI. DISCUSSION

We have seen that neural network variational wave functions capture, in detail, the physics of a semiclassical spherical geometry that emerges in the mini-BMN model (2.1) at large ν. Away from the semiclassical limit, the spherical geometry either abruptly or gradually collapses towards a new state. In Fig. 8, we saw that in the "supersymmetric" sector, this new state was characterized by an increase in both the expectation value and quantum mechanical variance of the radius as $\nu \to 0$. To understand the physics of this process, and to start thinking about the nature of the collapsed state as $\nu \to 0$, it is helpful to consider the string theoretic embedding of the model.

The mini-BMN model can be realized in string theory as the description of N D-particles in an AdS$_4$ spacetime. Let us review some aspects of this realization [32]. The parameter

$$\frac{1}{\nu^3} \sim g_{\mathrm{s}} \left(\frac{L_{\mathrm{AdS}}}{L_{\mathrm{s}}} \right)^3. \tag{6.1}$$

Here, L_{AdS} is the AdS radius, L_{s} is the string length, and g_{s} is the string coupling. The proportionality in Eq. (6.1) depends on the volume, in units of the string length, of internal cycles wrapped by the branes in the compactification down to AdS$_4$. In particular, the mass of a single D-particle is proportional to $1/g_{\mathrm{s}}$ times the wrapped internal volume. The strength of the gravitational backreaction of N coincident D-particles is then controlled by $G_N \times N/g_{\mathrm{s}}$. Here, $G_N \sim g_{\mathrm{s}}^2$ is the four-dimensional Newton constant, where we have suppressed a factor of the volume of the compactification manifold. Therefore, if we keep the AdS radius fixed in string units, gravitational backreaction becomes important when $g_{\mathrm{s}} N \sim N/\nu^3 \gtrsim 1$. Up to factors of the volume of compactification cycles, this is equivalent to the statement that the dimensionless 't Hooft coupling $\lambda = N/\nu^3$, introduced below Eq. (1.1), becomes large.

For $N/\nu^3 \lesssim 1$, then, the D-particles can be treated as light probes on the background AdS spacetime. The fuzzy sphere configuration describes a polarization of the D-particles into spherical "dual giant gravitons." From the string theory perspective, this polarization is driven by the fourform flux $\Omega \sim 1/L_{\mathrm{AdS}}$ supporting the background AdS$_4$ spacetime. Together with the discussion in the previous paragraph on the strength of the gravitational interaction, we can write the heuristic relation $N/\nu^3 \sim$ gravity/flux. At large ν, the flux wins out, and semiclassical fuzzy spheres can exist; however, at small ν, gravitational forces cause the spheres to collapse. The entanglement and emergent locality that we have described in this paper is that of the polarized spheres, whose excitations are described by the usual gauge fields and transverse scalar fields of string theoretic D-branes.

For $N/\nu^3 \gg 1$, it is possible that the strongly interacting, collapsed D-particles will develop a geometric "throat," in the spirit of the canonical holographic correspondence [1]. It is not well understood when such a throat would be captured by the mini-BMN matrix quantum mechanics. The variational wave functions that we have developed here provide a new window into this problem. In particular, we hope to investigate the small ν collapsed state in more detail in the future, with the objective of revealing any entanglement associated with emergent local dynamics in the throat spacetime. If the emergent dynamics includes gravity, there are two potentially interesting complications. First, the entanglement of bulk fields may be entwined with entanglement due to the "stringy" degrees of freedom that seem to be manifested in the Bekenstein-Hawking entropy of black holes as well as in the Ryu-Takayanagi formula [68–71]. Second, and perhaps relatedly, it may become crucial to understand the edge-mode contribution to the entanglement, which we have avoided in our discussion here [72,73].

More generally, the methods we have developed will be applicable to a wide range of quantum problems of interest in the holographic correspondence. The benefit of the variational neural network approach is direct access to properties of the zero-temperature quantum mechanical state. Optimizing the numerical methods and variational ansatz further, and with more computational power, it should not be difficult to work with larger values of N. In addition to understanding the emergence of spacetime from first principles, it should also be possible to study, for example, the microstates and dynamics of quantum black holes.

The code used in the paper is available online [74].

[1] J. M. Maldacena, *The Large N Limit of Superconformal Field Theories and Supergravity*, Int. J. Theor. Phys. **38**, 1113 (1999).

[2] J. Polchinski, *Introduction to Gauge/Gravity Duality*, in *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics (TASI 2010). String Theory and Its Applications: From meV to the Planck Scale:*

Boulder, Colorado, USA, 2010, edited by M. Dine, T. Banks, and S. Sachdev (World Scientific, 2010) pp. 3–46, https://doi.org/10.1142/8153.

