Editors' Suggestion

Deep finite temperature bootstrap

V. Niarchos,[1,*] C. Papageorgakis,[2,†] A. Stratoudakis,[1,‡] and M. Woolley[2,§]

[1]*ITCP and CCTP, Department of Physics, University of Crete, 71003 Heraklion, Greece*

[2]*Centre for Theoretical Physics, Department of Physics and Astronomy Queen Mary University of London, London E1 4NS, United Kingdom*

(Received 6 September 2025; accepted 26 November 2025; published 16 December 2025)

We introduce a novel method to bootstrap crossing equations in conformal field theory (CFT) and apply it to finite temperature theories on $S^1 \times \mathbb{R}^{d-1}$. The proposed approach does not rely on positivity constraints and does not employ uncontrolled truncation schemes. Instead, we capture the contribution of an infinite number of operators in conformal block expansions using suitable functions, which are bootstrapped (numerically) together with a finite number of exposed CFT data. Our approach at finite temperature employs three key ingredients: (1) the Kubo-Martin-Schwinger (KMS) condition, (2) thermal dispersion relations, and (3) neural networks that model spin-dependent tail functions within the conformal block expansions. We test the efficiency of the new method in the case of generalized free fields and use it to perform a preliminary bootstrap analysis of double-twist thermal data in holographic CFTs.

DOI: 10.1103/qrjx-w4md

I. INTRODUCTION AND SUMMARY

A. The need for an efficient primal bootstrap

Some of the most successful techniques in the modern conformal bootstrap program rely on the study of feasibility conditions, reformulated suitably as convex (semi-definite programming) optimization problems. Most notably, in the analysis of the crossing equations of four-point correlation functions in unitary conformal field theories (CFTs), positivity constraints are used to exclude assumptions and obtain allowed regions for selected CFT data. This approach has been remarkably useful because it cleverly bypasses the seemingly impossible problem of solving explicitly the crossing equations, which involves dealing with continuous families of equations for an infinite set of data.

There are, however, several reasons why one would like to go beyond the feasibility analysis and solve the *primal bootstrap* problem that aims to reconstruct full solutions of the bootstrap equations. For example, if there is an island of allowed parameters, it would be useful to know how consistent correlation functions look like in the bulk of the island in order to gain better intuition about potential assumptions and guided searches for theories of interest. In addition, there are many physically relevant contexts, where the positivity conditions, necessary for the feasibility analysis, are altogether absent. Notable cases include: finite-temperature theories, theories with defects or boundaries, higher-point crossing equations, etc.

To make the direct (numerical) solution of the bootstrap equations more tractable, one typically truncates the sum over an infinite number of contributions to a finite subset of low-energy CFT data, which (one hopes) are the dominant contributions to the equations of interest. In a drastic truncation, only the low-lying CFT data are considered. In more nuanced, *soft* truncations, an approximation scheme of the contribution of the truncated high-energy data is also employed. Although in the past such schemes have produced results with notable accuracy in the context of specific examples, so far there has been no general, systematic methodology for truncations in the conformal bootstrap program, and the pertinent difficulties of such methodologies are well known (see, e.g., [1–10] and the review of [11] for related discussions).

To address this situation, we introduce a new framework for the primal bootstrap, which is not based on traditional hard or soft truncations and involves a more systematic control over the employed approximations, compared to previous schemes. No positivity conditions are assumed, the approach is general and applies to many different contexts. In this work, we center the presentation

*Contact author: niarchos@physics.uoc.gr
†Contact author: c.papageorgakis@qmul.ac.uk
‡Contact author: astratoudakis@physics.uoc.gr
§Contact author: mitchell.woolley@qmul.ac.uk

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

exclusively around the analysis of thermal two-point functions of scalar operators in finite-temperature CFTs on $S^1 \times \mathbb{R}^{d-1}$. The case of four-point correlation functions at zero temperature will be discussed in a companion paper [12].

B. The strategy

A typical conformal bootstrap problem involves a vector of sum rules of the form

$$\sum_J \sum_\Delta \vec{\mathcal{X}}_{\Delta,J} = 0, \tag{1.1}$$

which contain an infinite number of contributions labeled by the scaling dimension Δ and spin J of the corresponding CFT operators. Our strategy for solving (1.1) involves the following three steps:

(1) We make an arbitrary choice of a spin cutoff J_* and express the "high"-spin[1] part of Eq. (1.1),

$$\sum_{J>J_*} \sum_\Delta \vec{\mathcal{X}}_{\Delta,J}, \tag{1.2}$$

using dispersion relations in terms of a discontinuity. This is the only place where we need to perform an approximation in our numerical implementations, and the corresponding error can be systematically suppressed by increasing the arbitrary value of J_*.

(2) The remaining, "low"-spin part of Eq. (1.1),

$$\sum_{J\leq J_*} \sum_\Delta \vec{\mathcal{X}}_{\Delta,J} \tag{1.3}$$

involves sums over a finite number of spins but an infinite number of scaling dimensions. At each spin, we freely choose to expose a finite number of contributions from specific operators at low scaling dimension, and the remaining combined contribution of all the other operators (with arbitrarily high scaling dimension) is expressed in terms of suitable "tail" functions.[2] In summary, this part contains a finite number of explicit CFT data, call them `D`, and a finite number of unknown functions, call them `F`. In this part, there is no approximation.

(3) We formulate a nonconvex optimization problem that varies over the unknowns (`D`, `F`) to determine solutions of (1.1). To efficiently vary over the space of unknown functions `F`, we use a representation based on multi-layer-perceptron (MLP) neural networks. More specifically, we use an architecture known as multi-branch MLPs.

The crucial new element of this approach is that the traditionally obscure tails of contributions from operators with arbitrarily high scaling dimension are not dropped or approximated in some fixed *ad hoc* manner. In the low-spin component of the bootstrap equations, the tails become part of the data we are bootstrapping dynamically. In the high-spin part, we employ dispersion relations. These features enable the method to remove critical limitations of previous truncation schemes in the literature. The method additionally allows a lot of freedom on the choice of assumptions one would like to implement or the specific choice of the exposed CFT data, the value of which one would like to compute explicitly. The latter does not require any prior knowledge of the structure of the solution or what constitutes a significant low-energy CFT datum. The exposed data are arbitrary.

This paper establishes the above approach in a specific physical context, demonstrates its implementation through a proof-of-concept example compared against known analytic results, and identifies error sources and potential technical challenges to guide future developments.

C. The challenge of finite temperature

At finite temperature, the one-point functions of many primary operators are nonvanishing. The goal of the finite-temperature bootstrap is to determine (or constrain) these thermal data assuming knowledge of the zero-temperature CFT data (scaling dimensions and three-point function coefficients). It has been suggested [13] that this might be achievable by studying general consistency conditions on thermal two-point functions. One of these conditions is the periodicity of two-point correlation functions on the periodic, thermal coordinate in Euclidean space, also known as the Kubo-Martin-Schwinger (KMS) condition. This is a nontrivial condition, because the operator product expansion (OPE) of the two-point functions is not automatically periodic and that imposes consistency conditions on the unknown thermal one-point functions that appear in the OPE.

At infinite spatial volume, the KMS condition is superficially very similar to the crossing equations of four-point correlation functions at zero-temperature, but the analysis exhibits special complications due to the following features:

(i) In the OPE expressions of the KMS condition, there are no positivity requirements on the OPE coefficients. As a result, it is unclear how to reformulate the analysis in terms of a convex-optimization, semi-definite programming problem for feasibility conditions.[3]

[1]"High" here refers to any spin above the arbitrary J_*, which does not have to be a large number.

[2]The precise realization of these contributions, as one-dimensional functions of a single radial space coordinate, will be explained in the context of the thermal bootstrap in the main part of the paper.

[3]It is worth noting, however, that in finite-temperature quantum mechanics positivity constraints were recently used in Ref. [14].

(ii) The OPE of thermal two-point functions does not share the familiar, quick convergence properties of conformal block expansions for four-point functions at zero-temperature (see, e.g., [15] for a relevant discussion). That complicates attempts to perform hard truncations.

(iii) Soft truncations—such as those estimating tail contributions from high-dimension operators using Tauberian theorems [15,16]—are unsatisfactory for the following reason: They reduce the problem to convex linear regression for finitely many unknown thermal OPE coefficients, yielding a unique solution even when the KMS condition admits multiple consistent solutions.[4] As a result, linear regression-based soft truncation schemes cannot detect multiple solutions and become intractable in generic problems, requiring large numbers of operators with explicit spectral knowledge. These difficulties motivate the analysis of the KMS condition without explicit truncation, using methods that are able to detect multiple solutions and can treat the high-energy CFT spectrum more economically.

D. Key results

In the main part of the paper, we explain how the general approach outlined in Sec. I B can be used to bypass the above-mentioned issues. We focus on the finite-temperature case at infinite spatial volume as a problem of intrinsic interest but also as a special concrete case for an explicit illustration of the capabilities and potential challenges of the proposed approach. The most significant contributions of the present paper can be summarized as follows:

Approximate KMS formulation: We formulate an approximate KMS condition, using suitable spin-cutoff-dependent thermal dispersion relations and consider thermal two-point functions at finite spatial separation, obtaining constraints on spin-dependent thermal OPE coefficients $a_{\Delta,J}$. In contrast, several recent approaches [15–19] focus on the properties of thermal two-point functions—and the associated KMS condition—at zero spatial separation that are sensitive only to spin-independent thermal OPE data a_Δ. More importantly, our work shifts the focus away from the exclusive computation of individual CFT data to suitable one-dimensional functions of a radial space coordinate r that repackage the information of an infinite tower of operators at increasing scaling dimension but fixed spin. Under suitable conditions, we argue that these functions exhibit a universal behavior in the vicinity of $r = \beta$ (where β is the inverse temperature). Knowledge of these functions leads to an approximate reconstruction of the thermal two-point function in the OPE convergence region.

Generalized free field testing: We perform explicit tests of the approach in the case of generalized free field (GFF) theories, where thermal one-point functions can be computed analytically. We also use this case to exhibit the systematic error introduced by the approximations in the high-spin part and demonstrate how one can systematically approach the exact KMS condition by increasing the arbitrary spin cutoff J_*.

Tail contribution bootstrap: By analyzing specific theories (GFF and holographic CFTs), we provide a proof-of-concept that tail contributions can indeed be bootstrapped, avoiding the aforementioned truncation issues as well as issues inherent in linear regression schemes.

Neural network architecture: Unknown optimizable functions are modeled using suitable Neural Networks (NNs). We present a specific sparse architecture (multi-branch MLPs) that involves separate subnets for different output functions.

Loss function analysis: We explore, and compare, several different loss functions designed to handle nonconvex optimization in diverse situations, with or without exposed CFT data. In addition, we discuss practical features of the searches we perform and of the outputs of optimization while also identifying potential difficulties.

Holographic CFT applications: As a nontrivial application, we discuss the bootstrap of holographic CFTs, perform preliminary computations of explicit double-twist CFT data, and compare it with the limited results in the literature. We emphasize that in this context, our approach has access to quantitatively new, spin-dependent, thermal double-twist data, which are not accessible to other methods based on the study of thermal two-point functions at zero-spatial separation.

Future $\mathcal{N} = 4$ SYM extension: We dedicate a section to summarizing key elements needed for future applications of our approach to $\mathcal{N} = 4$ SYM holography. We emphasize the role of operator-mixing and the ensuing logarithmic contributions in the thermal OPE and set up several necessary elements for the implementation of our approach in this context.

E. Plan of the paper

In Sec. II we set the notation and introduce the necessary background material for the analysis of the main text. This includes a summary of the relevant conformal block expansions of thermal two-point functions and the KMS condition. In Sec. III, we collect useful material about the specific form of the finite-temperature dispersion relations that are employed in our approach and relegate several pertinent technical details to Appendix A. In Sec. IV, we formulate an approximate version of the KMS condition, the solution of which is the main target of the paper. The

[4]Holographic CFTs provide explicit examples where the KMS condition exhibits continuous families of solutions.

neural network representation of the tail functions is introduced in this section, with the technical details deferred to Appendix C. Other important elements introduced in Sec. IV are three different versions of loss functions that are used throughout the paper to find approximate solutions to the KMS condition.

The main applications of the proposed formalism appear in Secs. V–VIII. Section V contains tests in the context of the generalized free field theories and Secs. VI and VII applications to holographic CFTs. Section VIII summarizes necessary material for future applications to the bootstrap of thermal two-point functions of half-BPS operators in the supergravity limit of 4d $\mathcal{N}=4$ SYM theory. Further useful material for these sections appears in Appendixes B and D–G.

We conclude in Sec IX with a brief summary of interesting open problems and promising directions for future research.

II. THERMAL BLOCK EXPANSIONS, KMS, AND ALL THAT

We study d-dimensional CFTs on $S^1 \times \mathbb{R}^{d-1}$. We parametrize the space with coordinates $x=(\tau,\vec{x})$, where τ describes the thermal circle with period β and $\vec{x}$ the spatial $\mathbb{R}^{d-1}$. This setup captures thermal physics at inverse temperature β in the infinite spatial volume limit. It can also be viewed as the high temperature limit of the theory on $S^1 \times S^{d-1}$.

In this section, we collect useful well-known facts about this context and set up a significant part of the notation we will be using in the main text.

A. Two-point functions and the KMS condition

For concreteness, we will focus on two-point functions of identical scalar operators ϕ,

$$g(\tau,|x|) \coloneqq \langle \phi(x)\phi(0)\rangle_\beta. \tag{2.1}$$

These functions depend separately on the Euclidean time τ and

$$|x| = \sqrt{\tau^2+\vec{x}^2} \tag{2.2}$$

because of the reduced $SO(d-1)$ symmetry of the background. It will be convenient to use this symmetry to set $\vec{x}=(\sigma,0,\ldots,0)$. In this frame, we define the following set of variables that will be used interchangeably in the main text:

$$z \coloneqq \tau+i\sigma, \qquad \bar{z} \coloneqq \tau-i\sigma, \tag{2.3}$$

and

$$z = rw, \qquad \bar{z} = rw^{-1}. \tag{2.4}$$

In this definition, w is a pure phase, but later it will be continued to the whole complex plane. To summarize, $|x|=r$ and the two-point function g depends on (τ,r) or equivalently on $(z,\bar{z})$.

The Kubo-Martin-Schwinger (KMS) condition states that the two-point function is invariant under the transformation $\tau \to \beta-\tau$,

$$g(\tau,r) = g(\beta-\tau,r). \tag{2.5}$$

Henceforth, we will set $\beta=1$ (without loss of generality). By using the invariance of the two-point function under the parity transformation $\sigma \to -\sigma$, we can further recast the KMS condition as a "crossing" equation,

$$g(z,\bar{z}) = g(1-z,1-\bar{z}). \tag{2.6}$$

The importance of this equation as a nontrivial consistency condition on thermal CFT data was first recognized by El-Showk and Papadodimas in [13].

B. Operator product expansions

In a two-point function $\langle\phi(x)\phi(0)\rangle_\beta$, we can use the operator product expansion (OPE) to express the correlation function as a series over thermal one-point functions. For an operator with scaling dimension Δ, the one-point function is proportional to the temperature raised to the power Δ, namely it behaves as $\beta^{-\Delta}$.[5] Moreover, since conformal descendants have vanishing thermal one-point functions, only the primaries in a conformal family contribute. This fact simplifies the form of the thermal conformal blocks. Putting everything together (and setting $\beta=1$), one ends up with the following conformal block expansion of thermal two-point functions of identical scalar operators [20]:

$$g(rw,rw^{-1}) = \sum_{\mathcal{O}_{\Delta,J}\in\phi\times\phi} a_{\mathcal{O}_{\Delta,J}} C_J^{(\nu)}\left(\frac{1}{2}(w+w^{-1})\right) r^{\Delta-2\Delta_\phi}, \tag{2.7}$$

where

$$a_{\mathcal{O}} \coloneqq \frac{f_{\phi\phi\mathcal{O}} b_{\mathcal{O}}}{c_{\mathcal{O}}} \frac{J!}{2^J(\nu)_J}, \qquad \nu \coloneqq \frac{d-2}{2}, \tag{2.8}$$

and $C_J^{(\nu)}(\eta)$ are Gegenbauer polynomials. The coefficients $f_{\phi\phi\mathcal{O}}$ are three-point function coefficients in the zero-temperature theory, and $b_{\mathcal{O}}$ are the thermal one-point coefficients. Δ, J represent the scaling dimension and spin of each contributing operator $\mathcal{O}$ in the OPE. This expansion

[5] As expected, the one-point functions of primary operators, other than the identity, vanish at zero temperature ($\beta\to\infty$) and is in agreement with the conformal Ward identities.

is convergent for $r < 1$. Earlier discussions of thermal two-point functions and OPEs have appeared in [21,22].

We note in passing that for later purposes, it will be convenient to write (2.7) as a power-series in $z, \bar{z}$. We can achieve that by expanding the Gegenbauer polynomials using the identity

$$C_J^{(\nu)}\left(\frac{1}{2}\left(\sqrt{\frac{z}{\bar{z}}}+\sqrt{\frac{\bar{z}}{z}}\right)\right)=\sum_{s=0}^{J} p_s(J) z^{-\frac{J}{2}+s}\bar{z}^{\frac{J}{2}-s}, \tag{2.9}$$

where

$$p_s(J)=\frac{\Gamma(J-s+\nu)\Gamma(s+\nu)}{\Gamma(J-s+1)\Gamma(s+1)}\frac{1}{\Gamma(\nu)^2}. \tag{2.10}$$

In this manner,

$$g(z,\bar{z})=\sum_{\mathcal{O}_{\Delta,J}\in\phi\times\phi} a_{\mathcal{O}_{\Delta,J}}\sum_{s=0}^{J} p_s(J) z^{h-\Delta_\phi+s}\bar{z}^{\bar{h}-\Delta_\phi-s}, \tag{2.11}$$

where

$$h\coloneqq\frac{\Delta-J}{2},\qquad \bar{h}\coloneqq\frac{\Delta+J}{2}. \tag{2.12}$$

Substituting the OPE (2.7) into the KMS condition (2.6) yields an equation of the form

$$\sum_{\mathcal{O}_{\Delta,J}\in\phi\times\phi} a_{\mathcal{O}_{\Delta,J}}\left[C_J^{(\nu)}\left(\frac{1}{2}(w+w^{-1})\right)r^{\Delta-2\Delta_\phi} - C_J^{(\nu)}\left(\frac{1}{2}(\tilde{w}+\tilde{w}^{-1})\right)\tilde{r}^{\Delta-2\Delta_\phi}\right]=0, \tag{2.13}$$

where we used the parametrization $z = rw$, $1-z=\tilde{r}\tilde{w}$. This equation requires both OPEs to be valid, namely,

$$r<1,\qquad \tilde{r}<1. \tag{2.14}$$

Following familiar nomenclature from the analysis of four-point functions at zero temperature, we will sometimes call the OPE expansion around $r=0$ the s-channel, and the OPE expansion around $\tilde{r}=0$ the t-channel.

Equation (2.13) can be viewed as an infinite set of nontrivial sum rules for the one-point function coefficients $b_{\mathcal{O}}$ (or, equivalently, the coefficients $a_{\mathcal{O}}$). These sum rules can be reformulated in many different ways, e.g., by acting on the lhs of (2.13) with arbitrary linear functionals. In the numerical implementations of later sections, we will employ sum rules that arise from the point-wise evaluation of the lhs on a grid of different points on the allowed region of OPE-convergence on the z-plane.

The goal of the thermal bootstrap program is to flesh out explicit constraints on individual one-point coefficients assuming knowledge about the zero-temperature CFT data (scaling dimensions and three-point function coefficients). As we discussed in the introduction, this task is complicated by the fact that the solutions to the KMS condition with a fixed spectrum of scaling dimensions is not, in general, unique and by the fact that hard truncations (even with some universal fixed approximation of the high-scaling dimension contributions) cannot recover this ambiguity. A proper analysis of the KMS condition requires the inclusion of the *full*, infinite set of contributions to the OPE expansions. In the next section, we discuss the first step toward the implementation of such an analysis.

III. OPE TAILS AND THERMAL DISPERSION RELATIONS

Dispersion relations for scalar two-point functions on $S^1\times\mathbb{R}^{d-1}$ were discussed for the first time, to our knowledge, in Ref. [23]. In Appendix A, we discuss a variety of dispersion relations that can be deduced either with a straightforward use of Cauchy's theorem or with the use of a thermal Lorentzian inversion formula [20]. In this section, we present, without detailed explanations, a subtracted dispersion relation that will be particularly useful in the following sections. For an explicit derivation of this dispersion relation we refer the reader to Appendix A.

A two-point function $g(z,\bar{z})$ of identical scalars ϕ at finite temperature, (2.1), can be expressed in terms of its discontinuity,

$$\mathrm{Disc}[g(z,\bar{z})]=-i(g(z+i\epsilon,\bar{z})-g(z-i\epsilon,\bar{z})) \tag{3.1}$$

across $z\in(-\infty,-1)\cup(1,\infty)$ as follows[6]:

$$g(rw,rw^{-1})=\sum_{J=0}^{J_*}\sum_{\Delta} a_{\Delta,J}C_J^{(\nu)}\left(\frac{1}{2}(w+w^{-1})\right)r^{\Delta-2\Delta_\phi}+2\left(\int_{-\infty}^{-r^{-1}}+\int_{r^{-1}}^{\infty}\right)dw'\mathcal{K}_{J_*}(w,w')\mathrm{Disc}[g(rw',rw'^{-1})], \tag{3.2}$$

where $J_*\geq 0$ is an arbitrary spin cutoff. As detailed in Appendix A, J_* needs to be selected above some number J_0 (related to the Regge behavior of the two-point function), otherwise Eq. (3.2) includes an extra arc contribution [see Eq. (A19) for details]. In what follows, we assume $J_*\geq J_0$. The J_*-dependent kernel $\mathcal{K}_{J_*}(w,w')$ has a universal contribution (independent of the spacetime dimension) $\mathcal{K}(w,w')$, and a spacetime-dependent part

[6]For notational convenience, here, and in the rest of the text, we are changing notation from $a_{\mathcal{O}_{\Delta,J}}\to a_{\Delta,J}$.

$$\mathcal{K}_{J_*}(w,w') \coloneqq \frac{1}{2}\mathcal{K}(w,w') - w'^{-1}(w'-w'^{-1})^{2\nu} \times \left[\sum_{J=0}^{J_*} K_J C_J^{(\nu)}\left(\frac{1}{2}(w+w^{-1})\right) F_J(w'^{-1})\right], \tag{3.3}$$

where

$$\mathcal{K}(w,w') \coloneqq \frac{1}{2\pi w'}\frac{w'^2-1}{(w'-w)(w'-w^{-1})}, \tag{3.4}$$

$$K_J \coloneqq \frac{\Gamma(J+1)\Gamma(\nu)}{4\pi\Gamma(J+\nu)}, \tag{3.5}$$

$$F_J(w) = w^{J+d-2}{}_2F_1\left(J+d-2, \frac{d}{2}-1, J+\frac{d}{2}, w^2\right). \tag{3.6}$$

In Eq. (3.2) the integrated discontinuity term on the second line is capturing the full contribution of operators with spin $J > J_*$. The contributions of the operators with $J \leq J_*$ are captured by the sum in the first term on the rhs of (3.2). This sum is a truncated version of the thermal OPE in spin. It still involves an infinite number of contibutions with arbitrarily high scaling dimension Δ at spins at or below J_*.

In these expressions, J_* is a free integer parameter. As an extreme choice, we can remove completely the truncated OPE term and allow the full two-point function $g(z,\bar{z})$ to be captured by the integrated discontinuity with the universal kernel $\frac{1}{2}\mathcal{K}(w,w')$ and a potential arc contribution that is detailed in Appendix A. On the other end of extreme choices, we can increase J_* to an arbitrarily high value in order to include a very large number of spins in the truncated OPE term. As we do that, the contribution of the integrated discontinuity decreases, and [as one can easily check from the explicit definition (3.3)] the kernel $\mathcal{K}_{J*}(w,w')$ suppresses more and more the integral away from the branch-cut endpoints at $z=-1$, 1.

To summarize, a subtracted thermal dispersion relation, (3.2), allows us to capture all the "large"-spin[7] contributions to the thermal block expansion in terms of the discontinuity of the two-point function. In the next section, we will use approximations of the discontinuity in order to approximate the "large"-spin contributions to the OPE. These will be the only approximations we implement to the KMS condition.

Now that we have a way to deal with the "large"-spin part, let us elaborate on the "low"-spin part of Eq. (3.2). Using a spin-dependent cutoff $\Delta_*(J)$, we can recast the OPE term on the rhs of Eq. (3.2) as

$$\sum_{J=0}^{J_*}\sum_{\Delta} a_{\Delta,J} C_J^{(\nu)}\left(\frac{1}{2}(w+w^{-1})\right) r^{\Delta-2\Delta_\phi} = \sum_{J=0}^{J_*}\sum_{\Delta\leq\Delta_*(J)} a_{\Delta,J} C_J^{(\nu)}\left(\frac{1}{2}(w+w^{-1})\right) r^{\Delta-2\Delta_\phi} + \sum_{J=0}^{J_*}\sum_{\Delta>\Delta_*(J)} a_{\Delta,J} C_J^{(\nu)}\left(\frac{1}{2}(w+w^{-1})\right) r^{\Delta-2\Delta_\phi}. \tag{3.7}$$

The first term on the rhs of this expression is a truly truncated OPE that involves a finite number of terms. The second term involves an infinite number of contributions from operators with arbitrarily high scaling dimension. We recast it in the form

$$\sum_{J=0}^{J_*}\sum_{\Delta>\Delta_*(J)} a_{\Delta,J} C_J^{(\nu)}\left(\frac{1}{2}(w+w^{-1})\right) r^{\Delta-2\Delta_\phi} = \sum_{J=0}^{J_*} A_{\Delta_*(J),J}(r) C_J^{(\nu)}\left(\frac{1}{2}(w+w^{-1})\right), \tag{3.8}$$

in terms of a finite new set of one-dimensional (tail) functions

$$A_{\Delta_*(J),J}(r) \coloneqq \sum_{\Delta>\Delta_*(J)} a_{\Delta,J} r^{\Delta-2\Delta_\phi}. \tag{3.9}$$

Notice that up to a subtraction of a finite number of contributions with $\Delta \leq \Delta_*(J)$, the tail functions $A_{\Delta_*(J),J}(r)$ are simply a Gegenbauer projection of the full thermal two-point functions to a specific spin [similar to Eq. (B3)]. In that sense, these functions are always well defined, and the sums in (3.9) are suitably convergent.

Putting everything together, we can write

$$g(rw,rw^{-1}) = \sum_{J=0}^{J_*}\sum_{\Delta\leq\Delta_*(J)} a_{\Delta,J} C_J^{(\nu)}\left(\frac{1}{2}(w+w^{-1})\right) r^{\Delta-2\Delta_\phi} + \sum_{J=0}^{J_*} A_{\Delta_*(J),J}(r) C_J^{(\nu)}\left(\frac{1}{2}(w+w^{-1})\right) + 2\left(\int_{-\infty}^{-r^{-1}} + \int_{r^{-1}}^{\infty}\right) dw' \mathcal{K}_{J_*}(w,w') \times \mathrm{Disc}[g(rw',rw'^{-1})]. \tag{3.10}$$

We emphasize that this is an exact relation with freely tunable scaling-dimension and spin cutoffs $\Delta_*(J)$, J_*, respectively. This relation forms the basis of the bootstrap approach proposed in this paper.

