

Learning holographic horizons

Vishnu Jejjala^{*}

*Mandelstam Institute for Theoretical Physics, School of Physics, NITheCS, and CoE-MaSS,
University of the Witwatersrand, Johannesburg, WITS 2050, South Africa*

Sukrut Mondkar[†]

*Harish-Chandra Research Institute, A CI of Homi Bhabha National Institute,
Chhatnag Road, Jhansi, Prayagraj (Allahabad) 211019, India
and Homi Bhabha National Institute, Training School Complex, Anushakti Nagar, Mumbai 400094, India*

Ayan Mukhopadhyay[‡]

*Instituto de Física, Pontificia Universidad Católica de Valparaíso,
Avenida Universidad 330, Valparaíso, Chile*

Rishi Raj[§]

*Laboratoire de Physique Théorique et Hautes Energies—LPTHE, Sorbonne Université and CNRS,
4 Place Jussieu, 75005 Paris, France*



(Received 17 November 2024; accepted 3 January 2025; published 21 January 2025)

We apply machine learning to understand fundamental aspects of holographic duality, specifically the entropies obtained from the apparent and event horizon areas. We show that simple features of only the time series of the pressure anisotropy, namely the values and half-widths of the maxima and minima, the times these are attained, and the times of the first zeroes can predict the areas of the apparent and event horizons in the dual bulk geometry at all times with a fixed maximum length (10) of the input vector. We also argue that the entropy functions are the measures of information that need to be extracted from simple one-point functions to reconstruct specific aspects of correlation functions of the dual state with the best possible approximations.

DOI: [10.1103/PhysRevD.111.026016](https://doi.org/10.1103/PhysRevD.111.026016)

I. INTRODUCTION

The central idea behind modern approaches to Big Data is to devise a predictive analytics by isolating and focusing on salient features within a large dataset. Often, we do not know what the important features are *a priori*. Machine learning can be successful at identifying subtle correlations within datasets and extracting relationships in order to make testable predictions. Unboxing an effective machine learning architecture may in fact teach us about the structure of data.

Spacetime is Big Data. Moreover, in certain settings, we have prior access to the most efficient data compression

protocol available for encoding a dataset. This is holography [1,2]. The gauge/gravity correspondence in asymptotically anti-de Sitter (AdS) spaces [3–5] and the matrix model for M-theory [6] supply examples where the degrees of freedom of a gravitational system are explicitly in codimension one. In this respect, the fundamental understanding of the emergence of semiclassical spacetime from a strongly coupled gauge theory living at the boundary is also one of the prominent areas of interdisciplinary research in theoretical physics (see [7,8] for reviews).

To date, there have been few investigations of the holographic correspondence in the context of machine learning [9–24]. The subject of this paper is to demonstrate that features of the bulk geometry can be machine learned from information contained purely within the dual quantum field theory. In particular, we address the notion of the entropies associated with the dynamical horizons of the emergent geometry corresponding to a typical nonequilibrium state in the dual field theory. This can be expected to provide new insights into some fundamental aspects of the holographic duality since the horizons are among the important gauge invariant features of a spacetime, and the physical

^{*}Contact author: v.jejjala@wits.ac.za

[†]Contact author: sukrutmondkar@hri.res.in

[‡]Contact author: ayan.mukhopadhyay@pucv.cl

[§]Contact author: rishi.raj@lpthe.jussieu.fr

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

understanding of the nonequilibrium entropies associated with the horizons should have important consequences for our understanding of the emergence of spacetime.

We particularly focus on homogeneous thermalization in this work as in this context the Ward identities ensure that the energy density is constant in the absence of any drive. Since the energy density sets the timescale of thermalization, we can understand how the pressure isotropization is encoded at the apparent and event horizons, particularly via the entropy functions by comparing different evolutions with the same constant energy density. Our studies are based on the Chesler–Yaffe method of construction of the dual holographic geometries via numerical relativity [25]. Instead of driving the system via a time-dependent anisotropic boundary metric as in [25], we set nontrivial initial conditions (which are implicitly generated by homogeneous quenches/driving) corresponding to a fixed energy density in the dual states.

In the language of numerical general relativity, the problem is thus the following. We set initial conditions for a four-dimensional asymptotically anti-de Sitter (aAdS₄) geometry which satisfies the vacuum Einstein equation with a negative cosmological constant. The resulting solution of the gravitational equations is dual to a nonequilibrium state which eventually thermalizes, mirroring the ubiquitous feature of gravity that any spacetime eventually settles down to a static AdS₄–Schwarzschild black hole. The irreversibility is captured by the growth of the areas of the apparent and event horizons that ultimately coincide, and these give two nonequilibrium entropy functions. From the asymptotic behavior of the aAdS₄ metric, one can read off the (time-dependent) energy-momentum tensor in the dual theory, particularly the pressure anisotropy [26,27].

Here, we show that we can use machine learning to reconstruct the apparent and event horizon areas from simple features of the time series of the pressure anisotropy *without* any explicit knowledge of the initial conditions. Indeed, because there is no independent analytic understanding of the growth of horizon entropies, machine learning is a novel tool for determining how boundary and bulk physics are correlated from a minimal set of inputs.

This problem is also of interest because it has been shown earlier that just with the data of the time-dependent event and apparent horizon areas one can explicitly construct simple spacetime metrics of the AdS–Vaidya type [28,29] which do not satisfy the Einstein equations but give excellent approximations to many other observables in the dual field theory such as correlation functions.¹

¹An AdS–Vaidya metric has a single nontrivial function, namely, a mass, which is a function of the boundary (field theory) coordinates. An analogue construction is a thermal density matrix with a spacetime dependent temperature. However, the latter is not equivalent to an AdS–Vaidya metric, which is more natural in holography from the point of view of the emergent bulk geometry.