[3] K. N. Anagnostopoulos, M. Hanada, J. Nishimura, and S. Takeuchi, *Monte Carlo Studies of Supersymmetric Matrix Quantum Mechanics with Sixteen Supercharges at Finite Temperature*, Phys. Rev. Lett. **100**, 021601 (2008).

[4] S. Catterall and T. Wiseman, *Black Hole Thermodynamics from Simulations of Lattice Yang-Mills Theory*, Phys. Rev. D **78**, 041502 (2008).

[5] E. Berkowitz, E. Rinaldi, M. Hanada, G. Ishiki, S. Shimasaki, and P. Vranas, *Precision Lattice Test of the Gauge/Gravity Duality at Large-N*, Phys. Rev. D **94**, 094501 (2016).

[6] L. Bombelli, R. K. Koul, J. Lee, and R. D. Sorkin, *A Quantum Source of Entropy for Black Holes*, Phys. Rev. D **34**, 373 (1986).

[7] M. Srednicki, *Entropy and Area*, Phys. Rev. Lett. **71**, 666 (1993).

[8] D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac, *Matrix Product State Representations*, Quantum Inf. Comput. **7**, 401 (2007).

[9] R. Orús, *A Practical Introduction to Tensor Networks: Matrix Product States and Projected Entangled Pair States*, Ann. Phys. (Amsterdam) **349**, 117 (2014).

[10] G. E. Hinton and R. R. Salakhutdinov, *Reducing the Dimensionality of Data with Neural Networks*, Science **313**, 504 (2006).

[11] Y. LeCun, Y. Bengio, and G. Hinton, *Deep Learning*, Nature (London) **521**, 436 (2015).

[12] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge, MA, 2016).

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, *Imagenet Classification with Deep Convolutional Neural Networks*, in *Proceedings of the 25th International Conference on Neural Information Processing Systems—Volume 1* (Curran Associates Inc., Red Hook, NY, USA, 2012), pp 1097–1105.

[14] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham *et al.*, *Mastering the Game of Go with Deep Neural Networks and Tree Search*, Nature (London) **529**, 484 (2016).

[15] S. D. Sarma, D.-L. Deng, and L.-M. Duan, *Machine Learning Meets Quantum Physics*, Phys. Today **72**, No. 3, 48 (2019).

[16] G. Carleo and M. Troyer, *Solving the Quantum Many-Body Problem with Artificial Neural Networks*, Science **355**, 602 (2017).

[17] D.-L. Deng, X. Li, and S. D. Sarma, *Quantum Entanglement in Neural Network States*, Phys. Rev. X **7**, 021021 (2017).

[18] X. Gao and L.-M. Duan, *Efficient Representation of Quantum Many-Body States with Deep Neural Networks*, Nat. Commun. **8**, 662 (2017).

[19] I. Glasser, N. Pancotti, M. August, I. D. Rodriguez, and J. I. Cirac, *Neural-Network Quantum States, String-Bond States, and Chiral Topological States*, Phys. Rev. X **8**, 011006 (2018).

[20] L. Dinh, D. Krueger, and Y. Bengio, *NICE: Non-linear Independent Components Estimation*, arXiv:1410.8516.

[21] D. J. Rezende and S. Mohamed, *Variational Inference with Normalizing Flows*, arXiv:1505.05770.

[22] L. Dinh, J. Sohl-Dickstein, and S. Bengio, *Density Estimation Using Real NVP*, arXiv:1605.08803.

[23] M. Germain, K. Gregor, I. Murray, and H. Larochelle, *MADE: Masked Autoencoder for Distribution Estimation*, Proc. Mach. Learn. Res., **37**, 881 (2015).

[24] D. P. Kingma, T. Salimans, and M. Welling, *Improving Variational Inference with Inverse Autoregressive Flow*, arXiv:1606.04934.

[25] G. Papamakarios, I. Murray, and T. Pavlakou, *Masked Autoregressive Flow for Density Estimation*, in *Adv. Neural Inf. Process. Syst.* (Curran Associates Inc., Red Hook, NY, USA, 2017), pp. 2338–2347.

[26] J. Carrasquilla, G. Torlai, R. G. Melko, and L. Aolita, *Reconstructing Quantum States with Generative Models*, Nat. Mach. Intelligence **1**, 155 (2019).

[27] Z.-Y. Han, J. Wang, H. Fan, L. Wang, and P. Zhang, *Unsupervised Generative Modeling Using Matrix Product States*, Phys. Rev. X **8**, 031012 (2018).

