Heavy quark potential in the quark-gluon plasma: Deep neural network meets lattice quantum chromodynamics

Shuzhe Shi,[1,*] Kai Zhou,[2,†] Jiaxing Zhao,[3] Swagato Mukherjee,[4] and Pengfei Zhuang[3]

[1]*Department of Physics, McGill University, Montreal, Quebec H3A 2T8, Canada*
[2]*Frankfurt Institute for Advanced Studies, Ruth Moufang Strasse 1, D-60438,*
Frankfurt am Main, Germany
[3]*Department of Physics, Tsinghua University, Beijing 100084, China*
[4]*Physics Department, Brookhaven National Laboratory, Upton, New York 11973, USA*

Bottomonium states are key probes for experimental studies of the quark-gluon plasma (QGP) created in high-energy nuclear collisions. Theoretical models of bottomonium productions in high-energy nuclear collisions rely on the in-medium interactions between the bottom and antibottom quarks. The latter can be characterized by the temperature (T) dependent potential, with real ($V_R(T, r)$) and imaginary ($V_I(T, r)$) parts, as a function of the spatial separation (r). Recently, the masses and thermal widths of up to $3S$ and $2P$ bottomonium states in QGP were calculated using lattice quantum chromodynamics (LQCD). Starting from these LQCD results and through a novel application of deep neural network, here, we obtain $V_R(T, r)$ and $V_I(T, r)$ in a model-independent fashion. The temperature dependence of $V_R(T, r)$ was found to be very mild between $T \approx$ 0–334 MeV. For $T =$ 151–334 MeV, $V_I(T, r)$ shows a rapid increase with T and r, which is much larger than the perturbation-theory-based expectations.

I. INTRODUCTION

In-medium modifications of quarkonium states, i.e., bound states of a heavy charm or bottom quark and its antiquark, are sensitive probes of the quark-gluon plasma (QGP) produced in high-energy nuclear collisions [1–13]. Sequential suppression patterns among the $\Upsilon(1S)$, $\Upsilon(2S)$, and $\Upsilon(3S)$ states have been observed in heavy ion collision experiments [14–17]. Theoretical understanding of these experimental observations relies on effective field theories (EFTs), which naturally lead to an open-quantum-system-based treatment of both open and hidden bottom states in QGP (for a recent review, see Ref. [12]). Owing to the large mass (m_b) and small relative velocity (v) of the bottom quark, there exists a hierarchy of scales at high temperature: $m_b \gg m_b v \gg m_b v^2$. Sequentially integrating out the scales larger than m_b and $m_b v$ from the QCD Lagrangian, one, respectively, arrives at the nonrelativistic QCD (NRQCD) [18] and potential nonrelativistic QCD (pNRQCD) [19] EFTs. If interactions between the color-singlet and color-octet states are neglected then the pNRQCD reduces to a theoretical description of quarkonia solely based on a potential between the heavy quark and antiquark. A potential-based description allows studies of quarkonia by employing Schrödinger-type equations [20–23]. One-loop hard thermal loop (HTL) perturbative QCD calculations [24,25], and later on pNRQCD calculations [26,27], show that at finite temperatures heavy quark potential becomes complex with a nonvanishing imaginary part. However, it is difficult to provide satisfactory descriptions of bound states arising out of strong interactions solely using perturbative expansions, and a nonperturbative treatment, such as the lattice quantum chromodynamics (LQCD), is called for.

A bound state of strong interaction is a nonperturbative problem, which is difficult to be completely or relevantly treated in a conventional perturbation theory. Therefore, a critical step in all EFT-based studies is to relate parameters of the EFT to the underlying fundamental theory, i.e., a model-independent determination of the heavy quark potential starting from nonperturbative QCD. In the static limit, the heavy quark potential can be extracted from the spectral functions of the thermal Wilson loop using nonperturbative LQCD calculations [28–31]. On the other hand, recent LQCD studies have led to quantification of the masses, thermal widths, and Bethe–Salpeter amplitudes (BSA) of up to $3S$ and $2P$ bottomonium states in QGP [32–34]. While the lattice QCD results of Refs. [28–31]

*shuzhe.shi@mcgill.ca
†zhou@fias.uni-frankfurt.de

were obtained based on thermal correlation functions that provide a well-defined static quark potential in some specific limits, *a priori*, there is no obvious reason to expect that the LQCD results [32–34] on the properties of in-medium bottomonia admit any consistent in-medium potential-based description. Even the fact that the vacuum bottomonia masses below the threshold can be reasonably well-described by Cornell-type potential is an empirical observation, and cannot be rigorously proven from first-principle QCD. In this work we will empirically investigate whether the LQCD results of Refs. [32–34] can be consistently described by an in-medium potential, $V_R(T, r)$ and $V_I(T, r)$. Furthermore, even if it turns out that LQCD results of Refs. [32–34] can be consistently described using some $V_R(T, r)$ and $V_I(T, r)$, there is no theoretical reason for these to agree with the static quark potential obtained from correlation functions of thermal Wilson loops [28–31]. Comparisons among these different in-medium potentials is an interesting study by itself.

As we shall see later, one-loop HTL-motivated functional forms of $V_R(T, r)$ and $V_I(T, r)$ are not compatible with these LQCD results. This observation calls for a model-independent treatment of the in-medium heavy quark potential. In this work, we introduce a model-independent deep-neural-network-based (DNN-based) method and determine the r and T-dependence of the in-medium heavy quark potential starting from the LQCD results [33] for the masses and thermal widths of up to $3S$ and $2P$ bottomonium states at various temperatures. The underlying idea is as follows: At a fixed T, various bottomonium states differ in sizes and their wave functions probe different distances. Knowledge of the masses and thermal widths of multiple bottomonium states thereby provide constraints on not only the strength of the real and imaginary parts of the bottom-antibottom interactions in QGP but also its r-dependence. Thus, LQCD results for the masses and thermal widths of multiple bottomonium states at different temperatures can be used to extract $V_R(T, r)$ and $V_I(T, r)$ and, presently, a DNN is probably the best tool to achieve this in an unbiased fashion. To this goal, we develop a new method to optimize the deep neural network coupled with the Schrödinger equation.

This manuscript is organized as follows. We discuss the details of vacuum potential determination in Sec. II, followed by a detailed description of the potential extraction method in Sec. III. Then, we show our results of complex-valued heavy flavor potential using DNNs that depend on both distance and temperature in Sec. IV. Finally we perform two consistency tests: (i) we compare the DNN potential to the potentials using two other parametrization schemes—temperature-independent DNNs and polynomials (Sec. V A); (ii) we compare the eigenstate wave functions with the corresponding lattice QCD results [34] of Bethe–Salpeter amplitude (Sec. V B). After the summary in Sec. VI, we provide Supplemental Materials (Appendix)

to discuss the connection between the imaginary energy and the width extracted in lattice QCD.

II. SCHRÖDINGER EQUATION AND VACUUM POTENTIAL

Bottomonium states can be described well by the reduced two-body time-independent Schrödinger equation, [35]

$$-\frac{\nabla^2}{m_b}\psi_n + [V_R(T, r) + iV_I(T, r)]\psi_n = E_n\psi_n. \quad (1)$$

Here, the heavy quark potential $V(T, r) = V_R(T, r) + iV_I(T, r)$ is complex-valued. Accordingly, the wave function ψ_n and the energy eigenvalues E_n for in-medium bottomonia are also complex-valued. Further, $V_I(T = 0, r) = 0$, $\mathrm{Re}[E_n] = m_n - 2m_b$, and $\mathrm{Im}[E_n] = -\Gamma_n$, where m_n and Γ_n are the mass and thermal width of the nth bottomonium state, respectively (see the Appendix for detailed discussions). Vacuum properties of up to $3S$ and $2P$ bottomonium states [36] were found to be reproduced well by the Cornell potential

$$V_R(T = 0, r) = -\frac{\alpha}{r} + \sigma r + B, \quad (2)$$

with $m_b = 6.00$ GeV, strong coupling $\alpha = 0.406$, string tension $\sigma = 0.221$ GeV^2, and bag constant $B = -2.53$ GeV. These parameters are determined by fitting the vacuum masses of the bottomonium states reported in the Particle Data Group booklet [36] and the wave function, to be compared with the BSA from lattice QCD calculations [34]. The latter is computed in a consistent way as to evaluate the mass and width [33]. Due to its r-dependence, the BSAs are more decisive, compared to the mass spectrum, in determining the interaction potential. Considering that the lattice results are for spin-averaged states—e.g., there is no
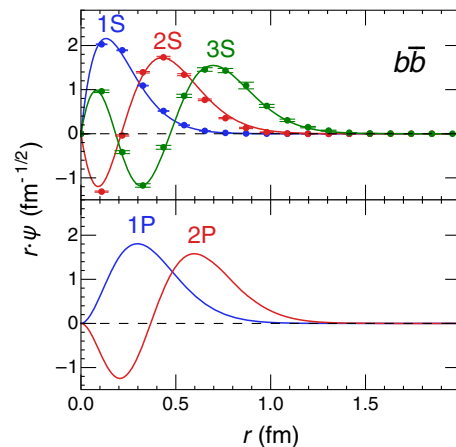

FIG. 1. Bottomonia wave functions in vacuum. Symbols in the upper panel are the Bethe–Salpeter amplitudes from lattice QCD calculation [34].

TABLE I. Best fit of the spin-averaged bottomonium mass spectrum (3).

	1S	2S	3S	1P	2P
Experiment (MeV)	9445	10017	10352	9891	10254
Model (MeV)	9449	10003	10356	9893	10258
Difference (MeV)	+4	−14	+4	+2	+4

distinction between $\eta_b(nS)$ and $\Upsilon(nS)$—we fit the parameters in a potential model by fitting the spin-averaged bottomonia mass spectrum

$$m_{\text{averaged}} \equiv \frac{1}{4}m_{\text{singlet}} + \frac{3}{4}m_{\text{triplet}}, \qquad (3)$$

as well as the BSA for 1S, 2S, and 3S states. The global parameter tuning technique employed here will be described later in Sec. III C. We present the bottomonium wave functions in Fig. 1, and correspondingly the mass in Table I. One can see the potential model agrees very well with the experimental results (with the maximum absolute difference being 14 MeV), as well as with the lattice results on the wave functions (BSAs).

