Deep Learning black hole metrics from shear viscosity

Yu-Kun Yan, 1 Shao-Feng Wu, 1, 2, Xian-Hui Ge, 1, 2 and Yu Tian 3, 4, 5, 2

1 Department of physics, Shanghai University, Shanghai, 200444, China
2 Center for Gravitation and Cosmology, Yangzhou University, Yangzhou 225009, China
3 School of Physics, University of Chinese Academy of Sciences, Beijing, 100049, China
4 Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing 100190, China
5 Center for Theoretical Physics, Massachusetts Institute of Technology, MA 02139, Cambridge, USA

Based on the AdS/CFT correspondence, we build up a simple deep neural network for the unsupervised learning of black-hole metrics from the complex frequency-dependent shear viscosity. The network architecture provides a discretized representation of the holographic renormalization group flow of shear response functions and is applicable for a large class of strongly coupled field theories. Given the existence of the horizon and guided by the continuity of spacetimes, we show that the Schwarzschild and Reissner-Nordström metrics can be learned accurately. Moreover, we illustrate that the generalization ability of the deep neural network can be excellent, which indicates that using the black hole spacetime as a hidden medium, a wide spectrum of shear viscosity can be generated from a narrow frequency range. Our work might not only suggest a data-driven way to study holographic transports, but also shed new light on the emergence mechanism of black hole spacetimes from field theories.

Introduction.—Renormalization group (RG) is a physical scheme to understand various emergent phenomena in the world through iterative coarse graining [1–4]. Deep learning (DL) algorithm is the core driving force of the recent wave of artificial intelligence [5]. It has been suggested that RG and DL might have a common logic [6] and their relation has attracted a lot of interest [7–14]. In particular, Mehta and Schwab [15] constructed a mapping between variational RG and deep neural network (DNN) based on restricted Boltzmann machines, from which they claimed that DL algorithms may implement a generalized RG scheme. Later, it was pointed out [16] that the standard renormalization process is essentially the feature extraction by supervised learning but cannot be generated by a general unsupervised learning unless the network structure is specific, such as the multi-scale entanglement renormalization ansatz (MERA) [17]. Furthermore, it was exhibited that by choosing the loss function to maximize the mutual information instead of the usual Kullback-Leibler divergence, the unsupervised learning can identify the relevant degrees of freedom and execute a RG process [18].

RG is believed as one of the key elements to understand the quantum gravity. In particular, by anti-de Sitter/conformal field theory (AdS/CFT) correspondence [19–22], a strongly coupled quantum critical theory in the d-dimensional spacetime is reorganized along the RG scale, inducing a classical theory of gravity in the d+1-dimensional AdS spacetime. RG is not the only connection between DL and gravity. Through the study of tensor networks [23–25], especially MERA, it has been realized that how the geometry is emergent from field theories usually involves the network and optimization, which are two important ingredients of DL.

Based on these connections, DNN is capable of providing a research platform for holographic duality [26–33]. One can expect two benefits at least: it is helpful to understand how the spacetime emerges and a data-driven phenomenological model can be built up for strongly coupled field theories. Work in the second respect was initiated by [28], where the inverse problem of the AdS/CFT is studied: how to reconstruct the spacetime metric from the given field theory data by the DNN which implements the AdS/CFT. Sequently, the so called AdS/DL correspondence is applied to learn the bulk metric from the lattice QCD data of the finite-temperature chiral condensate. Interestingly, the emergent metric exhibits both a black hole horizon and an IR wall with finite height, signaling the crossover of the QCD thermal phases [29].

In the prototype of the AdS/DL (i.e. the first numerical experiment in [28]), the architecture of the DNN is according to the discretized equation of motion of the real $\phi^4$ theory minimally coupled to the Einstein gravity, the data is the one-point function and the conjugate source with a label determined by the near-horizon scale field, the target is the metric of the Schwarzschild black hole, a key technique is to devise the regularization that ensures the continuity of the learned metric, and the DNN performs better near the boundary than near the horizon where the relative error is around 30%. In [33], it has been attempted to learn the Reissner-Nordström (RN) metric by AdS/DL but the mean square error (MSE) lies in the range from $\mathcal{O}(10^{-3})$ to $\mathcal{O}(10^{-1})$. Importantly, it is revealed that the form of the regularization term must be fine-tuned for different metrics. This suggests that the DNN cannot find the target metric if it is unknown previously, because one cannot judge which is closer to the target metric under different regularizations.