[28] D. Wu, L. Wang, and P. Zhang, *Solving Statistical Mechanics Using Variational Autoregressive Networks*, Phys. Rev. Lett. **122**, 080602 (2019).

[29] Y.-Z. You, Z. Yang, and X.-L. Qi, *Machine Learning Spatial Geometry from Entanglement Features*, Phys. Rev. B **97**, 045153 (2018).

[30] K. Hashimoto, S. Sugishita, A. Tanaka, and A. Tomiya, *Deep Learning and the AdS/CFT Correspondence*, Phys. Rev. D **98**, 046019 (2018).

[31] H.-Y. Hu, S.-H. Li, L. Wang, and Y.-Z. You, *Machine Learning Holographic Mapping by Neural Network Renormalization Group*, arXiv:1903.00804.

[32] C. T. Asplund, F. Denef, and E. Dzienkowski, *Massive Quiver Matrix Models for Massive Charged Particles in AdS*, J. High Energy Phys. 01 (2016) 055.

[33] D. E. Berenstein, J. M. Maldacena, and H. S. Nastase, *Strings in Flat Space and pp Waves from N = 4 Super Yang-Mills*, J. High Energy Phys. 04 (2002) 013.

[34] T. Anous and C. Cogburn, *Mini-BFSS in Silico*, Phys. Rev. D **100**, 066023 (2019).

[35] N. Itzhaki, J. M. Maldacena, J. Sonnenschein, and S. Yankielowicz, *Supergravity and the Large N Limit of Theories with Sixteen Supercharges*, Phys. Rev. D **58**, 046004 (1998).

[36] J. Madore, *The Fuzzy Sphere*, Classical Quantum Gravity **9**, 69 (1992).

[37] J. Hoppe, *Diffeomorphism Groups, Quantization and SU(∞)*, Int. J. Mod. Phys. A **04**, 5235 (1989).

[38] B. de Wit, J. Hoppe, and H. Nicolai, *On the Quantum Mechanics of Supermembranes*, Nucl. Phys. **B305**, 545 (1988).

[39] D. P. Jatkar, G. Mandal, S. R. Wadia, and K. P. Yogendran, *Matrix Dynamics of Fuzzy Spheres*, J. High Energy Phys. 01 (2002) 039.

[40] K. Dasgupta, M. M. Sheikh-Jabbari, and M. Van Raamsdonk, *Matrix Perturbation Theory for M Theory on a PP Wave*, J. High Energy Phys. 05 (2002) 056.

[41] R. C. Myers, *Dielectric Branes*, J. High Energy Phys. 12 (1999) 022.

[42] A. Yu. Alekseev, A. Recknagel, and V. Schomerus, *Brane Dynamics in Background Fluxes and Noncommutative Geometry*, J. High Energy Phys. 05 (2000) 010.

[43] J. McGreevy, L. Susskind, and N. Toumbas, *Invasion of the Giant Gravitons from Anti-de Sitter Space*, J. High Energy Phys. 06 (2000) 008.

[44] R. C. Myers, *Non-Abelian Phenomena on D Branes*, Classical Quantum Gravity **20**, S347 (2003).

[45] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevX.10.011069 for full technical details.

[46] Z. Lu, H. Pu, F. Wang, Z. Hu, and L. Wang, *The Expressive Power of Neural Networks: A View from the Width*, in *Advances in Neural Information Processing Systems 30*, edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Curran Associates, Inc., Red Hook, NY, USA, 2017), pp. 6231–6239.

[47] T. Azuma, S. Bal, K. Nagao, and J. Nishimura, *Non-perturbative Studies of Fuzzy Spheres in a Matrix Model with the Chern-Simons Term*, J. High Energy Phys. 05 (2004) 005.

[48] P. Castro-Villarreal, R. Delgadillo-Blando, and B. Ydri, *A Gauge-Invariant UV-IR Mixing and the Corresponding Phase Transition for U(1) Fields on the Fuzzy Sphere*, Nucl. Phys. **B704**, 111 (2005).

[49] R. Delgadillo-Blando, D. O'Connor, and B. Ydri, *Geometry in Transition: A Model of Emergent Geometry*, Phys. Rev. Lett. **100**, 201601 (2008).

[50] S. Iso, Y. Kimura, K. Tanaka, and K. Wakatsuki, *Non-commutative Gauge Theory on Fuzzy Sphere from Matrix Model*, Nucl. Phys. **B604**, 121 (2001).

[51] L. D. Paniak and R. J. Szabo, *Instanton Expansion of Noncommutative Gauge Theory in Two Dimensions*, Commun. Math. Phys. **243**, 343 (2003).