III. METHODOLOGY: HOW A DNN LEARNS THE POTENTIAL FROM THE SPECTRUM

We will move on to discuss our model-independent methodology of the potential reconstruction with DNNs coupled to a Schrödinger equation (1), which is summarized in the flow chart of Fig. 2. We represent the real and imaginary potentials by the DNN—a multistep iterative function composition scheme to approximate mapping between two functions in a smooth and unbiased manner [37,38]. We optimize the network parameters by

minimizing the chi-square function, i.e., the uncertainty-weighted distance,

$$\chi^2 = \sum_{T,n} \frac{(m_{T,n} - m_{T,n}^{\text{LQCD}})^2}{(\delta m_{T,n}^{\text{LQCD}})^2} + \frac{(\Gamma_{T,n} - \Gamma_{T,n}^{\text{LQCD}})^2}{(\delta\Gamma_{T,n}^{\text{LQCD}})^2}. \quad (4)$$

The summation runs over six temperature points, $T \in \{0, 151, 173, 199, 251, 334\}$ MeV, and five bottomonium states, $n \in \{1S, 2S, 3S, 1P, 2P\}$, and the LQCD values were taken from Ref. [33]. We used gradient descent with the back-propagation optimization technique, which is based on the derivatives of the cost function with respect to the network parameters. This technique requires knowledge of the explicit functional relationship between the cost function and the DNN output. However, such a relationship in our problem is implicit. We overcame this challenge of gradient evaluation through the perturbative solution of the Schrödinger equation with respect to a small change of $V(T, r)$. Moreover, we invoked Bayesian inference for the uncertainty quantification, whereby the posterior distribution of the network parameters was evaluated. To the best of our knowledge, the current method is developed for the first time here. In the rest of this section, we provide all the details about this method, including a general introduction to DNN (Sec. III A), the parameter optimization algorithm (Secs. III B and III C), uncertainty estimation using Bayesian Inference (Sec. III D), and a closure test to justify our methodology and assess its reliability (Sec. III E).

A. General introduction of deep neural network

According to the universal approximation theorem [37,38], a DNN can generally provide an unbiased, yet flexible enough, parametrization to approximate arbitrary
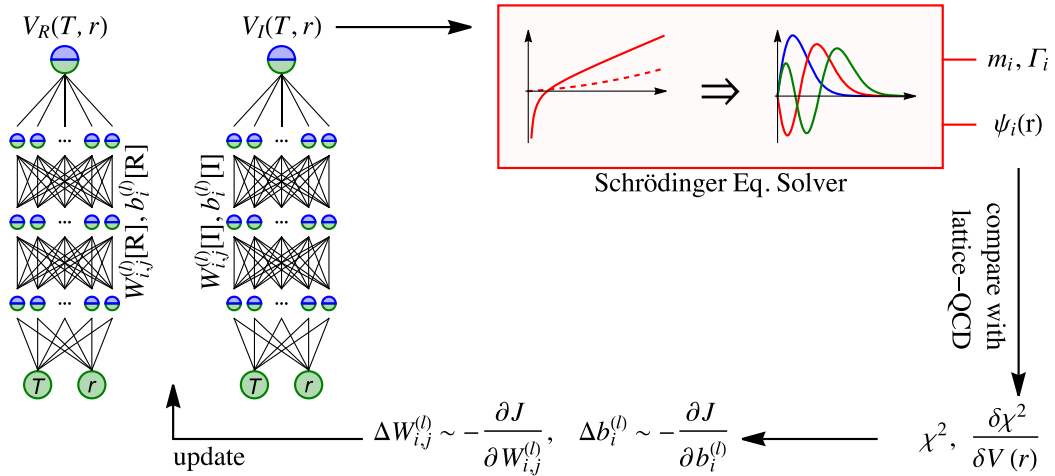

FIG. 2. Flow chart of the potential reconstruction scheme—using generalized back-propagation to optimize parameters in the deep neural networks coupled with a Schrödinger equation.

TABLE II. Common choices of activation functions.

Name	Description	$\sigma(z)$
linear	Linear	z,
tanh	Hyperbolic tangent	$\tanh(z)$,
relu	Rectified linear units	$\max(0, z)$,
elu	Exponential linear units	$\begin{cases} \exp(z) - 1 & z < 0, \\ z & z \geq 0. \end{cases}$

functional relations. Algorithms based on DNNs have been deployed to address various physics problems, e.g., determining the parton distribution function [39,40], reconstructing the spectral function [41–43], identifying phase transition [44–47], assisting lattice field theory calculation [48–51], evaluating centrality for heavy ion collisions [52–54], estimating parameters under detector effects [55,56], and speeding up hydrodynamics simulation [57].

To express, approximately, an $\mathbb{R}_n \to \mathbb{R}_m$ function mapping between independent variables $\mathbf{x} = \{x_1, \ldots, x_n\}$ and dependent variables $\mathbf{y} = \{y_1, \ldots, y_m\}$, $\mathbf{y} = \mathbf{y}(\mathbf{x})$, the DNN constructs a functional form by composing iteratively N simple building blocks (also called a *layer* representing a vector-to-vector function). Each layer performs a linear transformation on the output from the preceding layer, followed by an elementwise nonlinear transformation dictated by the *activation function* $\sigma^{(l)}(z)$:

$$a_i^{(l)} = \sigma^{(l)}(z_i^{(l)}), \qquad z_i^{(l)} \equiv b_i^{(l)} + \sum_j W_{ij}^{(l)} a_j^{(l-1)}, \quad (5)$$

for $i = 1, \ldots, n^{(l)}$ and $l = 1, \ldots, N$, while $a_i^{(0)} \equiv x_i$ stands for the input variables. The final layer gives the model output, which composes all the successive layers and defines the approximate function formula, $\tilde{\mathbf{y}}(\mathbf{x}; \{W_{ij}^{(l)}, b_i^{(l)}\}) = \mathbf{a}^{(N)}$. This compositional way of parametrization renders the DNN a universal function approximator, able to fit any continuous function to an arbitrary accuracy given enough hidden units.

In the language of deep learning, N is referred to as the *depth* of the network, $\mathbf{x}$ the *input layer*, $\mathbf{y}$ the *output layer*, $\mathbf{a}^{(l)}$ the lth layer, and $n^{(l)}$ the *width* of the lth layer. The intermediate layers with $1 \leq l < N$ are called *hidden layers*. The activation functions are nonlinear functions that modulate the function behavior. We list the common choices of activation functions in Table II. The choice of N, $n^{(l)}$, and $\sigma^{(l)}$ are *hyperparameters* of the model. Meanwhile, $W_{ij}^{(l)}$ and $b_i^{(l)}$ are respectively called *weights* and *biases*. They are parameters to be tuned during training by minimizing the *cost function*, which characterize the distance between the approximation formula $\tilde{\mathbf{y}}(\mathbf{x})$ and the corresponding true function $\mathbf{y}(\mathbf{x})$. The process of parameter optimization is called *model training* in deep learning.

The power of the DNN comes from its advanced parameter training method, called *gradient descent* via *back-propagation*. It updates the parameters according to the gradient of the cost function:

$$\Delta\boldsymbol{\theta} \equiv \boldsymbol{\theta}^{[k+1]} - \boldsymbol{\theta}^{[k]} \sim -\boldsymbol{\nabla}_{\boldsymbol{\theta}} J(\boldsymbol{\theta}), \qquad (6)$$

where $\boldsymbol{\theta}$ is the abbreviation of all the parameters, i.e., $\boldsymbol{\theta} \equiv \{W_{ij}^{(l)}, b_i^{(l)}\}$, and the superscript $[k]$ stands for the kth training step. While different optimization schemes take different exact relations between $\Delta\boldsymbol{\theta}$ and $\boldsymbol{\nabla}_{\boldsymbol{\theta}} J$, there is a common feature—noting that $\boldsymbol{\nabla}_{\boldsymbol{\theta}} J = 0$ when the cost function reaches its minimum, the training iterations eventually stop when closing to such a point. In this work, we adopt the Adam optimization method [58], an acceleration method based on gradient descent.

In the most typical case for regression problems, the cost function is defined as the summation of the mean-square-error and a regularizer, where the latter is introduced to avoid overfitting manifested with unreasonably large values of parameters:

$$J(\boldsymbol{\theta}) = \frac{1}{2} \sum_{\mathbf{x} \in \text{data set}} |\tilde{\mathbf{y}}(\boldsymbol{\theta}, \mathbf{x}) - \mathbf{y}(\mathbf{x})|^2 + \frac{\lambda}{2} \boldsymbol{\theta} \cdot \boldsymbol{\theta}, \qquad (7)$$

and

$$\frac{\partial J}{\partial \theta_i} = \sum_{\mathbf{x} \in \text{data set}} (\tilde{\mathbf{y}}(\boldsymbol{\theta}, \mathbf{x}) - \mathbf{y}(\mathbf{x})) \cdot \frac{\partial \tilde{\mathbf{y}}(\boldsymbol{\theta}, \mathbf{x})}{\partial \theta_i} + \lambda \theta_i, \quad (8)$$

where $(\cdot)$ is the inner product of all dimensions of $\mathbf{y}$.

The calculation of $\boldsymbol{\nabla}_{\boldsymbol{\theta}} J$ can be computationally very expensive in general, or even undoable, for arbitrary parametrized functions. However, computing $\boldsymbol{\nabla}_{\boldsymbol{\theta}} J$ is straightforward and efficient for the DNN via the back-propagation algorithm, thanks to its simple functional building block (5). This is one of the major advantages of DNNs. We denote that

$$u_i^{(l)} = \sigma'^{(l)}(z_i^{(l)}), \qquad z_i^{(l)} \equiv b_i^{(l)} + \sum_j W_{ij}^{(l)} a_j^{(l-1)}, \quad (9)$$

with $\sigma'(z) \equiv \frac{d\sigma(z)}{dz}$ to be the derivative of the activation function with respect to its argument. From Eq. (5) one can obtain the derivatives, for any input data point $\mathbf{x}$, as

$$\frac{\partial a_i^{(l)}}{\partial b_i^{(l)}} = u_i^{(l)}, \qquad (10)$$

$$\frac{\partial a_i^{(l)}}{\partial W_{i,j}^{(l)}} = a_j^{(l-1)} u_i^{(l)}, \qquad (11)$$

$$\frac{\partial a_i^{(l)}}{\partial a_j^{(l-1)}} = W_{i,j}^{(l)} u_i^{(l)}. \qquad (12)$$

Then, the derivative at each layer can be calculated using the back-propagation iterations:

$$\frac{\partial y_k}{\partial b_i^{(l)}} = \frac{\partial y_k}{\partial a_i^{(l)}} u_i^{(l)}, \tag{13}$$

$$\frac{\partial y_k}{\partial W_{i,j}^{(l)}} = a_j^{(l-1)} \frac{\partial y_k}{\partial b_i^{(l)}}, \tag{14}$$

$$\frac{\partial y_k}{\partial a_i^{(l-1)}} = \sum_j W_{j,i}^{(l)} \frac{\partial y_k}{\partial b_j^{(l)}}, \tag{15}$$

where the iteration begins by $\partial y_k/\partial a_i^{(N+1)} = \delta_{i,k}$. With these, one can obtain the $\boldsymbol{\nabla}_\theta J$ and then update the parameters accordingly.