IV. KMS CONDITION 2.0

We can use Eq. (3.10) to reformulate the KMS condition (2.6) in terms of a finite number of CFT data, a finite

[7]To avoid potential confusions, we repeat that "large"-spin in this context refers to spins $J > J_*$. Similarly, "low"-spin refers to spin $J \leq J_*$. The value of the cutoff J_* remains arbitrary in this language.

number of tail functions and the discontinuity. Our ultimate goal is to approximate the contribution of the discontinuity and use the resulting approximate crossing equation to bootstrap the explicit low-lying CFT data, as well as the corresponding tail functions. To our knowledge, this is the first general attempt to dynamically determine entire OPE tails (at fixed spin) using bootstrap methods.

A. A reformulation of the exact KMS condition

Employing the decomposition (3.10), the *exact* KMS condition

$$g(z,\bar{z}) = g(1-z,1-\bar{z}) \tag{4.1}$$

becomes (adopting the notation $z = rw$, $1-z = \tilde{r}\,\tilde{w}$)

$$\begin{aligned}
&\sum_{J=0}^{J_*}\sum_{\Delta\le\Delta_*(J)} a_{\Delta,J}\left[C_J^{(\nu)}\left(\frac{1}{2}(w+w^{-1})\right)r^{\Delta-2\Delta_\phi} - C_J^{(\nu)}\left(\frac{1}{2}(\tilde{w}+\tilde{w}^{-1})\right)\tilde{r}^{\Delta-2\Delta_\phi}\right]\\
&\quad+\sum_{J=0}^{J_*}\left[A_{\Delta_*(J),J}(r)C_J^{(\nu)}\left(\frac{1}{2}(w+w^{-1})\right) - A_{\Delta_*(J),J}(\tilde{r})C_J^{(\nu)}\left(\frac{1}{2}(\tilde{w}+\tilde{w}^{-1})\right)\right]\\
&\quad+\mathbb{T}_{\mathrm{Disc}}[J_*;rw,rw^{-1}] - \mathbb{T}_{\mathrm{Disc}}[J_*;\tilde{r}\,\tilde{w},\tilde{r}\tilde{w}^{-1}] = 0,
\end{aligned} \tag{4.2}$$

where

$$\mathbb{T}_{\mathrm{Disc}}[J_*;rw,rw^{-1}] \coloneqq 2\left(\int_{-\infty}^{-r^{-1}} + \int_{r^{-1}}^{\infty}\right)dw'\mathcal{K}_{J_*}(w,w')\mathrm{Disc}[g(rw',rw'^{-1})] \tag{4.3}$$

is the contribution of the integrated discontinuity. Equation (4.2) imposes a necessary condition on $g(z,\bar{z})$ but is not practically useful unless we find a sensible way to approximate the discontinuity and the tail functions $A_{\Delta_*(J),J}(r)$. This is our next goal.

B. Approximations

1. Capturing the discontinuity

A common way to approximate discontinuities of correlation functions is through the use of the first few terms of the OPE in the crossed channel. In several examples, and in different contexts, it has been observed that such approximations can work well [20,24–34]. A positive aspect of our setup is that the systematic error introduced by such an *ad hoc* approximation can be reduced by increasing J_*. Indeed, the discontinuity appears inside an integral with kernel $\mathcal{K}_{J_*}$, which naturally suppresses the contributions of high-twist operators in the crossed-channel expansion of the discontinuity (as we will see more clearly in a moment). The suppressing effects of $\mathcal{K}_{J_*}$ can also ameliorate another problem.

At finite temperature, a well-known difficulty in the above-mentioned approximation scheme for the discontinuity stems from the need to use the OPE simultaneously in both the s- and t-channels. The convergence properties of the corresponding OPEs require both $r<1$ and $\tilde{r}<1$. As a result, when we integrate the discontinuity along $z\in(-\infty,-1)\cup(1,\infty)$, a large part of the integrals in (4.2) lies outside the region of convergence of the crossed channel in the approximation of $\mathrm{Disc}[g]$. Specifically, it is not possible to use the crossed-channel OPE to approximate the discontinuity for $z\in(-\infty,-2)\cup(2,\infty)$.[8] This difficulty was also noted in the context of the Lorentzian OPE inversion of thermal two-point functions in [20].

To address this problem, we can restrict by fiat the integrals in (3.10) over $rw'\in(-\infty,-1)\cup(1,\infty)$ [or equivalently $w'\in(-\infty,-r^{-1})\cup(r^{-1},\infty)$] to the subregion of common s- and t-channel convergence $w'\in(-2r^{-1},-r^{-1})\cup(r^{-1},2r^{-1})$. This introduces an additional systematic error because we are ignoring the integral of the discontinuity in the excised w'-region, but this error is naturally suppressed by the kernel $\mathcal{K}_{J_*}$ and becomes increasingly smaller as one increases J_*. As a result, $\mathcal{K}_{J_*}$ can be used to suppress at the same time both errors introduced by the range of integration and by the use of hard truncations of the OPE in the modeling of the discontinuity.

With these caveats in place, it will be useful to flesh out the precise way in which we approximate the contribution of the discontinuities in Eq. (3.10) and the corresponding crossing equation (4.2).

Along the right truncated branch cut $[z' \coloneqq rw'\in(1,2)]$, we use the KMS condition and the expressions in (2.11) to write

[8] We note in passing that this problem does not arise for similar schemes in zero-temperature four-point crossing equations that employ dispersion relations with integrated double discontinuities [12].

$$\begin{aligned}\mathrm{Disc}[g(z',\bar{z}')] &= \mathrm{Disc}[g(1-z',1-\bar{z}')] \\ &\simeq \sum_{\mathcal{O}_{\Delta,J}\text{truncated}} a_{\Delta,J} \sum_{s=0}^{J} p_s(J)\mathrm{Disc}[(1-z')^{h_{\mathcal{O}}-\Delta_\phi+s}](1-\bar{z}')^{\bar{h}_{\mathcal{O}}-\Delta_\phi-s} \\ &= -2 \sum_{\mathcal{O}_{\Delta,J}\text{truncated}} a_{\Delta,J} \sum_{s=0}^{J} p_s(J) \sin\left[\pi\left(\frac{\tau_{\mathcal{O}}}{2}-\Delta_\phi+s\right)\right](z'-1)^{\frac{\tau_{\mathcal{O}}}{2}-\Delta_\phi+s}(1-\bar{z}')^{\frac{\tau_{\mathcal{O}}}{2}-\Delta_\phi+J-s}. \end{aligned} \tag{4.4}$$

Here, we have approximated the discontinuity using a truncated OPE in the crossed channel, where the sum $\sum_{\mathcal{O}_{\Delta,J}\text{truncated}}$ runs over a subset of leading-twist operators. In the third line we used the explicit expression for the discontinuity of a generic power and reformulated the quantities $h=\frac{\Delta-J}{2}$, $\bar{h}=\frac{\Delta+J}{2}$ in terms of the twist $\tau=\Delta-J$. Similarly, for the integral along the left truncated branch cut $(rw' \in (-2,-1))$ we set

$$\begin{aligned}\mathrm{Disc}[g(z',\bar{z}')] &= \mathrm{Disc}[g(1+z',1+\bar{z}')] \\ &\simeq \sum_{\mathcal{O}_{\Delta,J}\text{truncated}} a_{\Delta,J} \sum_{s=0}^{J} p_s(J)\mathrm{Disc}[(1+z')^{h_{\mathcal{O}}-\Delta_\phi+s}](1+\bar{z}')^{\bar{h}_{\mathcal{O}}-\Delta_\phi-s} \\ &= 2 \sum_{\mathcal{O}_{\Delta,J}\text{truncated}} a_{\Delta,J} \sum_{s=0}^{J} p_s(J) \sin\left[\pi\left(\frac{\tau_{\mathcal{O}}}{2}-\Delta_\phi+s\right)\right](-z'-1)^{\frac{\tau_{\mathcal{O}}}{2}-\Delta_\phi+s}(1+\bar{z}')^{\frac{\tau_{\mathcal{O}}}{2}-\Delta_\phi+J-s}. \end{aligned} \tag{4.5}$$

It is straightforward to numerically check that the kernel $\mathcal{K}_{J_*}$ suppresses the integrands along the left and right branch cuts away from the points $z' \simeq -1$ and $z' \simeq 1$. We can, therefore, see explicitly in Eqs. (4.4) and (4.5) how the contributions of higher-twist operators are suppressed.

In summary, after the implementation of the above approximations, we obtain an approximate version of $\mathbb{T}_{\mathrm{Disc}}[J_*, rw, rw^{-1}]$ in Eq. (4.3), which will be denoted from now on as $\mathbb{T}^{(\mathrm{approx})}_{\mathrm{Disc}}[J_*, \{a\}; rw, rw^{-1}]$. This function involves integrals along the truncated branch cuts $(rw' \in (-2,-1) \cup (1,2))$ and the approximate discontinuity, which involves a subset of the unknown coefficients $a_{\mathcal{O}}$. The notation $\{a\}$ refers collectively to this dependence.

2. Capturing the tail functions with Neural Networks

The second ingredient that we need to tackle efficiently in the crossing equation (3.10) is a flexible, generic way to model the multiple tail functions $A_{\Delta_*(J),J}(r)$. We propose the use of neural networks (NNs) for this purpose because of their versatility in expressing generic functions and because modern machine learning libraries, like `PyTorch` or `TensorFlow`, allow us to run efficiently nonconvex optimization problems involving such representations. In the applications below, we represent the tail functions using multi-branch MLPs. These are neural network architectures with an initial layer connected to $\frac{J_*}{2}+1$ subnets,[9] each consisting of two hidden layers. Each subnet is learning a single corresponding tail function $A_J(r)$. A graphical depiction of the architecture and a detailed description of the technical characteristics of the NNs that we used can be found in Appendix C. Throughout our computations, we employed networks that have tens of thousands of optimizable parameters. We will collectively call the vector of NN parameters $\vec{\theta}$ and denote the NN representations of the tail functions $A_{\Delta_*(J),J;\theta}(r)$.

We note in passing that the use of fully connected feedforward NNs with multiple output channels (instead of multi-branch MLPs) was much less efficient, yielding losses that were several orders of magnitude higher than the multi-branch MLPs.

3. Features of the approximate KMS condition

We are now in prime position to write down an approximate KMS condition that is amenable to numerical analysis and goes beyond the traditional hard truncation schemes or soft truncation schemes that are based on *ad hoc* tail approximations. The form of the new approximate KMS condition is

[9] J_* is an even integer in our setup.

$$\sum_{J=0}^{J_*}\sum_{\Delta\le\Delta_*(J)} a_{\Delta,J}\left[C_J^{(\nu)}\left(\frac{1}{2}(w+w^{-1})\right)r^{\Delta-2\Delta_\phi}-C_J^{(\nu)}\left(\frac{1}{2}(\tilde{w}+\tilde{w}^{-1})\right)\tilde{r}^{\Delta-2\Delta_\phi}\right]$$
$$+\sum_{J=0}^{J_*}\left[A_{\Delta_*(J),J;\boldsymbol{\theta}}(r)C_J^{(\nu)}\left(\frac{1}{2}(w+w^{-1})\right)-A_{\Delta_*(J),J;\boldsymbol{\theta}}(\tilde{r})C_J^{(\nu)}\left(\frac{1}{2}(\tilde{w}+\tilde{w}^{-1})\right)\right]$$
$$+\mathbb{T}^{(\text{approx})}_{\text{Disc}}[J_*,\{a\};rw,rw^{-1}]-\mathbb{T}^{(\text{approx})}_{\text{Disc}}[J_*,\{a\};\tilde{r}\,\tilde{w},\tilde{r}\tilde{w}^{-1}]=0. \tag{4.6}$$

The coefficients $\{a\}$ that appear in $\mathbb{T}^{(\text{approx})}_{\text{Disc}}$ may or may not be part of the truncated conformal block expansions in the first line. If they do not appear in the first line with spin $J\le J_*$, then their contributions are necessarily part of the tail functions approximated by NNs on the second line.

It is useful to highlight the following features of Eq. (4.6):

(1) In the exact KMS condition (4.2), one can freely tune the parameter J_*. The presence of the approximations in (4.6) can pose some limitations. For example, when $J_*=0$, only a single tail for scalar operators needs to be included, which makes the second line of (4.6) relatively simple with a single unknown function. However, this is also where the approximation in $\mathbb{T}^{(\text{approx})}_{\text{Disc}}$ performs worst. As we increase J_*, the approximation in $\mathbb{T}^{(\text{approx})}_{\text{Disc}}$ improves, but the second line involves an increasing number of unknown functions, which complicates the numerical analysis of (4.6). This suggests that an intermediate, relatively low value of J_* may be numerically optimal. For example, $J_*=10$ would involve the presence of six tail functions, which are amenable to efficient optimization.

(2) The scaling dimension cutoffs $\Delta_*(J)$ controlling the number of terms that are included explicitly in the first line of (4.6) are free parameters. Their choice does not affect the accuracy of (4.6). When an operator is included in the truncated OPE of the first line of (4.6), we will say that *the operator is exposed*.

(3) The presence of the tail functions in our formulation allows us to bootstrap the contribution of an infinite number of operators in the thermal block expansion. As we remarked in the introduction, in any other truncation scheme based on a finite number of unknown parameters, the problem reduces to a linear regression problem with a unique solution. This immediately clashes with the fact that the KMS condition can have an infinite number of solutions and obstructs one's ability to discover them. By including unknown optimizable tail functions in our formulation, we obtain a more flexible framework that evades this problem. Our approach leads to nonconvex optimization problems, typically involving a rich landscape of many different solutions.

(4) In guided searches for specific solutions, the discontinuity in (4.6) plays an important role. Since discontinuities capture a more fundamental part of the correlation function (compared to the full correlation function), one can view $\mathbb{T}^{(\text{approx})}_{\text{Disc}}$ as a source for the exposed unknown coefficients $a_{\Delta,J}$ and the unknown tail functions $A_{\Delta_*(J),J;\boldsymbol{\theta}}(r)$.

We will revisit several specific aspects of points (3) and (4) when discussing holographic CFTs in Sec. VII. In that context, the KMS condition is known to have an infinite number of solutions corresponding to bottom-up holographic CFTs that are dual to arbitrary higher-derivative theories of gravity in anti–de Sitter (AdS).

4. KMS as a nonconvex optimization problem

To solve Eq. (4.6) numerically, we discretize the values of $(z,\bar{z})=(rw,rw^{-1})$ on a grid of finite points and evaluate the lhs of (4.6) on this grid to form a vector $\vec{F}(\vec{a},\vec{\theta})$ that depends algebraically on the (exposed) unknown parameters $a_{\Delta,J}$ and the NN parameters $\vec{\theta}$ that appear in (4.6). Then, we choose a non-negative loss function $\mathcal{L}(a,\vec{\theta})$ and perform nonconvex optimization to determine the configurations

$$(\vec{a}_*,\vec{\theta}_*)=\text{argmin}\,\mathcal{L}(\vec{a},\vec{\theta}), \tag{4.7}$$

which minimize the loss in the KMS condition. We perform the optimization using `Adam`, an optimizer based on the stochastic gradient descent that incorporates adaptive learning rates and momentum.

The choice of a point-wise evaluation of Eq. (4.6) is obviously one of the many possibilities one can employ. Another common choice in such problems includes the evaluation of derivatives at the crossing-symmetric point $z=\frac{1}{2}$. In this paper, we choose point-wise evaluation because it is a natural approach when one tries to determine directly the tail functions $A_{\Delta_*(J),J}(r)$. We work with a uniform grid on the z-plane that covers most of the common convergence region of the two conformal block expansions in the crossing equation. The grid of 243 points that we used throughout the paper appears in Fig. 1. It covers the region $|z|<0.95$, $|1-z|<0.95$.

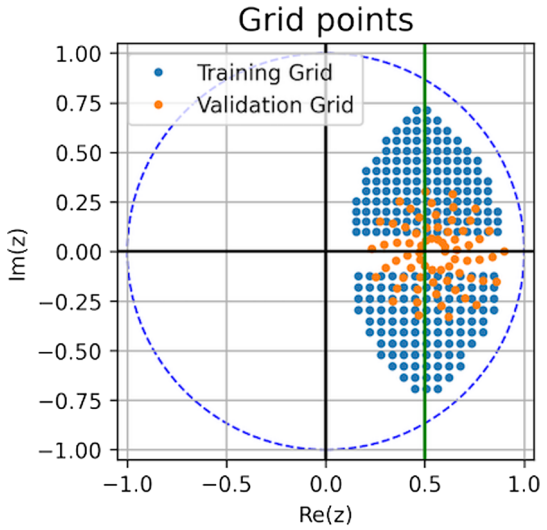


FIG. 1. Blue points represent a grid of 243 points on the complex z-plane used for optimization during NN training. The grid avoids the real axis and covers points in the region $|z| < 0.95$, $|1 - z| < 0.95$. The orange points represent the validation grid.

We implemented several different types of loss functions. We found that the following ones were the most promising.

Mean absolute loss $\mathcal{L}_{\overline{\text{abs}}}$. In this case, we define

$$\mathcal{L}_{\overline{\text{abs}}}(\vec{F}) \coloneqq \frac{1}{N}\sum_{i=1}^{N}|F_i|, \tag{4.8}$$

where N is the number of grid points and F_i the components of the KMS vector $\vec{F}$. We can apply this loss in generic situations. We optimize it by varying all the unknown parameters $(\vec{a}, \vec{\theta})$. One limitation of this loss function is that good minimum values cannot be determined *a priori*.

Another, related, common choice in this class of loss functions is the root mean square loss. We did not employ it extensively in this work.

Dot-product loss with no optimizable $a_{\Delta,J}$ *coefficients* $\mathcal{L}_{\text{dot}(0)}$. In some cases, we may want to optimize over a set of tail functions without varying any $a_{\Delta,J}$ coefficients. This can happen, for example, because we insert into the KMS condition some information about the $a_{\Delta,J}$ coefficients, and there are no other exposed optimizable a-coefficients outside the tail functions. To denote the absence of optimizable a-coefficients we use the subscript (0).

In such situations, the KMS vector $\vec{F}$ takes the form

$$\vec{F} = \vec{g}_\theta - \vec{h}, \tag{4.9}$$

where $\vec{h}$ is a fixed, known vector, and $\vec{g}_\theta$ is the vector that contains the optimizable NN parameters $\vec{\theta}$. We define

$$\mathcal{L}_{\text{dot}(0)}(\vec{F}(\vec{\theta})) \coloneqq 1 - \frac{|\vec{g}_\theta \cdot \vec{h}|}{|\vec{g}_\theta||\vec{h}|} + \left|1 - \frac{\vec{g}_\theta \cdot \vec{h}}{|\vec{h}|^2}\right|. \tag{4.10}$$

The first term, $1 - \frac{|\vec{g}_\theta \cdot \vec{h}|}{|\vec{g}_\theta||\vec{h}|}$, is 1 minus the absolute value of the cosine of the angle between the vectors $\vec{g}_\theta$ and $\vec{h}$. It is, therefore, manifestly non-negative. Its presence aims to suppress the component of $\vec{g}_\theta$ normal to $\vec{h}$. The square in the second term aims to equate the component of $\vec{g}_\theta$ parallel to $\vec{h}$ with $\vec{h}$. Indeed, in generic configurations of $\vec{\theta}$,

$$\vec{g}_\theta = c\vec{h} + \epsilon\hat{h}_\perp, \qquad \vec{h} \cdot \hat{h}_\perp = 0, \qquad |\hat{h}_\perp| = 1 \tag{4.11}$$

and

$$\begin{aligned}\mathcal{L}_{\text{dot}(0)}(\vec{F}(\vec{\theta})) &= \left(1 - \frac{1}{\sqrt{1 + \frac{\epsilon^2}{c^2|\vec{h}|^2}}}\right) + |1 - c| \\ &\underset{\epsilon \text{ small}}{\simeq} \frac{\epsilon^2}{2c^2|\vec{h}|^2} + |1 - c|.\end{aligned} \tag{4.12}$$

The goal of the optimization is to satisfy the KMS condition by minimizing $\frac{\epsilon}{c|\vec{h}|}$ simultaneously with $|1 - c|$.

Dot-product loss with a single optimizable $a_{\Delta,J}$ *coefficient* $\mathcal{L}_{\text{dot}(1)}$. In other situations, there will be a single exposed coefficient, call it a, to be optimized together with a number of tail functions. In that case, the form of the KMS vector is

$$\vec{F}(a, \vec{\theta}) = a\vec{f} + \vec{g}_\theta - \vec{h}, \tag{4.13}$$

where $\vec{f}$ is the crossed thermal block for the exposed operator with a thermal one-point function proportional to a, $\vec{g}_\theta$ is the sum of the contributions that depend on the NN parameters $\vec{\theta}$, and $\vec{h}$ a fixed vector of known contributions. Instead of varying simultaneously a and $\vec{\theta}$ in a loss function that contains both, we can minimize the loss function,

$$\mathcal{L}_{\text{dot}(1)}(\vec{F}(\vec{\theta})) \coloneqq 1 - \frac{|(\vec{g}_\theta - \vec{h}) \cdot \vec{f}|}{|\vec{g}_\theta - \vec{h}||\vec{f}|}, \tag{4.14}$$

with respect to the NN parameters $\vec{\theta}$ only, and at the very end set

$$a = -\frac{(\vec{g}_{\theta_*} - \vec{h}) \cdot \vec{f}}{|\vec{f}|^2}, \tag{4.15}$$

where $\vec{\theta}_* = \text{argmin}\, \mathcal{L}(\vec{F}(\vec{\theta}))$. The minimization of the loss (4.14) orients the vector $\vec{g}_\theta - \vec{h}$ parallel to the known vector $\vec{f}$, and Eq. (4.15) fixes the coefficient a to satisfy the KMS condition in the direction of $\vec{f}$.

This prescription is an obvious modification of the previous approach with loss (4.10). Compared to the absolute loss, here, we do not vary over the unknown OPE coefficient a during the optimization process. It is determined at the end of the computation by solving exactly one of the components of the KMS condition.

We used the subscript (1) in the notation of this loss to denote the presence of a single optimizable coefficient a. One can use this approach to determine multiple one-point coefficients exposing them sequentially one by one or by accordingly extending the prescription to handle multiple unknown one-point coefficients simultaneously.

V. GENERALIZED FREE FIELDS

In this section, we consider the theory of generalized free fields (GFFs), which can be solved analytically. As such, this theory provides a useful initial testing ground for the approach proposed above.

In the GFF CFT of a scalar primary field ϕ with scaling dimension Δ_ϕ (in any spacetime dimension $d \geq 2$), the thermal two-point function of ϕ can be expressed as a sum over images,

$$g(z,\bar{z}) = \sum_{m=-\infty}^{\infty} \frac{1}{[(m-z)(m-\bar{z})]^{\Delta_\phi}}. \tag{5.1}$$

The conformal block expansion of this expression receives contributions only from the identity operator ($\Delta = 0, J = 0$) and the double-twist operators $[\phi\phi]_{n,J}$ (with scaling dimensions $\Delta_{n,J} = 2\Delta_\phi + 2n + J$ and even spin $J = 2\ell$, $\ell = 0, 1, \ldots$). The double-twist operators are conformal primary composites of two ϕ's with J uncontracted spacetime derivatives and n Laplacians $\Box = \partial_\mu \partial^\mu$. The conformal block expansion of (5.1) reads

$$g(rw, rw^{-1}) = r^{-2\Delta_\phi} + \sum_{n=0}^{\infty}\sum_{\ell=0}^{\infty} a_{n,2\ell} \times C_{2\ell}^{(\nu)}\left(\frac{1}{2}(w + w^{-1})\right) r^{2(n+\ell)}, \tag{5.2}$$

with

$$a_{n,J} = 2\zeta(2\Delta_\phi + 2n + J)\frac{(J+\nu)(\Delta_\phi)_{J+n}(\Delta_\phi - \nu)_n}{n!(\nu)_{J+n+1}}, \tag{5.3}$$

and the Pochhammer symbol $(a)_n = \frac{\Gamma(a+n)}{\Gamma(a)}$.

In this case, the discontinuity that arises from the crossed t-channel expansion is particularly simple: *only the identity operator contributes*. As a result,

$$\mathbb{T}^{(\text{approx})}_{\text{Disc}}[J_*; rw, rw^{-1}] = 4\sin(\pi\Delta_\phi)\left[\int_{r^{-1}}^{2r^{-1}} dw'\, \mathcal{K}_{J_*}(w,w')(rw'-1)^{-\Delta_\phi}(1-rw'^{-1})^{-\Delta_\phi} - \int_{-2r^{-1}}^{-r^{-1}} dw'\, \mathcal{K}_{J_*}(w,w')(-rw'-1)^{-\Delta_\phi}(1+rw'^{-1})^{-\Delta_\phi}\right] \tag{5.4}$$

is a fixed function of the spacetime coordinates that does not involve any nontrivial thermal one-point functions. In Eq. (5.4), we assumed that the external scaling dimension Δ_ϕ is not an integer. When Δ_ϕ is an integer, the branch cut reduces to a pole, and the discontinuity receives δ-function contributions.

In what follows, we proceed to test the proposed approach for a specific randomly chosen case of a generalized free field with $\Delta_\phi = 1.68$ in $d = 4$ spacetime dimensions. We observed similar results for other values of Δ_ϕ, but do not report them explicitly here.

A. Tests of the approximate KMS condition

First, we want to examine how closely the approximate KMS condition captures the exact version as a function of J_*. For that purpose, we substitute the exact GFF solution (5.3) into several versions of the approximate KMS condition and evaluate the loss functions we formulated above. Low loss values would demonstrate that our scheme is based on the "correct" equations. This is important. Recall that one of the major issues in generic truncation schemes has been the inability to control the systematic error introduced by the truncation.

1. Loss on the analytic solution without exposed operators

In our first test, we consider the approximate KMS condition

$$r^{-2\Delta_\phi}-\tilde{r}^{-2\Delta_\phi}+\sum_{J=0}^{J_*}\left[A_{\Delta_*(J),J;\boldsymbol{\theta}}(r)C_J^{(\nu)}\left(\frac{1}{2}(w+w^{-1})\right)-A_{\Delta_*(J),J;\boldsymbol{\theta}}(\tilde{r})C_J^{(\nu)}\left(\frac{1}{2}(\tilde{w}+\tilde{w}^{-1})\right)\right]$$
$$+\mathbb{T}^{(\text{approx})}_{\text{Disc}}[J_*;rw,rw^{-1}]-\mathbb{T}^{(\text{approx})}_{\text{Disc}}[J_*;\tilde{r}\,\tilde{w},\tilde{r}\tilde{w}^{-1}]=0. \tag{5.5}$$

We expose only the contribution of the identity in the conformal block expansion and keep everything else inside the tail functions $A_{\Delta_*(J),J;\boldsymbol{\theta}}(r)$ with a very low scaling dimension cutoff $\Delta_*(J)$ placed at the unitarity bound.