In this letter, we will address the technical problems and extend the physical range of AdS/DL. We also hope to look for the insights on the emergence mechanism of spacetimes. Our strategies are as follows. First of
all, consider that AdS/CFT is almost customized for the computation of the transports of strongly coupled quantum critical systems at finite temperatures \[\text{[34]}\] and in particular, the application of holography is anchored partially in the calculation of the shear viscosity \[\text{[35]}\]. Thereby we will adapt the complex frequency-dependent shear viscosity as the given field theory data. Second, we propose to build up the DNN according to the holographic RG flow of shear response function. This flow plays the essential role in the well-known holographic membrane paradigm \[\text{[36, 37]}\], which interpolates the standard AdS/CFT correspondence and classical black hole membrane paradigm smoothly \[\text{[38, 39]}\]. Third, our goal is to learn the black hole metric accurately. By selecting suitable coordinates, the metric at horizon can be fixed as zero. This will reduce the learning difficulties. Fourth, the system error in \[\text{[28]}\] comes from adding labels fixed as zero. This will reduce the learning difficulties.

Fifth, we still use the regularization to guide the algorithm, indicating that there are no artificially given labels. Fifth, we still use the regularization to guide the network finding a continuous metric. However, our training process has two stages and the regularization is only required in the first stage. So we can choose any regularization term as long as it can induce a smaller loss in the second stage. Finally, we will discuss possible extensions and physical implications.

From RG flow to DNN.—Suppose that an isotropic and translation-invariant strongly coupled field theory is dual to the 3+1-dimensional classical Einstein gravity, which allows a static black-hole solution with the metric ansatz

\[
ds^2 = -g_{tt}(r)dt^2 + g_{rr}(r)dr^2 + g_{xx}(r)dz^2.
\]

When the black-hole background is perturbed by time-dependent sources, the shear mode \((\partial g)_{xx} = h(r)e^{-i\omega t}\) of the gravitational wave is controlled by the equation of motion

\[
\frac{1}{\sqrt{-g}} \partial_t (\sqrt{-g} g^{tt} \partial_t h) + g^{tt} \omega^2 h = 0.
\]

In the Hamiltonian form, it can be written as

\[
\Pi = -\sqrt{-g} g^{tt} \partial_t h,
\]

\[
\partial_r \Pi = \sqrt{-g} g^{tt} g^{rr} \partial_r \omega^2 h,
\]

where \(\Pi\) is the momentum conjugate to the massless field \(h\). Consider the foliation in the \(r\)-direction and define the shear response function on each cutoff surface

\[
\chi = \frac{\Pi}{i\omega h}.
\]

Substituting Eq. \([3]\) into Eq. \([1]\), one can obtain a flow equation

\[
\partial_r \chi - i\omega \sqrt{\frac{g_{xx}}{g_{tt}}} \left( \frac{\chi^2}{2g_{xx}} - 
\]

Note that this equation has been derived in \[\text{[36]}\] where the DC limit is focused\textsuperscript{1} We will study the frequency-dependent behavior.

Applying the regularity of \(\chi\) on the horizon, one can read off the horizon value of \(\chi\) directly

\[
\chi(r_h) = g_{xx}(r_h),
\]

where \(r_h\) is the horizon radius. Taking Eq. \([7]\) as the IR boundary condition, the flow equation can be integrated to UV. However, it should be pointed out that the response function \(\chi\) on the UV boundary is not equal to the shear viscosity \(\eta\) of the boundary field theory. In the Supplementary Material (SM), we will clarify the relationship between them using the Kubo formula of the complex frequency-dependent viscosity tensor \[\text{[43, 44]}\] and the holographic renormalization of the Einstein-Maxwell theory \[\text{[45]}\].