Model Setup—In practice, we employed two four-layered networks to represent $V_R(T,r)$ and $V_I(T,r)$. Each network contains a two-dimensional (2D) input layer, $\boldsymbol{a}^{(0)} = \{T, r\}$, and a one-dimensional output layer, $a^{(4)} = V_{R/I}$. The intermediate hidden layers $\boldsymbol{a}^{(1)}, \dots, \boldsymbol{a}^{(3)}$ were chosen to consist of $\{32, 16, 32\}$ and $\{16, 16, 16\}$ neurons for the networks corresponding to $V_R(T,r)$ and $V_I(T,r)$, respectively. We adopted the elu activation function, i.e., $\sigma(z) = \exp(x) - 1$ for $z < 0$ and $\sigma(z) = z$ for $z \geq 0$, for all the hidden layers, and a linear identity function, i.e., $\sigma(z) = z$, in the output layer. With $V_R(T,r)$ and $V_I(T,r)$ as inputs, represented by the DNNs described above, we solved Eq. (1) to obtain the masses and thermal widths. We take the regularizers to be $\lambda_R = 10^{-3}$ and $\lambda_I = 10^{-2}$. By dividing or multiplying λ by a factor of two, we have tested that the results are insensitive to the choice of λ.

B. Back-propagation for DNNs coupled with Schrödinger equations

In this work, however, we employ the DNNs to approximate the functional relation between (T, r), as the input layer, and (V_R, V_I), as the output layer, without knowing the "true" values of V_R and V_I to train the parameters. Instead, we further invoke the Schrödinger equation solver to convert the DNN-constructed potentials, $V_R(T,r)$ and $V_I(T,r)$, into the corresponding mass and width of different bound states, considering their availability from lattice QCD. The cost function is set to be

$$J(\boldsymbol{\theta}) = \frac{1}{2}\chi^2(\boldsymbol{\theta}) + \frac{\lambda}{2}\boldsymbol{\theta} \cdot \boldsymbol{\theta} \tag{16}$$

to train the parameters of the DNNs, where the chi-square function χ^2 is the uncertainty-weighted summation of the squared difference between mass and width from the Schrödinger equation, $m_{T,i}$ and $\Gamma_{T,i}$, and those from lattice

QCD, $m_{T,i}^{\text{LQCD}}$ and $\Gamma_{T,i}^{\text{LQCD}}$. In the most generic form, χ^2 can be expressed as

$$\chi^2 = \sum_{T,i,j}(R_{ij}^{(T)}\Delta m_{T,i}\Delta m_{T,j} + I_{ij}^{(T)}\Delta\Gamma_{T,i}\Delta\Gamma_{T,j}$$
$$+ 2M_{ij}^{(T)}\Delta m_{T,i}\Delta\Gamma_{T,j}), \tag{17}$$

where $\Delta m_{T,i} \equiv m_{T,i} - m_{T,i}^{\text{LQCD}}$ is the difference between the potential model and the lattice result for the mass of the ith state at temperature T, and likewise for the width $\Delta\Gamma_{T,i} \equiv \Gamma_{T,i} - \Gamma_{T,i}^{\text{LQCD}}$. R_{ij}, I_{ij}, and M_{ij} are the symmetric covariance matrices. In this work, we neglect the correlation between different quantities, hence $R_{ij}^{(T)} = (\delta m_{T,i}^{\text{LQCD}})^{-2} \cdot \delta_{ij}$, $I_{ij}^{(T)} = (\delta\Gamma_{T,i}^{\text{LQCD}})^{-2} \cdot \delta_{ij}$, and $M_{ij}^{(T)} = 0$.

Computing the parameter gradient is generally complicated if one is not able to find the explicit functional form between $\boldsymbol{\theta}$ and the cost function. In this system, however, the gradient $\boldsymbol{\nabla}_\theta\chi^2$ can be computed explicitly via perturbation treatment on the Schrödinger equation. One can express the eigenvalue problems, before and after a perturbation of the potential, respectively, as

$$\left(\frac{\hat{p}^2}{2m} + V(r)\right)|\psi_i\rangle = E_i|\psi_i\rangle,$$
$$\left(\frac{\hat{p}^2}{2m} + V(r) + \delta V(r)\right)|\psi_i'\rangle = (E_i + \delta E_i)|\psi_i'\rangle. \tag{18}$$

Up to the leading order in δV expansion, perturbation theory yields that

$$\delta E_i = \langle\psi_i|\delta V(r)|\psi_i\rangle \tag{19}$$

and

$$|\psi_i'\rangle = |\psi_i\rangle + \sum_{j\neq i}\frac{\langle\psi_j|\delta V(r)|\psi_i\rangle}{E_i - E_j}|\psi_j\rangle. \tag{20}$$

The former relation is also referred to as the Hellmann–Feynman theorem. Noting that both E_i and $V(r)$ can be complex, we separate the real and imaginary parts:

$$\delta m_i = \langle\psi_i|\delta V_R(r)|\psi_i\rangle,$$
$$\delta\Gamma_i = -\langle\psi_i|\delta V_I(r)|\psi_i\rangle. \tag{21}$$

In particular, for local perturbations

$$\delta V(r) = v\delta(r - r_k), \tag{22}$$

one can obtain

$$\frac{\delta E_i}{\delta v} = |\psi_i(r_k)|^2, \tag{23}$$

which leads to the functional derivative of complex eigenvalues with respect to the complex potential,

$$\frac{\delta m_i}{\delta V_R(r)} = -\frac{\delta \Gamma_i}{\delta V_I(r)} = |\psi_i(r)|^2,$$

$$\frac{\delta m_i}{\delta V_I(r)} = \frac{\delta \Gamma_i}{\delta V_R(r)} = 0. \quad (24)$$

With such relations, we obtain the gradients of the χ^2,

$$\frac{\partial \chi^2}{\partial \theta_{R,n}} = \sum_{T,i,k} \frac{\partial \chi^2}{\partial m_{T,i}} \frac{\partial V_R(T, r_k)}{\partial \theta_{R,n}} |\psi_i(T, r_k)|^2 dr,$$

$$\frac{\partial \chi^2}{\partial \theta_{I,n}} = -\sum_{T,i,k} \frac{\partial \chi^2}{\partial \Gamma_{T,i}} \frac{\partial V_I(T, r_k)}{\partial \theta_{I,n}} |\psi_i(T, r_k)|^2 dr, \quad (25)$$

and of the cost function,

$$\frac{\partial J}{\partial \theta_{R,n}} = \sum_{T,i} \left\{ \left[\sum_k \frac{\partial V_R(T, r_k)}{\partial \theta_{R,n}} |\psi_i(T, r_k)|^2 dr \right] \right.$$

$$\left. \times \sum_j [R_{i,j}^{(T)} \Delta m_{T,j} + M_{ij}^{(T)} \Delta \Gamma_{T,j}] \right\} + \lambda \theta_{R,n},$$

$$\frac{\partial J}{\partial \theta_{I,n}} = -\sum_{T,i} \left\{ \left[\sum_k \frac{\partial V_I(T, r_k)}{\partial \theta_{I,n}} |\psi_i(T, r_k)|^2 dr \right] \right.$$

$$\left. \times \sum_j [I_{i,j}^{(T)} \Delta \Gamma_{T,j} + M_{ij}^{(T)} \Delta m_{T,j}] \right\} + \lambda \theta_{I,n}, \quad (26)$$

where dr is the discrete step size in the distance r. Eventually, we develop the back-propagation scheme for the DNNs coupled with a Schrödinger equation.

While a DNN here can be viewed to provide an unbiased and robust special parametrization for the potentials, one could in principle take any other arbitrary parametrization scheme. The above perturbative analyses for the gradient evaluation could hold for an arbitrary parametrized form of potentials. Suppose the potentials are functions of parameter $\boldsymbol{\theta}$, $V_{R/I}(\boldsymbol{\theta}; T, r)$; then Eq. (26) would remain valid.

C. Fitting quark mass and vacuum potential

The b-quark mass and bottomonia vacuum potential are determined by fitting both the experimental results on the bottomonia masses and their Bethe–Salpeter amplitudes from the lattice calculation [34]. To match the lattice result of both Bethe–Salpeter amplitudes and the mass spectrum, one can set the loss function as

$$J = \frac{\mu}{2} \sum_i w_i (m_i - m_i^{\exp})^2$$

$$+ \frac{\nu}{2} \sum_{i,j} w_{ij} (\psi_i(r_j) - \psi_i^{BS}(r_j))^2, \quad (27)$$

where local weights w_i and w_{ij} account for the data uncertainties, and where μ and ν set the relative weight between the mass difference and the wave function difference in the fitting. In practice, we employ a global weighting with $w_i = (10 \text{ MeV})^{-2}$ and $w_{ij} = (10 \text{ MeV})^{-1}$, and take $\mu = \nu = 1$. From the perturbation theory we find the derivative of the loss function with respect to a potential parameter θ_n:

$$\frac{\partial J}{\partial \theta_n} = \mu \sum_i w_i (m_i - m_i^{\exp}) \sum_k \frac{\partial V(r_k)}{\partial \theta_n} |\psi_i(r_k)|^2 dr$$

$$+ \nu \sum_{i,j} \left[w_{ij}(\psi_i(r_j) - \psi_i^{BS}(r_j)) \right.$$

$$\left. \times \sum_{i' \neq i} \sum_k \frac{\partial V(r_k)}{\partial \theta_n} \frac{\psi_i(r_k)\psi_{i'}(r_k)}{m_i - m_{i'}} \psi_{i'}(r_j) dr \right]. \quad (28)$$

Similarly, one can analyze the linear response against a perturbation in m_b as

$$\left(\frac{\hat{p}^2}{m_b} + V(r) \right) |\psi_i\rangle = E_i |\psi_i\rangle,$$

$$\left(\frac{\hat{p}^2}{m_b + \delta m_b} + V(r) \right) |\psi_i'\rangle = (E_i + \delta E_i)|\psi_i'\rangle, \quad (29)$$

which leads to

$$\delta E_i = \frac{\delta m_b}{m_b} \left(\int V(r)|\psi_i(r)|^2 dr - E_i \right), \quad (30)$$

$$|\psi_i'\rangle = |\psi_i\rangle + \frac{\delta m_b}{m_b} \sum_{j \neq i} \frac{\langle \psi_j | V(r) | \psi_i \rangle}{E_i - E_j} |\psi_j\rangle, \quad (31)$$

and further arrives at the m_b-derivative of the loss function,

$$\frac{\partial J}{\partial m_b} = \frac{\mu}{m_b} \sum_i \left[w_i (m_i - m_i^{\exp}) \right.$$

$$\left. \times \left(4m_b - m_i + \sum_k V(r_k)|\psi_i(r_k)|^2 dr \right) \right]$$

$$+ \frac{\nu}{m_b} \sum_{i,j} \left\{ w_{ij} \sum_{i' \neq i} \left[\frac{\sum_k V(r_k)\psi_i(r_k)\psi_{i'}(r_k) dr}{m_i - m_{i'}} \right. \right.$$

$$\left. \left. \times \psi_{i'}(r_j)(\psi_i(r_j) - \psi_i^{BS}(r_j)) \right] \right\}. \quad (32)$$

We note that the wave-function-driven potential extraction method is also discussed in Ref. [59].