We evaluate Eq. (5.5) on the $(z,\bar{z})$-grid of Fig. 1 with 243 points,[10] producing a 243-dimensional vector $\vec{F}$ of algebraic equations. In addition, we formulate a loss function that quantifies how far this vector is from the zero-vector. Following the discussion of Sec. IV B 4, we consider two options. The first is the mean absolute loss,

$$\mathcal{L}_{\overline{\text{abs}}}(\vec{F})=\frac{1}{N}\sum_{i=1}^{N}|F_i|. \tag{5.6}$$

The second is the dot-product loss $\mathcal{L}_{\text{dot}(0)}$ of Eq. (4.10),

$$\mathcal{L}_{\text{dot}(0)}(\vec{F})=1-\frac{|\vec{g}_\theta\cdot\vec{h}|}{|\vec{g}_\theta||\vec{h}|}+\left|1-\frac{\vec{g}_\theta\cdot\vec{h}}{|\vec{h}|^2}\right|, \tag{5.7}$$

where $\vec{g}_\theta$ is the vector of grid evaluations of the tail contributions in Eq. (5.5), and $\vec{h}$ is the vector of grid evaluations of the function

$$h(z,\bar{z})=-\{r^{-2\Delta_\phi}-\tilde{r}^{-2\Delta_\phi}+\mathbb{T}^{(\text{approx})}_{\text{Disc}}[J_*;rw,rw^{-1}]$$
$$-\mathbb{T}^{(\text{approx})}_{\text{Disc}}[J_*;\tilde{r}\,\tilde{w},\tilde{r}\tilde{w}^{-1}]\}. \tag{5.8}$$

Using the data of the exact solution (5.3) we can compute numerically the analytic values of all the tail functions. The values of the respective loss functions are presented in Table I for even $J_*=0,2,4,6,8$. As expected, the result exhibits visible improvement with increasing J_* both for $\mathcal{L}_{\overline{\text{abs}}}$ and $\mathcal{L}_{\text{dot}(0)}$. At $J_*=8$, we observe loss values at the order of 10^{-4} and 10^{-10}, respectively.

[10]In the grid of Figure 1, we avoided, on purpose, points on the real axis, which correspond to two-point functions at zero spatial separation σ. At those points, operators with different spin at the same scaling dimension make indistinguishable contributions to the two-point function. Notice, however, that the region around the real axis, which is not sampled during training, is sampled via the orange points during validation. In addition, the grids above and below the real axis are chosen with a slight displacement that makes them inequivalent under the symmetry $\sigma\to-\sigma$.

2. Loss on the analytic solutions with a single exposed operator

It is also instructive to repeat the test of Sec. V A 1 by exposing a number of operators. For example, we can write an approximate KMS condition where the leading scalar operator is exposed and all other CFT data are inside the tail functions. The leading scalar operator cannot be the double-twist operator $[\phi\phi]_{0,0}$ because this operator contributes a constant in the KMS condition, which cancels out after crossing. Interestingly, the next leading scalar operator $[\phi\phi]_{1,0}$ alone is subtle for the following reason.

The contribution of a double-twist operator $[\phi\phi]_{n,2\ell}$ to the crossed conformal block expansion in d- dimensions is

$$a_{n,2\ell}\left[r^{2n+J}C_J^{(\nu)}\left(\frac{1}{2}(w+w^{-1})\right)-\tilde{r}^{2n+J}C_J^{(\nu)}\left(\frac{1}{2}(\tilde{w}+\tilde{w}^{-1})\right)\right]. \tag{5.9}$$

For $(n,2\ell)=(1,0),(0,2)$, we observe the peculiar identity[11]

$$r^2C_0^{(\nu)}\left(\frac{1}{2}(w+w^{-1})\right)-\tilde{r}^2C_0^{(\nu)}\left(\frac{1}{2}(\tilde{w}+\tilde{w}^{-1})\right)$$
$$=\frac{1}{\nu(2\nu+1)}\left[r^2C_2^{(\nu)}\left(\frac{1}{2}(w+w^{-1})\right)-\tilde{r}^2C_2^{(\nu)}\left(\frac{1}{2}(\tilde{w}+\tilde{w}^{-1})\right)\right]$$
$$=r(w+w^{-1})-1, \tag{5.10}$$

which implies that the scalar operator $[\phi\phi]_{1,0}$ comes together with the spin-2 $[\phi\phi]_{0,2}$ operator in the KMS condition to contribute the single term

$$(a_{1,0}+\nu(2\nu+1)a_{0,2})(r(w+w^{-1})-1). \tag{5.11}$$

Notice that the combination

$$a_{\Delta=2\Delta_\phi+2}\coloneqq a_{1,0}+\nu(2\nu+1)a_{0,2} \tag{5.12}$$

is exactly the same type of combination of spin-dependent coefficients that appears in the zero-spatial-separation analysis of Ref. [15]. Accordingly, when we choose to expose the operators $[\phi\phi]_{1,0},[\phi\phi]_{0,2}$, the approximate KMS condition (in $d=4$ dimensions) becomes

[11]We have not observed similar identities for other operators at generic $(z,\bar{z})$ points.

$$r^{-2\Delta_\phi} - \tilde{r}^{-2\Delta_\phi} + (a_{1,0} + 3a_{0,2})(r(w + w^{-1}) - 1) + \sum_{J=0}^{J_*} \left[A_{\Delta_*(J),J;\boldsymbol{\theta}}(r) C_J^{(\nu)} \left(\frac{1}{2}(w + w^{-1}) \right) - A_{\Delta_*(J),J;\boldsymbol{\theta}}(\tilde{r}) C_J^{(\nu)} \left(\frac{1}{2}(\tilde{w} + \tilde{w}^{-1}) \right) \right] + \mathbb{T}^{(\text{approx})}_{\text{Disc}}[J_*; rw, rw^{-1}] - \mathbb{T}^{(\text{approx})}_{\text{Disc}}[J_*; \tilde{r}\,\tilde{w}, \tilde{r}\tilde{w}^{-1}] = 0, \tag{5.13}$$

with $\Delta_*(J=0) = \Delta_*(J=2) = 2\Delta_\phi + 2$ and all other $\Delta_*(J)$ at the unitarity bound.

As we discussed in Sec. IV B 4, we can bootstrap this equation using the dot-loss $\mathcal{L}_{\text{dot}(1)}$ of Eq. (4.14), where the optimizable parameters are only the NN parameters $\vec{\theta}$. The value of $a_{1,0} + 3a_{0,2}$ is determined at the end of the computation by Eq. (4.15). More specifically, in this context

$$\mathcal{L}_{\text{dot}(1)}(\vec{F}(\vec{\theta})) := 1 - \frac{|(\vec{g}_\theta - \vec{h}) \cdot \vec{f}|}{|\vec{g}_\theta - \vec{h}||\vec{f}|}, \tag{5.14}$$

where again $\vec{g}_\theta$ is the vector of grid evaluations of the tail contributions in Eq. (5.13), $\vec{h}is$ the vector of grid evaluations of the function $h(z, \bar{z})$ in Eq. (5.8), and $\vec{f}$ is the vector of grid evaluations of the function

$$f(z, \bar{z}) = r(w + w^{-1}) - 1. \tag{5.15}$$

In the second column of Table II, we present the corresponding numerical values of $\mathcal{L}_{\text{dot}(1)}$ on the grid of Fig. 1, when we insert the analytic values of the tail functions for $J_* = 2, 4, 6, 8$. The choice $J_* = 0$ is not viable here as we are exposing a spin-2 operator. In the third column of Table II, we list the values of $a_{1,0} + 3a_{2,0}$ that are deduced from the analytic tail functions and Eq. (4.15), which follows from the approximate KMS condition (5.13). Once again, it is satisfying to observe that the accuracy of the approximate KMS condition increases for increasing J_*, achieving $\mathcal{L}_{\text{dot}(1)}$ of the order of 10^{-9} at $J_* = 8$. We also notice a clear convergence of the approximate-KMS-derived value of $a_{1,0} + 3a_{2,0}$ toward the analytic value 15.06013.

TABLE I. The second column presents the mean absolute loss, Eq. (5.6), of the approximate KMS condition for the analytic solution with varying J_* values. The third column presents the dot loss, Eq. (5.7), for the analytic solution. The spacetime dimension is $d = 4$, the scaling dimension of the external operator is $\Delta_\phi = 1.68$, and the grid used is depicted in Fig. 1.

J_*	$\mathcal{L}_{\overline{\text{abs}}}$	$\mathcal{L}_{\text{dot}(0)}$
0	0.1177	0.0125
2	0.0300	1.3×10^{-5}
4	0.0067	1.6×10^{-7}
6	0.0013	3.3×10^{-9}
8	0.0002	0.8×10^{-10}

TABLE II. The second column presents the values of the loss function $\mathcal{L}_{\text{dot}(1)}$, Eq. (5.14), for the analytic values of the tails in the GFF theory. The third column presents the value of the $a_{1,0} + 3a_{2,0}$ combination that follows from the approximate KMS condition according to Eq. (4.15). As J_* increases, there is clear convergence toward the analytic result listed at the bottom of the table.

J_*	$\mathcal{L}_{\text{dot}(1)}$	$a_{1,0} + 3a_{2,0}$ from dot-loss
2	9.5×10^{-6}	15.16582
4	1.4×10^{-6}	15.07614
6	7.8×10^{-8}	15.06252
8	3.5×10^{-9}	15.06049
		Exact value: 15.06013

B. Tail bootstrap

In this section, we assume no input of thermal data from the analytic solution of the GFF theory and try to recover it by bootstrapping the approximate KMS condition (5.5). We set $J_* = 6$, we do not expose any operators, and we bootstrap the four tail functions $A_0(r), A_2(r), A_4(r), A_6(r)$. However, before proceeding to the actual computation, we need to make a preparatory comment about the training grid and a related addition to the loss functions.

1. Constraining the tail asymptotics with $\mathcal{L}_{\text{BC}}[A]$

We are optimizing by evaluating the approximate KMS condition on the grid of Fig. 1. This grid covers most of the common region of convergence of the s- and t-channel expansions, but not all of it. The OPE-convergence region is defined by the inequalities $|z| < 1$, $|1 - z| < 1$, whereas the grid of Figure 1 covers points in the region $|z| < 0.95$, $|1 - z| < 0.95$. We excluded points near the boundary because they were numerically more challenging.

With this choice, initial experimentation shows that the landscape of minima for the original loss functions $\mathcal{L}_{\overline{\text{abs}}}$ and $\mathcal{L}_{\text{dot}(0)}$ is complicated with many low-loss configurations, some of which are clearly distinct from the analytic GFF solution. For example, we frequently observed low-loss configurations with tail functions exhibiting $r \to 1$ asymptotics that are inconsistent with the general analysis of

Appendix B, which is based on the KMS condition. This is a clear indication that by not training on the full OPE-convergence region, we are failing to extract the complete information of the KMS condition, which in turn allows many more optimal configurations beyond the GFF solution.

To guide the search more efficiently, and toavoid spurious minima, it is helpful to supplement our optimization problem with additional information. An alternative to enlarging the training grid is to add a condition that enforces the KMS constraint governing the asymptotic behavior of tail functions as $r \to 1$.

In Appendix B, we use the KMS condition to argue that the tail functions $A_{2\ell}(r)$ in thermal two-point functions of identical scalars exhibit a universal singularity structure in the limit $r \to 1^-$ when the external scalar scaling dimension Δ_ϕ obeys the inequality $\Delta_\phi \geq \frac{d-1}{2}$. Focusing on this regime, we can use Eq. (B6) to approximate the numerical value of the optimizable tail functions $A_{2\ell}(r)$ at a value $r = r_0$ close to $r = 1$ in a theory-independent way. In the numerical results reported below, we used $r_0 = 0.9999$. In order to recover accurately the universal asymptotics of Eq. (B6), we need to go exponentially close to $r = 1$. Even at $r_0 = 0.9999$, the application of (B6) introduces a systematic error. For example, when $\Delta_\phi = 1.68$ in $d = 4$ dimensions, the exact GFF values of $A_J(0.9999)$ (for $J = 0,\ 2,\ 4,\ 6$) are (83.044,250.765,413.139,573.387) and the corresponding values predicted by Eq. (B6) are (84.822,254.467,424.111,593.756).

Practically, we enforced the asymptotic boundary condition at r_0 by adding to the loss functions a new term $\mathcal{L}_{\rm BC}[A]$,

$$\mathcal{L} \longrightarrow \mathcal{L} + \mathcal{L}_{\rm BC}[A]. \tag{5.16}$$

We found through preliminary testing that $\mathcal{L}_{\rm BC}[A]$ can be implemented most efficiently in the following manner. For starters, we set

$$\mathcal{A}_{J;\theta}(r) \coloneqq \operatorname{arcsinh}\left(\frac{A_{J,\theta}(r)}{2}\right),$$
$$\mathcal{A}_{J;\rm asymp}(r) \coloneqq \operatorname{arcsinh}\left(\frac{A_{J,\rm asymp}(r)}{2}\right), \tag{5.17}$$

where $A_{J,\theta}(r)$ are the optimizable tail functions, and $A_{J,\rm asymp}(r)$ are the functions obtained by the asymptotic formula (B6). With these specifications, we defined

$$\mathcal{L}_{\rm BC}[A] \coloneqq \left(\frac{1}{\frac{J_*}{2}+1}\sum_{\ell=0}^{\frac{J_*}{2}}(\mathcal{A}_{2\ell;\theta}(r_0) - \mathcal{A}_{2\ell;\rm asymp}(r_0))^2\right)^2$$
$$+ \frac{1}{\frac{J_*}{2}+1}\sum_{\ell=0}^{\frac{J_*}{2}}(\mathcal{A}_{2\ell;\theta}(r_0) - \mathcal{A}_{2\ell;\rm asymp}(r_0))^2 + \frac{1}{2}\sum_{\ell=0}^{1}|\mathcal{A}_{2\ell;\theta}(r_0) - \mathcal{A}_{2\ell;\rm asymp}(r_0)|. \tag{5.18}$$

In this expression, the role of the arcsinh transformation should be clear. It allows us to renormalize the large numbers that appear near the $r \to 1$ singularity, and as such it facilitates a more efficient optimization of the neural networks.

The specific structure of the loss function in (5.18) (with a quartic, a quadratic, and a linear term) is less obvious. We observed that with single-power losses the optimization was still unstable toward different types of configurations with comparable low loss, despite the removal of the previous spurious minima with the wrong asymptotics. The variation was especially pronounced for the spin-0 and spin-2 tail functions.[12] The specific form in (5.18) was found to reduce this instability. We will argue momentarily that (5.18) not only reduces the instability but also pushes the result toward the GFF configuration.[13]

[12]That is also the origin of the last term in (5.18), which involves only the spin-0 and spin-2 tail functions.

[13]Unfortunately, (5.18) does not solve completely the problem of spurious minima. We will return to this issue below.

2. Bootstrap with $\mathcal{L}_{\overline{\rm abs}}$

Using the loss function $\mathcal{L}_{\overline{\rm abs}}(\vec{F}) + \mathcal{L}_{\rm BC}[A]$ we obtained the results depicted in Fig. 2. The plots in Fig. 2 present averages and standard deviations based on the 10 lowest-loss configurations that resulted after 50 K epochs for 1 K independent runs on the QMUL Apocrita cluster (with the same neural network hyperparameters). It is useful to highlight the following observations:

(a) The optimization yields low-loss configurations with loss comparable to that of the analytic GFF solution ($1.93 \times 10^{-3} \pm 1.06 \times 10^{-4}$ vs 1.29×10^{-3} for GFF).

(b) The leading low-loss configurations reproduce the analytic, higher-spin tail functions A_4, A_6 rather accurately and without significant variation. In contrast, there is significant variation in the profiles of the predicted functions A_0, A_2 with mean curves that are substantially far from the analytic GFF solution. Nevertheless, we observe that the obtained A_0, A_2 configurations have a total contribution to the crossing equation, which is comparable to that of the analytic GFF solution. This is apparent from the plot of the

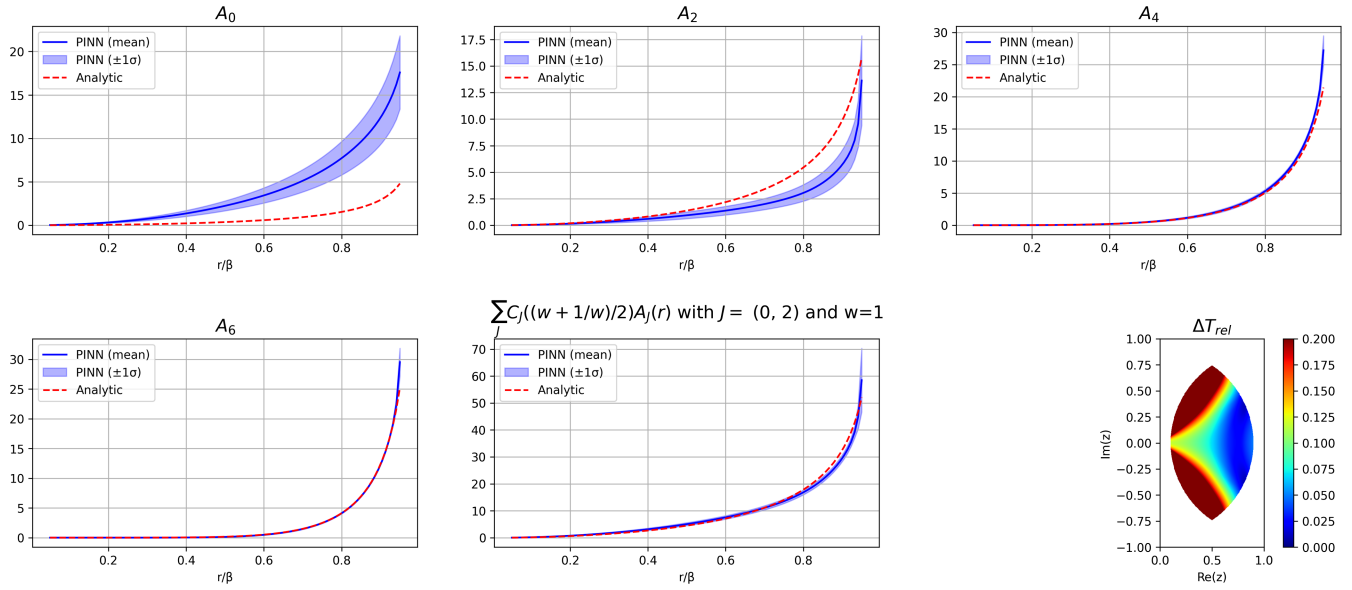


FIG. 2. Plots depicting the results obtained with $\mathcal{L}_{\overline{\text{abs}}} + \mathcal{L}_{\text{BC}}[A]$ in the $d = 4$ GFF theory with $\Delta_\phi = 1.68$ and $J_* = 6$ without any exposed operators and asymptotic boundary conditions for the tail functions at $r = 0.9999$. The results are based on the 10 lowest-loss configurations within a pool of 1 K independent training runs for 50 K epochs. The mean loss of these configurations is $1.93 \times 10^{-3} \pm 1.06 \times 10^{-4}$, whereas the absolute loss of the analytic GFF configuration is 1.29×10^{-3}. The first 4 plots depict (in blue) the mean and 1σ deviation of the predicted tail functions. The dashed red curves represent the exact, analytic result of the GFF theory. The middle plot on the second line depicts the combined contribution of the two leading tail functions, A_0 and A_2, to the conformal block expansion of the thermal correlator for $w = 1$ and its comparison against the analytic expression (in red). The final plot at the bottom right is a heatmap of the relative difference (5.20) between the predicted and analytic values in the training region.

combined conformal block contribution

$$T_{0+2}(r,w) \coloneqq \sum_{J=0,2} A_J(r) C_J\left(\frac{1}{2}(w+w^{-1})\right) \tag{5.19}$$

for $w = 1$ and from the heatmap of the relative difference

$$\Delta T_{\text{rel}} \coloneqq \frac{|T_{0+2;\text{predicted}}(r,w) - T_{0+2;\text{analytic}}(r,w)|}{|T_{0+2;\text{predicted}}(r,w)| + |T_{0+2;\text{analytic}}(r,w)|} \tag{5.20}$$

between predicted and analytic tails inside the training region of Fig. 1. The $w = 1$ curve in particular, shows that our setup has no trouble reproducing the spin-independent data $a_\Delta = \sum_J A_J(r) C_J(1)$ that appear in the bootstrap at zero spatial separation, [15,18,19].

The persistence of a landscape of many, roughly degenerate, low-loss A_0, A_2 configurations [despite the addition of (5.18)] implies that we are still not extracting the full information of the KMS condition. We noted previously, in the expression (5.11), that there is an analytic ambiguity between the data $a_{1,0}$ and $a_{0,2}$ that cannot be resolved in the KMS condition that we are studying. We have not found, and do not expect, any similar, exact, analytic ambiguities that involve the full tail functions A_0 and A_2, so what we observe here is a numerical accident, whose origin lies in the form of the landscape of minima of the loss function. Indeed, the specific details of the loss function can influence such features, and we will obtain further evidence of this statement in the next Sec. V B 3.

(c) Although the mean predicted A_0, A_2 curves are far from the corresponding analytic GFF curves, some of the low-loss runs produced configurations that are close to the analytic GFF result. To illustrate this point, we present the predictions of the 10th in order-of-loss run in Fig. 12 in Appendix G. This observation exhibits the fact that the optimization algorithm can indeed sample low-loss configurations in the vicinity of the GFF solution.

(d) We have also checked that the low-loss configurations exhibit comparable low validation loss on the grid of orange points in Fig. 1. This implies that we can satisfy well the KMS condition in the heart of the region of OPE convergence (around $z = 0.5$), even at points which were not included in the optimization grid. In sharp contrast, the loss rises significantly outside the training region, toward the boundary $\{|z| = 1 \cup |1 - z| = 1\}$, for the optimal configurations that deviate from the GFF solution.

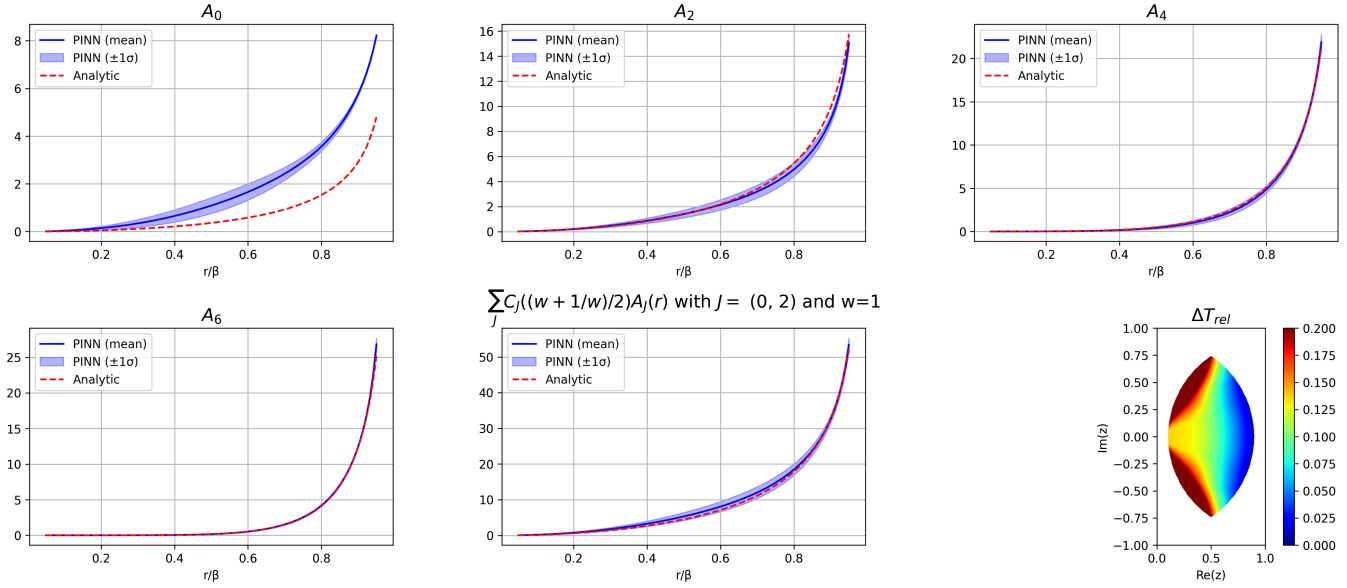


FIG. 3. Plots of the predicted tail functions obtained with $\mathcal{L}_{\text{dot}(0)} + \mathcal{L}_{\text{BC}}$ for $J_* = 6$ in the $d = 4$ GFF theory for $\Delta_\phi = 1.68$. The obtained results are based on the 10 lowest-loss configurations for 1 K independent runs of 50 K epochs. The mean loss, $3.54 \times 10^{-9} \pm 6.4 \times 10^{-10}$, should be compared to the loss of the analytic GFF solution 3.35×10^{-9}. The middle plot in the second line represents the combined contribution to the two-point function of the A_0, A_2 tails at $w = 1$. The heatmap depicts the relative difference (5.20) inside the training region.

3. *Bootstrap with* $\mathcal{L}_{\text{dot}(0)}$

Next, we repeat the same exercise using the loss function $\mathcal{L}_{\text{dot}(0)}(\vec{F}) + \mathcal{L}_{\text{BC}}[A]$. The statistics collected from the 10 lowest-loss configurations from a pool of 1 K independent runs with 50 K epochs are depicted in Fig. 3.

In this case, we observe that

(a) The optimization with the dot-loss (5.7) is more efficient compared to the absolute loss optimization, producing visibly closer configurations to those of the analytic GFF solution. This is apparent when one compares the A_0, A_2 plots and the corresponding heatmaps in Figs. 2 and 3. We additionally observe that the dot-loss-optimized NNs achieve a very low loss with values $3.54 \times 10^{-9} \pm 6.4 \times 10^{-10}$, which are close to the loss of the analytic GFF solution 3.35×10^{-9}.

(b) The issue with the numerical ambiguity of the A_0, A_2 tail functions still remains, but it has improved compared to Fig. 2.

(c) Similar to the case of the absolute loss, we can also verify here that the dot-loss optima satisfy the KMS condition well in the orange validation grid of Fig. 1 around $z = 0.5$ but do poorly near the boundary of the OPE convergence region for configurations that deviate from the GFF solution. For example, the explicit validation of our configurations in a ring region $(\{0.975 < |z| < 0.99\} \cup \{0.975 < |1 - z| < 0.99\})$ gave losses of the order of 10^{-3} when the corresponding loss of the analytic GFF solution was of the order of 10^{-7}.

To obtain a better understanding of the situation, we explored how these results are affected when we incorporate additional information from the exact solution of the GFF theory—for example, a low-lying CFT datum or the exact value of the tail functions inside the training region (instead of the approximate universal asymptotic value near $r = 1$).