In the metric ansatz \([1]\), \(g_{xx}\) can be fixed as \(r^2\) without loss of generality but \(g_{tt}\) is independent with \(g_{rr}\) in general. However, there are many black holes which share the feature \(g_{tt}g_{rr} = 1\), indicating that the radial pressure is the negative of the energy density \[\text{[10]}\]. For simplicity, we will study this situation at first and return to the more general case latter. For the simple situation, the metric ansatz can be reduced to

\[
ds^2 = \frac{1}{z^2} \left[-f(z)dt^2 + \frac{1}{f(z)} dz^2 + dz^2\right],
\]

where we have selected the coordinates so that the horizon is located at \(z = 1\) and the boundary at \(z = 0\). Accordingly, Eq. \([6]\) is simplified as

\[
\chi' + \frac{i\omega}{f} \left(z^2 \chi^2 - \frac{1}{z^2}\right) = 0.
\]

\textsuperscript{1} In the Wilsonian formulation, the flow equation can be retrieved as the \(\beta\)-functions of double-trace couplings \[\text{[10,42]}\].
We will build up a DNN according to this flow equation. A schematic diagram of the DNN is plotted in FIG. 1. The $N$ deep layers are located by discretizing the radial direction

$$z(n) = z_b + n \Delta z, \quad \Delta z = \frac{z_h - z_b}{N}, \quad (10)$$

where $z_b$ is the UV boundary, $z_h$ is the IR boundary, and the integer $n$ belong to $[1, N]$. The trainable weights of the network represent the discretized metrics. The input of the network is the given value of the response at IR boundary. The output is the response at UV boundary. The data is transferred from the $N$th layer (IR) to the 1th layer (UV). The upgrade rule of the response data between layers is determined by the discretized representation of Eq. (9)

$$\begin{align*}
\text{Re} \chi (z + \Delta z) &= \text{Re} \chi (z) + \Delta z \frac{2 \omega z^2}{f(z)} \times \\
\text{Im} \chi (z + \Delta z) &= \text{Im} \chi (z) + \Delta z \frac{\omega}{f(z)} \times \\
&\left[ \frac{1}{z^2} + z^2 (\text{Im} \chi (z))^2 - z^2 (\text{Re} \chi (z))^2 \right].
\end{align*} \quad (11)$$

Here we have separated the discretized flow equation into real and imaginary parts for the convenience in DL. We train this neural network to learn the mapping of frequencies to boundary response functions. Once training successfully, we can extract the discretized black hole metrics from the trained neural network.

The loss function we choose is the $L^2$-norm

$$L_{\text{DNN}} = \sum_{\text{data}} |\chi(\omega, z_b) - \tilde{\chi}(\omega, z_b)|^2, \quad (12)$$

up to a regularization term, if existed. Here $\chi$ denotes the boundary response generated by the DNN with the target metric and $\tilde{\chi}$ is the learned result.

We need a regularization term which can guide the DNN to find a continuous metric with the existence of a horizon. The form of the regularization term can be arbitrary as long as it can reduce the final loss. In practice, our regularization term can be specified as

$$L_{\text{REG}} = c_1 \sum_{n=1}^{N-1} \frac{1}{z(n)^2} |f(z(n+1)) - f(z(n))|^2 + c_3 |f(z(N)) - 0|^2, \quad (13)$$

where the two parts are designed for the continuity and the horizon, respectively. They involve three hyperparameters $c_1, c_2$ and $c_3$.

Data, training, and results.—We specify the discretized RG flow and hence the DNN by setting $z_b = 0.01$, $z_h = 0.99$, and $N = 10$. Using Eq. (11) with the IR boundary condition $\chi(z_h) = 1$ and the Schwarzschild metric

$$f(z) = 1 - z^3 \quad (14)$$

or the RN metric

$$f(z) = 1 - z^3 - \frac{q^2 z^3}{4} + \frac{q^2 z^4}{4}. \quad (15)$$

where $q$ is the charge density, we generate 2000 data $(\omega, \chi(\omega, z_b))$ from $\omega = 0$ to $\omega = 1$ with even spacing, see Fig. 2. The training set and validation set account for $90\%$ and $10\%$, respectively. Obviously, we haven’t labeled the data, so our learning algorithm is unsupervised.