Therefore, deducing the growth of the entropies from simple features of the time series of pressure anisotropy can lead us to predict other observables such as correlation functions of the dual state in holographic theories, while simplifying numerical holography, and opening the door to new insights not only into the dynamics of strongly coupled quantum many-body systems but also into the fundamentals of the holographic duality as discussed in the concluding section.

The plan of the paper is as follows. In Sec. II, we discuss the setup which we use to generate the data for the holographic bulk horizons. In Sec. III, we discuss the training and performance of the neural networks for predictions of the time-dependent areas of the bulk horizons. Finally, in Sec. IV, we conclude with discussions on the implications of our results.

II. HOLOGRAPHIC SETUP

In the limit of strong coupling and large rank of the gauge group, holography maps the nonequilibrium dynamics of quantum field theories to classical gravitational dynamics in asymptotically AdS spacetimes in one higher spatial dimension [3–5]. In this section, we describe the setup of the dual gravitational problem, which will subsequently be used to generate data for our machine learning task. To do this, we solve the Einstein equations starting from a nonequilibrium asymptotically AdS metric, which represents the out-of-equilibrium initial state of the dual field theory. (See [30,31] for some excellent reviews on this topic.) By holographic renormalization [26,27,32], we obtain the expression for the stress tensor of the dual quantum theory.

The bulk metric settles down eventually to a planar AdS–Schwarzschild black hole which is dual to the finite temperature state in the gauge theory. We obtain the entropies by locating the event and apparent horizons in the numerically obtained bulk metrics and computing their time dependent areas.

A typical homogeneous metric of an asymptotically AdS₄ black brane in Eddington–Finkelstein coordinates (with flat Minkowski metric on the boundary) is

$$ds^2 = 2drdt - A(r, t)dt^2 + S(r, t)^2(e^{-B(r, t)}dx^2 + e^{B(r, t)}dy^2), \quad (1)$$

where r is the bulk radial coordinate (the boundary of spacetime is at $r \rightarrow \infty$), t is time coordinate on the boundary, which is a null coordinate in the bulk, and x and y are boundary spatial coordinates. In what follows, we will denote the bulk coordinates by capital Latin indices and boundary coordinates by Greek indices. The metric is written in terms of A , B , and S , which are functions of the radial coordinate and time. We focus on AdS₄ as this supplies the simplest examples; it is straightforward to generalize to higher dimensions.

A special case of (1) is the Schwarzschild–AdS₄ black brane metric for which

$$A(r, t) = \frac{r^2}{L^2} - \frac{ML^2}{r}, \quad B(r, t) = 0, \quad S(r, t) = \frac{r}{L}, \quad (2)$$

where M is the ADM mass [33], which is related to Hawking temperature T by the relation

$$M = \frac{64}{27} L^3 \pi^3 T^3, \quad (3)$$

with L being the radius of AdS₄. The event and apparent horizons coincide at the position $r_H = L^{4/3} M^{1/3}$. We will use units in which $L = 1$ since only dimensionless combinations with L relate to physical constants and variables in the dual three dimensional conformal field theory.

The boundary metric is identified with the physical background metric on which the dual field theory lives. Requiring that it is a codimension one Minkowski space implies that the functions A , B , and S should behave as²

$$A(r, t) = r^2 + \frac{a_n(t)}{r} + \dots, \quad B(r, t) = \frac{b_n(t)}{r^3} + \dots, \\ S(r, t) = r - \frac{1}{8r} + \dots, \quad (4)$$

at large r . The normalizable modes $a_n(t)$ and $b_n(t)$ give the energy density ε and the pressure anisotropy ($P_x - P_y$) in the dual state, respectively [26,27]. Explicitly, the expectation value of the dual energy-momentum tensor $T^{\mu\nu}$ obtained from holographic renormalization has the following nonvanishing components:

$$\hat{T}^{tt} = \hat{\varepsilon} = -2a_n(t), \quad \hat{T}^{xx} = \hat{P}_x = -a_n(t) - 3b_n(t), \\ \hat{T}^{yy} = \hat{P}_y = -a_n(t) + 3b_n(t), \quad (5)$$

where $\hat{T}^{\mu\nu}$ is defined via $T^{\mu\nu} = \kappa^{-1} \hat{T}^{\mu\nu}$ with $\kappa^{-1} = (8\pi G_N)^{-1}$ (in units $L = 1$ and G_N being Newton's gravitational constant) identified with N^2 up to a numerical constant. As expected in a conformal field theory, the energy-momentum tensor is traceless. The functions $a_n(t)$ and $b_n(t)$ are determined by the initial conditions for the bulk metric as discussed below.

The bulk Einstein equations in the characteristic form are [30]:

$$S'' = -\frac{1}{4} S(B')^2, \quad \dot{S}' = \frac{3S}{2} - \frac{\dot{S}S'}{S}, \\ \dot{B}' = -\frac{\dot{S}B'}{S} - \frac{\dot{B}S'}{S}, \quad A'' = \frac{4\dot{S}S'}{S^2} - \dot{B}B', \quad \dot{S} = \frac{\dot{S}A'}{2} - \frac{\dot{B}^2 S}{4}. \quad (6)$$

²Here we have fixed a residual gauge freedom by setting the coefficient of r in the expansion of $A(r, t)$ to zero.

Above, the prime denotes the radial derivative, and the dot denotes the derivative along the outgoing radial null geodesic, e.g.,

$$\dot{f}(r, t) := \partial_t f + \frac{A(r, t)}{2} f'(r, t). \quad (7)$$

The first four expressions in (6) are *radial* evolution equations that are solved for the functions S , $\dot{S}$, $\dot{B}$, and A in this order at every time step. The last equation for $\dot{S}$ is a constraint equation implying the conservation of the dual energy-momentum tensor (5) in background Minkowski metric. The latter simply states that $a_n(t)$ is a constant.