Supplying the exact value of a CFT datum. The coefficient $a_{1,0}$ would be the leading scalar datum that contributes to the KMS condition, but as we noted in (5.11), it contributes together with $a_{0,2}$ through the combination $a_{1,0} + 3a_{0,2}$ (in $d = 4$). As a result, we fixed $a_{1,0} + 3a_{0,2}$ to its analytic value ($\simeq$15.06) and repeated the exercise of Fig. 3 (with the same approximate asymptotic value for the tail functions at $r = 0.9999$). The results are reported in Fig. 4. When compared to Fig. 3 we observe further improvement toward the configurations of the analytic GFF solution, but part of the A_0, A_2 ambiguity and a visible discrepancy between the predicted A_0 and the analytic GFF A_0 still remain.

Supplying the value of the tail functions at an intermediate radius. As an alternative, we replaced the input of the universal asymptotic behavior of the tail functions in the vicinity of $r = 1$ with exact information about the value of the tail functions at some radius r_i inside the training region. No further information from the exact solution was used (e.g., no analytic values of low-lying CFT data). This constraint was implemented with a simple quadratic loss

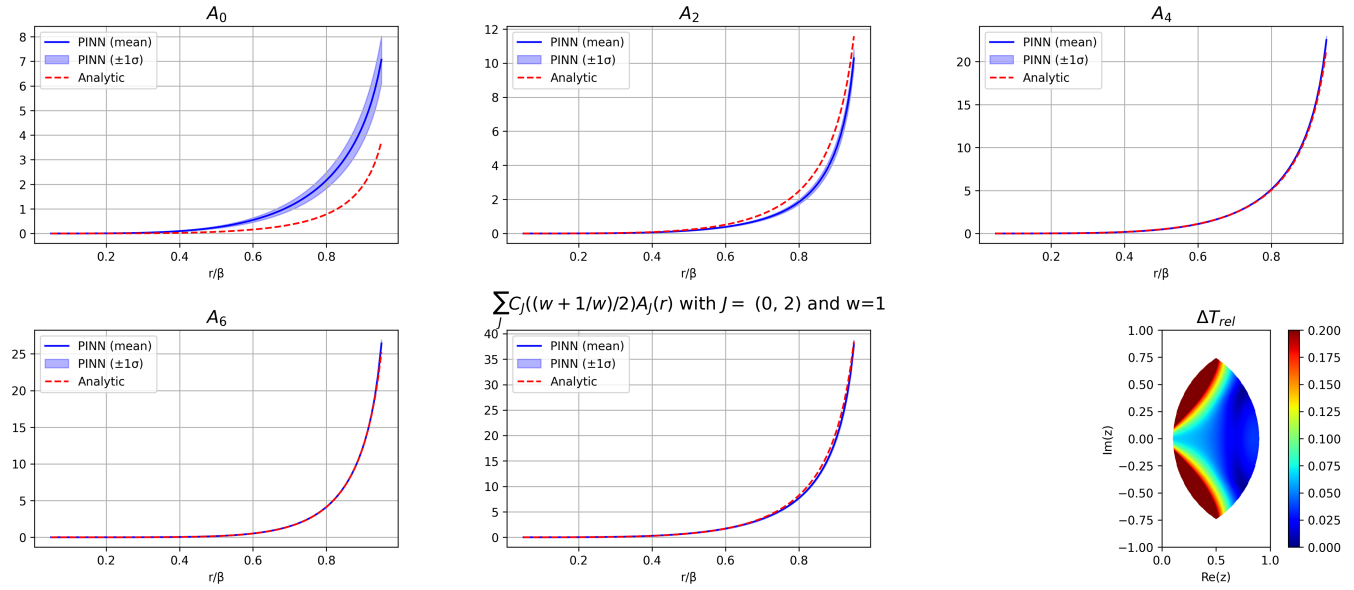


FIG. 4. Results obtained with the loss $\mathcal{L}_{\text{dot}(0)} + \mathcal{L}_{\text{BC}}$ performing the same runs as in Fig. 3 but with the additional input of the analytic value $a_{1,0} + 3a_{0,2} \simeq 15.06$.

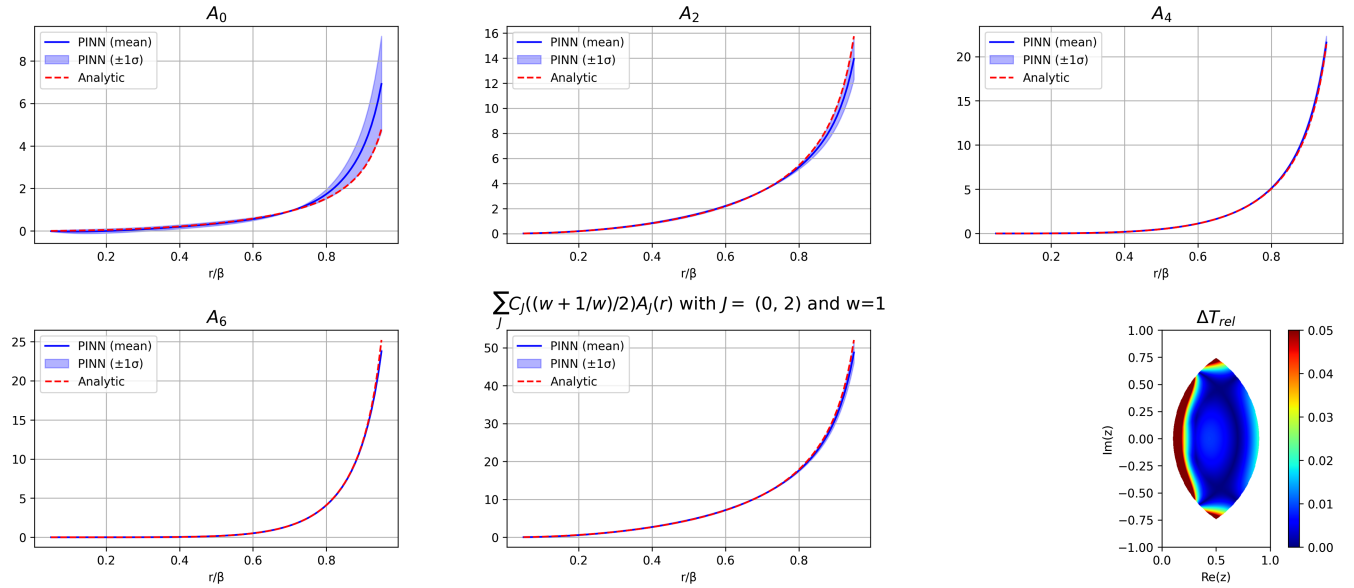


FIG. 5. Results obtained with the dot-loss $\mathcal{L}_{\text{dot}(0)}$ performing the same runs as in Fig. 3 after replacing the condition on the asymptotic values of the tail functions at $r = 0.9999$ with a condition that sets the values of the tail functions at $r_i = 0.7$ to the analytic values of the GFF solution. Notice that the maximum of the color bar scale in the heatmap is now set to 0.05 compared to the higher value of 0.20 in the previous Figs. 3 and 4.

and can be viewed as a set of sum-rules that are specific to the GFF theory.

The results obtained in this fashion (for $r_i = 0.7$) are reported in Fig. 5. In this case, the loss of the predicted configurations is $2.56 \times 10^{-8} \pm 1.21 \times 10^{-8}$ (vs 3.35×10^{-9} for the analytic GFF solution). Although the dot-loss is slightly higher than the one observed in the runs of Fig. 3, we notice that there is now much better agreement with the analytic tail functions, and the A_0, A_2 variation has been suppressed significantly without the need to perform any finetuning of the loss similar to the one reported in the context of Eq. (5.18).

For this implementation, the precise choice of the intermediate radius r_i does not matter. As long as it is not too close to 0 or 1, e.g., when $r_i \in [0.3, 0.7]$, we found comparable results in all the checks we performed.

It is interesting to ask what happens if we impose generic, non-GFF values $A_J(r_i)$. We might expect that such generic input is problematic; for example, we might expect that it leads to incorrect asymptotic behavior near

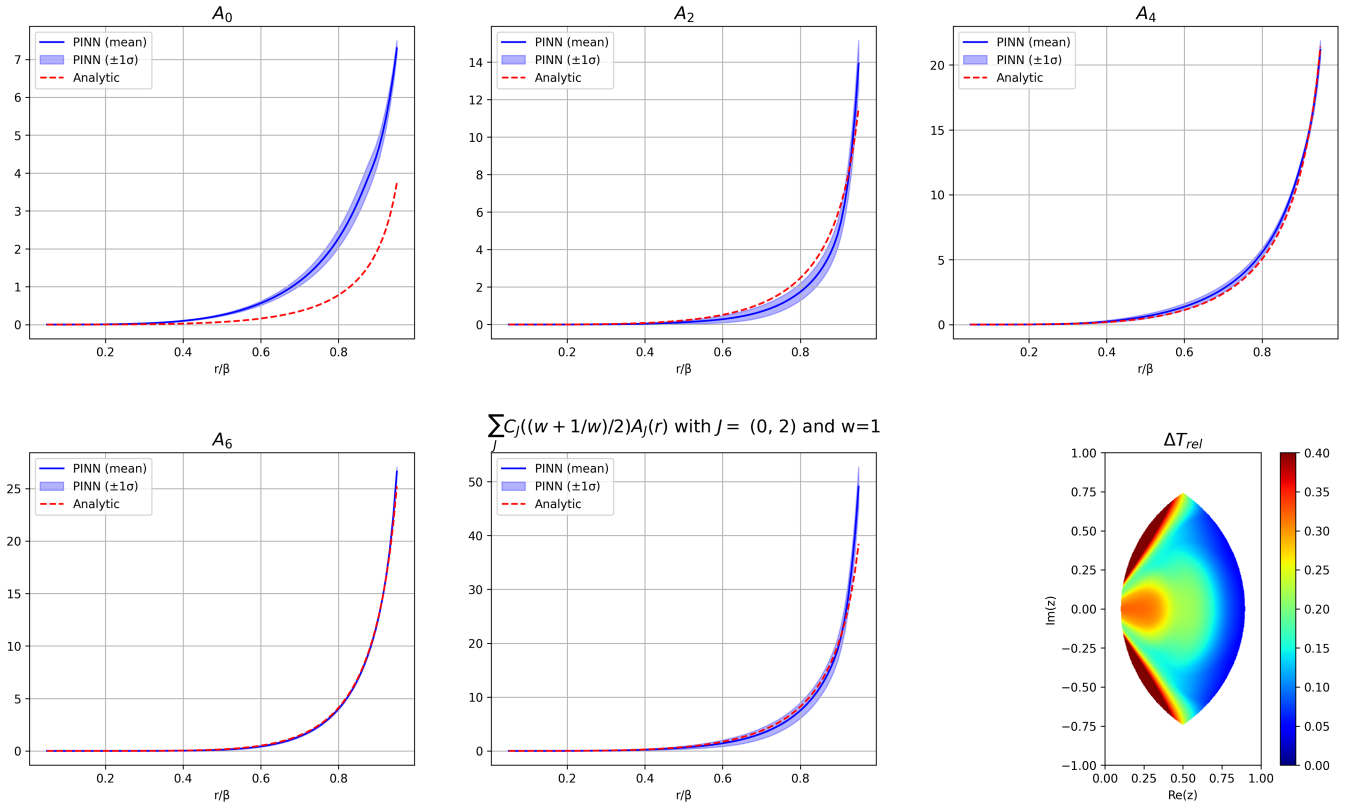


FIG. 6. Results obtained with the loss $\mathcal{L}_{\text{dot}(1)} + \mathcal{L}_{\text{BC}}$ and the CFT datum $a_{1,0} + 3a_{0,2}$ exposed. $\mathcal{L}_{\text{BC}}$ in Eq. (5.18) is set once again at $r = 0.9999$ using information about the asymptotic behavior of the tails in the vicinity of $r = 1$. The mean loss of the predicted functions is $6.12 \times 10^{-7} \pm 1.26 \times 10^{-7}$, with the loss of the analytic solution at 7.78×10^{-8}. The predicted coefficient is $a_{1,0} + 3a_{0,2} = 13.29 \pm 2.82$, with the analytic value at 15.06013. The first 4 plots (from top left to right) depict the predicted tail functions. The 5th plot and the heatmap provide information about the combined contribution of the tail functions A_0, A_2 to the conformal block expansion at $w = 1$ and the full training region, respectively.

$r = 1$, failing to produce two-point functions that satisfy the KMS condition sufficiently well everywhere in the OPE-convergence region. We will explore this question in more detail in the next subsection. There, we will make the additional observation that non-GFF values $A_J(r_i)$ also lead to enhanced instability toward many low-loss configurations, in contrast to the relative stability of the configurations in Fig. 5. This observation forms the basis of a potential strategy to recover the GFF $A_J(r_i)$ values from our analysis without any input from the exact GFF solution.

C. $a_{\Delta,J}$ bootstrap

Finally, we would like to expose one low-lying operator and use the dot-loss $\mathcal{L}_{\text{dot}(1)}$ to optimize the tail functions and recover the unknown exposed CFT datum from Eq. (4.15). For the reasons explained above, we expose the operators $[\phi\phi]_{1,0}$, $[\phi\phi]_{0,2}$ and compute the combination of coefficients $a_{1,0} + 3a_{0,2}$ that appears naturally in the KMS condition. Once again, we set $d = 4$, $\Delta_\phi = 1.68$, and collect the 10 lowest-loss configurations for the loss $\mathcal{L}_{\text{dot}(1)} + \mathcal{L}_{\text{BC}}$ in 1K independent runs with 50K epochs. For $\mathcal{L}_{\text{BC}}$ we continue to use Eq. (5.18).

Fixing the universal asymptotic behavior of the tail functions at $r = 0.9999$, we obtained the results summarized in Fig. 6. In analogy to the results of the previous subsection, the optimization in this setup struggles to accurately capture the leading tails, A_0, A_2, but better recovers their combined contribution to the conformal block expansion. The analytic value of $a_{1,0} + 3a_{0,2} = 15.06013$ is within the predicted 1σ range 13.29 ± 2.82, but this prediction is clearly affected by the difficulty of recovering the GFF profile of the tail functions A_0, A_2.

For reference, in Fig. 7 we also present the result of the same computation when the asymptotic value of the tail functions at $r = 0.9999$ are replaced by their analytic values at $r = 0.7$. Similar to the previous subsection, we notice that this information allows us to recover very accurately the analytic solution, and the predicted value of $a_{1,0} + 3a_{0,2}$ is now 15.0647 ± 0.0291, which is much closer to the analytic one at 15.06013.

Let us now examine what happens when, instead of the GFF values $\vec{A}_{\text{GFF}} \coloneqq \{A_J(0.7)|_{\text{GFF}}\}$, we impose a random vector $\vec{A}$ of tail values at $r = 0.7$. We performed several stochastic runs on the QMUL Apocrita cluster for $\vec{A}$

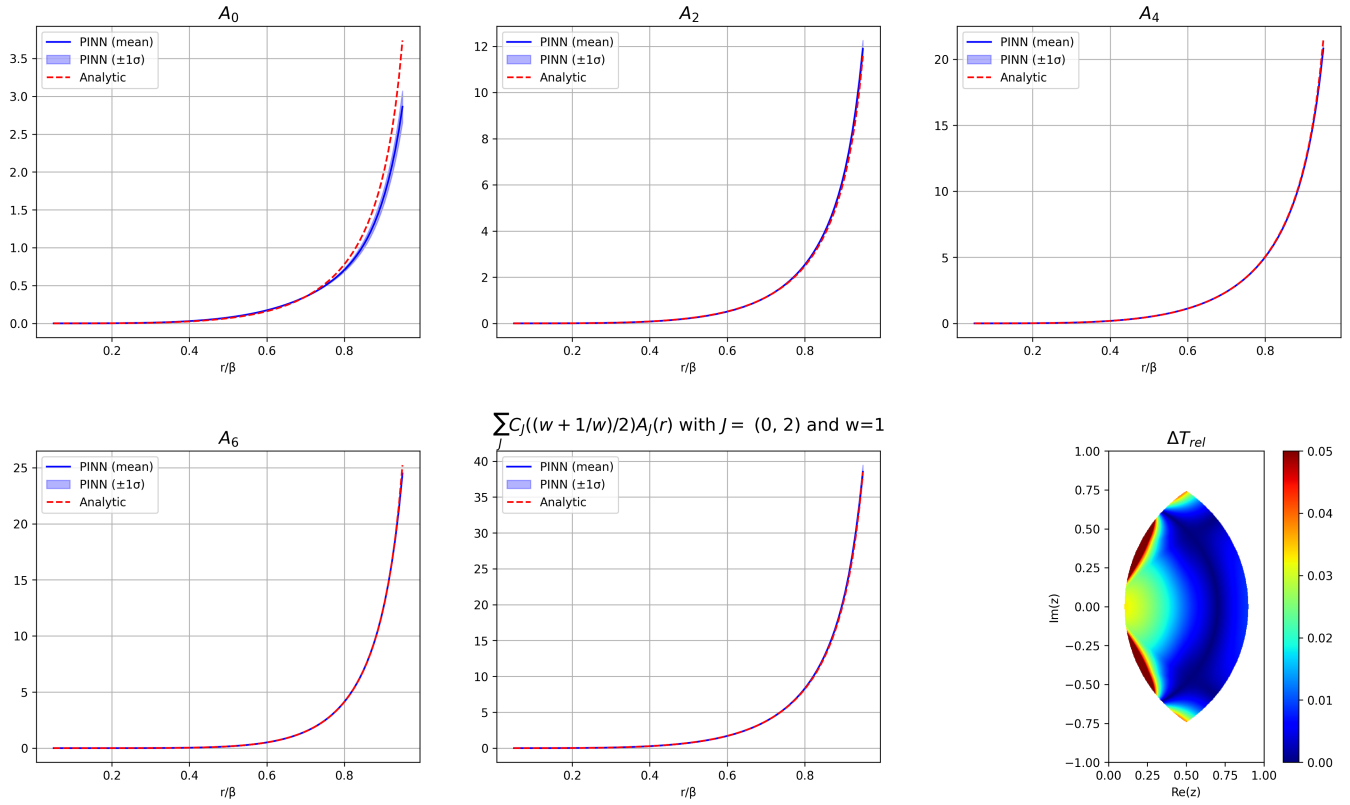


FIG. 7. Results obtained with the dot-loss $\mathcal{L}_{\mathrm{dot}(1)}$ and the CFT datum $a_{1,0}+3a_{0,2}$ exposed. In this case, the conditions on the asymptotic values of the tail functions at $r=0.9999$ have been replaced by conditions that set the values of the tail functions at $r=0.7$ to the analytic values of the GFF solution. The mean loss of the predicted functions is $9.49\times10^{-9}\pm6.27\times10^{-10}$ with the loss of the analytic solution at 7.78×10^{-8}. The predicted coefficient is $a_{1,0}+3a_{0,2}=15.0647\pm0.0291$, with the analytic value at 15.06013. The first 4 plots (from top left to right) depict the predicted tail functions. The 5th plot and the heatmap provide information about the combined contribution of the tail functions A_0, A_2 to the conformal block expansion at $w=1$ and the training region, respectively.

away from $\vec{A}_{\mathrm{GFF}}$, observing that we could still recover low-loss configurations inside the training region. The resulting configurations had two features: (1) mean curves that deviate significantly from the GFF curves and (2) considerable standard deviations. A typical example for $\vec{A}=\vec{A}_{\mathrm{GFF}}+\kappa|\vec{A}_{\mathrm{GFF}}|\vec{v}_{\mathrm{random}}$, with factor $\kappa=0.5$ and random unit 4-vector $\vec{v}_{\mathrm{random}}=(0.327696,-0.176708,-0.382838,-0.845473)$ is presented in Fig. 8.

The accumulated evidence supports the picture that there is an island of four-vectors $\vec{A}$ around $\vec{A}_{\mathrm{GFF}}$ with stochastic optimization results that exhibit small standard deviations. The size of the island correlates with the size of the chosen standard deviation cutoff. This picture opens up an exciting prospect: the possibility to extract an approximation of $\vec{A}_{\mathrm{GFF}}$ (and corresponding predictions like those of Fig. 7, which reproduce the GFF solution at high accuracy) from an analysis solely within the limited training region of Fig. 1, without having to address the numerical difficulties near the boundary of the OPE convergence region. We will not pursue a full implementation of this strategy in this section, but we will consider a closely related approach in the context of holographic CFTs in Sec. VII.

D. Brief summary

The above exercises in the context of the GFF theory provide a number of useful lessons about the approach we have introduced, including some of the potential challenges that arise during the implementation. Most notably, we can see explicitly in the above examples that tail functions can be bootstrapped, but the relevant optimization is subtle if the training grid is restricted and does not allow us to extract the full information of the KMS condition.

To address this problem one has two options: (1) extend the training grid to cover the OPE-convergence region as much as possible and explore how this affects the optimization outcome, or (2) supplement the optimization of the KMS condition on a fixed restricted grid, like the one in Fig. 1, with additional assumptions. For any option that can extract the full information of the KMS condition (at fixed discontinuity), we would expect to recover a unique solution. In the context of the GFF theory in this section, apart from a constant corresponding to the OPE coefficient $a_{0,0}$ and the fact that the KMS condition cannot distinguish the coefficients $a_{1,0}$ and $a_{0,2}$, that solution should be the GFF solution. A similar statement was

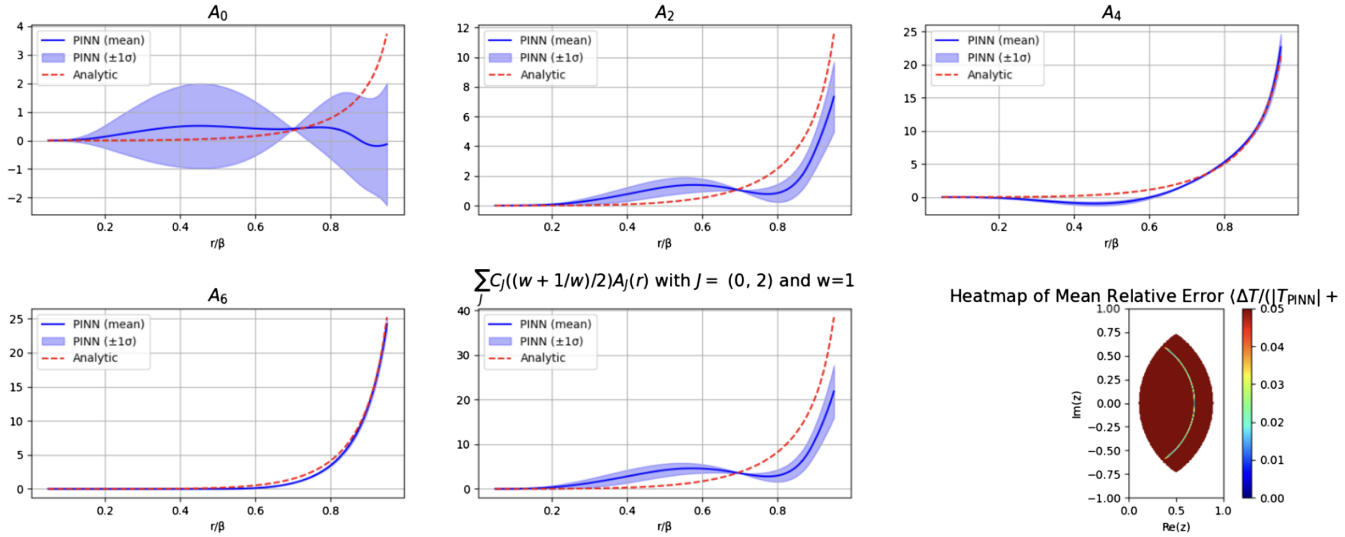


FIG. 8. Results obtained with the dot-loss $\mathcal{L}_{\text{dot}(1)}$ and the CFT datum $a_{1,0} + 3a_{0,2}$ exposed. In this case, we impose conditions on the tail functions at $r = 0.7$ (as in Fig. 6) but to values that differ from the analytic values of the GFF solution. The mean loss of the predicted functions is $6.64 \times 10^{-9} \pm 8.67 \times 10^{-9}$ (to be compared against the loss of the analytic solution at 7.78×10^{-8}). The predicted coefficient, $a_{1,0} + 3a_{0,2} = 28.784 \pm 4.083$, is now significantly farther away from the analytic value at 15.06013. The first 4 plots (from top left to right) depict the predicted tail functions. The 5th plot and the heatmap provide information about the combined contribution of the tail functions A_0, A_2 to the conformal block expansion at $w = 1$ and the training region, respectively. The thin yellowish line within the heatmap represents the $r = 0.7$ curve, where we impose the condition on $A_J(0.7)$. Since these values are relatively close to the GFF values, ΔT_{rel} is suppressed there.

shown at zero-spatial separation in the recent analyses of Refs. [18,19].

In the present work, numerical limitations did not allow us to make significant progress with option (1), but we hope to revisit this point in the future. In the context of option (2), we explored several concrete avenues. We observed that universal information about the behavior of tail functions near $r = 1$ can improve the stability of the optimization. When combined with proper loss functions, it nudges the result toward the GFF solution and recovers zero-spatial separation data but leaves room for configurations that are not KMS-invariant outside the training region. At the same time, we observed that extra information inside the training region can be very restrictive when it comes from the exact GFF solution. Reading this statement backwards, suggests the exciting possibility of recovering this extra information inside the training region by demanding that the optimization problem produces a stable, (numerically) unique minimum. We have numerical evidence for this proposal but not a proper, complete argument that proves it.

In the following sections, we develop these observations further in the context of holographic CFTs. This context is also interesting for another reason. One of the key simplifying features of the GFF theory is that it has a fixed universal contribution to the discontinuity (that comes only from the identity operator). In generic cases, $\mathbb{T}_{\text{Disc}}$ will not be a fixed function; it will be a function of, in principle, unknown CFT data. Holographic CFTs provide a setup where we can start discussing the challenges of theories with more complicated discontinuities.

VI. HOLOGRAPHIC CFTS: SETUP

In this section, we set up the KMS condition for thermal two-point functions of scalar operators, $\langle\phi\phi\rangle_\beta$, in holographic CFTs with a classical (super)gravity dual. We work at leading order in the large-c limit and assume that the scaling dimension Δ_ϕ of the operator ϕ is a positive, noninteger, real number above the unitarity bound $(\Delta_\phi \geq \frac{d-2}{2})$. We focus on situations where the thermal one-point function of the operator ϕ vanishes, namely, $\langle\phi\rangle_\beta = 0$.

In Appendix D, we review, for the convenience of the reader, basic properties of the large-c counting for CFTs on $S^1 \times \mathbb{R}^{d-1}$, while in Appendix E, we outline the subtleties of the case with integer Δ_ϕ. The latter will be useful in Sec. VIII, where we present a preliminary discussion of the $d = 4$ $\mathcal{N} = 4$ SYM case in the context of our approach.

The interest in holographic CFTs is obvious because of their connection to gravity and black holes. Beyond this connection, such theories also pose a useful testing ground for novel bootstrap methodologies at finite temperature since two-point functions computed holographically in *any* two- or higher-derivative classical theory of gravity have the correct analytic structure and automatically solve the KMS condition. As a result, it is interesting to ask how one could recover such solutions independently within a

bootstrap method. As we discussed in the introduction, current numerical techniques based on hard or soft truncations of the spectrum in the KMS condition are not suitable for such studies due to well-known limitations. The approach proposed in this paper does not have those limitations, and it is therefore interesting to explore how far it can take us in this holographic CFT context.