We train the network in two stages. First, the initial weights are randomly selected from $(0, 2)$. The loss function is given by the sum of Eq. (12) and Eq. (13). We will adopt the RMSProp optimizer [47]. Second, the initial weights of the DNN will be replaced by the trained weights of the first stage. The loss function is re-set as Eq. (12) without the regularization. Then the network will be trained again with the optimizer Adam [48].

After the training of each stage, one can read the loss, extract the weights, and calculate their error. It can be found that after the second stage of training, the performance of the DNN is usually improved. In particular, the loss (without regularizations) of the first stage can be reduced by several orders of magnitude. In addition, turning the regularization factors in the first stage can improve the performance of the DNN in the second stage. With these in mind, we will scan the parameter space of regularization factors carefully and train the DNN by two stages. At last, the results with minimum loss will be collected. Note that more details of the training scheme will be given in the SM. Importantly, we perform some statistical average to suppress the fluctuation due to the randomized initialization of the network, by which the predictions of the DNN are defined.

In TABLE S.1 of the SM, we list the final training reports after two stages and the statistical average of various numerical experiments in the letter. Among others, it is shown that from the data with $\omega \in [0, 1]$, the Schwarzschild and RN metrics can be learned with high accuracy: the mean relative error (MRE) is around
are generated using the metrics learned from the narrowband data, which are learned from the data with narrow ranges B and C. The left half represents the MRE of the metrics which are learned from the data with narrow ranges B and C. The right half represents the MRE of the wideband data, which are generated using the metrics learned from the narrowband data.

$O(0.1\%)^2$. The target and learned metrics have been plotted in FIG. (3a).

Hence, we almost naively select the frequency range of the data as $\Delta \omega = 1$. One important question in DL is how well does the model generalize? To proceed, we consider different datasets with the narrow frequency range $\Delta \omega = 10^{-2}$ and keep each of them with 2000 data. Interestingly, we find that both Schwarzschild and RN metrics still can be well learned, although the error will increase when the frequency window is close to zero and especially when the charge density is large. This is shown by the MRE of the metrics learned from two typical windows, see the right half of FIG. (3b). Furthermore, it suggests that the generalization ability of the DNN can be excellent. Indeed, in the left half of FIG. (3b), we illustrate that using the metric learned from the data with $\Delta \omega = 10^{-2}$, one can generate the data with $\Delta \omega = 1$ very accurately: in the best performing example, the MRE of the generated data can reach the order of one million! We also note that the examples with relatively large errors in FIG. (3b) can be expected because the DC limit of shear viscosity is determined solely by the physics at horizon, and for the extremal RN black hole with $q = \sqrt{12}$, the IR CFT associated with the AdS$_2 \times \mathbb{R}^2$ geometry dominates the low-frequency physics $^{49, 50}$. Similarly, we do not expect that the DNN can learn well from a very high frequency window, where the UV CFT should dominate.$^3$

Two metric components.— A more general black hole metric has two independent metric components $g_{tt}$ and $g_{rr}$. From Eq. (6), one can find that they appear in the form of the joint factor $g_{rr}/g_{tt}$. Therefore, the DNN can be applied to learn the joint factor but in general each of them cannot be learned separately from the shear response. Nevertheless, if there are other ways to determine one, the other can be obtained by the DNN. For example, there is evidence that the entanglement plays important role in weaving the spacetime $^{51, 52}$. Among others, it has been shown that the holographic entanglement entropy $S(l)$ can be used to fix the bulk metric wherever the extremal surface reaches $^{53}$. Here $l$ is the scale of the boundary entangling region on which the bulk extremal surface is anchored $^{52}$. Under the present metric ansatz, the holographic entanglement entropy only depends on $g_{rr}$, so it can complement to the shear viscosity to determine two metric components.