D. Uncertainty quantification with Bayesian inference

We invoke the Bayesian inference to estimate the uncertainties of the DNN reconstructed potentials.

Bayesian inference is a statistical paradigm that utilizes the (experimental) data to constrain model parameters using probability statements. Based on the Bayes' theorem, the posterior distribution over the model parameters (conditional on the observed data) is proportional to the product of the likelihood given the observed data and the prior distribution of the parameters,

$$\text{Posterior}(\boldsymbol{\theta}|\text{data}) \propto L(\boldsymbol{\theta}|\text{data}) \times \text{Prior}(\boldsymbol{\theta}). \quad (33)$$

The likelihood function of the parameters given the observed data specifies the chance that those data appear under the model with the taken parameters, which, due to the central limit theorem, can be expressed in Gaussian form with the chi-square shown naturally:

$$L(\boldsymbol{\theta}|\text{data}) = P(\text{data}|\boldsymbol{\theta}) \propto \exp[-\chi^2(\boldsymbol{\theta})/2]. \quad (34)$$

The prior distribution, in principle, reflects our naive belief in the model parameters, while in practice we take a Gaussian prior distribution accounting for the quadratic regularizers introduced in the cost (16),

$$\text{Prior}(\boldsymbol{\theta}) \propto \exp\left[-\frac{\lambda}{2}\boldsymbol{\theta}\cdot\boldsymbol{\theta}\right], \quad (35)$$

which represents our relative "ignorance" about $\boldsymbol{\theta}$ and also acts as a nonlocal regulator to account for correlated distributions of the to-be-determined target (potential) values in the language of Bayesian statistics.

With the above we thus obtain the posterior distribution over the parameters to be

$$\text{Posterior}(\boldsymbol{\theta}|\text{data}) = N_0 \exp\left[-\frac{\chi^2(\boldsymbol{\theta})}{2} - \frac{\lambda}{2}\boldsymbol{\theta}\cdot\boldsymbol{\theta}\right], \quad (36)$$

with N_0 being a constant normalization factor to ensure $\int \text{Posterior}(\boldsymbol{\theta})d^N\boldsymbol{\theta} = 1$. To estimate the uncertainty of $V(T,r)$ for any given T and r, we allow the variation of parameters away from their optimal values, and the probability of accepting such a parameter set, as well as the corresponding potential, is determined by the posterior function

$$P(V_{\boldsymbol{\theta},T,r})dV = \text{Posterior}(\boldsymbol{\theta}|\text{data})d^N\boldsymbol{\theta}. \quad (37)$$

With a sufficient number of potential samples following the above distribution, we can estimate the credible interval of the potential at each of the (T,r) points.

Sampling high-dimensional parameters is tricky by itself. In a most direct way, one sample M points in the parameter space according to a flat distribution, denoted as $\{\boldsymbol{\theta}_i\}$, so that each of them corresponds to the volume element $d^N\boldsymbol{\theta}_i = M^{-1}$, and the corresponding potential represents a point in the histogram with the weighting

$$w_i = P(V_{\boldsymbol{\theta}_i,T,r})dV_i = \frac{\text{Posterior}(\boldsymbol{\theta}_i|\text{data})}{M}. \quad (38)$$

However, there are $\sim 10^3$ parameters $\boldsymbol{\theta}$ in a DNN, and a majority of points in the parameter space correspond to a vanishing posterior. Thus, computing the likelihood distribution according to Eq. (38) is computationally expensive.

In principle, the most efficient way would be to sample $\{\boldsymbol{\theta}_i\}$ according to Posterior($\boldsymbol{\theta}|$data), hence the volume element $d^N\boldsymbol{\theta}_i = M^{-1}\text{Posterior}^{-1}(\boldsymbol{\theta}_i|\text{data})$, and the corresponding potential is of the weighting

$$w_i = P(V_{\boldsymbol{\theta}_i,T,r})dV_i = \frac{1}{M}. \quad (39)$$

Nevertheless, the posterior function is generally unknown or unable to be represented in a simple way, hence it is not possible to sample according to the posterior function. A practical method is importance sampling (see, e.g., Ref. [60]), which samples $\{\boldsymbol{\theta}_i\}$ according to a reference distribution $\tilde{P}(\boldsymbol{\theta})$, hence $d^N\boldsymbol{\theta}_i = M^{-1}\tilde{P}^{-1}(\boldsymbol{\theta}_i)$, and performs reweighting on each sample by assigning an extra weight determined by the ratio of the posterior to the reference, $w_i = \text{Posterior}(\boldsymbol{\theta}_i)/\tilde{P}(\boldsymbol{\theta}_i)/M$.

The computational efficiency would be higher when the reference distribution $\tilde{P}(\boldsymbol{\theta})$ is close to the Posterior. For general systems, one usually invokes variational inference [61] or the Bayesian Neural Network [62] to find $\tilde{P}(\boldsymbol{\theta})$. In the work, however, we are able to make use of the underlying physics to construct the reference distribution. According to the first-order perturbation theory, the posterior is a nondiagonal normal distribution around the optimal parameter set ($\boldsymbol{\theta}^{\text{opt}}$), and we adopt it to be the reference distribution

$$\tilde{P}(\boldsymbol{\theta}) = (2\pi)^{-N_\theta/2}\sqrt{\det[\Sigma^{-1}]}$$
$$\times \exp\left[-\frac{\Sigma_{ab}^{-1}}{2}(\theta_a - \theta_a^{\text{opt}})(\theta_b - \theta_b^{\text{opt}})\right], \quad (40)$$

with the inverse covariance matrix given by

$$\Sigma_{ab}^{-1} \equiv \frac{\partial^2 J(\boldsymbol{\theta})}{\partial\theta_a\partial\theta_b} = \lambda\delta_{ab} + \frac{1}{2}\frac{\partial^2\chi^2(\boldsymbol{\theta})}{\partial\theta_a\partial\theta_b}, \quad (41)$$

where

$$\frac{1}{2}\frac{\partial^2\chi^2(\boldsymbol{\theta})}{\partial\theta_{R,a}\partial\theta_{R,b}} = \sum_{T,i,j}R_{i,j}^{(T)}\frac{\partial m_{T,i}}{\partial\theta_{R,a}}\frac{\partial m_{T,j}}{\partial\theta_{R,b}}, \quad (42)$$

which is similar for $\frac{1}{2}\frac{\partial^2\chi^2(\boldsymbol{\theta})}{\partial\theta_{R,a}\partial\theta_{I,b}}$ and $\frac{1}{2}\frac{\partial^2\chi^2(\boldsymbol{\theta})}{\partial\theta_{I,a}\partial\theta_{I,b}}$.

To sample parameters according to the reference distribution $\tilde{P}(\boldsymbol{\theta})$, one needs to solve the eigenvalues and

eigenstates of the inverse covariance matrix for the parameters Σ_{ab}^{-1}. There are $\sim 10^3$ parameters in the DNNs, and solving the eigenvalues and eigenvectors for Σ_{ab}^{-1} are generally expensive. Fortunately, $\frac{\partial^2 \chi^2(\boldsymbol{\theta})}{\partial \theta_a \partial \theta_b}$ are large matrices (of dimension $\sim 10^3$) constructed by multiplying low rank [rank $\leq 30 = 6$(temperatures) $\times$ 5(states)] matrices. One can conclude there are, at most, 30 nonvanishing eigenvalues for such a large matrix. We obtain these nonvanishing eigenvalues and the corresponding eigenvectors by employing the power method, while the rest eigenvectors, corresponding to the highly degenerated zero eigenvalue, are obtained by applying the Gram–Schmidt orthogonalization procedure. Such a procedure is also referred to as the *principal component analysis* (PCA) in deep learning.

In our method, the aleatoric (statistical) uncertainty is naturally encoded inside the posterior, since the χ^2 is considered the lattice data error and/or correlations. Meanwhile the epistemic (systematic) uncertainty is also manifested. On one hand, our "ignorance" of the network parameters is reflected in the prior which is consistent with the regularizer used in the cost, and on the other hand, the limits of the model—limited energy levels for a quantum system would only manifest partial information of the interaction—can also retain our methodology via the generalized back-propagation through the Schrödinger equation.

E. Method validation: Closure test with known potentials

To justify the above described method of potential reconstruction with DNNs, in this section we perform closure tests to assess the reliability of the methodology. We start by assuming a "ground-truth" potential, taking the known HTL formulae

$$V_R(T,r) = \frac{\sigma}{\mu_D}(2 - (2 + \mu_D r)e^{-\mu_D r}) - \alpha\left(\mu_D + \frac{e^{-\mu_D r}}{r}\right) + B,$$
(43)

$$V_I(T,r) = -\frac{\sqrt{\pi}}{4}\mu_D T \sigma r^3 G_{2,4}^{2,2}\left(\begin{array}{c}-\frac{1}{2},-\frac{1}{2}\\\frac{1}{2},\frac{1}{2},-\frac{3}{2},-1\end{array}\middle|\frac{\mu_D^2 r^2}{4}\right) - \alpha T \phi(\mu_D r),$$
(44)

where the Debye-screening mass, μ_D, is a function of temperature, G the Meijer-G function, and

$$\phi(x) = 2\int_0^\infty \frac{z\,dz}{(1+z^2)^2}\left(1 - \frac{\sin(xz)}{xz}\right).$$
(45)

We note that the imaginary potential is consistent with the hard thermal loop calculation [24,63]. Adopting such a potential, we further solve the Schrödinger equations at six
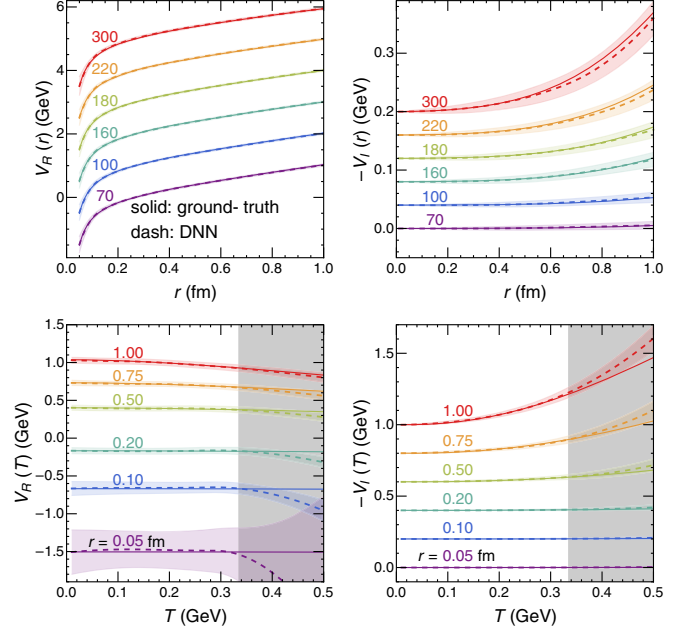

FIG. 3. Closure test on applying deep neural networks to reconstruct potential functions. The upper panels show the r-dependence at different temperature points, while the lower panels show the T-dependence at different distances. The left (right) panels plot the real (imaginary) potential. The solid lines represent the "ground-truth" formulae [(43)–(44)], while the dashed lines with the uncertainty band are the potential reconstructed by the DNN given only the mass and width for the first five bottomonia states. The gray shaded area in the lower panels indicates the extrapolation region.

different temperatures, $\{0, 151, 173, 199, 251, 334\}$ MeV, to generate a set of pseudodata—the mass and width for 1S, 2S, 3S, 1P, and 2P states. This set of pseudodata is then fed into the above described method to train the DNN for potential reconstruction. With the comparison to the assumed "ground-truth" input potential, one can therefore assess the reliability and robustness of the method. In Fig. 3, we compare the DNN reconstructed potential with the "ground-truth" values systematically. The comparison is shown for different temperatures (not limited to pseudodata generation temperatures) and distances.