A. Spectra and discontinuities

In holographic CFTs, and at leading order in the large-c limit, scalar operators behave as generalized free fields coupled to the energy-momentum sector. The latter is intricately connected to the classical properties of the dual planar black hole solution. More specifically, the spectrum of operators that is expected to contribute to the conformal block expansion of a scalar one-point function $\langle\phi\phi\rangle_\beta$ comprises

(i) The identity operator.
(ii) The energy-momentum tensor $T_{\mu\nu}$.
(iii) The double-twist operators $[\phi\phi]_{n,J}$, with $J = 2\ell$ for $\ell = 0, 1, 2, \ldots$.
(iv) The multi-trace energy-momentum tensor operators $[T^k]_J$. These are composite operators with k insertions of the energy-momentum tensor and an arbitrary number of index contractions. Conformal primaries with derivative insertions are also possible but do not contribute at leading order in the large-c limit on $S^1 \times \mathbb{R}^{d-1}$. We review the argument in Appendix D.

TABLE III. List of the operators contributing to the conformal block expansion of the thermal two-point function $\langle\phi\phi\rangle_\beta$ in holographic CFTs. The scaling dimension of the operator ϕ is Δ_ϕ.

Operator	Scaling dimension Δ	Spin J	$a_{\Delta,J}$
Identity	0	0	$a_{\mathbf{1}} = 1$
$T_{\mu\nu}$	d	2	a_T
$[\phi\phi]_{n,J}$	$2(\Delta_\phi + n + \ell)$	$0 \le 2\ell$	$a_{n,J}$
$[T^k]_J$	dk	$0 \le 2\ell \le 2k$	$a_J^{(k)}$

The scaling dimensions, the spin, and the notation for the corresponding $a_{\Delta,J}$ coefficients for each of the above contributions is summarized in Table III.

Similar to the GFF case, the double-twist operators have vanishing discontinuity. The approximate contributions of the remaining operators to $\mathbb{T}_{\text{Disc}}$ can be summarized as follows. For concreteness, and for more direct reference to the case analyzed in the next section, we fix the spacetime dimension to $d = 4$. The *identity* contribution is

$$\mathbb{T}^{(\text{approx})(\mathbf{1})}_{\text{Disc}}[J_*; rw, rw^{-1}] = 4\sin(\pi\Delta_\phi)\left[\int_{r^{-1}}^{2r^{-1}} dw' \mathcal{K}_{J_*}(w,w')(rw'-1)^{-\Delta_\phi}(1-rw'^{-1})^{-\Delta_\phi} - \int_{-2r^{-1}}^{-r^{-1}} \mathcal{K}_{J_*}(w,w')(-rw'-1)^{-\Delta_\phi}(1+rw'^{-1})^{-\Delta_\phi}\right]. \tag{6.1}$$

The *energy-momentum* contribution is

$$\mathbb{T}^{(\text{approx})(T)}_{\text{Disc}}[J_*; rw, rw^{-1}] = 4a_T\sum_{s=0}^{2} p_s(J)\sin(\pi(-\Delta_\phi+s))\left[\int_{r^{-1}}^{2r^{-1}} dw' \mathcal{K}_{J_*}(w,w')(rw'-1)^{1-\Delta_\phi+s}(1-rw'^{-1})^{3-\Delta_\phi-s} - \int_{-2r^{-1}}^{-r^{-1}} dw' \mathcal{K}_{J_*}(w,w')(-rw'-1)^{1-\Delta_\phi+s}(1+rw'^{-1})^{3-\Delta_\phi-s}\right]. \tag{6.2}$$

The contribution of the *multi-trace energy-momentum tensors* $[T^k]_J$ is

$$\mathbb{T}^{(\text{approx})([T^k]_J)}_{\text{Disc}}[J_*; rw, rw^{-1}] = -4(-1)^{\frac{J}{2}} a_J^{(k)} \sum_{s=0}^{J} p_s(J)\sin(\pi(-\Delta_\phi+s)) \left[\int_{r^{-1}}^{2r^{-1}} dw' \mathcal{K}_{J_*}(w,w')(rw'-1)^{2k-\frac{J}{2}-\Delta_\phi+s}(1-rw'^{-1})^{2k+\frac{J}{2}-\Delta_\phi-s} - \int_{-2r^{-1}}^{-r^{-1}} dw' \mathcal{K}_{J_*}(w,w')(-rw'-1)^{2k-\frac{J}{2}-\Delta_\phi+s}(1+rw'^{-1})^{2k+\frac{J}{2}-\Delta_\phi-s}\right]. \tag{6.3}$$

B. Expectations from holography

In the AdS/CFT correspondence, the finite-temperature CFT on $S^1 \times \mathbb{R}^{d-1}$ is captured by a planar, asymptotically AdS_{d+1} black hole solution. A scalar conformal primary operator ϕ in the CFT corresponds in the dual gravitational description to a scalar field Φ. By studying the bulk-boundary propagator of the field Φ, one can extract the holographic two-point function $\langle\phi\phi\rangle_\beta$ on the boundary. Different classical gravitational theories in the bulk (e.g., theories with arbitrary higher-derivative terms) are characterized by different black hole solutions that produce different holographic two-point functions of scalar operators. In this context, holography becomes an apparatus that produces an infinite class of consistent solutions to the KMS condition.

Holographic computations of this type have been performed by many authors in the past. Recent papers, related to the present discussion, include [19,35–42]. The conformal-block expansion of the holographic correlators reveals interesting properties, some of which are worth summarizing here.

First, the contribution of the energy-momentum tensor comes with the coefficient a_T, which is related to the thermal one-point function of the energy-momentum tensor b_T through the equation

$$a_T = -\frac{2\Delta_\phi}{(d-2)(d-1)}\frac{\Gamma(\frac{d}{2})}{2\pi^{\frac{d}{2}}}\frac{b_T}{C_T}. \tag{6.4}$$

Here, C_T is the two-point function coefficient of the energy-momentum tensor, which is proportional to the c Weyl anomaly coefficient of the CFT. At the same time, b_T is connected to the entropy density S of the d-dimensional CFT (and of the corresponding $(d+1)$-dimensional planar black hole) via the relation

$$S = -b_T T^{d-1}. \tag{6.5}$$

Additionally, the conformal block expansion of the holographic thermal two-point function reveals contributions from the multi-traces $[T^k]_J$ and the double-twist operators $[\phi\phi]_{n,J}$. Interestingly, the multi-trace $[T^k]_J$ data, $a_J^{(k)}$, can be deduced from a relatively straightforward analysis of the asymptotic behavior of the bulk-boundary propagator near the AdS boundary [35]. In contrast, the double-twist data $a_{n,J}$ require the analysis of the bulk-boundary propagator in the full geometry, including information from the vicinity of the black hole horizon. This information is relatively harder to extract (see [36] for a recent relevant computation) and is not always readily available. As a result, it is very interesting to explore if it is possible to extract the double-twist data from the multi-trace energy-momentum data with an independent bootstrap computation.

Another noteworthy feature of the holographic two-point function is the universality of the lowest-twist multi-trace energy-momentum data $a_{2k}^{(k)}$, originally observed in [35]. As one varies the couplings in the bulk gravitational action, the data $a_J^{(k)}$ vary as well, but not all of them are independent. The data of the lowest-twist operators $[T^k]_{2k}$ (with twist $\tau_k = \Delta_k - J = 4k - 2k = 2k$) exhibit a fixed relation with a_T: once a_T is fixed, so are all the lowest-twist multi-trace coefficients $a_{2k}^{(k)}$ for $k \geq 2$. For example, in $d = 4$

$$\begin{aligned} a_4^{(2)} &= \frac{\Delta_\phi(7\Delta_\phi^2 + 6\Delta_\phi + 4)}{201600(\Delta_\phi - 2)}\left(\frac{120}{\Delta_\phi}a_T\right)^2, \\ a_6^{(3)} &= \frac{\Delta_\phi(1001\Delta_\phi^4 + 3575\Delta_\phi^3 + 7310\Delta_\phi^2 + 7500\Delta_\phi + 3024)}{10378368000(\Delta_\phi - 3)(\Delta_\phi - 2)}\left(\frac{120}{\Delta_\phi}a_T\right)^3 \end{aligned} \tag{6.6}$$

etc., irrespective of what the bulk gravitational action is. In Appendix F, we collect further explicit formulas for the lowest-twist data and some exact results for the nonlowest-twist data in the case of Einstein gravity.

VII. HOLOGRAPHIC CFTS: AN APPLICATION

In holographic CFTs we can perform two general types of exercises. We can either fix the discontinuity using input from independent gravitational computations and bootstrap the rest of the data, or we can limit the input from gravity and ask within the thermal bootstrap itself how to constrain the discontinuity as well.

In the first case, which is more straightforward, we expect that fixing the discontinuity of a holographic CFT within our setup, will allow an approximate reconstruction of the thermal two-point functions of the holographic theory. This amounts to a (numerical) derivation of double-twist data from (multi-trace) energy-momentum data. We will test this hypothesis in Sec. VII A.

The second question concerns the constraints we can derive about thermal data of generic holographic CFTs without fixing the discontinuity. This should also be possible within our approach, but is more demanding. We do not address this aspect fully in this work but provide some preliminary comments in Sec. VII B.

A. Fixed discontinuity

In this section, we consider the two-point function of a scalar operator with scaling dimension $\Delta_\phi = 1.5$ in $d = 4$ spacetime dimensions.[14] We analyze the approximate KMS condition (4.6), setting $J_* = 6$ (similar to the GFF analysis in previous sections).

1. Discontinuity from Einstein gravity data

As a specific example, we will consider thermal data in a holographic CFT as having an Einstein gravitational dual description. For such a theory, our first task is to find an approximation for the corresponding two-point function discontinuity.

From the previous Sec. VI, we recall that the discontinuity receives contributions only from the energy-momentum sector. There is an infinite number of such contributions, but for the purposes of this section, we will model $\mathbb{T}^{(\text{approx})}_{\text{Disc}}$ using only the contributions of the identity operator, the energy-momentum tensor and the multi-trace energy-momentum tensors

$$[T^2]_0, \quad [T^2]_2, \quad [T^2]_4, \quad [T^3]_4, \quad [T^3]_6, \quad [T^4]_8 \tag{7.1}$$

according to Eqs. (6.1)–(6.3). In this list, we include all the operators that contribute to the discontinuity up to maximum twist $\tau_{\max} = 8$; the twist is defined as $\tau := \Delta - J$. This is an *ad hoc* truncation.[15] By including higher-twist operators one can probe the systematic error of this truncation, but we will not attempt to do so in this paper. Our primary goal is to fix some approximation of the discontinuity and examine the solutions of the corresponding approximate KMS condition.

To fix the discontinuity with these specifications, we use the thermal OPE coefficients for the operators (7.1), as dictated by Einstein gravity in AdS/CFT through, a holographic computation devised in [35] and described in Appendix F to find

$$\begin{aligned} &a_{T,\text{GR}} = 1.21761, \quad && a^{(2)}_{0,\text{GR}} = -1.37668, \quad && a^{(2)}_{2,\text{GR}} = 1.58848, \quad && a^{(2)}_{4,\text{GR}} = -4.05945, \\ &a^{(3)}_{4,\text{GR}} = 1.77035, \quad && a^{(3)}_{6,\text{GR}} = 8.52362, \quad && a^{(4)}_{8,\text{GR}} = -15.9641. \end{aligned} \tag{7.2}$$

Using this information, we want to determine the double-twist datum $a_{1,0} + 3a_{0,2}$, similar to the exercises we performed in the GFF theory in Sec. V. Therefore, in the low-spin part of the approximate KMS condition (4.6) we expose only the energy-momentum contribution and the contributions of the double-twist operators $[\phi\phi]_{1,0}, [\phi\phi]_{0,2}$; the contribution of all other operators is included in the four tail functions A_0, A_2, A_4, A_6.

Repeating the optimization of Sec. V, with a loss of the form $\mathcal{L}_{\text{KMS}} + \mathcal{L}_{\text{asymptotic BC}}$, we encounter immediately the same technical issues we observed in the GFF case: a landscape of many configurations with comparable low loss. The problem arises for both types of KMS losses, $\mathcal{L}_{\text{KMS}} = \mathcal{L}_{\overline{\text{abs}}}$ or $\mathcal{L}_{\text{dot}(1)}$, and any type of loss $\mathcal{L}_{\text{BC}}$ for the implementation of the asymptotic boundary condition for the tails near $r = 1$. In fact, in the present case we could not even identify an analog of (5.18) to partially stabilize the outcome of the optimization. For those reasons, and motivated by the observations at the end of Sec. V, we would like to implement a strategy that identifies a suitable vector of constraints on $A_J(r_i)$ (for $J = 0, 2, 4, 6$ at some intermediate radius r_i) leading to an almost unique minimum in our optimization. We will propose that this minimum is the true, unique solution of our approximate KMS condition everywhere in the OPE-convergence region.

We proceed to explain how to implement this strategy.

2. Shooting method

In Sec. V C, we observed that by specifying the four-vector $\vec{A} := \{A_J(r_i)\}$ we obtained many KMS minima with comparable low-loss, unless $\vec{A}$ was fixed at $\vec{A}_{\text{GFF}}$ with the corresponding values of the tail functions of the exact GFF solution. In that case, the optimization yielded an almost unique minimum that approximated well the analytic GFF solution.

Using the same criterion of optimization stability, we would like to search for a similar vector $\vec{A}_*$ that replaces the GFF vector of Sec. V C with one suitable for the holographic CFT corresponding to the prescribed approximate discontinuity. In principle, we can achieve this goal with a shooting method where we scan over optimizations with generic $\vec{A}$ vectors until we find values that yield tail curves with low standard deviation. Anticipating that $\vec{A}_*$ is in the vicinity of $\vec{A}_{\text{GFF}}$, we can restrict the scan in a neighborhood

[14] This particular half-integer value of Δ_ϕ was chosen in order to have a direct comparison with double-twist data approximated with different methods in Ref. [19]. We note that we can obtain similar results in generic, nonhalf-integer, scaling dimensions.

[15] Unlike the truncations to the OPE expansion in common truncation schemes for crossing equations, here the situation is under better control because (as we noted in Sec. IV B) the truncation occurs within the discontinuity and the kernel $\mathcal{K}_{J_*}$ naturally suppresses higher-twist contributions.

of $\vec{A}_{\rm GFF}$. To assist this search, we can also use as another criterion the proximity of the mean asymptotic values of the tails $A_J(r)$ in the vicinity of $r = 1$ to the expected universal values of Eq. (B6).

Unfortunately, such searches can be very expensive; they require a cluster run for each candidate $\vec{A}$, and $\vec{A}$ has to be sampled in a four-dimensional domain [a $(\frac{J_*}{2}+1)$-dimensional domain, in general]. Consequently, the results we obtained in this way were not very accurate, leaving a lot of room for technical improvement. That motivated us to look for a more efficient implementation of this approach.

3. Soft intermediate-r constraints with a ReLU loss

As an alternative, consider an optimization problem, where one attempts to satisfy at the same time the following three objectives: (a) the approximate KMS condition with a fixed $\mathbb{T}_{\rm Disc}$ on the grid of Fig. 1 (b) the universal asymptotics at $r = 0.9999$, and (c) a constraint that the values of $A_J(r_i)$ remain within a fixed neighborhood U of a prescribed vector $\vec{A}_0$. We can select $\vec{A}_0$ at will, but for the rest of the discussion, we set $\vec{A}_0 = \vec{A}_{\rm GFF}$. This is motivated by the fact that in the holographic case of interest, it is natural to start the search by exploring deviations of the tail functions at r_i away from the GFF values $A_J(r_i)|_{\rm GFF}$. In this setup, we can perform a less expensive, more focused analysis of the effects of intermediate-r constraints and examine how the landscape of minima changes for neighborhoods U of different sizes. Unlike the shooting method of the previous subsection, here, $A_J(r_i)$ are not rigidly fixed throughout the optimization process. They are allowed to vary, and the algorithm can choose preferable values if these lead to configurations with lower KMS loss.

We can implement (c) straightforwardly using a rectified linear unit (ReLU) loss function in the following manner. Similar to (5.17), it is convenient to impose conditions on the non-linear transformation of the tail functions

$$\mathcal{A}_J(r) \coloneqq \operatorname{arcsinh}\left(\frac{A_J(r)}{2}\right) \tag{7.3}$$

instead of $A_J(r)$ directly to avoid having to train on large absolute values. Then, we can constrain the allowed values of each $\mathcal{A}_J(r_f)$ within a range of the GFF values using the mean ReLU loss[16]

$$\mathcal{L}_{\rm ReLU} = \frac{1}{\frac{J_*}{2}+1}\sum_J \mathrm{ReLU}(|\mathcal{A}_J(r_i) - \mathcal{A}_J(r_i)|_{\rm GFF}| - \mathtt{p}\times|\mathcal{A}_J(r_i)|_{\rm GFF}|). \tag{7.4}$$

[16]This choice is not unique. Higher powers of the ReLU function can also be employed. Although we do not expect the results to depend significantly on this choice, it would be useful to explore further the efficiency of alternative powers.

The ReLU function is defined as

$$\mathrm{ReLU}(x) = \max(0, x). \tag{7.5}$$

In this prescription, the factor $\mathtt{p}$ is an arbitrary real positive number that controls the size of the range of the region of allowed variation of $A_J(r_i)$ around the values of the central vector $A_J(r_i)|_{\rm GFF}$. $\mathtt{p} = 0$ penalizes any variation away from the GFF values $A_J(r_i)|_{\rm GFF}$, but $\mathtt{p} > 0$ penalizes only those configurations that deviate more than $\mathtt{p}\times|\mathcal{A}_J(r_i)|_{\rm GFF}|$ from the GFF values.

Putting everything together, the total loss function in the specific approach of this subsection has three components,

$$\mathcal{L} = \mathcal{L}_{\rm KMS} + \mathcal{L}_{\rm ReLU} + \mathcal{L}_{\rm BC}, \tag{7.6}$$

where $\mathcal{L}_{\rm KMS}$ is either the absolute or the dot-pinn loss, $\mathcal{L}_{\rm ReLU}$ is the loss in Eq. (7.4), and $\mathcal{L}_{\rm BC}$ is a loss that enforces the asymptotic behavior of the tails near $r = 1$, e.g., $r = 0.9999$. In this case, there is no need for convoluted functions like the one in Eq. (5.18). The simple quadratic loss

$$\mathcal{L}_{\rm BC} = \frac{1}{\frac{J_*}{2}+1}\sum_J (\mathcal{A}_J(r_f) - \mathcal{A}_J(r_f)|_{\rm GFF})^2 \tag{7.7}$$

works efficiently.

By analogy to the shooting method, where the goal was to find parameters yielding a stable and (numerically) unique optimization outcome, our main goal here is to likewise identify a special value $\mathtt{p}_*$, where the outcome of the optimization is similarly stable with small standard deviations. It is not immediately obvious that such a value exists, but here is what we can expect based on the analysis so far.

In the holographic case, the solution we are searching for is different from GFF, and therefore, fixing $A_J(r_i)$ to the GFF values is not expected to yield low-loss configurations with small standard deviations. In fact, the loss of these configurations is expected to be relatively high because we are forcing unnatural conditions at r_i together with the universal boundary conditions near $r = 1$. As we increase $\mathtt{p}$, we allow the values $A_J(r_i)$ to vary and the algorithm to explore a larger set of lower-loss configurations. With increasing $\mathtt{p}$, the loss is expected to continue to decrease, but eventually it will saturate and settle toward the value we would recover by doing an unconstrained optimization in the absence of the ReLU loss. Again, in this regime of large enough $\mathtt{p}$, the outcome of the optimization includes many low-loss configurations, and the standard deviations of statistical runs will be accordingly large. At intermediate values of $\mathtt{p}$, however, the standard deviations could be smaller and there could be a picture where they achieve a minimum value for a certain value of $\mathtt{p}$. Mirroring the discussion of the previous section, we would like to

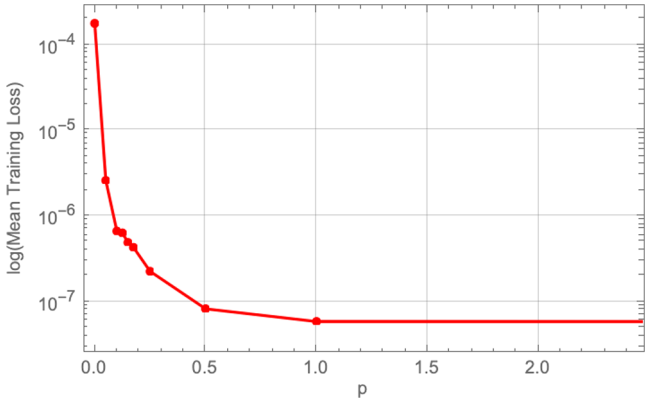


FIG. 9. Mean training loss for independent ReLU runs at different values of the factor p. The mean is computed for the 10 lowest-loss configurations of each cluster run. The y-axis scale is logarithmic. There is a concentration of data points near the value $\mathtt{p} = 0.15$, where we observed outcomes with the smallest statistical variation.

conjecture that this most stable configuration captures the true solution of the complete KMS condition. Could such a point exist? Possibly yes, because as we increase p away from $\mathtt{p} = 0$, there will be a point where the neighborhood U becomes sufficiently large to include the values $A_J(r_i)$ corresponding to the solution we are looking for. For those values, the standard deviations are expected to be minimal, and it is natural to expect that the corresponding configurations will dominate the ReLU optimization. Explicit computation in the next section shows that this picture is indeed correct and that allows us to identify a critical value $\mathtt{p}_*$ of stable optimization, which is conjectured to represent an approximation of the unique solution to the complete KMS condition.

4. Preliminary results on a double-twist datum

We are now in position to return to the analysis of the KMS condition in four-dimensional holographic CFTs. We fix the discontinuity using the (multi-trace) energy-momentum data (7.2), and set up the approximate KMS condition with $J_* = 6$ to compute the double-twist datum $a_{1,0} + 3a_{0,2}$. We use the dot-loss in the KMS optimization and the ReLU approach to determine a critical factor $\mathtt{p}_*$.

Let us examine what happens to the outcome of the optimization as we vary the factor p. 1 K independent runs were performed on the QMUL Apocrita cluster for 50 K epochs with different values of p from 0 to 10. The mean training loss of the 10 lowest-loss configurations as a function of p is presented in Fig. 9. At $\mathtt{p} = 0$ the loss takes high values at the order of 10^{-4} and very quickly drops (in accordance with the expectations at the end of the previous subsection) by three orders of magnitude to an almost constant value.

Figure 10 presents the explicit output of the tails for three different values of $\mathtt{p} \in \{0, 0.15, 10\}$. We observe the features anticipated in the previous subsection. There is visible variation of the output at $\mathtt{p} = 0$, which then decreases as p increases up to some point around $\mathtt{p} = 0.15$. Then the variation of the output increases again producing significantly deformed curves. We performed multiple cluster runs around the value $\mathtt{p} = 0.15$ in order to determine the factor p of minimum instability, which was identified in the vicinity of $\mathtt{p} = 0.15$.

Once p is fixed, we can read off the corresponding value of the double-twist parameter $a_{10} + 3a_{02}$ using Eq. (4.15). Specifically, we found

$$\mathtt{p} = 0.125\colon\ a_{10} + 3a_{02} = 9.81 \pm 0.35, \tag{7.8}$$

$$\mathtt{p} = 0.150\colon\ a_{10} + 3a_{02} = 10.09 \pm 0.37, \tag{7.9}$$

$$\mathtt{p} = 0.175\colon\ a_{10} + 3a_{02} = 10.58 \pm 0.46. \tag{7.10}$$

Adopting the conjecture that the configuration at $\mathtt{p}_* \sim 0.15$ is a sensible approximation of the sought-after solution of the KMS condition for this holographic CFT, we can use (7.9) to make a preliminary estimate of the double-twist datum $a_{10} + 3a_{02}$.

We note in passing that another way to perform the computation is without enforcing an asymptotic condition on the tail function near $r = 1$. That removes a potential systematic error from the explicit, subtle specification of the tail functions in that region. In that case, we remove the $\mathcal{L}_{\mathrm{BC}}$ contribution to the loss function $\mathcal{L}$ in (7.6) and set

$$\mathcal{L} = \mathcal{L}_{\mathrm{KMS}} + \mathcal{L}_{\mathrm{ReLU}}. \tag{7.11}$$

Repeating the exercise with this loss, we found that we could still observe the same qualitative features as in Fig. 10, namely, the existence of a value of $\mathtt{p}_*$ with increased stability. That value was approximately $\mathtt{p}_* \simeq 0.20$, yielding

$$\mathtt{p}_* = 0.20\colon\ a_{10} + 3a_{02} = 9.37 \pm 0.44. \tag{7.12}$$

It is worth noting here that the asymptotic model of Ref. [19] predicts the following value for $a_{10} + 3a_{02}$ [see formula (2.29) in that paper for $m = 1$]:

$$[a_{10} + 3a_{02}]_{[19]} = 7.54. \tag{7.13}$$

In the more recent work, [43], the predicted value is

$$[a_{10} + 3a_{02}]_{[20]} = 7.686. \tag{7.14}$$

There is significant difference between these results and Eq. (7.10) (at the order of 20%–30%), but this is not completely unexpected. The estimates (7.9) or (7.12), contain several systematic errors, whose effect has not been properly analyzed at this stage. These include

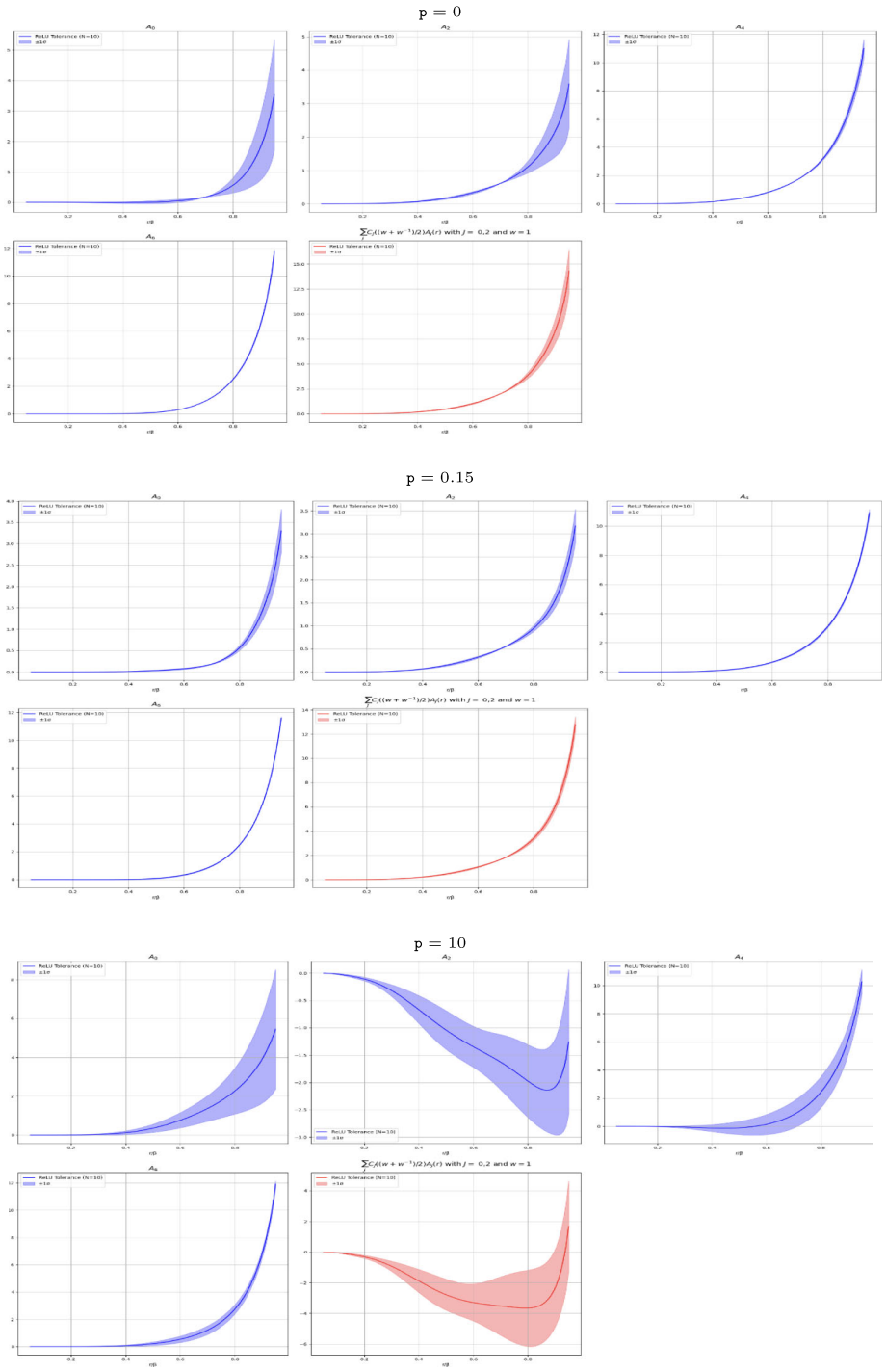


FIG. 10. Output of the ReLU optimization for three different values of $\mathtt{p} \in \{0, 0.15, 10\}$. In each case, the first four plots represent the functions A_0, A_2, A_4, A_6. The last plot in red presents the combined contribution of A_0 and A_2 to the conformal block expansion at $w = 1$.