Conclusion and discussion.—With simple DL, we studied an inverse problem of AdS/CFT: given the complex frequency-dependent shear viscosity of boundary field theories at finite temperatures, whether the metrics of bulk black holes can be extracted? We showed that the Schwarzschild and RN metrics can be learned by the DNN with high accuracy. The network architecture can be taken as a discretized representation of the holographic RG flow of the shear response, hence being a new application of the holographic membrane paradigm. We emphasize that it is universal for any isotropic and translation-invariant field theories which are dual to the $3+1$-dimensional Einstein gravity. The extensions to the symmetry-breaking theories, the higher spacetime dimensions, and the modified theories of gravity should be plausible. Among others, we note that the wave equation of the shear mode has been built up for a general isotropic theory without translational symmetry $^{54}$. The equation involves a radially varying function describing the graviton mass. Accordingly, one can construct the RG flow and the DNN, where the graviton mass is encoded into new trainable weights. It would be interesting to see whether the DNN can learn the metric and the mass simultaneously. In addition to various extensions, there are two open questions which deserve to be mentioned. (i) Is there a better ansatz for the regularization term? Note that the regularization in this

$^2$ The MSE is around $O(10^{-6})$.

$^3$ We have checked that the performance of the DNN is declined when the frequency of the window is high enough. In fact, it has been observed in $^{63, 64}$ that the retarded Green function for the shear stress operator at the infinite frequency is determined by the energy density. We thank Matteo Baggioli for the discussion on this point.
work is not to prevent overfitting as usual in machine learning. Instead, it is a guide to the minimum loss. We might need a deeper physical understanding of the regularization. (ii) How to realize the continuous RG flow by the DNN? Apparently, a direct increase in the number of layers requires very powerful computing capabilities. A more efficient method might be to apply the recently proposed DNN models of ordinary differential equations [58]. We ultimately hope that our work could suggest a data-driven way to study holographic transports.

Moreover, we found that the complete black hole metric from IR to UV can be well learned from the data with narrow frequency ranges. We also have checked that randomly deleting several data points in our numerical experiments will not qualitatively change the performance of the DNN. These two facts indicate that the shear viscosity encodes the spacetime in a very different way from the entanglement entropy. The latter probes shear viscosity encodes the spacetime in a very different way from the entanglement entropy. The latter probes shear viscosity encodes the spacetime in a very different way from the entanglement entropy.

Perhaps we can describe this difference coarsely as follows: the non-local observable (entanglement entropy) on the boundary probes the bulk spacetime locally, while the local observable (shear viscosity) probes non-locally.

Furthermore, this non-locality leads to the excellent generalization ability of the DNN, which should be important in the application to the experimental data that may be collected only in a part of the spectrum. Theoretically, from the perspective of machine learning, it usually implies that the data is highly structured. This structure is often important but obscure due to the black-box problem of machine learning. However, here the structure is nothing but the higher-dimensional black hole spacetime. Also, it suggests that the strongly coupled field theory with gravity dual could exhibit another feature of the hologram in addition to encoding the higher dimension: the local (a small piece of the hologram) can reproduce the whole, see the schematic diagram FIG. 4. for this insight.

Acknowledgments.—We are grateful to Matteo Baggioli, Yongcheng Ding, Wei-Jia Li, Tomi Ohtsuki, and Fu-Wen Shu for helpful discussions. SFW and XHG are supported by NSFC grants No. 11675097 and No. 11875184, respectively. YT is supported partially by NSFC grants No. 11675097 and No. 11875184, and by the “Strategic Priority Research Program of the Chinese Academy of Sciences” with Grant No. XDB23030000.

4 Another possibility is that the network has some symmetry, see [50] for an example.
5 For example, using the generative adversarial network (GAN), the approximate statistical predictions have been made recently in the string theory landscape [50], where the accurate extrapolation capability has been exhibited on simulating Kähler metrics. It was speculated that this is the first evidence of Reid’s fantasy: all Calabi-Yau manifolds with fixed dimension are connected.
FREQUENCY-DEPENDENT SHEAR VISCOSITY

Compared to the shear viscosity at the zero frequency limit, the study of frequency-dependent counterparts is rare. So let’s begin from reviewing the definition of the full complex frequency-dependent viscosity tensor. In [43], the Kubo formulas for the stress-stress response function at zero wavevector is derived from first principles. Then the frequency-dependent viscosity tensor is defined. The approach given in [43] involves the response to a uniform external strain and a microscopic Hamiltonian. In [44], an alternative field-theory approach is proposed, by which the Ward identity of viscosity coefficients in [43] is retrieved and extended. Here we will follow [44] to give the definition of the frequency-dependent shear viscosity.