[52] F. Lizzi, R. J. Szabo, and A. Zampini, *Geometry of the Gauge Algebra in Noncommutative Yang-Mills Theory*, J. High Energy Phys. 08 (**2001**) 032.

[53] B. de Wit, *Supersymmetric Quantum Mechanics, Super-membranes and Dirichlet Particles*, Nucl. Phys. B, Proc. Suppl. **56B**, 76 (1997).

[54] K. N. Anagnostopoulos, T. Azuma, K. Nagao, and J. Nishimura, *Impact of Supersymmetry on the Nonperturbative Dynamics of Fuzzy Spheres*, J. High Energy Phys. 09 (2005) 046.

[55] B. Ydri, *Impact of Supersymmetry on Emergent Geometry in Yang-Mills Matrix Models II*, Int. J. Mod. Phys. A **27**, 1250088 (2012).

[56] H. Casini, M. Huerta, and J. A. Rosabal, *Remarks on Entanglement Entropy for Gauge Fields*, Phys. Rev. D **89**, 085012 (2014).

[57] J. Lin and D. Radičević, *Comments on Defining Entanglement Entropy*, arXiv:1808.05939.

[58] W. Donnelly, *Decomposition of Entanglement Entropy in Lattice Gauge Theory*, Phys. Rev. D **85**, 085004 (2012).

[59] W. Donnelly, *Entanglement Entropy and Nonabelian Gauge Symmetry*, Classical Quantum Gravity **31**, 214003 (2014).

[60] It should, nonetheless, be possible to identify meaningful SU(N) "edge modes" that would reproduce the edge-mode contribution of the emergent Maxwell field. This is an especially interesting question in light of the fact that the microscopic SU(N) gauge symmetry also acts as an area-preserving diffeomorphism on the emergent fields in Eq. (4.11). This aim is left for future work.

[61] We can find a gauge transformation $U \in$ SU(N) mapping any matrices X^i into this perpendicular gauge as follows. We are looking for $\tilde{X}^i = UX^iU^{-1}$, such that $\tilde{X}^i - \nu J^i \in V_\perp$, which means that $\sum_i \mathrm{tr}([Y, J^i]^\dagger (\tilde{X}^i - \nu J^i)) = 0$ for any Hermitian matrix Y. Equivalently, $\sum_i \mathrm{tr}(J^i[Y, \tilde{X}^i]) = 0$ for any Y. This result is achieved by numerically finding the U that maximizes the overlap $\sum_i \mathrm{tr}(J^i UX^iU^{-1})$.

[62] S. R. Das, *Degrees of Freedom in Two-Dimensional String Theory*, Nucl. Phys. B, Proc. Suppl. **45**, 224 (1996).

[63] S. A. Hartnoll and E. Mazenc, *Entanglement Entropy in Two Dimensional String Theory*, Phys. Rev. Lett. **115**, 121602 (2015).

[64] D. Dou and B. Ydri, *Entanglement Entropy on Fuzzy Spaces*, Phys. Rev. D **74**, 044014 (2006).

[65] J. L. Karczmarek and P. Sabella-Garnier, *Entanglement Entropy on the Fuzzy Sphere*, J. High Energy Phys. 03 (2014) 129.

[66] S. Okuno, M. Suzuki, and A. Tsuchiya, *Entanglement Entropy in Scalar Field Theory on the Fuzzy Sphere*, Prog. Theor. Exp. Phys. **2016**, 023B03 (2016).

[67] H. Z. Chen and J. L. Karczmarek, *Entanglement Entropy on a Fuzzy Sphere with a UV Cutoff*, J. High Energy Phys. 08 (2018) 154.

[68] L. Susskind and J. Uglum, *Black Hole Entropy in Canonical Quantum Gravity and Superstring Theory*, Phys. Rev. D **50**, 2700 (1994).

[69] T. M. Fiola, J. Preskill, A. Strominger, and S. P. Trivedi, *Black Hole Thermodynamics and Information Loss in Two-Dimensions*, Phys. Rev. D **50**, 3987 (1994).

[70] E. Bianchi and R. C. Myers, *On the Architecture of Spacetime Geometry*, Classical Quantum Gravity **31**, 214002 (2014).

[71] T. Faulkner, A. Lewkowycz, and J. Maldacena, *Quantum Corrections to Holographic Entanglement Entropy*, J. High Energy Phys. 11 (2013) 074.

[72] W. Donnelly and L. Freidel, *Local Subsystems in Gauge Theory and Gravity*, J. High Energy Phys. 09 (2016) 102.

[73] D. Harlow, *The Ryu-Takayanagi Formula from Quantum Error Correction*, Commun. Math. Phys. **354**, 865 (2017).

[74] https://github.com/hanxzh94/matrix-model.