(i) The errors in the approximation of the discontinuity. These could be estimated by increasing J_* and the number of multi-trace energy-momentum tensor contributions.

(ii) The errors in the asymptotic boundary condition for the tails near $r = 1$ using Eq. (B6). These are not present in (7.14), which is visibly closer to the estimates of Refs. [19,43].

(iii) The imprecise determination of $\mathtt{p}$. In this context, it would also be useful to check the efficiency of the optimization using another variant of the ReLU loss, (7.4), e.g., a quadratic $(\mathrm{ReLU})^2$ loss.

It is also worth noting here that an independent estimation of $a_{10}+3a_{02}$ directly from the numerical solution of the PDE in AdS in Ref. [19] yields

$$[a_{10}+3a_{02}]_{\text{holographic-PDE[19]}} \in [8.84, 10.97]. \tag{7.15}$$

Both our estimates, (7.9) and (7.12), are comparable to this result.

We again stress that the estimates (7.9) and (7.12) are based on the conjecture that the $\mathtt{p}_*$-configurations approximate the solution of the complete KMS condition. This remains a conjecture that needs to be better understood. For example, it would be interesting to establish whether a suitable $\mathtt{p}_*$ value always exists in problems with fixed discontinuity.

B. Dynamic discontinuity

So far, we have been dealing exclusively with situations where the discontinuity is fixed to a specific form and does not involve any unknown thermal OPE coefficients. In generic problems, the discontinuity contains unknown thermal OPE coefficients that need to be included and take active part in the optimization. Some of these coefficients can either be exposed or can be treated as implicit parts of the tail functions. At the end of the day, the optimization should constrain these data and accordingly modify the allowed form of the discontinuity.

In the holographic CFTs discussed in this and the previous section, it is the data of the energy-momentum sector that contribute to the discontinuity. We know that the KMS condition cannot completely fix these data, but there are certain restrictions they have to obey, most notably the universality relations of the lowest-twist multi-trace energy-momentum tensors observed in [35]. As a preliminary study in this context, we attempted to investigate if fixed discontinuities with arbitrarily fixed (multi-trace) energy-momentum data lead, after optimization, to configurations with lowest-losses that depend on the details of the discontinuity. We did not detect any such dependence and therefore no signs that discriminate between the discontinuities that obey the universality relations and those that do not. For this check, we used the absolute loss combined with a loss for the universal asymptotic boundary condition. As we remarked above, this approach allows for a lot of spurious minima, which could have contaminated our analysis. It would be useful to repeat this analysis with a more proper treatment, e.g., the ReLU loss prescription, which eliminates the spurious minima. It is also important to further develop a version of the ReLU loss approach when the data entering the approximation of the discontinuity are optimizable. We intend to return to this aspect in future work.

On a more general note, it is possible that in some theories the spectrum is such that the KMS condition fixes the discontinuity and the corresponding two-point functions uniquely. The results of Ref. [16] suggest this may be the case for the $O(N)$ vector models in $d=3$ dimensions. It would be interesting to explore this case as well with the methods introduced in this paper.

VIII. PRELIMINARY DISCUSSION OF $\mathcal{N}=4$ SYM THEORY IN THE SUPERGRAVITY REGIME

As explained in Appendix E, thermal two-point functions of (scalar) operators with integer scaling dimensions in holographic CFTs are subtle because of the presence of logarithmic contributions to the conformal block expansion. The source of the logarithms are poles in certain three-point function coefficients, which are canceled between the respective contributions of degenerate multi-trace energy-momentum operators and double-twist operators. In this section, we discuss the case of integer scaling dimensions in some detail, laying out the form of the OPEs, the contributions of the discontinuity, and the corresponding KMS condition.

The primary example that we have in mind are thermal two-point functions of half-BPS superconformal primary operators $\phi = S_p$ (with scaling dimension $\Delta_\phi = p$, $p=2,3,\ldots$) in the supergravity limit of 4d $\mathcal{N}=4$ SYM theory. The aim of the present discussion is to set the stage for future work in this direction.

Similar to Sec. VI, we assume that the contributing operators to the thermal two-point function $\langle\phi\phi\rangle_\beta$ are the identity, the energy-momentum tensor, the double-twist operators $[\phi\phi]_{n,J}$ and the multi-trace energy-momentum tensor operators $[T^k]_J$ for $k\geq 2$ and $J=2\ell\leq 2k$.

In Appendix E, we recall that when a multi-trace energy-momentum tensor $[T^k]_J$ is degenerate with a double-twist operator $[\phi\phi]_{n,J}$, namely, when

$$2k = p+n+\frac{J}{2}, \tag{8.1}$$

its contribution to the thermal OPE gets modified to

$$g(rw,rw^{-1}) \supset (a'_{[T^k]_J}\log(z\bar{z}) + (a_{[T^k]_J}+a_{[\phi\phi]_{n,J}}))C_J^{(\nu)}\left(\frac{1}{2}(w+w^{-1})\right)r^{4k-2p}. \tag{8.2}$$

The meaning of the coefficient $a'_{[T^k]_J}$ is explained in Eq. (E9). Clearly, in this case the KMS condition cannot track the coefficients $a_{[T^k]_J}$ and $a_{[\phi\phi]_{n,J}}$ independently, but it can in principle track the combinations $a'_{[T^k]_J}$ and $a_{[T^k]_J}+a_{[\phi\phi]_{n,J}}$.

A. Discontinuities

When the scaling dimension Δ_ϕ of the external operator is an integer p, many of the discontinuities simplify and receive only δ-function contributions. The double-twist operators have vanishing discontinuity. As a result, the only potential contributions arise from the identity operator and the energy-momentum sector. Here, we list the explicit form of these contributions.

Identity. The identity operator contributes the following discontinuity:

$$\mathbb{T}^{(\mathbf{1})}_{\text{Disc}}[J_*; rw, rw^{-1}] = \frac{4\pi r^{-p}}{\Gamma(p)}\{(-1)^{p-1}\partial_{w'}^{p-1}[\mathcal{K}_{J_*}(w,w')(1-rw'^{-1})^{-p}]|_{w'=r^{-1}} - \partial_{w'}^{p-1}[\mathcal{K}_{J_*}(w,w')(1+rw'^{-1})^{-p}]|_{w'=-r^{-1}}\}. \tag{8.3}$$

The energy-momentum tensor. The contribution to the discontinuity from the energy-momentum tensor, with thermal OPE coefficient a_T, is

$$\begin{aligned}\mathbb{T}^{(T)}_{\text{Disc}}[J_*; rw, rw^{-1}] = 4\pi a_T \sum_{s=0}^{2} \frac{r^{-p+s+1}}{\Gamma(p-s-1)}\{&(-1)^{p+s}\partial_{w'}^{p-s-2}[\mathcal{K}_{J_*}(w,w')(1-rw'^{-1})^{3-p-s}]|_{w'=r^{-1}} \\ &- \partial_{w'}^{p-s-2}[\mathcal{K}_{J_*}(w,w')(1+rw'^{-1})^{3-p-s}]|_{w'=-r^{-1}}\}.\end{aligned} \tag{8.4}$$

When the Γ-functions in the denominator have poles, the corresponding term does not contribute.

Multi-trace energy-momentum tensors. There are two different types of discontinuities in this case. We noted in Eq. (2.11) that we can recast the conformal block expansion as a series in powers of z and $\bar{z}$. In this form, multi-trace energy-momentum operators that are not degenerate with double-twist operators contribute to the conformal block expansion in the following way:

$$g(z,\bar{z}) \supset \sum_{[T^k]_{2\ell}\in\phi\times\phi} a^{(k)}_{2\ell} \sum_{s=0}^{2\ell} p_s(2\ell) z^{2k-\ell-p+s}\bar{z}^{2k+\ell-p-s}. \tag{8.5}$$

On the other hand, multi-trace energy-momentum operators that are degenerate with double-twist operators have a modified conformal block expansion, see Eq. (8.2), which yields contributions of the following form:

$$\begin{aligned}g(z,\bar{z}) \supset \sum_{[T^k]_{2\ell}} &(a'^{(k)}_{2\ell}\log(z\bar{z}) + (a^{(k)}_{2\ell} + a_{n,2\ell})) \\ &\times \sum_{s=0}^{2\ell} p_s(2\ell) z^{2k-\ell-p+s}\bar{z}^{2k+\ell-p-s}\Big|_{2k=p+n+\ell}.\end{aligned} \tag{8.6}$$

Here, for the thermal OPE coefficients, we are using the notation of Table III.

The first type of contributions in Eq. (8.5) exhibit a nonvanishing discontinuity when a power of z is negative, namely, when

$$2k - p - \ell \le -1 - s \le -1. \tag{8.7}$$

Since $J = 2\ell \le 2k \Leftrightarrow \ell \le k$, this implies

$$k \le k + (k-\ell) + s \le p - 1. \tag{8.8}$$

For example, this condition excludes contributions from the multi-trace energy-momentum operators for $p = 2$, where $k \le 1$ cannot be satisfied. For $p = 3$, on the other hand, there is a contribution from the double-trace operator $[T^2]_4$, which is equal to

$$\begin{aligned}&\mathbb{T}^{([T^2]_4)}[J_*; rw, rw^{-1}] \\ &= 4\pi a^{(2)}_4 \frac{(1-r^2)^3}{r}[\mathcal{K}_{J_*}(w, r^{-1}) - \mathcal{K}_{J_*}(w, -r^{-1})].\end{aligned} \tag{8.9}$$

The second type of contributions, in Eq. (8.6), have a nonvanishing discontinuity that arises from logarithmic terms. Notice that (8.6) requires

$$2k - p - \ell = n \ge 0, \tag{8.10}$$

which is complementary to (8.7). The total approximate contribution to the discontinuity from such operators is

$$\mathbb{T}^{(\text{approx})(\log)}_{\text{Disc}}[J_*; rw, rw^{-1}] = -4\pi \sum_{k=2}^{\infty} \sum_{\ell=0}^{k} \sum_{s=0}^{2\ell} a'^{(k)}_{2\ell} p_s(2\ell) \Bigg\{ - \int_{-2r^{-1}}^{-r^{-1}} dw' \, \mathcal{K}_{J_*}(w, w')(1 + rw')^{2k-\ell-p+s}(1 + rw'^{-1})^{2k+\ell-p-s} + \int_{r^{-1}}^{2r^{-1}} dw' \, \mathcal{K}_{J_*}(w, w')(1 - rw')^{2k-\ell-p+s}(1 - rw'^{-1})^{2k+\ell-p-s} \Bigg\}_{2k-p-\ell \geq 0}. \tag{8.11}$$

In conclusion, some multi-trace energy-momentum operators contribute with inverse power discontinuities via (8.5) and the rest with log discontinuities via (8.6). As we increase the integer scaling dimension p of the external operator, more and more operators exchange logarithmic discontinuities with inverse power discontinuities.

B. KMS condition

An (approximate) KMS condition along the lines of Eq. (4.6) can now be formulated for any thermal two-point function of half-BPS superconformal primary operators S_p. For example, for $p = 2$, and only with the contributions of the identity and energy-momentum tensor exposed, we obtain

$$\begin{aligned} & r^{-4} + a_T C_2^{(1)} \left(\frac{1}{2}(w + w^{-1}) \right) + \sum_{\ell=0}^{\frac{J_*}{2}} A_{\Delta_*(2\ell),2\ell;\boldsymbol{\theta}}(r) C^{(1)}_{2\ell} \left(\frac{1}{2}(w + w^{-1}) \right) \\ & \quad + \mathbb{T}^{(\mathbf{1})}_{\text{Disc},p=2}[J_*; rw, rw^{-1}] + \mathbb{T}^{(\log)}_{\text{Disc},p=2}[J_*; rw, rw^{-1}] + \mathbb{T}^{(T)}_{\text{Disc},p=2}[J_*; rw, rw^{-1}] \\ & = \tilde{r}^{-4} + a_T C_2^{(1)} \left(\frac{1}{2}(\tilde{w} + \tilde{w}^{-1}) \right) + \sum_{\ell=0}^{\frac{J_*}{2}} A_{\Delta_*(2\ell),2\ell;\boldsymbol{\theta}}(\tilde{r}) C^{(1)}_{2\ell} \left(\frac{1}{2}(\tilde{w} + \tilde{w}^{-1}) \right) \\ & \quad + \mathbb{T}^{(\mathbf{1})}_{\text{Disc},p=2}[J_*; \tilde{r}\,\tilde{w}, \tilde{r}\tilde{w}^{-1}] + \mathbb{T}^{(\log)}_{\text{Disc},p=2}[J_*; \tilde{r}\,\tilde{w}, \tilde{r}\tilde{w}^{-1}] + \mathbb{T}^{(T)}_{\text{Disc},p=2}[J_*; \tilde{r}\,\tilde{w}, \tilde{r}\tilde{w}^{-1}]. \end{aligned} \tag{8.12}$$

Higher p cases can be treated in a similar fashion using the formulas of the previous subsection.

We intend to return to a detailed analysis of these equations in future work.

IX. OUTLOOK

In this work, we provided a new framework for the bootstrap of crossing equations in CFTs focusing, as a concrete example, on the KMS condition at finite temperature. The proposed approach combines two complementary elements: (1) modeling of the high-spin contributions to the thermal OPE with discontinuities in a controlled approximation that employs thermal dispersion relations, and (2) capturing low-spin, but arbitrarily high-scaling-dimension contributions via tail functions represented by neural networks. After testing and calibrating this approach on generalized free fields, we applied it to holographic CFTs and provided a preliminary prediction for a thermal one-point coefficient of a simple double-twist operator.

Moving forward, we would like to better understand how to reliably extract the complete information of the KMS condition from everywhere within the simultaneous s and t OPE convergence region. Since it is numerically challenging to set up an optimization scheme near the boundary of this region, it is useful to obtain a better understanding of the properties of the tail functions in the bulk of the OPE convergence region and how these are related to features of the optimization. The analysis of the GFF case led us to conjecture that the analytic solution (which obeys the KMS condition everywhere in the OPE convergence region) can be recovered from the KMS optimization within a subregion of the OPE convergence, when the evaluation of the tail functions $A_J(r_i)$ at a generic bulk point r_i is constrained to take a (numerically) unique value that stabilizes the output of the optimization. This conjectured relation between special values of $A_J(r_i)$, the stability of the optimization, and the exact KMS solutions needs to be better understood. The existence of such special values was also observed in the case of holographic CFTs, where we used the above conjecture as a guiding principle to identify approximations of physical thermal two-point functions.

In addition, when $\Delta_\phi \geq \frac{d-1}{2}$, we derived a universal constraint on the $r \to 1$ asymptotics of the tail functions using the KMS condition. Imposing this constraint as an asymptotic boundary condition—lying outside of the bulk convergence region used for NN training—had a positive effect in restricting the space of low-loss configurations. It would be interesting to understand what happens when the above inequality is not satisfied.

On a more conceptual level, the proposed approach is shifting the focus away from individual CFT data to suitable tail functions and discontinuities. It would be very interesting to study the general properties of tail functions (as formulated in this paper) and use them to reformulate CFT constraints, or use theory-specific constraints that are imposed on the tails to guide the bootstrap search towardp specific CFTs.

As an immediate next step to the analysis presented in this paper, one could explore in further detail the more general bootstrap with a fully dynamical discontinuity. It would be interesting to study holographic CFTs in this manner and demonstrate the universality relations of [35]. Further natural directions include going beyond the (super) gravity limit and applying our methods to other CFTs, e.g., $O(N)$ vector models.

Our approach can be extended beyond the KMS bootstrap for CFTs at finite-temperature. In fact, the bootstrap program in many different directions can be reformulated and studied anew. For instance, in the four-point bootstrap for zero-temperature CFTs, our prescription can supplement standard methods based on positivity constraints [12]. There are many physically interesting setups where positivity constraints are absent (or hard to identify), where modeling the crossing equations with tails and discontinuities would be a source of new information. Interesting setups of this type beyond finite-temperature physics include the physics of defects, the five-point bootstrap [44–46], and CFTs in the presence of boundaries. We intend to return and address many of the above open problems in the near future.

Note added. As we were finalizing the draft, we became aware of the paper [43], which partially overlaps with our results.

ACKNOWLEDGMENTS

We would like to thank Kyriakos Papadodimas, Andrei Parnachev, and Andreas Stergiou for useful discussions. The work of C. P. was partially supported by the Science and Technology Facilities Council (STFC) Consolidated Grant No. ST/X00063X/1, "Amplitudes, Strings & Duality." M. W. was funded by a STFC studentship. This research utilized Queen Mary's Apocrita HPC facility, supported by QMUL Research-IT (Doi: 10.5281/zenodo.438045).

DATA AVAILABILITY

The data that support the findings of this article are not publicly available upon publication because it is not technically feasible, and/or the cost of preparing, depositing, and hosting the data would be prohibitive within the terms of this research project. The data are available from the authors upon reasonable request.

APPENDIX A: COMMENTS ON THERMAL DISPERSION RELATIONS

The material in this appendix collects useful facts about dispersion relations for scalar two-point functions in thermal CFTs. This is a partial review of the MSc thesis [47] of one of the authors (A. S.). Similar results were previously obtained in Ref. [23].

1. Dispersion relations from Cauchy's theorem

Consider the thermal two-point function $g(z,\bar{z}) = \langle\phi(x)\phi(0)\rangle_\beta$ of a single scalar primary ϕ (in the same notation as in Sec. II, where $z = rw$). In what follows, we set, for convenience, $\beta = 1$. Let $G(z,\bar{z})$ denote a "subtracted" version of g that improves the large-w behavior of the function, e.g., $G(z,\bar{z}) = f(z,\bar{z})g(z,\bar{z})$ or $G(z,\bar{z}) = g(z,\bar{z}) - f(z,\bar{z})$, for a suitably chosen function f.[17] The subtraction allows us to drop arc contributions in the following discussion and streamline the presentation.

By using Cauchy's theorem

$$G(rw, rw^{-1}) = \frac{1}{2\pi i}\oint_{C_w} dw' \frac{1}{w'-w} G(rw', rw'^{-1}), \tag{A1}$$

one can derive the dispersion relation

$$\begin{aligned} G(rw, rw^{-1}) = &\int_{-r}^{r} dw'\, K(w,w')\mathrm{Disc}[G(rw', rw'^{-1})] \\ &+ \int_{-\infty}^{-r^{-1}} dw'\, K(w,w')\mathrm{Disc}[G(rw', rw'^{-1})] \\ &+ \int_{r^{-1}}^{\infty} dw'\, K(w,w')\mathrm{Disc}[G(rw', rw'^{-1})], \end{aligned} \tag{A2}$$

which involves three contributions from integrals around the branch cuts $(-r, r)$, $(-\infty, -r^{-1})$ and (r^{-1}, ∞), respectively. By default, $r < 1$. All three terms involve a common kernel in this expression

$$K(w,w') := \frac{1}{4\pi}\left(\frac{1}{w'-w} + \frac{1}{w'-w^{-1}}\right). \tag{A3}$$

We assumed that, after a suitable choice of subtraction, potential arc contributions from infinity do not contribute in the contour-deformation arguments that led to this result.

Equivalently, changing variables $w' \to w'^{-1}$ in the inner branch-cut integral along $(-r, r)$, and assuming the invariance $G(rw', rw'^{-1}) = G(rw'^{-1}, rw')$, we can reformulate (A2) as an integral solely over the "outer" branch cuts as

$$\begin{aligned} G(rw, rw') = &\left(\int_{-\infty}^{-r^{-1}} + \int_{r^{-1}}^{\infty}\right) dw'\, \mathcal{K}(w,w') \\ &\times \mathrm{Disc}[G(rw', rw'^{-1})], \end{aligned} \tag{A4}$$

with kernel

$$\mathcal{K}(w,w') := \frac{1}{2\pi w'}\frac{w'^2-1}{(w'-w)(w'-w^{-1})}. \tag{A5}$$

[17]The function $f(z,\bar{z})$ is symmetric under $z \leftrightarrow \bar{z}$ but not necessarily under $z \to -z, \bar{z} \to -\bar{z}$.

In this derivation of the dispersion relation, the kernel is trivially spacetime-dimension independent.

The dispersion relation (A4) can be generalized as follows. Consider a meromorphic function $\mathcal{F}(r,w,w')$ that has N poles at $w'=w_n$ $(n=1,\ldots,N)$ anywhere in the w'-plane except at $w'=w$. A variant of Eq. (A1) is

$$G(rw,rw^{-1})=\frac{1}{2\pi i}\oint_{C_w}dw'\left(\frac{1}{w'-w}+\mathcal{F}(r,w,w')\right)G(rw',rw'^{-1}). \tag{A6}$$

Repeating the previous contour argument, one can easily derive the generalized dispersion relation

$$\begin{aligned}G(rw,rw^{-1})=&-\sum_{n=1}^{N}\mathrm{Res}_{w'\to w_n}\{\mathcal{F}(r,w,w')G(rw',rw'^{-1})\}\\&+\oint_{arc}\frac{dw'}{2\pi i}\left[\left(\frac{1}{w'-w}+\mathcal{F}(r,w,w')\right)G(rw',rw'^{-1})\right]\\&+\left(\int_{-\infty}^{-r^{-1}}+\int_{r^{-1}}^{\infty}\right)dw'\left\{\left(\mathcal{K}(w,w')+\frac{1}{2\pi}\mathcal{F}(r,w,w')-\frac{1}{2\pi w'^2}\mathcal{F}(r,w,w'^{-1})\right)\right.\\&\left.\times\,\mathrm{Disc}[G(rw',rw'^{-1})]\right\}.\end{aligned} \tag{A7}$$

For completeness, in this case we have explicitly included a potential arc contribution at $|w'|\to\infty$ in the second line of this expression. If G is suitably subtracted so that the $\frac{1}{w'-w}G(rw',rw'^{-1})$ term does not contribute and $\mathcal{F}$ is reasonably well behaved at infinity, we can safely set the arc contribution to zero. Alternatively, the arc contribution can be made to vanish if the auxiliary function $\mathcal{F}$ is chosen such that $\mathcal{F}\sim-\frac{1}{w'}$ at large $|w'|$.

Special cases. We note two special cases:

(i) For $\mathcal{F}=0$ we recover trivially (A4).

(ii) Ref. [23] chose

$$\mathcal{F}(r,w,w')=-\frac{1}{2(w'-w)}+\frac{1}{2(w'+w)}-\frac{1}{w'}. \tag{A8}$$

2. Dispersion relations from the Lorentzian OPE inversion formula

The dispersion relations (A7) can also be derived from the Lorentzian OPE inversion formula [20]. We note here that this is unlike the corresponding situation for the dispersion relations in zero-temperature four-point correlation functions, where the direct application of Cauchy's theorem yields the Bissi-Dey-Hansen dispersion relation [48] in terms of the single-discontinuity, whereas a derivation via the Lorentzian OPE inversion formula [49] yields the Carmi-Caron Huot dispersion relation [32]. The latter is expressed in terms of the double-discontinuity of the four-point correlation function.

Here, we summarize the derivation of a dispersion relation for the subtracted thermal two-point function from the Lorentzian OPE inversion formula following [23]. This is the main dispersion relation used in the main text. The resulting formula does not follow from a direct application of (A7).

First, let us recall the spectral sum decomposition of the two-point function $g(z,\bar{z})$ and the Lorentzian OPE inversion formula for $g(z,\bar{z})$, [20],

$$\begin{aligned}g(rw,rw^{-1})=&\sum_{J=0}^{\infty}\oint_{-\epsilon-i\infty}^{-\epsilon+i\infty}\frac{d\Delta}{2\pi i}a(\Delta,J)\\&\times C_J^{(\nu)}\left(\frac{1}{2}(w+w^{-1})\right)r^{\Delta-2\Delta_\phi},\end{aligned} \tag{A9}$$

$$\begin{aligned}a(\Delta,J)=&K_J\int_0^1\frac{dr}{r}r^{2\Delta_\phi-\Delta}\left\{(-1)^{d+1}\int_{-r}^{r}\frac{dw}{w}(w-w^{-1})^{2\nu}F_J(w)\mathrm{Disc}[g(rw,rw^{-1})]\right.\\&+\int_{-\infty}^{-r^{-1}}\frac{dw}{w}(w-w^{-1})^{2\nu}F_J(w^{-1})\mathrm{Disc}[g(rw,rw^{-1})]\\&\left.+\int_{r^{-1}}^{\infty}\frac{dw}{w}(w-w^{-1})^{2\nu}F_J(w^{-1})\mathrm{Disc}[g(rw,rw^{-1})]\right\}\\&+\theta(J_0-J)a_{\mathrm{arc}}(\Delta,J).\end{aligned} \tag{A10}$$

The coefficient K_J is

$$K_J = \frac{\Gamma(J+1)\Gamma(\nu)}{4\pi\Gamma(J+\nu)}, \tag{A11}$$

precisely as in Eq. (3.5) in the main text. In the last line of (A10) we included the arc contributions for $J \leq J_0$. J_0 has a value related to the Regge behavior of the discontinuity, [20].