It should be obvious from the nested characteristic form of the gravitational equations that given the radial profile of B [i.e., $B(r, t = 0)$] at an initial time and the initial value of a_n (which turns out to be a constant), we obtain unique solutions for $S(r, t)$, $B(r, t)$, and $A(r, t)$ satisfying the asymptotic behavior (4), and thus a unique bulk metric. The constraint equation evaluated at several radial locations can be used to monitor the accuracy of numerics.³

For numerical simulations, it is convenient to use the radial coordinate $z = L^2/r$. We set the range of z between 0 and L which includes the apparent horizon (and thus the event horizon as well). We also set $L = 1$ and $a_n = -1$ (i.e., the energy density $\hat{\varepsilon}$ to 2 at all times).

We use the following family of profiles of $B(z, t = 0)$ for specifying the initial conditions:

$$B(z, t = 0) = z^3 (\mathcal{R}_1 e^{-200(z-0.2)^2} + \mathcal{R}_2 e^{-200(z-0.5)^2} \\ + \mathcal{R}_3 e^{-100(z-0.8)^2} + N(z)) \\ N(z) = \sum_{i=1}^{50} \mathcal{N}_{1i} \sin(\mathcal{N}_{2i} z + \mathcal{N}_{3i}) \quad (8)$$

where $\mathcal{R}_{1,2,3}$ are (pseudo)random numbers between 0 and 1 and $z = L^2/r$. $\mathcal{N}_{1i}$ are drawn from a Gaussian distribution with mean zero and standard deviation of 0.01. $\mathcal{N}_{2i}$ and $\mathcal{N}_{3i}$ are (pseudo)random numbers between 0 and 2π . These profiles are superpositions of three Gaussians centered at 0.2, 0.5, and 0.8, respectively, with added Gaussian noise $N(z)$. By varying the parameters $\mathcal{R}_{1,2,3}$, $\mathcal{N}_{1i}$, $\mathcal{N}_{2i}$, and $\mathcal{N}_{3i}$ with $i = 1, \dots, 50$, we generate various initial conditions. Thus the total number of these parameters specifying the initial conditions is 153.

We solve the set of nested ordinary differential equations (6) in the radial coordinate using the pseudospectral method [30,34]. In the pseudospectral method, the radial coordinate is discretized using a Chebyshev grid with a certain choice of the number of grid points, n . In our simulations, we have chosen $n = 60$. The time updates are

³For the purpose of numerics, it is useful to further redefine variables so that they take finite values at the boundary.

performed using the fourth order Adams–Bashforth method. This method requires knowing the values of the functions to be time evolved at previous four time steps. Therefore, for the first nine time steps, we use smaller time steps and perform numerical integration of the interpolated functions. From the tenth time step onwards, the fourth order Adams–Bashforth method is used for time stepping with the value of time step $\delta t = 0.001$. Finally, the time-dependent pressure anisotropy $[b_n(t)]$ in the corresponding state of the dual theory can be readily extracted from the numerical solution.

A. Event and apparent horizons

In thermal equilibrium, the entropy of the dual field theory is calculated from the area of the event horizon via the Bekenstein–Hawking formula: $S = \frac{A_{\text{EH}}}{4G}$. However, in out-of-equilibrium settings, the entropy in the dual field theory can be given by the area of the apparent horizon, which is the outermost trapped surface [25,35,36]. The area of the event horizon, which satisfies the second law strictly [37,38], enumerates the spacetime degrees of freedom.⁴ The apparent horizon is always interior to (or coincident with) the event horizon. As well, the apparent horizon is defined locally on a given spacelike slice, whereas the event horizon is a global property of the spacetime, and therefore harder to compute.

The position of the event horizon (r_{EH}) is obtained from the solution to the equation [31]

$$\frac{d}{dt} r_{\text{EH}}(t) - \frac{A(r_{\text{EH}}(t), t)}{2} = 0, \quad (9)$$

subject to the boundary condition $r_{\text{EH}}(t \rightarrow \infty) = M_f^{1/3}$, where M_f is the ADM mass of the *final* equilibrated black brane. The position of the apparent horizon (r_{AH}) is determined by solving the algebraic equation [31]

$$\dot{S}(r_{\text{AH}}, t) = 0. \quad (10)$$

The entropy density of the dual field theory system, which is proportional to the area of either the event or the apparent horizons as appropriate, is given by

$$s_*(t) = \frac{1}{2\pi} S(r_*(t), t)^2. \quad (11)$$

In the numerical code, we use the variable $z = \frac{1}{r}$ instead of r for the radial bulk coordinate since z has finite range unlike r .

We can use various initial conditions given by $B(z, t = 0)$ and $a_n(t = 0)$ and solve the Einstein equations. [Recall that

⁴A generalized second law can be argued to hold for general null congruences [39,40].

$a_n(t)$ turns out to be a constant.] Thus we obtain the boundary anisotropy $b_n(t)$ time series data. Since the dual conformal field theory has no intrinsic scale, we can set $a_n = -1$ so that the dual energy density is given by $\hat{\epsilon} = 2$ without any loss of generality. We design the neural network to predict the areas of the event and apparent horizons (at all times) from this boundary anisotropy time series data.

III. MACHINE LEARNING ENTROPY

To predict the areas of event and apparent horizons, we employ *deep neural networks*. (We briefly summarize neural networks in Appendix A. See [41,42] for excellent contemporary reviews.) The training data for these predictions are the time series for the boundary pressure anisotropy, $b_n(t)$.

As mentioned in the previous section, we have used various initial conditions for $B(z, t = 0)$ with 153 parameters while holding the energy density fixed (in the appropriate units) to 2. The latter ensures that all the initial conditions in the dataset thermalize at approximately the same time $t_{\text{th}} \approx 2$. Since the energy density remains constant throughout the evolution, it sets the final temperature and therefore the timescale for thermalization. After the thermalization time, the event and apparent horizons undergo a ringdown before coinciding with the horizon of the final static black hole. This ringdown is governed by the quasinormal mode of the final black hole which also governs the time-dependent pressure anisotropy. Therefore, the areas of the event and apparent horizons, and the pressure anisotropy are sensitive to the details of the initial conditions only prior to the thermalization time. So, it is of interest to understand how the pressure anisotropy encodes the bulk horizon areas only before the thermalization time.