From Fig. 3, it is evident that the DNN reconstructed potentials (dashed lines) are almost identical to the "ground-truth" potentials (solid lines) over the physically relevant range in r. Slight deviation and a relatively large uncertainty band at $r = 0.05$ fm are related to the facts that there the wave functions are small and the energy eigenvalues are insensitive to the potential values at this small range. In particular, note that although the DNN is trained by fitting only the energy eigenvalues (pseudodata) generated at the aforementioned discretized temperature points, the DNN reconstructed potentials show smooth behavior along with temperatures and agree nicely with the "ground-truth" potential values in the interpolation region.
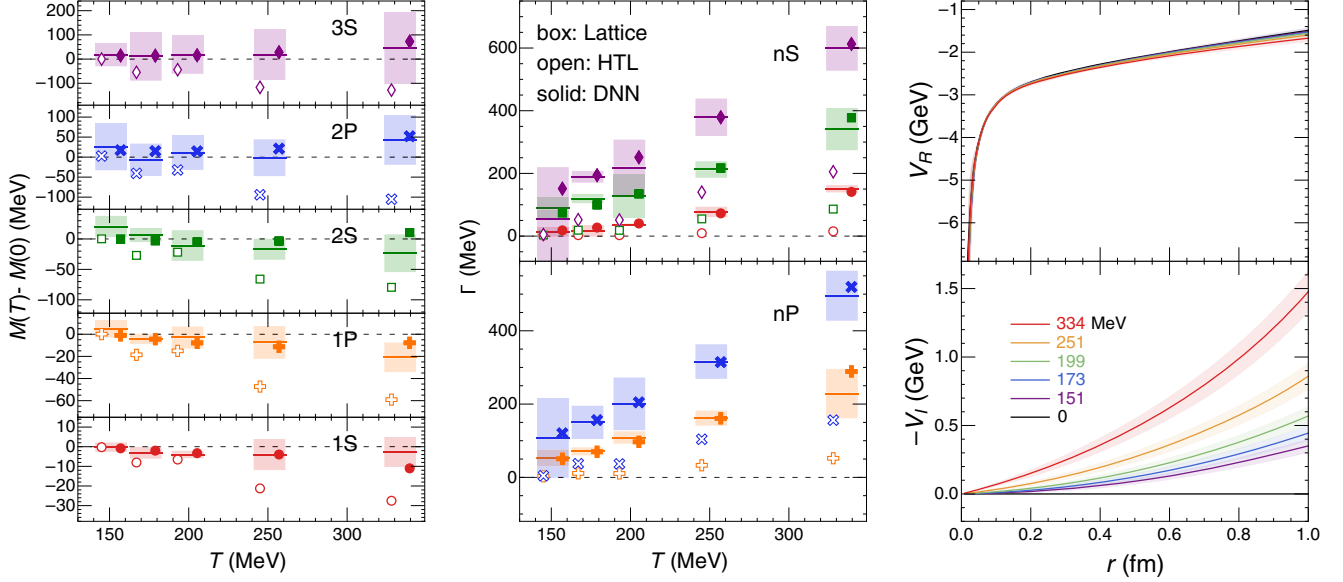
FIG. 4. Left and middle: In-medium mass shifts with respect to the vacuum mass (left) and the thermal widths (middle) of different bottomonium states obtained from fits to LQCD results of Ref. [33] (lines and shaded bands) using weak-coupling-motivated functional forms [64] (open symbols) and DNN-based optimization (solid symbols). The points are shifted horizontally for better visualization. $\Upsilon(1S)$, $\chi_{b_0}(1P)$, $\Upsilon(2S)$, $\chi_{b_0}(2P)$, and $\Upsilon(3S)$ states are represented by red circles, orange pluses, green squares, blue crosses, and purple diamonds, respectively. Right: The DNN reconstructed real (top) and imaginary (bottom) parts of the heavy quark potential at temperatures $T = 0$ (black), 151 (purple), 173 (blue), 199 (green), 251 (orange), and 334 (red) MeV. The uncertainty bands represent the 68%(1σ) confident region.

Not surprisingly, the DNN potential starts to diverge from the "ground-truth" values in the extrapolation region ($T > 334$ MeV), which is also captured by the increasing uncertainty band from the Bayesian analysis, as indicated by the gray shaded area in the lower panels of Fig. 3.

IV. HEAVY QUARK POTENTIAL FROM DNN

With the computation framework established and tested, we move on to discuss the extraction of heavy quark potential from LQCD results. We begin with pointing out the inadequacy of the weak-coupling-motivated functional form of the potential to consistently describe the LQCD results for bottomonia masses and thermal widths [33]. For this purpose, we chose the functional form proposed in Ref. [64]. This incorporates one-loop HTL-based functional forms of V_I and of color-electric screening, in addition to a vacuum potential satisfying Gauss's law [see Eqs. (43) and (44)]. Taking this functional form for the potential, we fix α, σ, and B by their vacuum values, and tune μ_D at different temperatures to fit the finite-temperature bottomonia masses and widths. We find the most optimal values to be $\mu_D = 0.01, 0.19, 0.17, 0.32, 0.37$ GeV for $T = 151, 173, 199, 251$, and 334 MeV, respectively, with the corresponding χ^2-per-data being $\{13.5, 159, 111, 154, 244\}/5$. As shown by the open symbols in Fig. 4 (left and middle), the one-loop HTL-motivated functional form of V_I and color-electric screening in V_R fail to simultaneously reproduce the LQCD

results for the mass shifts and the thermal widths of bottomonium. This failure might be due to the missing contributions from the color-magnetic scale, which is normally beyond the scope of a conventional perturbation theory [65].

The failure of the only known analytic form to describe the LQCD results necessitates a model-independent extraction of $V(T, r)$ using an adequate unbiased parametrization. To achieve this, we devised the above outlined method by coupling a Schrödinger equation with DNNs. Using this setup, we optimized the DNNs' parameters and achieved good agreement with the LQCD results [33]. The optimized fitting for the mass shifts and thermal widths are shown by the solid symbols in Fig. 4 (left and middle), with the corresponding χ^2-per-data point to be 16.5/30. The T- and r-dependence of the real (top) and imaginary (bottom) are shown in Fig. 4 (right). We see signs that with increasing temperature $V_R(T, r)$ becomes flatter at large r, as expected from the color screening effect. However, the temperature dependence of $V_R(T, r)$ is very mild between $T \approx 151$–334 MeV, and closely approximates its vacuum counterpart. In the same temperature range, $V_I(T, r)$ show significant monotonic increase both with temperature and distance. In the succeeding section, we also performed similar analyses using temperature-independent DNNs and polynomials in r to represent the functional form of the potential. We obtained consistent results with our original implementation. While the r-dependence could be retrieved by simpler parametrizations, the T-dependence is nontrivial
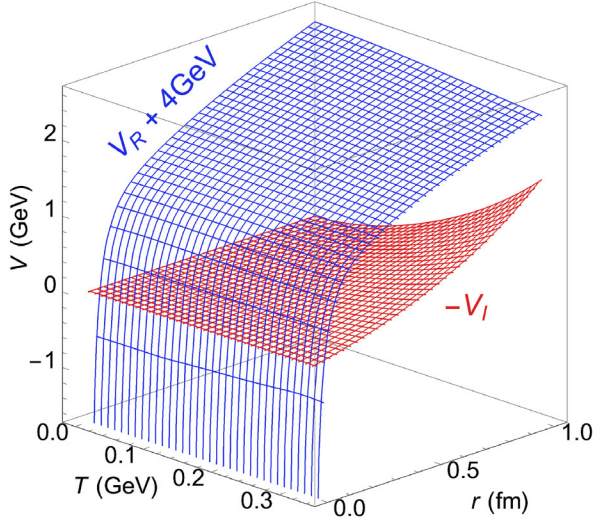
FIG. 5. Real (blue) and imaginary (red) parts of interaction potentials versus temperature T and quark-antiquark distance r, extracted via DNNs.

to be captured in an unbiased manner. Using a DNN to represent the potentially nonlinear temperature and distance dependence, we reconstructed the potential with reliable uncertainty for the temperatures in the region $T \in [0, 334]$ MeV. These results are shown in Fig. 5.

V. CONSISTENCY TESTS

A. Temperature-independent parametrizations with DNNs or polynomials

In order to examine the consistency of potentials obtained in Sec. IV from aspects of parametrizations, we performed two independent tests with two different parametrization schemes both being temperature-independent: (a) the DNNs with only distance r to be the input argument and (b) polynomial parametrization, of the real and imaginary potentials. The flow chart of such model training is similar to what has been discussed in Sec. III and shown in Fig. 2. The only modification one needs to apply is to replace the $(T, r) \to V_{R/I}(T, r)$ DNNs (the left upper corner of Fig. 2) by the $(r) \to V_{R/I}(r)$ DNNs for the parametrization scheme (a), while for scheme (b) one must replace it by the polynomial parametrization as follows:

$$V_R(r) = \sum_{i=-1}^{3} c_{R,i} r^i, \qquad V_I(r) = -\sum_{i=1}^{3} c_{I,i} r^i. \quad (46)$$

For the polynomial parametrization, we have taken into account a physical conjecture that $V_I(r)$ vanishes when $r \to 0$. Also, we applied our prior belief that higher-order coefficients shall not be large, hence we employed the regularizer $J_{\text{reg}} = \sum_i \lambda_i (c_{R,i}^2 + c_{I,i}^2)$, with $\lambda_{i \leq 1} = 0$, $\lambda_2 = 1000/\text{GeV}^6$, and $\lambda_3 = 5000/\text{GeV}^8$. In the T-independent
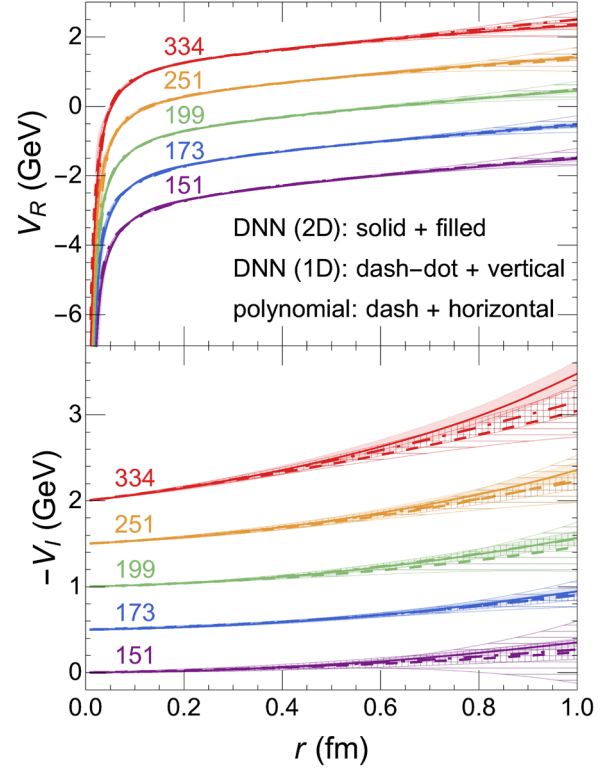

FIG. 6. Real (top) and imaginary (bottom) interaction potentials versus quark distance r extracted by T-dependent DNNs (also known as DNN(2D), solid line with filled uncertainty bands), T-independent DNNs (also known as DNN(1D), dash-dotted line with vertical hashed uncertainty bands) and polynomial parametrizations (dashed line with horizontal hashed uncertainty bands). Different colors respectively represent temperatures $T = 151$ (purple), 173 (blue), 199 (green), 251 (orange), and 334 (red) MeV which are also ordered from bottom to top. For better visualization, the curves are shifted vertically. The error bands represent the 68% (1σ) confidence interval.