Following [23], we rewrite (A9) by separating out the contribution of a truncated s-channel expansion,

$$\begin{aligned} g(rw, rw^{-1}) = &\sum_{J=0}^{J_*} \sum_{\Delta} a_{\Delta,J} C_J^{(\nu)}\left(\frac{1}{2}(w+w^{-1})\right) r^{\Delta-2\Delta_\phi} \\ &+ \sum_{J>J_*}^{\infty} \int_{-\epsilon-i\infty}^{-\epsilon+i\infty} \frac{d\Delta}{2\pi i} a(\Delta, J) C_J^{(\nu)} \\ &\times \left(\frac{1}{2}(w+w^{-1})\right) r^{\Delta-2\Delta_\phi}. \end{aligned} \tag{A12}$$

Notice that the separated first term is truncated only in spin. It still includes an infinite sum over scaling dimensions. J_* is an arbitrary cutoff on the spin (it can be less, equal, or larger than J_0). We want to insert $a(\Delta, J)$ from the Lorentzian OPE inversion formula (A10) into the last term of (A12) and obtain an expression in terms of the discontinuity of g. It is clear without computation that this manipulation will produce a term that involves an integral of Disc$[g]$ multiplied by a modified kernel, which should be contrasted with the expression in (A7), which involves the fixed kernel $\mathcal{K}(w, w')$ in Eq. (A5).

The last term in (A12) is

$$\mathcal{I} = \sum_{J>J_*}^{\infty} \int_{-\epsilon-i\infty}^{-\epsilon+i\infty} \frac{d\Delta}{2\pi i} a(\Delta, J) C_J^{(\nu)}\left(\frac{1}{2}(w+w^{-1})\right) r^{\Delta-2\Delta_\phi}. \tag{A13}$$

Substituting $a(\Delta, J)$ in the form provided by the Lorentzian OPE inversion formula in (A10), exchanging the J-sum and Δ-integral with the r, w-integrals in (A10), yields [20] the expression

$$\begin{aligned} \mathcal{I} = 2\left(\int_{-\infty}^{-r^{-1}} + \int_{r^{-1}}^{\infty}\right) dw' \mathcal{K}_{J_*}(w, w') \\ \times \text{Disc}[g(rw', rw'^{-1})] + \mathcal{I}_{\text{arc}}, \end{aligned} \tag{A14}$$

where

$$\begin{aligned} \mathcal{K}_{J_*}(w, w') &= w'^{-1}(w' - w'^{-1})^{2\nu}\left[\sum_{J=J_*+1}^{\infty} K_J C_J^{(\nu)}\left(\frac{1}{2}(w+w^{-1})\right) F_J(w'^{-1})\right] \\ &= \frac{1}{2}\mathcal{K}(w, w') - w'^{-1}(w' - w'^{-1})^{2\nu}\left[\sum_{J=0}^{J_*} K_J C_J^{(\nu)}\left(\frac{1}{2}(w+w^{-1})\right) F_J(w'^{-1})\right]. \end{aligned} \tag{A15}$$

$\mathcal{K}(w, w')$ is the kernel that appears in the standard thermal dispersion relation (A5). We note that this derivation requires the mathematical identity

$$\mathcal{K}(w, w') = 2w'^{-1}(w' - w'^{-1})^{2\nu}\left[\sum_{J=0}^{\infty} K_J C_J^{(\nu)}\left(\frac{1}{2}(w+w^{-1})\right) F_J(w'^{-1})\right], \tag{A16}$$

which is valid in any spacetime dimension.[18] The cases $d = 2$ and $d = 4$ provide a simple verification of these expressions.

$\mathcal{I}_{\text{arc}}$ is a potential contribution from a_{arc} in (A10),

$$\mathcal{I}_{\text{arc}} = \sum_{J>J_*}^{\infty} \int_{-\epsilon-i\infty}^{-\epsilon+i\infty} \frac{d\Delta}{2\pi i} \theta(J_0 - J) a_{\text{arc}}(\Delta, J) C_J^{(\nu)}\left(\frac{1}{2}(w+w^{-1})\right) r^{\Delta-2\Delta_\phi}. \tag{A17}$$

This contribution requires $J \leq J_0$ to be nonvanishing. Therefore, if $J_* < J_0$, then

$$\mathcal{I}_{\text{arc}} = \sum_{J=J_*+1}^{J_0} \int_{-\epsilon-i\infty}^{-\epsilon+i\infty} \frac{d\Delta}{2\pi i} a_{\text{arc}}(\Delta, J) C_J^{(\nu)}\left(\frac{1}{2}(w+w^{-1})\right) r^{\Delta-2\Delta_\phi}, \tag{A18}$$

but whenever $J_* \geq J_0$, it vanishes automatically. In the main text we present formulas with the implicit assumption $J_* \geq J_0$.

[18]Notice that the lhs of (A16) is manifestly spacetime-dimension-independent, but the rhs involves a series over spacetime-dimension-dependent terms.

To summarize, in this subsection we have shown that

$$g(rw, rw^{-1}) = \sum_{J=0}^{J_*} \sum_{\Delta} a_{\Delta,J} C_J^{(\nu)}\left(\frac{1}{2}(w+w^{-1})\right) r^{\Delta-2\Delta_\phi} + 2\left(\int_{-\infty}^{-r^{-1}} + \int_{r^{-1}}^{\infty}\right) dw'\, \mathcal{K}_{J_*}(w,w') \times \mathrm{Disc}[g(rw', rw'^{-1})] + \mathcal{I}_{\mathrm{arc}}. \tag{A19}$$

APPENDIX B: UNIVERSAL FEATURES OF THE TAIL FUNCTIONS IN THE VICINITY OF $r=1$

In this appendix, we argue that, under certain conditions, the tail functions $A_J(r)$ exhibit a universal behavior in the limit $r \to 1$, follows from the KMS condition. Information about this universal behavior can be used to facilitate better optimization results when we vary tail functions in search of KMS solutions.

The conformal block expansion of a thermal two-point function of identical scalar operators is

$$g(rw, rw^{-1}) = r^{-2\Delta_\phi} + \sum_{\ell=0}^{\infty} A_{2\ell}(r) C_{2\ell}^{(\nu)}(\xi(w)), \tag{B1}$$

where $\xi(\omega) = \frac{1}{2}(w+w^{-1})$. Using the orthogonality of the Gegenbauer polynomials

$$\int_{-1}^{1} d\xi (1-\xi^2)^{\nu-\frac{1}{2}} C_n^{(\nu)}(\xi) C_m^{(\nu)}(\xi) = h_n^{(\nu)} \delta_{nm}, \qquad h_n^{(\nu)} = \frac{2^{1-2\nu}\pi\Gamma(n+2\nu)}{n!(n+\nu)\Gamma(\nu)^2}, \tag{B2}$$

we obtain

$$A_{2\ell}(r) = \frac{1}{h_{2\ell}^{(\nu)}} \int_{-1}^{1} d\xi (1-\xi^2)^{\nu-\frac{1}{2}} C_{2\ell}^{(\nu)}(\xi) \times [g(rw(\xi), rw(\xi)^{-1}) - r^{-2\Delta_\phi}]. \tag{B3}$$

The function $g(z,\bar{z})$ diverges at $z=-1, 1$. The divergence is captured by the KMS condition through the identity contribution in the crossed channel

$$g(z,\bar{z}) \sim \frac{1}{|1-z|^{2\Delta_\phi}} = \frac{1}{(1-2r\xi+r^2)^{\Delta_\phi}} \quad \text{for } z \to 1, \tag{B4}$$

and

$$g(z,\bar{z}) \sim \frac{1}{|1+z|^{2\Delta_\phi}} = \frac{1}{(1+2r\xi+r^2)^{\Delta_\phi}} \quad \text{for } z \to -1. \tag{B5}$$

This behavior introduces a potential $r \to 1$ divergence in the functions $A_{2\ell}(r)$ in (B3) from the regions of the ξ-integral near $\xi = 1$ and $\xi = -1$ as $r \to 1$. If there is such a divergence, we expect the functions

$$\tilde{A}_{2\ell}(r) = \frac{1}{h_{2\ell}^{(\nu)}} \int_0^1 d\xi (1-\xi^2)^{\nu-\frac{1}{2}} C_{2\ell}^{(\nu)}(\xi) \frac{1}{(1-2r\xi+r^2)^{\Delta_\phi}} + \frac{1}{h_{2\ell}^{(\nu)}} \int_{-1}^0 d\xi (1-\xi^2)^{\nu-\frac{1}{2}} C_{2\ell}^{(\nu)}(\xi) \frac{1}{(1+2r\xi+r^2)^{\Delta_\phi}} \tag{B6}$$

to exhibit the same behavior near the corresponding point as $A_{2\ell}(r)$.

In fact, we can maintain the correct form of the divergence and simplify things even further by noticing that in the vicinity of $\xi \sim \pm 1$

$$C_{2\ell}^{(\nu)}(\xi) \sim f_{2\ell}^{(\nu)} := \frac{2^{1-2\nu}\sqrt{\pi}\Gamma(2\ell+2\nu)}{(2\ell)!\Gamma(\nu)\Gamma(\nu+\frac{1}{2})}. \tag{B7}$$

The quantity

$$\tilde{\tilde{A}}_{2\ell}(r) = \frac{f_{2\ell}^{(\nu)}}{h_{2\ell}^{(\nu)}} \left(\int_0^1 d\xi \frac{(1-\xi^2)^{\nu-\frac{1}{2}}}{(1-2r\xi+r^2)^{\Delta_\phi}} + \int_{-1}^0 d\xi \frac{(1-\xi^2)^{\nu-\frac{1}{2}}}{(1+2r\xi+r^2)^{\Delta_\phi}}\right) = \frac{2(2\ell+\nu)\Gamma(\nu)}{\sqrt{\pi}\Gamma(\frac{1}{2}+\nu)} \int_0^1 d\xi \frac{(1-\xi^2)^{\nu-\frac{1}{2}}}{(1-2r\xi+r^2)^{\Delta_\phi}} \tag{B8}$$

is also expected to have the same singularity structure as $A_{2\ell}(r)$.

The expression (B8) implies that $A_{2\ell}(r)$ is divergent at $r=1$ when

$$\nu - \frac{1}{2} - \Delta_\phi \le -1 \Leftrightarrow \Delta_\phi \ge \nu + \frac{1}{2}. \tag{B9}$$

In the narrow window, $\nu \leq \Delta_\phi < \nu + \frac{1}{2}$, it implies instead that the tail functions $A_{2\ell}(r)$ are regular at $r = 1$, but this is not in immediate contradiction with the fact that $g(z, \bar{z})$ diverges at $z \to \pm 1$. Indeed, at $z = \pm 1$, the series

$$g(1,1) = 1 + \sum_{\ell=0}^{\infty} A_{2\ell}(1) C_{2\ell}^{(\nu)}(1) = 1 + \sum_{\ell=0}^{\infty} \frac{\Gamma(2\ell + 2\nu)}{(2\ell)!\Gamma(2\nu)} A_{2\ell}(1) \tag{B10}$$

can still diverge for finite $A_{2\ell}(1)$.

When the tail functions are divergent, the above argument also implies that the divergence is universal, independent of the details of the CFT, and directly related to the identity contribution in the crossed channel. We can verify the condition for the divergence, (B9), and determine its type in the GFF case as follows.

For GFFs

$$A_{2\ell}(r) = r^{2\ell} \sum_{n=0}^{\infty} a_{n,2\ell} r^{2n}, \tag{B11}$$

with

$$a_{n,J} = 2\zeta(2\Delta_\phi + 2n + J) \frac{(J+\nu)(\Delta_\phi)_{J+n}(\Delta_\phi - \nu)_n}{n!(\nu)_{J+n+1}}. \tag{B12}$$

At leading order in the large-n limit (at fixed J),

$$a_{n,J} \sim \frac{2(J+\nu)\Gamma(\nu)}{\Gamma(\Delta_\phi)\Gamma(\Delta_\phi - \nu)} (n(n+J))^{\Delta_\phi - \nu - 1}, \tag{B13}$$

and the series

$$A_{2\ell}(1) = \sum_{n=0}^{\infty} a_{n,2\ell} \tag{B14}$$

diverges when

$$2\Delta_\phi - 2\nu - 2 \geq -1 \Leftrightarrow \Delta_\phi - \nu - \frac{1}{2} \geq 0, \tag{B15}$$

reproducing the inequality in (B9).

Moreover, the asymptotic behavior of the a-coefficients in (B13) suggests that in GFF the functions $A_{2\ell}(r)$ diverge in the following way as $r \to 1^-$:

$$A_{2\ell}(r) \sim \frac{2(2\ell+\nu)\Gamma(\nu)}{\Gamma(\Delta_\phi)\Gamma(\Delta_\phi - \nu)} \frac{\Gamma(2\Delta_\phi - 2\nu - 1)}{(1-r^2)^{2\Delta_\phi - 2\nu - 1}} r^{2\ell}. \tag{B16}$$

APPENDIX C: NEURAL NETWORKS

As we mentioned in the main text, the tail functions $A_{\Delta_*(J),J}(r)$ are modeled by an artificial neural network. In this appendix, we provide more details regarding the architecture we employed, as well as the reasoning behind it.

1. GFF

The functions we are trying to capture are defined in Eq. (3.9). For double-twist operators, they reduce to

$$A(r)_{n_*(J),J} = \sum_{n=n_*(J)+1}^{\infty} a_{n,J} r^{2n+J}. \tag{C1}$$

Optimizing a model to directly capture (C1) can be challenging. For example, the model could learn the wrong small r behavior, i.e., the wrong starting power, as that information resides only within Eq. (C1) and not within any of the loss functions we use. Similarly, the model could also fail to capture the $r \to 1$ behavior that is implied by the KMS condition,

$$A(r)_{n_*(J),J} \propto (1-r^2)^{1+2\nu-2\Delta_\phi}, \tag{C2}$$

as described in Appendix B. In addition, the values of the tail functions close to $r \to 1$ can be several orders of magnitude larger than those at other points, which can cause severe numerical instabilities.

For all these reasons, we define our neural network model as

$$A(r)_{n_*(J),J} = 2r^{2n_*(J)+J+2}(1-r^2)^{1+2\nu-2\Delta_\phi} \sinh\{\mathrm{NN}(r^2)\}, \tag{C3}$$

where $\mathrm{NN}(r)$ is the trainable neural network. In this expression, the prefactor implements the correct low and high r behavior, as well as the boundary condition $A(0)_{n_*(J),J} = 0$. The power of r in the prefactor is chosen to be such that the desired $\mathrm{NN}(r^2)$ behaves as a constant at low r. The hyperbolic sine function is introduced for regularization purposes. For small output values, it simply behaves like $\mathrm{NN}(r^2)$, whereas for larger outputs it behaves like $\exp[\mathrm{NN}(r^2)]$. In that way, the huge output variation needed to capture $A(r)_{n_*(J),J}$ is achieved without a huge variation of the trainable model's output. Finally, the model's input is always squared in order to enforce the $r \to -r$ symmetry that the GFF $A(r)_{n_*(J),J}$ have by construction.

As for the trainable $\mathrm{NN}(r^2)$, the model consists of two main parts: the backbone and the subnetworks. The backbone is a fully connected deep neural network (DNN) that performs the first processing of the input. Consequently, the

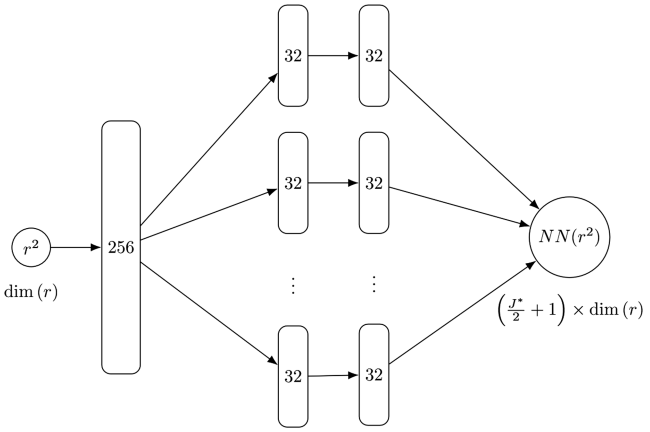


FIG. 11. Neural network model used for the GFF runs. This architecture sums a total of $512 + (\frac{J_*}{2}+1)9313$ trainable parameters.

output is passed to $\frac{J_*}{2}+1$ subnetworks. These subnets aim to capture the contribution of each spin, i.e., to represent the sums (C1) for different spins. Between layers, on both backbone and subnets, the inputs are passed through hyperbolic tangent activation functions. Finally, each subnet yields a single value leading to the full model, outputting a $\frac{J_*}{2}+1$ sized vector. A visual representation of such an architecture is given in Fig. 11.

Another point worth highlighting is the number of NN trainable parameters. A model with our proposed architecture involves a total of

$$\mathcal{P} = 2n_b + (\mathcal{B}-1)n_b(n_b+1) + \left(\frac{J_*}{2}+1\right) \times (n_s(n_b+1) + (\mathcal{S}-1)n_s(n_s+1) + n_s + 1) \quad \text{(C4)}$$

parameters, where

(i) n_b is the number of nodes per backbone layer,
(ii) $\mathcal{B} > 0$ is the number of backbone layers,
(iii) n_s is the number of nodes per subnet layer, and
(iv) $\mathcal{S} > 0$ is the number of subnet layers.

In our applications, we observed that we were able to efficiently bootstrap the KMS condition with relatively small models (for GFF, $n_b = 256$, $\mathcal{B} = 1$, $n_s = 32$, and $\mathcal{S} = 2$). In most cases, our models had a few tens of thousands parameters and individual runs could be performed on a laptop with execution times of about 45 minutes for 50 K epochs. It is possible that larger models might be able to yield better results at the cost of requiring greater computational power to train.

2. Holographic CFTs

In the case of holographic CFTs, we used a very similar model, with some minor modifications. Here, we want to model sums of the form

$$A_J(r) = \sum_{n=0}^{\infty} a_{n,J} r^{2n+J} + \sum_{k=\frac{J}{2}}^{\infty} a_J^{(k)} r^{kd-2\Delta_\phi} \quad \text{(C5)}$$

for a some appropriate lower boundary truncation characterized by some $\Delta_*(J)$.

The first difference from the GFF model comes from the existence of the energy-momentum data in the spectrum. These terms break the $r \to -r$ symmetry of the GFF case, and therefore, we should not square the input before passing it to the neural network. Second, some changes have to be made to the starting powers in the prefactor. The purpose here is to once again factor out the leading powers for the tails at small r. This is not as straightforward as in the GFF case since the leading power for each J can change as a function of Δ_ϕ, the spacetime dimension d, and the exposed operators. In general, the model can be defined as

$$A(r)_{\Delta_*(J),J} = 2r^{Q(\Delta_*,J,\Delta_\phi,d)}(1-r^2)^{1+2\nu-2\Delta_\phi} \sinh\{\mathrm{NN}(r)\}, \quad \text{(C6)}$$

where the power

$$Q(\Delta_*(J),J,\Delta_\phi,d) = \min\{2(n_*+1)+J, d(k_*+1)-2\Delta_\phi\}, \quad \text{(C7)}$$

with n_*, k_* being the maximum n, k of exposed operators for a given $\Delta_*(J)$. The rest of the model specifications, such as the architecture and the number of trainable parameters, was the same as in the GFF case.

APPENDIX D: LARGE-c SCALING OF CORRELATION FUNCTIONS AT FINITE TEMPERATURE

The organization of the large-c expansion of correlation functions in CFT is considerably different at finite temperature compared to the zero-temperature case. Although this is well known, we are not aware of a detailed discussion of the relevant properties in the literature. Therefore, for the convenience of the reader, we include here a short review of pertinent material.

Motivated by large-N gauge theories, we set $c := N^2$ and use a common terminology in terms of single- and multi-trace operators (which refers to operators with specific large-N scaling properties).

We consider the thermal two-point function $\langle\phi\phi\rangle_\beta$ of a single-trace operator ϕ and write schematically its conformal block expansion as

$$\langle\phi(x_1)\phi(x_2)\rangle_\beta = \sum_{\mathcal{O}} \frac{\langle\phi\phi\mathcal{O}\rangle_{\beta=0}}{\langle\mathcal{O}\mathcal{O}\rangle_{\beta=0}} \langle\mathcal{O}(x_2)\rangle_\beta, \quad \text{(D1)}$$

with an appropriate sum over single- and multi-trace conformal primaries $\mathcal{O}$.

1. *Case I:* $\langle\phi\rangle_\beta = 0$

In this case, $\langle\phi(x_1)\phi(x_2)\rangle_\beta$ has no disconnected contribution and scales at leading order in the large-N limit as N^0. We proceed to discuss in detail the leading large-N scaling of each potential contribution on the rhs of (D1). To make the notation lighter, we will denote zero-temperature correlators by dropping the β-subscript,

$$\langle\cdots\rangle \coloneqq \langle\cdots\rangle_{\beta=0}.$$

$\mathcal{O} = \phi^2$ We have

$$\langle\phi\phi\phi^2\rangle \sim N^0, \qquad \langle\phi^2\phi^2\rangle \sim N^0. \tag{D2}$$

Since ϕ has no thermal vev, the leading $\mathcal{O}(N^2)$ contribution from large-N factorization vanishes, yielding

$$\langle\phi^2\rangle_\beta \sim N^0. \tag{D3}$$

Consequently,

$$\frac{\langle\phi\phi\phi^2\rangle}{\langle\phi^2\phi^2\rangle}\langle\phi^2(x_2)\rangle_\beta \sim N^0. \tag{D4}$$

$\mathcal{O}_{k,n} = \phi\partial^k\Box^n\phi$ For the rest of the double-twist operators, we have similarly

$$\langle\phi\phi\mathcal{O}_{k,n}\rangle \sim N^0, \quad \langle\mathcal{O}_{k,n}\mathcal{O}_{k,n}\rangle \sim N^0, \quad \langle\mathcal{O}_{k,n}\rangle_\beta \sim N^0, \tag{D5}$$

yielding

$$\frac{\langle\phi\phi\mathcal{O}_{k,n}\rangle}{\langle\mathcal{O}_{k,n}\mathcal{O}_{k,n}\rangle}\langle\mathcal{O}_{k,n}(x_2)\rangle_\beta \sim N^0. \tag{D6}$$

$\mathcal{O} = \phi^n_\partial$ for $n \geq 3$ By ϕ^n_∂ we denote compactly any n-trace (conformal primary) operator with or without derivatives. Since the leading-order large-N scaling of the (connected) correlation $\langle\phi\phi\phi^n_\partial\rangle$ is N^{-n} [the same as that of the connected $(n+2)$-point function $\langle\phi(x_1)\phi(x_2)\prod_{i=1}^n\phi(y_i)\rangle$], we obtain

$$\langle\phi\phi\phi^n_\partial\rangle \sim N^{-n}, \quad \langle\phi^n_\partial\phi^n_\partial\rangle \sim N^0, \quad \langle\phi^n_\partial\rangle_\beta \sim N^\alpha, \ \alpha < n, \tag{D7}$$

yielding

$$\frac{\langle\phi\phi\phi^n_\partial\rangle}{\langle\phi^n_\partial\phi^n_\partial\rangle}\langle\phi^n_\partial(x_2)\rangle_\beta \sim N^{\alpha-n}, \tag{D8}$$

which is subleading to N^0. Hence, n-trace operators with $n \geq 3$ (beyond the double-twist operators) *cannot appear* at leading order in the large-N limit in the conformal block expansion (D1).

$\mathcal{O} = T_{\mu\nu}$ In this case,

$$\langle\phi\phi T\rangle \sim N^{-1}, \qquad \langle TT\rangle \sim N^0, \qquad \langle T\rangle_\beta \sim N, \tag{D9}$$

yielding

$$\frac{\langle\phi\phi T\rangle}{\langle TT\rangle}\langle T(x_2)\rangle_\beta \sim N^0. \tag{D10}$$

Similar behavior is exhibited by any single-trace operator that has a nonvanishing thermal vev.

$\mathcal{O} = T^n$ In this case,

$$\langle\phi\phi T^n\rangle \sim N^{-n}, \quad \langle T^nT^n\rangle \sim N^0, \quad \langle T^n\rangle_\beta \sim N^n, \tag{D11}$$

yielding

$$\frac{\langle\phi\phi T^n\rangle}{\langle T^nT^n\rangle}\langle T^n(x_2)\rangle_\beta \sim N^0. \tag{D12}$$

$\mathcal{O} = T^n\phi^k$ with or without derivatives In this case, the leading large-N factorization contribution to the thermal vev of the operator vanishes, yielding (as in similar cases above) a contribution to the conformal block expansion that is subleading. Hence, such operators do not contribute to the leading order. A special case are multi-trace energy-momentum operators with derivatives, the contributions of which we dropped in the main text.

2. *Case II:* $\langle\phi\rangle_\beta \neq 0$

In this case, $\langle\phi(x_1)\phi(x_2)\rangle_\beta$ has a disconnected contribution that scales at leading order in the large-N limit as N^2. We now discuss in detail how to reproduce both the disconnected $\mathcal{O}(N^2)$ contribution and the leading $\mathcal{O}(N^0)$ connected contributions in the conformal block expansion.

$\mathcal{O} = \phi^2$ We have

$$\langle\phi\phi\phi^2\rangle \sim N^0, \qquad \langle\phi^2\phi^2\rangle \sim N^0. \tag{D13}$$

Since ϕ has a nonvanishing thermal vev,

$$\langle\phi^2\rangle_\beta \sim N^2, \tag{D14}$$

and the leading $\mathcal{O}(N^2)$ contribution from large-N factorization is reproduced as expected from the term

$$\frac{\langle\phi\phi\phi^2\rangle}{\langle\phi^2\phi^2\rangle}\langle\phi^2(x_2)\rangle_\beta \sim N^2. \tag{D15}$$

$\mathcal{O}_{k,n} = \phi\partial^k\Box^n\phi$ Since the leading $\mathcal{O}(N^2)$ contribution to the large-N factorization of $\langle\mathcal{O}_{k,n}\rangle_\beta$ is annihilated by the derivatives, for the rest of the double-twist operators we have

$$\langle\phi\phi\mathcal{O}_{k,n}\rangle \sim N^0, \qquad \langle\mathcal{O}_{k,n}\mathcal{O}_{k,n}\rangle \sim N^0, \qquad \langle\mathcal{O}_{k,n}\rangle_\beta \sim N^0, \tag{D16}$$

yielding

$$\frac{\langle\phi\phi\mathcal{O}_{k,n}\rangle}{\langle\mathcal{O}_{k,n}\mathcal{O}_{k,n}\rangle}\langle\mathcal{O}_{k,n}(x_2)\rangle_\beta \sim N^0. \tag{D17}$$

$\mathcal{O} = \phi^n$ for $n \geq 3$ Here, we do not include any derivatives. We obtain

$$\langle\phi\phi\phi^n\rangle \sim N^{-n}, \quad \langle\phi^n\phi^n\rangle \sim N^0, \quad \langle\phi^n\rangle_\beta \sim N^n, \tag{D18}$$

yielding

$$\frac{\langle\phi\phi\phi^n\rangle}{\langle\phi^n\phi^n\rangle}\langle\phi^n(x_2)\rangle_\beta \sim N^0, \tag{D19}$$

which contributes to the leading $\mathcal{O}(N^0)$ result. In contrast, n-trace products of ϕ with derivatives will not contribute at leading order because the leading $\mathcal{O}(N^n)$ contribution to their vev vanishes.