Although the apparent horizon can be causally determined unlike the event horizon, we are interested only in using discrete time series of the pressure anisotropy for predicting the area of the bulk horizons. Therefore, it is necessary to use the discrete time series of the pressure anisotropy up to a sufficiently long time for the predictions to be sufficiently accurate. We find that the optimal truncation of the discrete time series is $t = t_0 + 3$, with t_0 being the initialization time, and this is a bit beyond the time $t_0 + t_{\text{th}} = t_0 + 2$ when the solutions thermalize.

We find that the predictions of the event horizon as well as the apparent horizon areas happen almost perfectly at *all times* starting from the initial time both with the inputs of the *long* and *short* (sparse) data for the time series of the pressure anisotropy in the dual boundary theory. More precisely, the long and short input vectors are as defined below.

The long input vector is the discrete time series of the pressure anisotropy $b_n(t_0 + k\delta t)$, where $t_0 = 0$ is the initial time, $\delta t = 0.01$, and

$$k = \left(10^{-4}, 10^{-7/2}, 10^{-3}, 10^{-5/2}, 10^{-2}, 10^{-3/2}, 10^{-1}, 10^{-1/2}, 1 - \sum_{i=-8}^{-1} 10^{i/2} \right)$$

for first nine time steps respectively and $k = 1$ from the tenth time step onward. (Although, we use time step $\delta t = 0.001$ for the simulations as mentioned in the previous section to generate elements of our dataset, for the long input vector, we only retain values of various dynamical quantities with $\delta t = 0.01$ after the ninth time stepping.) The full time series is truncated to $t = 3$ which is well past the thermalization time scale. The *short input vector* contains a handful of only ten features of the pressure anisotropy time series which are

- (1) $b_{n,\max}$: maximum value that $b_n(t)$ takes;
- (2) $t_{\max}$: time t when $b_{n,\max}$ is attained;
- (3) $b_{n,\min}$: minimum value that $b_n(t)$ takes;
- (4) $t_{\min}$: time t when $b_{n,\min}$ is attained;
- (5) full width at half maximum;

(6) full width at half minimum; and
 (7) times corresponding to first four zeroes of $b_n(t)$.
 Therefore, the length of the *short input vector* is 10, which is significantly shorter than the number of parameters in the initial condition (8), which is 153 as discussed in the previous section. Note also that the full time series is truncated to $t = 3$.

Our main result is that training the neural network with only the handful of simple features of $b_n(t)$ time series in the short input vector suffices to generate predictions comparable to the predictions made by the neural network trained on the almost full $b_n(t)$ time series data, namely the long input vector. The neural network predicts both the event and apparent horizon areas (almost) perfectly at all times in both cases. Figure 1 gives representative plots for