For a theorist, the viscosity can be measured by sending a gravitational wave through the system [60]. Suppose that the two-dimensional flat space is perturbed by \( \delta g \). Then the subscript \( r \) and the subscript \( a \) below indicate that we have invoked the closed time-path formalism to discuss the right hand of Eq. (S.1) into two parts,

\[
\delta \langle T^{ij}(x) \rangle_r = -\frac{1}{2} \int d^3 x' \lambda^{ijkl}(x-x') \delta g_{rkl}(x') - \frac{1}{2} \int d^3 x' \eta^{ijkl}(x-x') \partial_t \delta g_{rkl}(x').
\]

The stress tensor can be derived by the variation of the generating functional with respect to the metric

\[
\langle T^{ij}(x) \rangle_r = \frac{2}{\sqrt{g}} \frac{\delta W}{\delta g_{aij}(x)}
\]

and the second variation leads to the retarded correlator

\[
G^{ij,kl}_{ra}(x) = \frac{4 \delta W}{\delta g_{aij}(x) \delta g_{rkl}(0)} = \delta^{ij}(T^{ij}) \delta(x) - \lambda^{ijkl}(x) - \partial_t \eta^{ijkl}(x).
\]

The elastic modulus is the stress response up to the zeroth-order in time derivatives, which can be determined by the constitutive relation of perfect fluid. In hydrodynamic expansion, one has

\[
\delta \langle T^{ij}(x) \rangle_r = -P \delta^{ik} \delta^{jl} + \frac{1}{2} \delta^{ij} \delta^{kl} \kappa^{-1} \delta g_{rkl}(x),
\]

where \( P \) is the pressure and \( \kappa^{-1} \) is the inverse compressibility. Then the elastic modulus can be given by

\[
\chi^{ijkl}(x) = \left[ P \left( \delta^{ik} \delta^{jl} + \delta^{il} \delta^{jk} \right) + \delta^{ij} \delta^{kl} \kappa^{-1} \right] \delta(x).
\]

The viscosity tensor can be decomposed as

\[
\eta^{ijkl}(x) = \zeta(x) \delta^{ij} \delta^{kl} + \eta(x) \left( \delta^{ik} \delta^{jl} + \delta^{il} \delta^{jk} - \delta^{ij} \delta^{kl} \right) + \eta^H(x) \left( \delta^{ik} \epsilon^{jl} - \delta^{il} \epsilon^{kj} \right),
\]

where we have assumed that the system is isotropic in two spatial dimensions. The coefficients \( \zeta, \eta, \eta^H \) denote the bulk, shear, and Hall viscosities, respectively. Substituting the last two equations into Eq. (S.4), one can obtain

\[
G^{12,12}_{ra}(x) = -\chi_{1212}(x) - \partial_t \eta_{1212}(x)
= -P \delta(x) - \partial_t \eta(x).
\]

In Fourier space, we have the Kubo formula of the complex frequency-dependent shear viscosity

\[
\eta(\omega) = \frac{G^{12,12}_{ra}(\omega) + P}{i \omega}.
\]

Note that this formula applies to any frequencies. In contrast, the shear viscosity in most literatures is introduced by the constitutive relationship of hydrodynamics, so strictly speaking its Kubo formula is only applicable to low frequencies, although its form happens to be the same as Eq. (S.9).
We proceed to bridge the shear viscosity $\eta(\omega)$ to the shear response $\chi(\omega)$. The essential procedure is to carry out the holographic renormalization [45]. Consider the bulk action of the Einstein-Maxwell theory

$$S_{\text{bulk}} = \int d^4x \sqrt{-g} \left[ R + 6 - \frac{1}{4} F^2 \right],$$  \hspace{1cm} (S.10)

where the AdS radius $L$ and the Newton constant $16\pi G_N$ are set to unity. The Reissner-Nordstrom solution can be written as

$$ds^2 = -r^2 f(r) dt^2 + \frac{1}{r^2 f(r)} dr^2 + r^2 d\Omega^2,$$

$$f(r) = 1 - \frac{r_+}{r} - (1 - \frac{r_+}{r}) \frac{q^2}{4r_+^3},$$  \hspace{1cm} (S.11)