DNNs, we used simpler network structures, i.e., $1 \times 16 \times 16 \times 1$ for the real potential, and $1 \times 4 \times 4 \times 1$ for the imaginary part, with regularizer $\lambda_R = 0.1$ for the former while $\lambda_I = 0.001$ is used for the latter, which are consistent with the complexity of the corresponding network structure as well. Similar to the preceding section, we used `elu` (`linear`) activation functions for the hidden(output) layers in the DNNs. We have also applied the $\lim_{r \to 0} V_I = 0$ conjecture by letting $V_I(r) = r V_I^{\text{DNN}}(r)$. Noting that lattice QCD simulation [33] provides independent sets of bottomonia mass and width at different temperatures, we perform the Bayesian analysis at each temperature point separately, and extract the optimal parameter set together with the corresponding likelihood functions. Again, the prior distribution is defined according to the regularizer. In Fig. 6, we compare the complex potentials obtained in such three schemes, and find nice agreement between them.

The polynomial parametrization scheme has in total eight coefficients. Hence, it is possible to list the optimal

TABLE III. Optimal values of the polynomial coefficients at different temperatures.

T (MeV)	151	173	199	251	334
$c_{R,-1}$	−0.41	−0.41	−0.41	−0.40	−0.39
$c_{R,0}$ (GeV)	−2.53	−2.53	−2.53	−2.54	−2.55
$c_{R,1}$ (GeV^2)	0.22	0.22	0.22	0.21	0.20
$10^3 \times c_{R,2}$ (GeV^3)	0.84	0.47	−0.31	−2.26	−0.14
$10^4 \times c_{R,3}$ (GeV^4)	0.82	0.43	−0.24	−2.82	1.38
$10 \times c_{I,1}$ (GeV^2)	0.20	0.31	0.34	0.65	1.46
$10^2 \times c_{I,2}$ (GeV^3)	0.34	0.58	0.68	1.08	0.76
$10^3 \times c_{I,3}$ (GeV^4)	0.40	0.96	0.98	1.28	1.00

values of the polynomial coefficients at different temperatures, as well as their marginal likelihood distribution, for better visualizing the fitting quality and the correlation between different parameters. We list the optimal coefficient set in Table III, and show the marginal likelihood distribution for $T = 151$ MeV in Fig. 7. One can see the strong correlation between "neighboring" coefficients.

In addition, we note that while the Schrödinger equation is solved within the range $r \in [0, 2]$ fm, the potentials can be well-constrained only within the range $r \leq 1$ fm. Such a limitation can be well understood: the wave functions of such bound states concentrate in the $r \leq 1$ fm region. According to the Hellmann–Feynman theorem, the mass spectrum is not sensitive to the potential in the $1 < r \leq 2$ fm region, and it can hardly constrain the potential for that region.

To conclude, we emphasize that although the polynomial parametrization provides a relatively simple picture of the distance dependence of potential, it has difficulties describing the temperature dependence without enough priors. Thus, one needs to generically employ an unbiased but robust parametrization scheme to obtain the 2D potential depending on both distance and temperature, for which
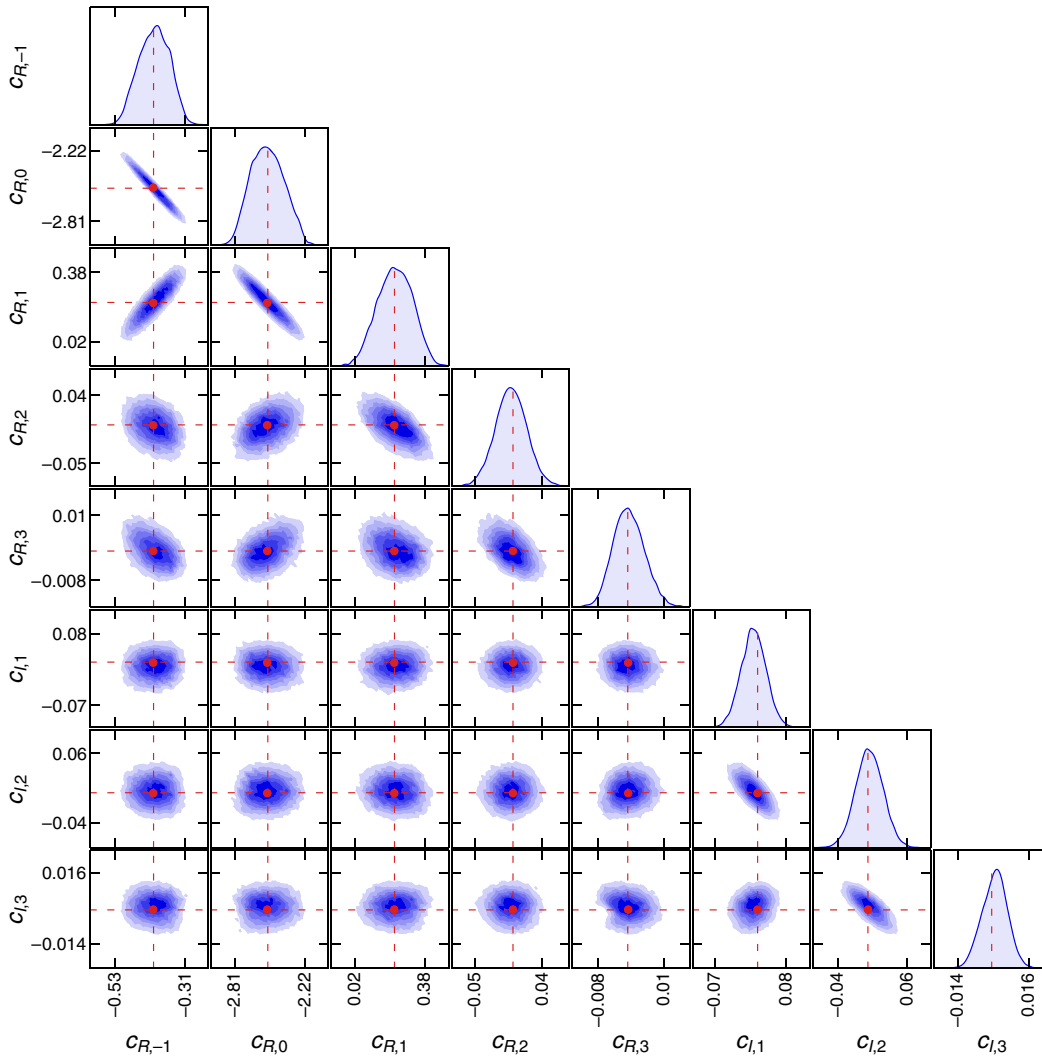

FIG. 7. Marginal likelihood distribution for the polynomial coefficients at $T = 151$ MeV. Red dots and dashed lines indicate the most optimal parameter set. The unit of the coefficients are listed in Table III.
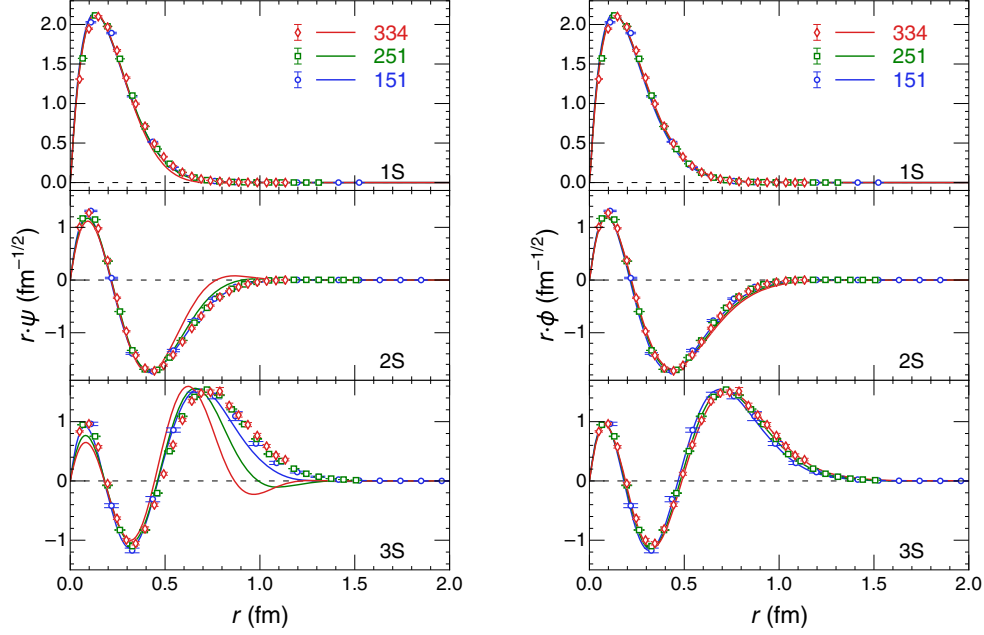
FIG. 8.   Left: Comparison of the real part of finite-temperature wave functions (curves) and Bethe–Salpeter amplitudes (symbols). Results at $T = 151$, 251, and 334 MeV are respectively colored in blue, green, and red. Right: Same as left but for the "pseudo" wave function obtained only from the real potential. See text for explanation.

DNNs provide the proper parametrization with moreover the well-developed optimization approach in practice.