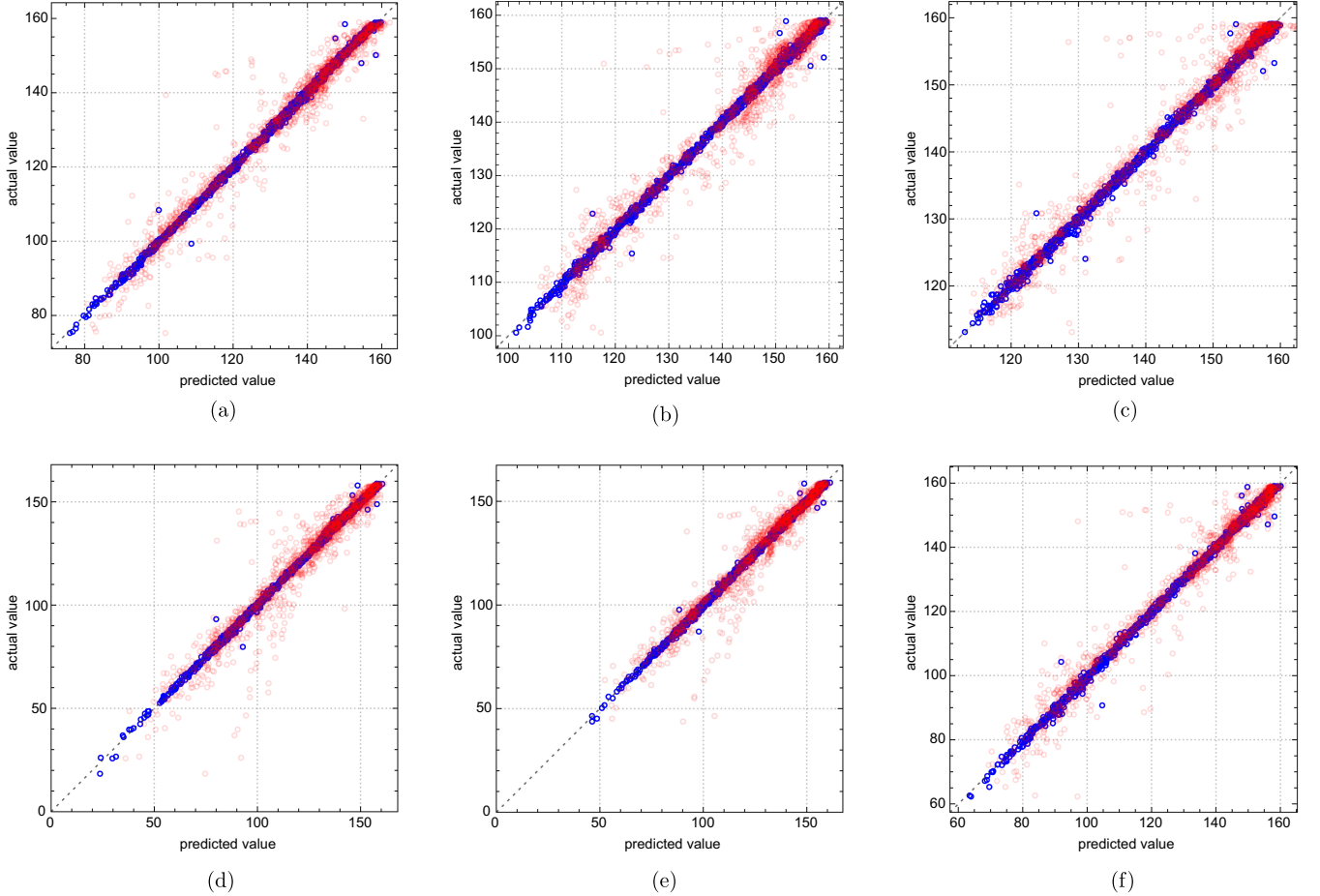


FIG. 1. Testing statistics for S_{EH} and S_{AH} with both *long input vector* (blue) as well as *short input vector* (red) at various times. Note that the scale is set by the (conserved) energy density which is the same for all the elements of the dataset used for machine learning, which makes it meaningful to compare the times across the dataset. (a) S_{EH} at $t = 0$. (b) S_{EH} at $t = 0.75$. (c) S_{EH} at $t = 1$. (d) S_{AH} at $t = 0$. (e) S_{AH} at $t = 0.75$. (f) S_{AH} at $t = 1$.

TABLE I. Various statistical measures assessing performance of neural network for S_{EH} prediction.

Time	Input vector	MD	Baseline	SD	R -squared	χ -squared
$t = 0$	Long	0.00422111	0.14546	0.818	0.998	7.77747
	Short	0.018707	0.14546	3.91	0.964	187.567
$t = 0.75$	Long	0.00323828	0.101198	0.686	0.998	5.1811
	Short	0.0159789	0.101198	3.41	0.952	128.494
$t = 1$	Long	0.00276327	0.0821333	0.617	0.998	4.063
	Short	0.0119103	0.0821333	2.85	0.952	85.4856

TABLE II. Various statistical measures assessing performance of neural network for S_{AH} prediction.

Time	Input vector	MD	Baseline	SD	R -squared	χ -squared
$t = 0$	Long	0.00681106	0.246200	1.10	0.998	19.3582
	Short	0.0452788	0.246200	7.54	0.930	1262.02
$t = 0.75$	Long	0.00582296	0.194025	1.02	0.998	13.5395
	Short	0.0326038	0.194025	6.02	0.942	627.272
$t = 1$	Long	0.0055171	0.180533	1.06	0.998	15.0186
	Short	0.0227724	0.180533	4.71	0.962	289.725

the predictions of the neural network for the event and apparent horizon areas at initial time, and at times 0.5 and 1 after the initial time, both when it is trained with the long and short input vectors, respectively.

To assess the performance of the neural network, we compute the following quantities over the test set:

$$\text{mean deviation (MD): } \text{Mean} \left[\frac{\text{Abs}[\text{actual} - \text{predicted}]}{\text{actual}} \right], \quad (12)$$

$$\text{baseline: } \text{Mean} \left[\frac{\text{Abs}[\text{actual} - \text{Mean}[\text{actual}]]}{\text{actual}} \right]. \quad (13)$$

The baseline (13) simply assigns the mean in the dataset as a universal prediction. The accuracy of the trained neural network is compared to this baseline. We also calculate the standard deviation (SD) between the actual value and the predicted value, R -squared, and χ -squared in the usual way. These statistics are reported in Tables I and II.

We have performed a more robust test of the performance of the neural networks with respect to the predictions when trained with short or long input vectors via k -fold cross validation. This has been reported in Appendix B.

IV. DISCUSSION AND OUTLOOK

In this work, we have demonstrated that neural networks can learn some aspects of bulk geometry given simply the time series of a one-point function (expectation value of an operator) as input. In particular, we have established that the areas of the event and apparent horizons can be learned

from the time series data of the pressure anisotropy in the boundary field theory. Furthermore, neural networks also predict the black hole entropies at all times from only a small number of features of the time series data, namely the values and half widths of the maxima and minima, the times the latter are attained, and the times of the first zeroes with the maximum length of the input vector set to 10 and with the time series truncated to a maximum value of time commensurate with the thermalization time.

Generally, in order to construct the bulk spacetime metric, one would need the boundary data and also initial conditions. In our work, we have considered a wide class of initial conditions, and we have demonstrated that machine learning of the event and apparent horizon entropies is possible with inputs of *only* compressed boundary data. These results are surprising and point to universality classes of bulk metrics and their dual states which can be characterized only by simple features of the time series of simple observables, namely one-point functions. Our work therefore indicates that neural networks are not only promising tools for addressing bulk reconstruction in AdS/CFT but also for obtaining new insights into holographic dynamics.

We argue that our approach can yield the holographic interpretation of the entropy functions obtained from the evolving event and apparent horizons in dynamical black hole geometries dual to nonequilibrium states. It is natural to think of entropy functions as determining simplistic statistical ensembles which can be constructed from simple inputs and which provide best approximations to all observables of a given state. In fact, a precise holographic interpretation of semiclassical spacetime itself as a minimum bias statistical ensemble has been advocated in [43].

The event and apparent horizon entropies can be naturally thought of as further simplifications of the emergent spacetime.

The first step towards developing such a precise statistical interpretation of nonequilibrium event and apparent horizon entropies would be to test if machine learning of these entropy functions is possible without a detailed knowledge of the two-derivative gravity theory (Einstein's theory coupled to a few bulk fields) governing the bulk dynamics in holographic duality. If the entropies can lead to a statistical ensemble which can approximate observables of the non-equilibrium state dual to the dynamical black hole geometry, the algorithm for obtaining such an ensemble should not depend on the specific details of the theory. Conversely, it should be possible to extract these entropy functions without detailed knowledge of the dual strongly interacting holographic theory.