with $A = \frac{q}{r_+}(1 - \frac{r_+}{r}) dt$. We need to write down the Gibbons-Hawking term and the counterterm

$$S_{\text{GH}} = -2 \int d^3x \sqrt{-\gamma} K,$$

$$S_{\text{ct}} = \int d^3x \sqrt{-\gamma} (-4 + R),$$  \hspace{1cm} (S.12, S.13)

where $K$ is the external curvature and $\gamma^{ab}$ is the induced metric. Near the boundary, the wave equation of the shear mode has the asymptotic solution

$$h = h^{(0)} + \frac{1}{r_+^2} h^{(2)} + \frac{1}{r_+^3} h^{(3)} + \cdots.$$  \hspace{1cm} (S.14)

Here $h^{(0)}$ is the source, $h^{(2)}$ is fixed by $h^{(0)}$ as $h^{(2)} = h^{(0)} \omega^2/2$, and $h^{(3)}$ relies on $h^{(0)}$ and the incoming boundary condition at the horizon.

Now we expand the bulk action, the Gibbons-Hawking term and the counterterm to the quadratic order of the shear mode,

$$S_{\text{bulk}}^{(2)} |_{\text{os}} + S_{\text{GH}}^{(2)} + S_{\text{ct}}^{(2)} = \int d^2x \int_{-\infty}^{\infty} d\omega \frac{1}{2} \left[ \left( -\frac{r_+}{\sqrt{T}} + 4r_+^3 \sqrt{T} - 2r_+^3 f - r_+^4 f' \right) \tilde{h} h - r_+^4 f \tilde{h} h' \right],$$  \hspace{1cm} (S.15)

where $\tilde{h}$ has the argument $-\omega$. Substituting the asymptotic solution $\text{(S.14)}$ and the blackening factor $\text{(S.11)}$ into Eq. $\text{(S.15)}$, we obtain the renormalized action:

$$S_{\text{ren}}^{(2)} = \int d^2x \int_{-\infty}^{\infty} d\omega \frac{1}{2} \left( -\frac{1}{2} P h^{(0)} h^{(0)} + \frac{3}{2} h^{(0)} h^{(3)} \right).$$  \hspace{1cm} (S.16)

Here we have used the pressure $P = r_+^3 + \frac{1}{4r_+^2} q^2$. Invoking the holographic dictionary, one can extract the retarded correlator from $S_{\text{ren}}^{(2)}$:

$$G_{t^a t^b}^{12,12}(\omega) = -P + 3 \frac{\tilde{h}^{(3)}}{h^{(0)}}.$$  \hspace{1cm} (S.17)

It would be interesting to express this formula under the Eddington-Finkelstein coordinates. Consider the shear mode $(\delta g)^{x_1}_{x_2} = H(z)e^{-i\omega v}$, where $v = t - z^*$ and $z^* = \int \frac{dz}{f(z)}$. Then we have the asymptotic expansion

$$H = h^{(0)} - i\omega h^{(0)} z + \left( -\frac{1}{3} i\omega^3 h^{(0)} + h^{(3)} \right) z^3 + \cdots = H^{(0)} + z H^{(1)} + z^3 H^{(3)} + \cdots,$$  \hspace{1cm} (S.18)

from which the retarded correlator can be rewritten as

$$G_{t^a t^b}^{12,12}(\omega) = -P + i\omega^3 + 3 \frac{H^{(3)}}{H^{(0)}}.$$  \hspace{1cm} (S.19)
Combining Eq. (S.9) and Eq. (S.17), we obtain

$$\eta(\omega) = \frac{3}{i\omega} \frac{h^{(3)}}{h^{(0)}}. \quad \text{(S.20)}$$

Expand the response function near the boundary

$$\chi \equiv \frac{\Pi}{i\omega h} = -\sqrt{-g^{rr}\partial_r h} = \frac{3}{i\omega} \frac{h^{(3)}}{h^{(0)}} - i\omega r + \cdots. \quad \text{(S.21)}$$

We finally obtain its relation to the shear viscosity

$$\eta(\omega) = \chi(\omega) + i\omega r|_{r \to \infty}. \quad \text{(S.22)}$$

Note that using $s = 4\pi$ and $\chi(0) = 1$, one can recover the well known result: $\eta(0)/s = 1/(4\pi)$.