B. Comparing the wave functions with the Bethe–Salpeter amplitude at finite temperature

In this subsection, we compare the finite-temperature wave functions with the corresponding BSAs from the lattice QCD calculation [34], which is obtained consistently with the masses and widths [33]. With such complementary information, the comparison serves as an independent test of the finite-temperature potential. We compare the real part of wave functions at different temperatures in Fig. 8 (left). We observe mild temperature dependence of the BSAs, while the wave functions are obviously different at higher temperature. We note that the real part of the interaction potentials show weak dependence on temperature, and the change of wave function is dominated by the imaginary potential.

As noted in Ref. [34], due to their nontrivial Euclidean time dependence, the BSAs at $T > 0$ fail to capture the thermal broadening of the states, and rather resemble the vacuum wave functions. Consequently, we solve the "pseudo" wave functions, denoted as ϕ, according to the real potential in Fig. 5,

$$-\frac{\nabla^2}{m_b}\phi_n + V_R(T, r)\phi_n = \tilde{E}_n\phi_n, \qquad (47)$$

and compare them with the BSA in Fig. 8 (right), and find excellent agreement, especially regarding the large-r tail at

different temperatures. Such comparisons serve as an independent test of the real part of the interaction potential at finite temperature. In particular, the tail behavior of the wave functions is sensitive to the flatness of the potential at $r \gtrsim 0.5$ fm. The excellent agreement shown in Fig. 8 (right), especially for the 3S state at all temperatures, confirms the weak screening effect observed in the real part of the potential.

It would be interesting to check the role of the complex wave functions. In the main content, we solve the complex wave functions according to the complex potential and obtain the mass and width from the complex energy eigenvalue. An alternative method is to treat the imaginary potential as a perturbation, and extract the wave function (ϕ_n) according to the real part of the potential, as in Eq. (47). As has been discussed above, ϕ_n's are equivalent to the BSAs in the lattice calculations, and we refer to it as the BSAs. Then, we compute the thermal width as the Bethe-Salpeter expectation of the imaginary potential,

$$\Gamma_{\text{BS}} \equiv -\int |\phi_{\text{BS}}(r)|^2 V_I(r) r^2 dr. \qquad (48)$$

In Fig. 9, we compare the Γ_{BS}, from the perturbative treatment, with the complete thermal width from the lattice result. We find that at lower temperatures, at which Γ are small, the perturbation results agree well with the complete ones, whereas, at high temperature, e.g., 334 MeV, Γ_{BS} are slightly, but systematically, higher than the lattice results.
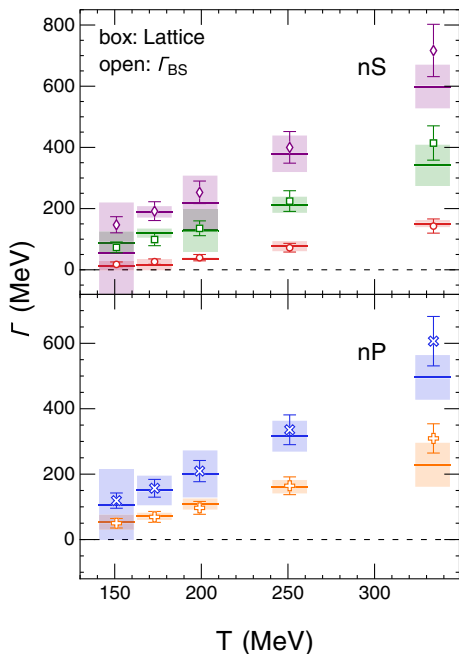
FIG. 9. Expectation of the thermal width based on the Bethe–Salpeter amplitudes.

VI. CONCLUSION AND DISCUSSION

In this work, we report a surprising empirical finding: LQCD results [33] for the masses and thermal widths of up to $3S$ and $2P$ bottomonium states in QGP admits a consistent quantum mechanical description based on an complex-valued potential and nonrelativistic Schrödinger equation. By coupling the Schrödinger equation to a DNN, we introduced a novel method for unbiased extractions of the real and imaginary parts of the heavy quark potential, and invoked Bayesian inference to quantify the potential uncertainties in a nonlocal fashion. With such a model-independent method, we obtained the empirical $V_R(T, r)$ and $V_I(T, r)$ for $r \lesssim 1$ fm and $T \lesssim 334$ MeV, which can successfully map the QCD spectrum of the lowest-lying bottomonium states in QGP to a quantum mechanical system.

The heavy quark potential obtained here renders an empirical mapping from the masses and thermal widths of bottomonium states at finite temperature to an effective quantum mechanics framework, based on LQCD calculations of the bottomonium state using a $2 + 1$ flavor dynamical gauge field background with nearly physical values of up, down, and strange quark masses. Direct quantitative comparison with extant LQCD calculations of static quark potentials [28–31] is difficult. However, our result for the heavy quark potential is qualitatively different from those potentials. Unlike the previous studies, the V_R obtained in this work show very little signs of color-electric Debye screening for $r \lesssim 1$ fm for the entire temperature range $T \in [0, 334]$ MeV. The V_I here is much larger in magnitude and increases more rapidly, both with T and r, than the one-

loop HTL-motivated extractions. On the other hand, it is reassuring that the potential obtained here is quantitatively consistent with the very recent LQCD calculations [66] on the peak position and the width of the Gaussian-form spectral function, as functions of the separation r and temperature T. Agreement—in the sense of the strongly-coupled behavior, as well as a large magnitude of imaginary potential—is observed in the comparison with heavy quark potentials computed in the T-matrix approach [67,68] and phenomenologically extracted from bottomonium data [69–72], despite some difference in the exact value. Meanwhile, it might be worth noting that large imaginary parts are also seen in Ref. [31], and are not ruled out by Refs. [29,30], due to the large errors for their results. It would be very interesting to see the phenomenological consequences [13] of this heavy quark potential, model-independently extracted from the nonperturbative LQCD calculations.

Further, we carried out detailed comparisons of the real parts of the wave functions with the BSAs obtained from LQCD calculations [34]. As noted in Ref. [34], due to their nontrivial Euclidean-time dependence, the BSAs at $T > 0$ fail to capture the thermal broadening of the states and resemble the vacuum wave functions. Our comparisons seem to support this picture. While the real parts of the actual wave functions show deviations from the BSAs at large r, the "pseudo" wave functions obtained using only V_R (with $V_I = 0$) reproduce the BSAs. Furthermore, we also find that the "pseudo" thermal widths, $\Gamma_{\mathrm{BS}}(T) = -\int |\psi_{\mathrm{BS}}(T, r)|^2 V_I(T, r) r^2 \mathrm{d}r \approx \Gamma(T)$, suggest that $V_I(T, r)$ might be considered as a perturbation on top of an approximately vacuumlike excitation. Based on our results, one might speculate that, for phenomenologically relevant temperatures $T \lesssim 334$ MeV, bottomonia are approximately vacuumlike excitations but of very short lifetimes that are inversely proportional to their large thermal widths. At high enough temperatures, we anticipate that this speculative picture would smoothly turn over to the more conventional picture based on quarkonia melting due to color-electric Debye screening [20,21,73] and perturbative Landau damping [24,25].

APPENDIX: ABOUT THE IMAGINARY ENERGY AND WIDTH OF THE SPECTRAL FUNCTION

In this work, we have assumed that the width obtained in Ref. [33] is the imaginary part of the energy eigenvalue, $\Gamma_n = |\text{Im}[E_n]|$. Such a relation is not obvious as the lattice QCD results assume a Gaussian form for the spectral function. In this appendix, we investigate the relation between the imaginary part of the energy and the Gaussian width.

Following the procedure in Ref. [74], we start the correlation as a spatial Dirac-δ function,

$$C^>(0, \mathbf{r}) = \delta^{(3)}(\mathbf{r}), \tag{A1}$$

and evolve the forward and backward propagator according to the Hamiltonian

$$\begin{cases} \hat{H} C^>(t, \mathbf{r}) = i\partial_t C^>(t, \mathbf{r}), & t > 0, \\ \hat{H}^\dagger C^>(t, \mathbf{r}) = i\partial_t C^>(t, \mathbf{r}), & t < 0. \end{cases} \tag{A2}$$

We note that the eigenfunctions of the Hamiltonian, $\{\psi_n\}$, form a complete set of the Hilbert space. Hence, we expand the Dirac-δ function in a series of the wave functions,

$$c_n \equiv \int d^3\mathbf{r}\, C^>(0, \mathbf{r}) \psi_n^*(\mathbf{r}) = \psi_n^*(0), \tag{A3}$$

and the time dependence of the correlation function can be expressed as the super position of different eigenmodes,

$$C^>(t, \mathbf{r}) = \begin{cases} \sum_n c_n e^{-iE_n t} \times \psi_n(\mathbf{r}), & t > 0, \\ \sum_n c_n^* e^{-iE_n^* t} \times \psi_n^*(\mathbf{r}), & t < 0, \end{cases} \tag{A4}$$

where $\sum_n$ denotes summation over all bound states, as well as the integral over the scattering continuum when applicable.

With these, we find that the spectral function

$$\rho(\omega) \equiv \int_{-\infty}^{+\infty} dt\, e^{i\omega t} C^>(t, 0) \tag{A5}$$
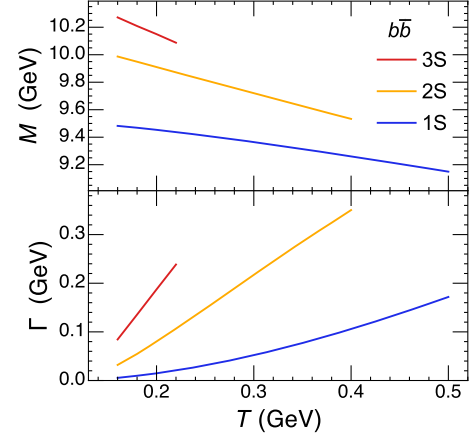

FIG. 10. Masses and thermal widths of bottomonium states by solving the bound-state problem of the Schrödinger Hamiltonian with the complex-valued potential listed in Ref. [64].

$$= \sum_n \frac{-2|\psi_n(0)|^2 \text{Im}[E_n]}{(\omega - \text{Re}[E_n])^2 + (\text{Im}[E_n])^2} \tag{A6}$$

takes the Lorentzian form, with the Lorentzian width being the imaginary energy eigenvalue. To guarantee that the amplitude $|\psi(t)\rangle$ decays rather than explodes, the imaginary energy should always be nonpositive, and

$$\Gamma_n^{\text{Lor}} = -\text{Im}[E_n]. \tag{A7}$$

On the other hand, in Refs. [32–34] the masses and widths are extracted under the assumption of a Gaussian spectral function,

$$\rho(\omega) \propto \sum_n \exp\left[-\frac{(\omega - M_n)^2}{2\Gamma_n^2}\right]. \tag{A8}$$

While the mass can be uniquely defined as the peak position, there is no obvious way to map the Lorentzian width with the Gaussian one. In this work, we take $\Gamma_n^{\text{Lor}} = \Gamma_n^{\text{Gau}}$, as they both represent the characteristic width. From a different point of view, one might otherwise match them according to the half-maximum of the spectral function. In the latter case, one would find $\Gamma_n^{\text{Lor}} = \sqrt{2\ln 2}\,\Gamma_n^{\text{Gau}}$, hence $\text{Im}[E_n] = -1.18\Gamma_n^{\text{Gau}}$. If taking the latter mapping, the decay width of the bottomonium states shall be multiplied by a factor of 1.18, and the extracted V_I shall increase by ∼18%.