Indeed the results of [29] strongly suggest that the entropy functions can be learned from simple boundary data without detailed knowledge of the underlying microscopic theory or the dual bulk gravitational theory. It was shown in [29] that a simple AdS–Schwarzschild geometry with a time-dependent mass (*viz.*, an AdS–Vaidya geometry) gives the *best* approximation to nonequilibrium causal correlation functions provided it reproduces the *exact* time-dependent apparent horizon entropy of the *exact* dual gravitational solution dual to the given state. Similarly, a simple AdS–Vaidya geometry gives the *best* approximation to coarse-grained features of non-equilibrium causal correlation functions provided it reproduces the *exact* time-dependent event horizon entropy of the *exact* dual gravitational solution dual to the given state.⁵ In both cases, the approximations are typically better than one percent on average.

Such AdS–Vaidya geometries are natural gravitational analogues of thermal density matrices with just time-dependent temperatures (although not exactly equivalent to the latter) as their construction requires only inputs of the entropy functions (and no knowledge of the dual gravitational theory). Our present work together with the results of [29] suggest that conversely the entropy functions can possibly be learned from simple one-point functions *without* knowledge of the dual bulk two-derivative gravitational theory, and thus such entropy functions can be interpreted as measures of information which need to be extracted from

⁵This work builds on earlier suggestions of [28] where AdS–Vaidya geometries were proposed as the starting point of approximating slow quenches. However, instead of giving exact event/apparent horizon entropies, these AdS–Vaidya geometries were supposed to reproduce exact one-point functions like pressure anisotropies. In [29], it was shown that better approximations are obtained if the AdS–Vaidya geometries are constructed such that exact event/apparent horizon entropies are reproduced. The numerical studies in [29] were based on methods developed in [44].

one-point functions in order to construct quantum statistical ensembles that give best possible approximations to the dual states, especially for reproducing the two-point correlation functions. Our approach can be useful to arrive at such a precise interpretation.

In the future, we would like to unbox the neural network in order to understand how learning takes place, for instance, by employing techniques like layerwise relevance propagation [45,46] and this should be important for the stated goal of interpreting the entropy functions. Because this work is a proof of principle, in which we can apply Big Data techniques to holography and machine learning salient features of spacetime dynamics from dual data, we do not have a need to optimize the network hyperparameters to improve performance. It would be good to do this, of course.

ACKNOWLEDGMENTS

We are grateful to Jessica Craven, Koji Hashimoto, Shivaprasad Hulyal, Dileep Jatkar, Lata Joshi, Arjun Kar, Tanay Kibe, David Mateos, and Harald Skarke for enlightening discussions on this and related work. We acknowledge the High-Performance Scientific Computing facility of the Harish-Chandra Research Institute, where we generated most of the data used for training and testing our neural networks. We thank the organizers and participants of String Data 2022, where aspects of this work were presented. V. J. is supported by the South African Research Chairs Initiative of the Department of Science and Innovation and the National Research Foundation. V. J. would also like to thank the Isaac Newton Institute for Mathematical Sciences for support and hospitality during the program “Black holes: bridges between number theory and holographic quantum information” during which work on this paper transpired; this work was supported by EPSRC Grant No. EP/R014604/1. The research of A. M. was partly supported by the center of excellence grants of the Ministry of Education of India.

DATA AVAILABILITY

The data that support the findings of this article are openly available [47].

APPENDIX A: FEEDFORWARD NEURAL NETWORKS

Consider the map

$$f_i: \mathbb{R}^{n_{i-1}} \rightarrow \mathbb{R}^{n_i}$$

$$\mathbf{v}_{i-1} \mapsto \mathbf{v}_i = \sigma(\mathbf{v}'_i), \quad (\text{A1})$$

where $\mathbf{v}'_i = W^{(i)} \cdot \mathbf{v}_{i-1} + \mathbf{b}_i$. The vector $\mathbf{b}_i \in \mathbb{R}^{n_i}$ is called a *bias vector* and the $n_i \times n_{i-1}$ matrix $W^{(i)}$ is called a *weight matrix*. The *activation function* σ acts nonlinearly element-wise on the components of the vector $\mathbf{v}'_i$. In our case, we

TABLE III. Average and standard deviation across 5-folds (of cross validation) of various statistical measures assessing performance of neural network for S_{EH} prediction.

		$t = 0$		$t = 0.75$		$t = 1$	
		Long	Short	Long	Short	Long	Short
Average	MD	0.0053267	0.022483	0.0045241	0.017409	0.0046980	0.012438
	Baseline	0.14975	0.14975	0.10403	0.10403	0.08427	0.08427
	SD	1.1996	4.836	1.0022	3.91	1.04	3.074
	R -squared	0.9952	0.9432	0.9944	0.9358	0.993	0.9458
	χ -squared	33.345	431.96	20.283	241.93	18.168	136.94
SD	MD	0.0040222	0.0052088	0.0024334	0.0035442	0.0018669	0.0010649
	Baseline	0.0024558	0.0024558	0.0014224	0.0014224	0.0011119	0.0011119
	SD	0.88969	1.3000	0.65153	0.92081	0.46326	0.36657
	R -squared	0.0061400	0.031885	0.0056833	0.030078	0.0064420	0.013442
	χ -squared	38.258	229.85	19.683	110.71	15.443	32.451

choose $\sigma(x) = x\Theta(x)$, where $\Theta(x)$ is the heaviside step function. (This is the rectified linear unit or ReLU activation.) Taking the vector $\mathbf{v}_{in} := \mathbf{v}_0 \in \mathbb{R}^{n_0}$ as the input vector, this procedure is iterated:

$$\mathbf{v}_\ell = f_\ell(f_{\ell-1}(\cdots(f_1(\mathbf{v}_0)\cdots))), \quad a_{out} := \sum_{m=1}^{n_\ell} \mathbf{v}_\ell^m. \quad (\text{A2})$$

In writing a_{out} , we have summed the components of the vector $\mathbf{v}_\ell$.