## TRAINING SCHEME AND REPORT

For all numerical experiments in this letter, we implement the same training scheme. The scheme has been introduced in the main text and here we add some details of turning hyperparameters. In two stages of training, we fix the batch size $c_{\text{bs}} = 512$, but the learning rate is changed even not once and selected by experience. The number of epoches in each stage is large enough so that the training will not stop until the validation loss is almost not reduced. The regularization factor $c_3$ is set as $15c_1$. We focus on turning the regularization factors $c_1$ and $c_2$ because the performance of the DNN is more sensitive to them than other hyperparameters. Initially, $c_1$ and $c_2$ are limited to some suitable ranges. Then we scan the two-dimensional parameter space. Considering the statistical fluctuation due to the randomized initialization of the network, we train 5 times for each set of regularization factors [67]. We gradually move and reduce the ranges of parameters. We also gradually reduce the step sizes. Thus, the scanning is concentrated around the $c_1$ and $c_2$ where the DNN produces smaller losses. We stop the scanning when the step sizes $\Delta c_1 < 10^{-5}$ and $\Delta c_2 < 0.01$. The optimal regularization factors are selected as the ones according to the minimum loss. In order to suppress the impact of random initialization on the performance of the DNN, we perform the statistical average as follows: train the DNN with the optimal $c_1$ and $c_2$ 100 times and average the top 5 results with the minimum loss. We read the average minimum loss and the average weights of the network, and then calculate the MRE of the average weights and the MRE of the wideband data generated by the average weights learned from narrowband data. These quantities are considered as the final performance of the DNN. We simply refer them as the "minimum loss", the "learned metric", the "MRE of learned metrics", and the "MRE of generated data", respectively. They have been exhibited in FIG. 3 of the main text and in TABLE S.1.

### TABLE S.1. Final training reports of various numerical experiments. We have three target metrics and each of them is learned from three datasets: A $\omega \in [0, 1]$, B $\omega \in [0.99, 1]$, and C $\omega \in [0, 0.01]$. We list two optimal regularization factors and three quantities which characterize the performance of the DNN.

| Target     | Data | Optimal $c_1$ | Optimal $c_2$ | Minimum loss | MRE of learned metrics | MRE of generated data |
|------------|------|---------------|---------------|--------------|------------------------|----------------------|
| Schwarzschild | A    | $2.6 \times 10^{-4}$ | 1.35 | $8 \times 10^{-13}$ | $1 \times 10^{-3}$ | /                     |
| Schwarzschild | B    | $1.0 \times 10^{-4}$ | 1.59 | $7 \times 10^{-13}$ | $3 \times 10^{-3}$ | $2 \times 10^{-6}$   |
| Schwarzschild | C    | $9.0 \times 10^{-5}$ | 0.91 | $4 \times 10^{-14}$ | $3 \times 10^{-2}$ | $1 \times 10^{-4}$   |
| RN q=2     | A    | $6.0 \times 10^{-5}$ | 0.50 | $1 \times 10^{-11}$ | $4 \times 10^{-3}$ | /                     |
| RN q=2     | B    | $1.0 \times 10^{-4}$ | 1.26 | $3 \times 10^{-11}$ | $8 \times 10^{-3}$ | $9 \times 10^{-6}$   |
| RN q=2     | C    | $1.4 \times 10^{-4}$ | 0.33 | $5 \times 10^{-14}$ | $6 \times 10^{-2}$ | $7 \times 10^{-4}$   |
| RN q=3     | A    | $4.6 \times 10^{-4}$ | 0.60 | $4 \times 10^{-12}$ | $1 \times 10^{-3}$ | /                     |
| RN q=3     | B    | $1.7 \times 10^{-3}$ | 0.95 | $2 \times 10^{-10}$ | $6 \times 10^{-3}$ | $6 \times 10^{-5}$   |
| RN q=3     | C    | $4.5 \times 10^{-4}$ | -1.39 | $1 \times 10^{-13}$ | $2 \times 10^{-1}$ | $2 \times 10^{-3}$   |