In this appendix, we have shown, analytically, that mass and thermal widths obtained by solving the evolution of correlators shall be equivalent to those from solving the energy bound states. One may wonder if such an equivalence would still hold in numerical procedures, especially given the possible uncertainty of fitting the peaks of the spectral function. To answer this question, we perform a numerical verification as follows. We start from the

complex-valued potential listed in Ref. [64], solve the bound-state problem for the Schrödinger Hamiltonian according to the inverse power method [75], and obtain the complex-valued energy eigenvalues for various bottomonium states. Their masses and thermal widths—being the real and imaginary part of the energy eigenvalues, respectively—are shown in Fig. 10. We find our results are

consistent with those of [64], which computes the masses and thermal widths from the correlator evolution as outlined in [74]. With both the analytical derivation and numerical verifications, we conclude that these two procedures are equivalent in computing the mass and thermal widths from the Schrödinger Hamiltonian with time-independent potentials.

---

[1] T. Matsui and H. Satz, Phys. Lett. B **178,** 416 (1986).

[2] F. Karsch, M. T. Mehr, and H. Satz, Z. Phys. C **37,** 617 (1988).

[3] J.-P. Blaizot and J.-Y. Ollitrault, Phys. Rev. Lett. **77,** 1703 (1996).

[4] P. Braun-Munzinger and J. Stachel, Phys. Lett. B **490,** 196 (2000).

[5] S. Digal, P. Petreczky, and H. Satz, Phys. Rev. D **64,** 094015 (2001).

[6] L. Grandchamp, R. Rapp, and G. E. Brown, Phys. Rev. Lett. **92,** 212301 (2004).

[7] T. Song, K. C. Han, and C. M. Ko, Phys. Rev. C **84,** 034907 (2011).

[8] X. Du and R. Rapp, Nucl. Phys. **A943,** 147 (2015).

[9] Y. Liu, B. Chen, N. Xu, and P. Zhuang, Phys. Lett. B **697,** 32 (2011).

[10] K. Zhou, N. Xu, Z. Xu, and P. Zhuang, Phys. Rev. C **89,** 054911 (2014).

[11] R. Katz and P. B. Gossiaux, Ann. Phys. (Amsterdam) **368,** 267 (2016).

[12] X. Yao, W. Ke, Y. Xu, S. A. Bass, and B. Müller, J. High Energy Phys. 01 (2021) 046.

[13] A. Islam and M. Strickland, Phys. Lett. B **811,** 135949 (2020).

[14] S. Chatrchyan *et al.* (CMS Collaboration), Phys. Rev. Lett. **107,** 052302 (2011).

[15] S. Chatrchyan *et al.* (CMS Collaboration), Phys. Rev. Lett. **109,** 222301 (2012); **120,** 199903(E) (2018).

[16] V. Khachatryan *et al.* (CMS Collaboration), Phys. Lett. B **770,** 357 (2017).

[17] A. M. Sirunyan *et al.* (CMS Collaboration), Phys. Rev. Lett. **120,** 142301 (2018).

[18] W. E. Caswell and G. P. Lepage, Phys. Lett. **167B,** 437 (1986).

[19] N. Brambilla, A. Pineda, J. Soto, and A. Vairo, Nucl. Phys. **B566,** 275 (2000).

[20] H. Satz, J. Phys. G **32,** R25 (2006).

[21] J. Zhao, K. Zhou, S. Chen, and P. Zhuang, Prog. Part. Nucl. Phys. **114,** 103801 (2020).

[22] H. W. Crater, J.-H. Yoon, and C.-Y. Wong, Phys. Rev. D **79,** 034011 (2009).

[23] X. Guo, S. Shi, and P. Zhuang, Phys. Lett. B **718,** 143 (2012).

[24] M. Laine, O. Philipsen, P. Romatschke, and M. Tassler, J. High Energy Phys. 03 (2007) 054.

[25] A. Beraudo, J. P. Blaizot, and C. Ratti, Nucl. Phys. **A806,** 312 (2008).

[26] N. Brambilla, J. Ghiglieri, A. Vairo, and P. Petreczky, Phys. Rev. D **78,** 014017 (2008).

[27] N. Brambilla, M. A. Escobedo, J. Ghiglieri, J. Soto, and A. Vairo, J. High Energy Phys. 09 (2010) 038.

[28] A. Rothkopf, T. Hatsuda, and S. Sasaki, Phys. Rev. Lett. **108,** 162001 (2012).

[29] Y. Burnier, O. Kaczmarek, and A. Rothkopf, Phys. Rev. Lett. **114,** 082001 (2015).

[30] Y. Burnier, O. Kaczmarek, and A. Rothkopf, J. High Energy Phys. 12 (2015) 101.

[31] D. Bala and S. Datta, Phys. Rev. D **101,** 034507 (2020).

[32] R. Larsen, S. Meinel, S. Mukherjee, and P. Petreczky, Phys. Rev. D **100,** 074506 (2019).

[33] R. Larsen, S. Meinel, S. Mukherjee, and P. Petreczky, Phys. Lett. B **800,** 135119 (2020).

[34] R. Larsen, S. Meinel, S. Mukherjee, and P. Petreczky, Phys. Rev. D **102,** 114508 (2020).

[35] It might be worth noting that the potential here is the effective potential for a Schrödinger equation.

[36] M. Tanabashi *et al.* (Particle Data Group), Phys. Rev. D **98,** 030001 (2018).

[37] M. Leshno and S. Schocken, Neural Netw. **6,** 861 (1993).

[38] A. Kratsios, Annals of Mathematics and Artificial Intelligence **89,** 435 (2021).

[39] S. Forte, L. s. Garrido, J. I. Latorre, and A. Piccione, J. High Energy Phys. 05 (2002) 062.

[40] L. D. Debbio, S. Forte, J. I. Latorre, A. Piccione, J. Rojo (T.N. Collaboration), J. High Energy Phys. 03 (2007) 039.

[41] L. Kades, J. M. Pawlowski, A. Rothkopf, M. Scherzer, J. M. Urban, S. J. Wetzel, N. Wink, and F. P. G. Ziegler, Phys. Rev. D **102,** 096001 (2020).

[42] M. Zhou, F. Gao, J. Chao, Y.-X. Liu, and H. Song, Phys. Rev. D **104,** 076011 (2021).

[43] S. Y. Chen, H. T. Ding, F. Y. Liu, G. Papp, and C. B. Yang, arXiv:2110.13521.

[44] J. Carrasquilla and R. G. Melko, Nat. Phys. **13,** 431 (2017).

[45] L.-G. Pang, K. Zhou, N. Su, H. Petersen, H. Stcker, and X.-N. Wang, Nat. Commun. **9,** 210 (2018).

[46] R. Wang, Y.-G. Ma, R. Wada, L.-W. Chen, W.-B. He, H.-L. Liu, and K.-J. Sun, Phys. Rev. Research **2,** 043202 (2020).

[47] L. Jiang, L. Wang, and K. Zhou, Phys. Rev. D **103,** 116023 (2021).

[48] K. Zhou, G. Endrődi, L.-G. Pang, and H. Stöcker, Phys. Rev. D **100**, 011501 (2019).

[49] D. Boyda, G. Kanwar, S. Racanière, D. J. Rezende, M. S. Albergo, K. Cranmer, D. C. Hackett, and P. E. Shanahan, Phys. Rev. D **103**, 074504 (2021).

[50] G. Kanwar, M. S. Albergo, D. Boyda, K. Cranmer, D. C. Hackett, S. Racanière, D. J. Rezende, and P. E. Shanahan, Phys. Rev. Lett. **125**, 121601 (2020).

[51] M. S. Albergo, G. Kanwar, and P. E. Shanahan, Phys. Rev. D **100**, 034515 (2019).

[52] M. Omana Kuttan, J. Steinheimer, K. Zhou, A. Redelbach, and H. Stoecker, Phys. Lett. B **811**, 135872 (2020).

[53] P. Thaprasop, K. Zhou, J. Steinheimer, and C. Herold, Phys. Scr. **96**, 064003 (2021).

[54] F. Li, Y. Wang, H. Lü, P. Li, Q. Li, and F. Liu, J. Phys. G **47**, 115104 (2020).

[55] A. Andreassen, S.-C. Hsu, B. Nachman, N. Suaysom, and A. Suresh, Phys. Rev. D **103**, 036001 (2021).

[56] M. O. Kuttan, K. Zhou, J. Steinheimer, A. Redelbach, and H. Stoecker, J. High Energy Phys. 21 (2020) 184.

[57] H. Huang, B. Xiao, Z. Liu, Z. Wu, Y. Mu, and H. Song, Phys. Rev. Research **3**, 023256 (2021).

[58] D. P. Kingma and J. Ba, arXiv:1412.6980.

[59] H. Xie, J.-G. Liu, and L. Wang, Phys. Rev. B **101**, 245139 (2020).

[60] R. Y. Rubinstein and D. P. Kroese, *Simulation and the Monte Carlo Method* (John Wiley & Sons, New York, 2016), Vol. 10.

[61] A. Graves, in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., New York,, 2011), pp. 2348–2356.

[62] C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra, arXiv:1505.05424.

[63] Y. Burnier, M. Laine, and M. Vepsalainen, J. High Energy Phys. 01 (2010) 054; 01 (2013) 180(E).

[64] D. Lafferty and A. Rothkopf, Phys. Rev. D **101**, 056010 (2020).

[65] K. Fukushima and N. Su, Phys. Rev. D **88**, 076008 (2013).

[66] D. Bala, O. Kaczmarek, R. Larsen, S. Mukherjee, G. Parkar, P. Petreczky, A. Rothkopf, and J. H. Weber, arXiv:2110.11659.

[67] S. Y. F. Liu and R. Rapp, Eur. Phys. J. A **56**, 44 (2020).

[68] S. Y. F. Liu and R. Rapp, Phys. Rev. C **97**, 034918 (2018).

[69] X. Du, R. Rapp, and M. He, Phys. Rev. C **96**, 054901 (2017).

[70] X. Du, S. Y. F. Liu, and R. Rapp, Phys. Lett. B **796**, 20 (2019).

[71] M. Strickland, Phys. Rev. Lett. **107**, 132301 (2011).

[72] A. Islam and M. Strickland, J. High Energy Phys. 03 (2021) 235.

[73] A. Mocsy and P. Petreczky, Phys. Rev. Lett. **99**, 211602 (2007).

[74] Y. Burnier, M. Laine, and M. Vepsalainen, J. High Energy Phys. 01 (2008) 043.

[75] H. W. Crater, J. Comput. Phys. **115**, 470 (1994).