Suppose we have a set of input vectors labeled by j that we use for training. To each of these input vectors $\mathbf{v}_{in,j}$ we associate a target value $\alpha_j \in \mathbb{R}$. We then consider a mean squared loss function that compares each α_j to its predicted value $a_{out,j}$. Backpropagation enables us to tune the elements of the weight matrices and bias vectors to minimize this loss. In this way, we have constructed an ℓ -hidden layer fully connected feedforward neural network that is trained to approximate α_j from the features encoded in $\mathbf{v}_{in,j}$ via stochastic gradient descent. The dimension n_i specifies the number of neurons in the i th hidden layer.

In our experiments, the architectures of the neural network are dependent on the vector $\mathbf{v}_{in}$. With long input vectors in $\mathbb{R}^{299}$, we use three hidden layers with 600 neurons each. With short input vectors in $\mathbb{R}^{10}$, we use three hidden layers with 20 neurons each. (The input vectors are described in Sec. III.) We use 70% of a 10000 data points for training, 15% for validation, and 15% for testing. Performance statistics are reported on the test set. The neural networks are implemented in *Mathematica* [48] and use Adam for optimization. We make our code available on Github [47].

APPENDIX B: k -FOLD CROSS VALIDATION

As a further test of the robustness of predictive power of our neural networks trained on both long and short input data, we perform k -fold cross validation. In this method, one randomly divides the entire dataset into k subsets (also called k folds). Then, one keeps one of these k folds as testing data and trains the neural network on the remaining $k - 1$ folds. One repeats this k times, each time choosing a different fold

TABLE IV. Average and standard deviation across 5-folds (of cross validation) of various statistical measures assessing performance of neural network for S_{AH} prediction.

		$t = 0$		$t = 0.75$		$t = 1$	
		Long	Short	Long	Short	Long	Short
Average	MD	0.0129314	0.040043	0.00998527	0.032521	0.00753043	0.029117
	Baseline	0.25751	0.25751	0.20153	0.20153	0.18673	0.18673
	SD	1.6724	6.702	1.3656	5.998	1.0708	5.804
	R -squared	0.9952	0.9466	0.9954	0.9438	0.9972	0.9434
	χ -squared	71.6847	1085.3	48.1872	754.47	28.3094	644.71
SD	MD	0.01067	0.0044043	0.00881007	0.0036941	0.00654729	0.0043984
	Baseline	0.010400	0.010400	0.0050024	0.0050024	0.0029318	0.0029318
	SD	1.31589	0.73291	1.16507	1.0039	0.777636	0.98571
	R -squared	0.00535724	0.012177	0.00650385	0.020535	0.00363318	0.020598
	χ -squared	77.777	325.87	65.9824	256.38	39.0392	215.72

as the testing dataset and the corresponding remaining $k - 1$ folds as the training dataset. Finally, one takes the average and standard deviation of all k runs' learning statistics. The most common choices of k are 5 or 10. We chose $k = 5$.

We find that the various statistical measures reported in Tables I and II do not vary much across 5-folds. In

particular R -squared stays ~ 1 and others also do not deviate significantly from their mean value across 5-folds which are close to their corresponding values reported in Tables I and II. We report the average values of each of the quantities of Tables I and II across 5-folds and also their standard deviation in the Tables III and IV.

-
- [1] G. 't Hooft, Dimensional reduction in quantum gravity, *Conf. Proc. C* **930308**, 284 (1993).
 - [2] L. Susskind, The World as a hologram, *J. Math. Phys. (N.Y.)* **36**, 6377 (1995).
 - [3] J. M. Maldacena, The large N limit of superconformal field theories and supergravity, *Adv. Theor. Math. Phys.* **2**, 231 (1998).
 - [4] S. S. Gubser, I. R. Klebanov, and A. M. Polyakov, Gauge theory correlators from noncritical string theory, *Phys. Lett. B* **428**, 105 (1998).
 - [5] E. Witten, Anti-de Sitter space and holography, *Adv. Theor. Math. Phys.* **2**, 253 (1998).
 - [6] T. Banks, W. Fischler, S. H. Shenker, and L. Susskind, M theory as a matrix model: A conjecture, *Phys. Rev. D* **55**, 5112 (1997).
 - [7] D. Harlow, TASI lectures on the emergence of bulk physics in AdS/CFT, *Proc. Sci. TASI2017 (2018)* 002 [arXiv:1802.01040].
 - [8] T. Kibe, P. Mandayam, and A. Mukhopadhyay, Holographic spacetime, black holes and quantum error correcting codes: A review, *Eur. Phys. J. C* **82**, 463 (2022).
 - [9] K. Hashimoto, S. Sugishita, A. Tanaka, and A. Tomiya, Deep learning and holographic QCD, *Phys. Rev. D* **98**, 106014 (2018).
 - [10] K. Hashimoto, S. Sugishita, A. Tanaka, and A. Tomiya, Deep learning and the AdS/CFT correspondence, *Phys. Rev. D* **98**, 046019 (2018).
 - [11] K. Hashimoto, AdS/CFT correspondence as a deep Boltzmann machine, *Phys. Rev. D* **99**, 106017 (2019).
 - [12] H.-Y. Hu, S.-H. Li, L. Wang, and Y.-Z. You, Machine learning holographic mapping by neural network renormalization group, *Phys. Rev. Res.* **2**, 023369 (2020).
 - [13] J. Tan and C.-B. Chen, Deep learning the holographic black hole with charge, *Int. J. Mod. Phys. D* **28**, 1950153 (2019).
 - [14] Y.-K. Yan, S.-F. Wu, X.-H. Ge, and Y. Tian, Deep learning black hole metrics from shear viscosity, *Phys. Rev. D* **102**, 101902 (2020).
 - [15] T. Akutagawa, K. Hashimoto, and T. Sumimoto, Deep learning and AdS/QCD, *Phys. Rev. D* **102**, 026020 (2020).
 - [16] K. Hashimoto, H.-Y. Hu, and Y.-Z. You, Neural ordinary differential equation and holographic quantum chromodynamics, *Mach. Learn. Sci. Technol.* **2**, 035011 (2021).
 - [17] M. Song, M. S. H. Oh, Y. Ahn, and K.-Y. Kim, AdS/deep-learning made easy: Simple examples, *Chin. Phys. C* **45**, 073111 (2021).
 - [18] K. Hashimoto, K. Ohashi, and T. Sumimoto, Deriving the dilaton potential in improved holographic QCD from the meson spectrum, *Phys. Rev. D* **105**, 106008 (2022).
 - [19] J. Lam and Y.-Z. You, Machine learning statistical gravity from multi-region entanglement entropy, *Phys. Rev. Res.* **3**, 043199 (2021).
 - [20] K. Li, Y. Ling, P. Liu, and M.-H. Wu, Learning the black hole metric from holographic conductivity, *Phys. Rev. D* **107**, 066021 (2023).
 - [21] P. Kumar, T. Mandal, and S. Mondal, Black holes and the loss landscape in machine learning, *J. High Energy Phys.* **10** (2023) 107.
 - [22] K. Hashimoto, K. Ohashi, and T. Sumimoto, Deriving the dilaton potential in improved holographic QCD from the chiral condensate, *Prog. Theor. Exp. Phys.* **2023**, 033B01 (2023).
 - [23] R. Katsube, W.-H. Tam, M. Hotta, and Y. Nambu, Deep learning metric detectors in general relativity, *Phys. Rev. D* **106**, 044051 (2022).
 - [24] C. Park, S. Kim, and J. H. Lee, Holography transformer, arXiv:2311.01724.
 - [25] P. M. Chesler and L. G. Yaffe, Horizon formation and far-from-equilibrium isotropization in supersymmetric Yang-Mills plasma, *Phys. Rev. Lett.* **102**, 211601 (2009).
 - [26] V. Balasubramanian and P. Kraus, A stress tensor for anti-de Sitter gravity, *Commun. Math. Phys.* **208**, 413 (1999).
 - [27] S. de Haro, S. N. Solodukhin, and K. Skenderis, Holographic reconstruction of space-time and renormalization in the AdS/CFT correspondence, *Commun. Math. Phys.* **217**, 595 (2001).
 - [28] S. Bhattacharyya and S. Minwalla, Weak field black hole formation in asymptotically AdS spacetimes, *J. High Energy Phys.* **09** (2009) 034.
 - [29] L. K. Joshi, A. Mukhopadhyay, F. Preis, and P. Ramadevi, Exact time dependence of causal correlations and non-equilibrium density matrices in holographic systems, *Phys. Rev. D* **96**, 106006 (2017).
 - [30] P. M. Chesler and L. G. Yaffe, Numerical solution of gravitational dynamics in asymptotically anti-de Sitter spacetimes, *J. High Energy Phys.* **07** (2014) 086.
 - [31] W. van der Schee, Gravitational collisions and the quark-gluon plasma, Ph.D. thesis, Utrecht University, 2014, arXiv:1407.1849.
 - [32] K. Skenderis, Lecture notes on holographic renormalization, *Classical Quantum Gravity* **19**, 5849 (2002).

- [33] R. L. Arnowitt, S. Deser, and C. W. Misner, Dynamical structure and definition of energy in general relativity, *Phys. Rev.* **116**, 1322 (1959).
- [34] J. Boyd, *Chebyshev and Fourier Spectral Method* (Springer, New York, 1989).
- [35] I. Booth, M. P. Heller, G. Plewa, and M. Spalinski, On the apparent horizon in fluid-gravity duality, *Phys. Rev. D* **83**, 106005 (2011).
- [36] N. Engelhardt and A. C. Wall, Decoding the apparent horizon: Coarse-grained holographic entropy, *Phys. Rev. Lett.* **121**, 211301 (2018).
- [37] S. W. Hawking, Gravitational radiation from colliding black holes, *Phys. Rev. Lett.* **26**, 1344 (1971).
- [38] R. M. Wald, The thermodynamics of black holes, *Living Rev. Relativity* **4**, 6 (2001).
- [39] R. Bousso, A covariant entropy conjecture, *J. High Energy Phys.* **07** (1999) 004.
- [40] R. Bousso, Z. Fisher, S. Leichenauer, and A. C. Wall, Quantum focusing conjecture, *Phys. Rev. D* **93**, 064044 (2016).
- [41] P. Mehta, M. Bukov, C.-H. Wang, A. G. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, A high-bias, low-variance introduction to machine learning for physicists, *Phys. Rep.* **810**, 1 (2019).
- [42] F. Ruehle, Data science applications to string theory, *Phys. Rep.* **839**, 1 (2020).
- [43] D. L. Jafferis, D. K. Kolchmeyer, B. Mukhametzhanov, and J. Sonner, Jackiw-Teitelboim gravity with matter, generalized eigenstate thermalization hypothesis, and random matrices, *Phys. Rev. D* **108**, 066015 (2023).
- [44] S. Banerjee, T. Ishii, L. K. Joshi, A. Mukhopadhyay, and P. Ramadevi, Time-dependence of the holographic spectral function: Diverse routes to thermalisation, *J. High Energy Phys.* **08** (2016) 048.
- [45] G. Montavon, A. Binder, S. Lapuschkin, W. Samek, and K.-R. Müller, Layer-wise relevance propagation: An overview, in *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning* (Springer, New York, 2019), pp. 193–209.
- [46] J. Craven, V. Jejjala, and A. Kar, Disentangling a deep learned volume formula, *J. High Energy Phys.* **06** (2021) 040.
- [47] V. Jejjala, S. Mondkar, A. Mukhopadhyay, and R. Raj, ML-holo, <https://github.com/sukrut123/ML-Holo> (2023).
- [48] W. R. Inc., *Mathematica*, Version 13.3, Champaign, IL, 2023, <https://www.wolfram.com/mathematica>.