$\mathcal{O} = T_{\mu\nu}$ In this case,

$$\langle\phi\phi T\rangle \sim N^{-1}, \qquad \langle TT\rangle \sim N^0, \qquad \langle T\rangle_\beta \sim N, \tag{D20}$$

yielding

$$\frac{\langle\phi\phi T\rangle}{\langle TT\rangle}\langle T(x_2)\rangle_\beta \sim N^0. \tag{D21}$$

Similar behavior is exhibited by any single-trace operator that has a nonvanishing thermal vev.

$\mathcal{O} = T^n$ In this case,

$$\langle\phi\phi T^n\rangle \sim N^{-n}, \quad \langle T^nT^n\rangle \sim N^0, \quad \langle T^n\rangle_\beta \sim N^n, \tag{D22}$$

yielding

$$\frac{\langle\phi\phi T^n\rangle}{\langle T^nT^n\rangle}\langle T^n(x_2)\rangle_\beta \sim N^0. \tag{D23}$$

$\mathcal{O} = T^n\phi^k$with or without derivatives Similar to the previous subsection, when there are derivatives the leading large-N factorization contribution to the thermal vev of the operator vanishes, yielding a contribution to the conformal block expansion that is subleading. However, operators of the schematic form $\mathcal{O} = T^n\phi^k$ without derivatives, can contribute to leading order since

$$\langle\phi\phi[T^n\phi^k]\rangle \sim N^{-n-k}, \qquad \langle[T^n\phi^k][T^n\phi^k]\rangle \sim N^0, \qquad \langle[T^n\phi^k]\rangle_\beta \sim N^{n+k}, \tag{D24}$$

yielding

$$\frac{\langle\phi\phi[T^n\phi^k]\rangle}{\langle[T^n\phi^k][T^n\phi^k]\rangle}\langle[T^n\phi^k](x_2)\rangle_\beta \sim N^0. \tag{D25}$$

APPENDIX E: INTEGER EXTERNAL SCALING DIMENSIONS IN HOLOGRAPHIC CFTS

In this section, we review the appearance of logarithms in the conformal block expansions of thermal two-point functions $\langle\phi\phi\rangle$ in holographic CFTs when the scaling dimension of the external operator Δ_ϕ is integer. The source of the logarithms are poles in the three-point function coefficients $C_{\phi\phi[T^k]_J}$ when Δ_ϕ is an integer. These poles are canceled by corresponding poles in the three-point function coefficients $C_{\phi\phi[\phi^2]_{n,J}}$ with double-twist operators. We review the mechanism behind this cancellation and how it leaves behind logarithmic contributions.

When $\Delta_\phi = p \in \mathbb{N}$, some energy-momentum multi-trace operators can be degenerate with double-twist operators. For example, let us assume that the energy-momentum multi-trace operator $\mathcal{O}_T \coloneqq [T^k]_J$ is degenerate with the double-twist operator $\mathcal{O}_\phi \coloneqq [\phi\phi]_{n,J}$. This implies

$$\Delta_{[T^k]_J} = 4k = 2p + 2n + J = \Delta_{[\phi\phi]_{n,J}} \Leftrightarrow 2k = p + n + \frac{J}{2}. \tag{E1}$$

Such operators contribute to the thermal conformal block expansion with terms of the form (up to a constant factor $\beta^{-\Delta}\frac{J!}{2^J(\nu)_J}$)

$$(f_{\phi\phi\mathcal{O}_T}g^{\mathcal{O}_T\mathcal{O}_T}b_{\mathcal{O}_T} + f_{\phi\phi\mathcal{O}_\phi}g^{\mathcal{O}_\phi\mathcal{O}_\phi}b_{\mathcal{O}_\phi})G^{\Delta_\phi}_{\Delta,J}(z,\bar{z}), \tag{E2}$$

where $G^{\Delta_\phi}_{\Delta,J}(z,\bar{z}) = C^{(\nu)}_J(\frac{1}{2}(w+w^{-1}))r^{\Delta-2\Delta_\phi}$ is the standard thermal conformal block. In (E2) $\Delta = \Delta_{\mathcal{O}_\phi} = \Delta_{\mathcal{O}_T}$.

The effect we want to discuss follows from the fact that the corresponding three-point function coefficients share a similar pole structure (see e.g., [41] Eqs. (B.4)–(B.6), [35], and Appendix F for concrete examples) of the form

$$f_{\phi\phi\mathcal{O}_T} \propto \frac{\lambda^{(\mathcal{O}_T)}_{\text{sing}}}{\Delta_\phi - p} + \lambda_{\phi\phi\mathcal{O}_T}, \tag{E3}$$

$$f_{\phi\phi\mathcal{O}_\phi} \propto \frac{\lambda^{(\mathcal{O}_\phi)}_{\text{sing}}}{\Delta_\phi - p} + \lambda_{\phi\phi\mathcal{O}_\phi}, \tag{E4}$$

with simple poles at integer values of Δ_ϕ.

To understand what happens when $\Delta_\phi = p$, we set $\varepsilon \coloneqq \Delta_\phi - p$ and discuss the limit $\varepsilon \to 0$. For starters, let us isolate any of the terms $z^{\frac{\Delta-J}{2}-\Delta_\phi+s}\bar{z}^{\frac{\Delta+J}{2}-\Delta_\phi-s}$ in the power expansion of

$$G^{\Delta_\phi}_{\Delta,J}(z,\bar{z}) = \sum_{s=0}^{J} p_s(J) z^{\frac{\Delta-J}{2}-\Delta_\phi+s} \bar{z}^{\frac{\Delta+J}{2}-\Delta_\phi-s}. \tag{E5}$$

When $\varepsilon \ll 1$, we can expand in the following manner:

$$\begin{aligned} z^{\frac{\Delta_{\mathcal{O}_T}-J}{2}-\Delta_\phi+s} \bar{z}^{\frac{\Delta_{\mathcal{O}_T}+J}{2}-\Delta_\phi-s} &= z^{2k-\frac{J}{2}-p+s-\varepsilon} \bar{z}^{2k+\frac{J}{2}-p-s-\varepsilon} \\ &\simeq z^{2k-\frac{J}{2}-p+s}(1-\varepsilon\log z)\bar{z}^{2k+\frac{J}{2}-p-s}(1-\varepsilon\log\bar{z})+\ldots, \\ z^{\frac{\Delta_{\mathcal{O}_\phi}-J}{2}-\Delta_\phi+s} \bar{z}^{\frac{\Delta_{\mathcal{O}_\phi}+J}{2}-\Delta_\phi-s} &= z^{n+s}\bar{z}^{n+J-s} = z^{2k-\frac{J}{2}-p+s}\bar{z}^{2k-p+\frac{J}{2}-s}+\ldots, \end{aligned} \tag{E6}$$

where the dots indicate terms with higher powers in ε. In the last line, and in the second equality, we used (E1). Then, as we take the limit $\varepsilon \to 0$, the corresponding term in the expression (E2) behaves as

$$\begin{aligned} &\left(\frac{\lambda^{(\mathcal{O}_T)}_{\text{sing}}}{\varepsilon}+\lambda_{\phi\phi\mathcal{O}_T}\right) g^{\mathcal{O}_T\mathcal{O}_T} b_{\mathcal{O}_T}(1-\varepsilon\log z) z^{2k-\frac{J}{2}-p+s}(1-\varepsilon\log\bar{z})\bar{z}^{2k-p+\frac{J}{2}-s} \\ &\quad+\left(\frac{\lambda^{(\mathcal{O}_\phi)}_{\text{sing}}}{\varepsilon}+\lambda_{\phi\phi\mathcal{O}_\phi}\right) g^{\mathcal{O}_\phi\mathcal{O}_\phi} b_{\mathcal{O}_\phi} z^{2k-\frac{J}{2}-p+s}\bar{z}^{2k-p+\frac{J}{2}-s}. \end{aligned} \tag{E7}$$

The cancellation of the divergent ε^{-1} terms requires

$$\lambda^{(\mathcal{O}_T)}_{\text{sing}} g^{\mathcal{O}_T\mathcal{O}_T} b_{\mathcal{O}_T} = -\lambda^{(\mathcal{O}_\phi)}_{\text{sing}} g^{\mathcal{O}_\phi\mathcal{O}_\phi} b_{\mathcal{O}_\phi}. \tag{E8}$$

After this cancellation, and the limit $\varepsilon \to 0$, the expression (E7) becomes

$$\begin{aligned} &(-\lambda^{(\mathcal{O}_T)}_{\text{sing}}\log(z\bar{z})+\lambda_{\phi\phi\mathcal{O}_T}) g^{\mathcal{O}_T\mathcal{O}_T} b_{\mathcal{O}_T} z^{2k-\frac{J}{2}-p+s}\bar{z}^{2k-p+\frac{J}{2}-s} \\ &+\lambda_{\phi\phi\mathcal{O}_\phi} g^{\mathcal{O}_\phi\mathcal{O}_\phi} b_{\mathcal{O}_\phi} z^{2k-\frac{J}{2}-p+s}\bar{z}^{2k-p+\frac{J}{2}-s} \\ &=(-\lambda^{(\mathcal{O}_T)}_{\text{sing}}\log(z\bar{z})+\lambda_{\phi\phi\mathcal{O}_T}) g^{\mathcal{O}_T\mathcal{O}_T} b_{\mathcal{O}_T} z^{n+s}\bar{z}^{n+J-s}+\lambda_{\phi\phi\mathcal{O}_\phi} g^{\mathcal{O}_\phi\mathcal{O}_\phi} b_{\mathcal{O}_\phi} z^{n+s}\bar{z}^{n+J-s}, \end{aligned} \tag{E9}$$

where in the second line we used (E1) to re-express the result in terms of the positive integer n. Notice that only the log term contributes to the discontinuity. Let us denote its coefficient as

$$a'_{\mathcal{O}_T} \coloneqq -\lambda^{(\mathcal{O}_T)}_{\text{sing}} g^{\mathcal{O}_T\mathcal{O}_T} b_{\mathcal{O}_T} \frac{J!}{2^J(\nu)_J}. \tag{E10}$$

Re-assembling the full contribution of the degenerate operators $[T^k]_J$, $[\phi\phi]_{n,J}$ to the conformal block expansion, we find that it takes the form

$$g(z,\bar{z}) \supset (a'_{[T^k]_J}\log(z\bar{z})+(a_{[T^k]_J}+a_{[\phi\phi]_{n,J}}))G^{\Delta_\phi}_{\Delta,J}(z,\bar{z}). \tag{E11}$$

APPENDIX F: HOLOGRAPHIC MULTI-TRACE ENERGY-MOMENTUM COEFFICIENTS $a^{(k)}_J$

One way to compute thermal one-point coefficients $a^{(k)}_J$ for the energy-momentum sector of a holographic theory is through the study of boundary "heavy-light" four-point correlators of the form

$$\langle\phi_H(0)\phi_L(z,\bar{z})\phi_L(1)\phi_H(\infty)\rangle \tag{F1}$$

at zero temperature, where the light operators ϕ_L have $\Delta_L \ll C_T$ and the heavy operators ϕ_H have $\Delta_H \sim C_T$. At large C_T, this correlator is argued to be equivalent to a bulk two-point function $\langle\Phi_L(x_1)\Phi_L(x_2)\rangle_{\text{BH}}$ of the dual light scalar in a black hole background created by the heavy operator. This is then dual to precisely the kind of boundary thermal two-point function $\langle\phi_L(x_1)\phi_L(x_2)\rangle_\beta$ studied in this work.

References [35,37] use this relation to compute the thermal one-point unknowns $a_{\mathcal{O}}$ for the energy-momentum sector with $\mathcal{O}=[T^k]_J$, which we denote $a^{(k)}_J$. These coefficients are found in terms of the parameters appearing in a static, spherically symmetric solution to a general higher-derivative gravity theory. Taking $d=4$ henceforth, such a solution has the form

$$ds^2 = r^2 f(r) dt^2 + \frac{dr^2}{r^2 h(r)} + r^2 \sum_{i=1}^{3} dx_i^2, \tag{F2}$$

with functions $f(r)$, $h(r)$, which have a near-boundary ($r \to \infty$) expansion

$$f(r) = 1 - \frac{1}{r^4} \sum_{i=0}^{\infty} \frac{f_{4i}}{r^{4i}}, \tag{F3}$$

$$h(r) = 1 - \frac{1}{r^4} \sum_{i=0}^{\infty} \frac{h_{4i}}{r^{4i}}. \tag{F4}$$

When these expansions are truncated to the r^{-4} term and we impose $f_0 = h_0$, we reduce to Einstein gravity in asymptotically AdS_4 spacetime. Reference [35] demonstrates that one needs to impose $f_0 = h_0$ for consistency with the boundary "heavy-light" conformal block decomposition and moreover, that leading-twist coefficients $a_{2k}^{(k)}$ universally depend only on f_0, i.e., they are agnostic to higher-derivative effects and depend only on Einstein gravity. We note, however, that additional nonminimal gravitational interaction terms in the AdS effective action will induce nonuniversal corrections to leading-twist coefficients, as studied in, e.g., [50–52].

While [35,41] report all $a_J^{(k)}$ up to $k = 2$ and leading twist unknowns $a_{2k}^{(k)}$ up to $k = 4$, our study uses certain subleading twist unknowns at $k = 3$. Using the calculation scheme described in [35], we find

$$a_6^{(3)} = \frac{\Delta(\Delta(\Delta(143\Delta(7\Delta+25)+7310)+7500)+3024)}{10378368000(\Delta-3)(\Delta-2)} f_0^3, \tag{F5}$$

$$\begin{aligned} a_4^{(3)} = {} & \frac{\Delta(\Delta(\Delta(\Delta(143\Delta(7\Delta-15)-2760)-2390)+2244)+2160)}{5189184000(\Delta-4)(\Delta-3)(\Delta-2)} f_0^3 \\ & + \frac{\Delta(\Delta+1)(\Delta+2)(\Delta(143\Delta+427)+540)}{43243200(\Delta-4)(\Delta-3)(\Delta-2)} f_0 f_4, \end{aligned} \tag{F6}$$

$$\begin{aligned} a_2^{(3)} = {} & \frac{\Delta(\Delta(\Delta(\Delta(\Delta(143\Delta(7\Delta-48)+12615)-3980)-6156)-11736)-1440)}{3459456000(\Delta-5)(\Delta-4)(\Delta-3)(\Delta-2)} f_0^3 \\ & + \frac{\Delta(\Delta+1)(\Delta(11\Delta(\Delta(26\Delta-43)+107)+4446)+6120)}{43243200(\Delta-5)(\Delta-4)(\Delta-3)(\Delta-2)} f_0 f_4 \\ & - \frac{\Delta(\Delta+1)(\Delta(\Delta(\Delta(143\Delta+101)+6200)+15912)+15840)}{86486400(\Delta-5)(\Delta-4)(\Delta-3)(\Delta-2)} f_0 h_4 \\ & + \frac{\Delta(\Delta+1)(\Delta+2)(\Delta+3)(\Delta+4)}{48048(\Delta-5)(\Delta-4)(\Delta-3)(\Delta-2)} f_8, \end{aligned} \tag{F7}$$

$$\begin{aligned} a_0^{(3)} = {} & \frac{(\Delta-8)\Delta(\Delta(\Delta(\Delta(143\Delta(7\Delta-48)+17115)-10460)+1584)+21384)+12960)}{10378368000(\Delta-6)(\Delta-5)(\Delta-4)(\Delta-3)(\Delta-2)} f_0^3 \\ & + \frac{(\Delta-8)\Delta(\Delta+1)(\Delta+2)(\Delta(\Delta(143\Delta-185)+1422)+3240)}{43243200(\Delta-6)(\Delta-5)(\Delta-4)(\Delta-3)(\Delta-2)} f_0 f_4 \\ & + \frac{\Delta(\Delta+1)(\Delta+2)(\Delta+3)((\Delta-4)\Delta+72)}{144144(\Delta-6)(\Delta-5)(\Delta-4)(\Delta-3)(\Delta-2)} (3f_8 - h_8). \end{aligned} \tag{F8}$$

Equation (F5) also appeared in [39].

APPENDIX G: ADDITIONAL PLOTS

In this appendix, we include a few additional plots (Fig. 12) that flesh out an example of a low-loss configuration obtained by optimizing $\mathcal{L}_{\overline{\text{abs}}}$ in the context of the GFF analysis of Sec. V B 2. In contrast to Fig. 2, where many optima exhibit A_0, A_2 functions that are displaced relative to the corresponding GFF analytic curves, the result in Fig. 12 is much closer to the analytic curves, demonstrating that the analytic GFF solution is part of the minima of $\mathcal{L}_{\overline{\text{abs}}}$ and the optimization can access it.

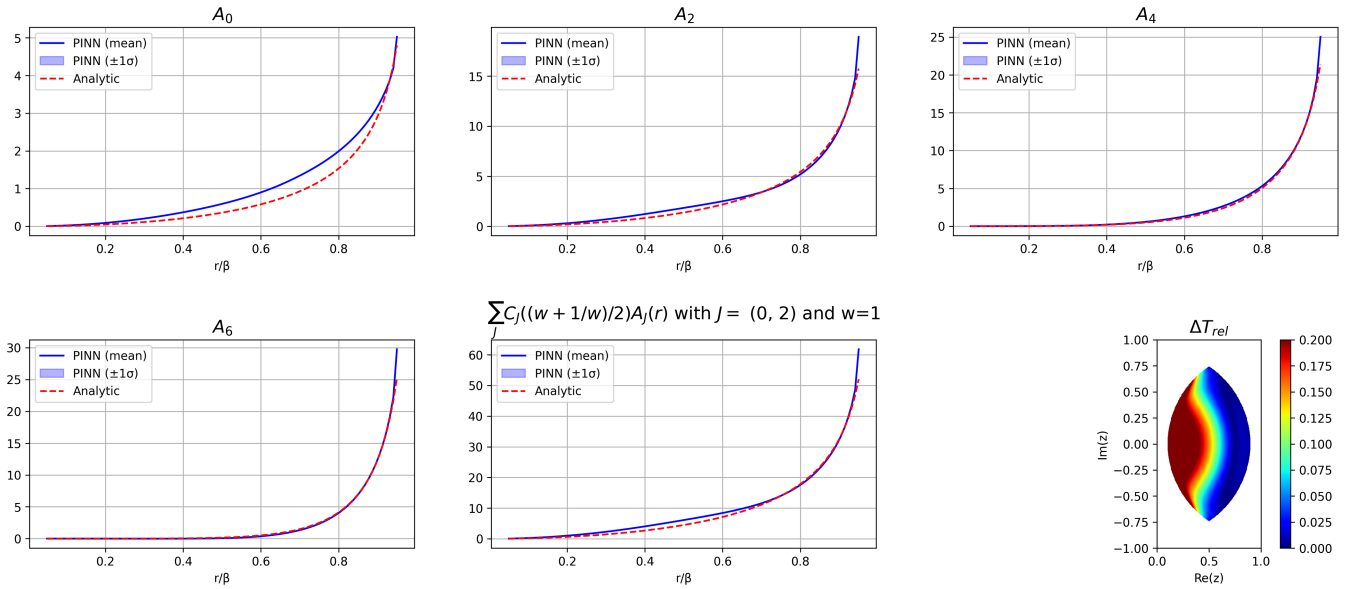


FIG. 12. Plots summarizing the results obtained in the $d = 4$ GFF theory for the 10th in order-of-loss run in a pool of 1K independent training runs with 50K epochs. We use $\Delta_\phi = 1.68$ without any exposed operators and asymptotic boundary conditions for the tail functions at $r = 0.9999$.

[1] Ferdinando Gliozzi, More constraining conformal bootstrap, Phys. Rev. Lett. **111**, 161602 (2013).

[2] Ferdinando Gliozzi and Antonio Rago, Critical exponents of the 3D Ising and related models from conformal bootstrap, J. High Energy Phys. 10 (2014) 042.

[3] Ferdinando Gliozzi, Pedro Liendo, Marco Meineri, and Antonio Rago, Boundary and interface CFTs from the conformal bootstrap, J. High Energy Phys. 05 (2015) 036; 12 (2021) 093(E).

[4] Ferdinando Gliozzi, Truncatable bootstrap equations in algebraic form and critical surface exponents, J. High Energy Phys. 10 (2016) 037.

[5] Ilya Esterlis, A. Liam Fitzpatrick, and David Ramirez, Closure of the operator product expansion in the non-unitary bootstrap, J. High Energy Phys. 11 (2016) 030.

[6] Wenliang Li, New method for the conformal bootstrap with OPE truncations, arXiv:1711.09075.

[7] Wenliang Li, Inverse bootstrapping conformal field theories, J. High Energy Phys. 01 (2018) 077.

[8] André Leclair and Joshua Squires, Conformal bootstrap for percolation and polymers, J. Stat. Mech. (2018) 123105.

[9] Gergely Kántor, Vasilis Niarchos, and Constantinos Papageorgakis, Conformal bootstrap with reinforcement learning, Phys. Rev. D **105**, 025018 (2022).

[10] V. Niarchos, C. Papageorgakis, P. Richmond, A. G. Stapleton, and M. Woolley, Bootstrability in line-defect CFTs with improved truncation methods, Phys. Rev. D **108**, 105027 (2023).

[11] David Poland, Slava Rychkov, and Alessandro Vichi, The conformal bootstrap: Theory, numerical techniques, and applications, Rev. Mod. Phys. **91**, 015002 (2019).

[12] Vasilis Niarchos, Constantinos Papageorgakis, and Andreas Stergiou, Deep Conformal Bootstrap (to be published).

[13] Sheer El-Showk and Kyriakos Papadodimas, Emergent spacetime and holographic CFTs, J. High Energy Phys. 10 (2012) 106.

[14] Minjae Cho, Barak Gabai, Joshua Sandor, and Xi Yin, Thermal bootstrap of matrix quantum mechanics, J. High Energy Phys. 04 (2025) 186.

[15] Enrico Marchetto, Alessio Miscioscia, and Elli Pomoni, Sum rules & Tauberian theorems at finite temperature, J. High Energy Phys. 09 (2024) 044.

[16] Julien Barrat, Enrico Marchetto, Alessio Miscioscia, and Elli Pomoni, Thermal bootstrap for the critical O(N) model, Phys. Rev. Lett. **134**, 211604 (2025).

[17] Julien Barrat, Bartomeu Fiol, Enrico Marchetto, Alessio Miscioscia, and Elli Pomoni, Conformal line defects at finite temperature, SciPost Phys. **18**, 018 (2025).

[18] Julien Barrat *et al.*, The analytic bootstrap at finite temperature, arXiv:2506.06422.

[19] Ilija Burić, Ivan Gusev, and Andrei Parnachev, Thermal holographic correlators and KMS condition, J. High Energy Phys. 09 (2025) 053.

[20] Luca Iliesiu, Murat Kologlu, Raghu Mahajan, Eric Perlmutter, and David Simmons-Duffin, The conformal bootstrap at finite temperature, J. High Energy Phys. 10 (2018) 070.

[21] Emanuel Katz, Subir Sachdev, Erik S. Sørensen, and William Witczak-Krempa, Conformal field theories at nonzero temperature: Operator product expansions, Monte Carlo, and holography, Phys. Rev. B **90**, 245109 (2014).
[22] William Witczak-Krempa, Constraining quantum critical dynamics: $(2+1)$D Ising model and beyond, Phys. Rev. Lett. **114**, 177201 (2015).
[23] Luis F. Alday, Murat Kologlu, and Alexander Zhiboedov, Holographic correlators at finite temperature, J. High Energy Phys. 06 (2021) 082.
[24] Luis F. Alday and Simon Caron-Huot, Gravitational S-matrix from CFT dispersion relations, J. High Energy Phys. 12 (2018) 017.
[25] Luis F. Alday, Johan Henriksson, and Mark van Loon, Taming the ϵ-expansion with large spin perturbation theory, J. High Energy Phys. 07 (2018) 131.
[26] Madalena Lemos, Pedro Liendo, Marco Meineri, and Sourav Sarkar, Universality at large transverse spin in defect CFT, J. High Energy Phys. 09 (2018) 091.
[27] Junyu Liu, Eric Perlmutter, Vladimir Rosenhaus, and David Simmons-Duffin, d-dimensional SYK, AdS loops, and 6j symbols, J. High Energy Phys. 03 (2019) 052.
[28] Simon Caron-Huot and Anh-Khoi Trinh, All tree-level correlators in $AdS_5 \times S_5$ supergravity: Hidden ten-dimensional conformal symmetry, J. High Energy Phys. 01 (2019) 196.
[29] Carlos Cardona, Sunny Guha, Surya Kiran Kanumilli, and Kallol Sen, Resummation at finite conformal spin, J. High Energy Phys. 01 (2019) 077.
[30] Wenliang Li, Closed-form expression for cross-channel conformal blocks near the lightcone, J. High Energy Phys. 01 (2020) 055.
[31] Soner Albayrak, David Meltzer, and David Poland, More analytic bootstrap: Nonperturbative effects and fermions, J. High Energy Phys. 08 (2019) 040.
[32] Dean Carmi and Simon Caron-Huot, A conformal dispersion relation: Correlations from absorption, J. High Energy Phys. 09 (2020) 009.
[33] Luis F. Alday, Johan Henriksson, and Mark van Loon, An alternative to diagrams for the critical O(N) model: Dimensions and structure constants to order $1/N^2$, J. High Energy Phys. 01 (2020) 063.
[34] Madalena Lemos, Balt C. van Rees, and Xiang Zhao, Regge trajectories for the (2, 0) theories, J. High Energy Phys. 01 (2022) 022.
[35] A. Liam Fitzpatrick and Kuo-Wei Huang, Universal lowest-twist in CFTs from holography, J. High Energy Phys. 08 (2019) 138.
[36] Enrico Parisini, Kostas Skenderis, and Benjamin Withers, The ambient space formalism, J. High Energy Phys. 05 (2024) 296.
[37] Robin Karlsson, Manuela Kulaxizi, Andrei Parnachev, and Petar Tadić, Leading multi-stress tensors and conformal bootstrap, J. High Energy Phys. 01 (2020) 076.
[38] Yue-Zhou Li, Zhan-Feng Mai, and H. Lü, Holographic OPE coefficients from AdS black holes with matters, J. High Energy Phys. 09 (2019) 001.
[39] Yue-Zhou Li, Heavy-light Bootstrap from Lorentzian inversion formula, J. High Energy Phys. 07 (2020) 046.
[40] Yue-Zhou Li and Hao-Yu Zhang, More on heavy-light bootstrap up to double-stresstensor, J. High Energy Phys. 10 (2020) 055.
[41] Chantelle Esper, Kuo-Wei Huang, Robin Karlsson, Andrei Parnachev, and Samuel Valach, Thermal stress tensor correlators near lightcone and holography, J. High Energy Phys. 11 (2023) 107.
[42] Nejc Čeplak, Hong Liu, Andrei Parnachev, and Samuel Valach, Black hole singularity from OPE, J. High Energy Phys. 10 (2024) 105.
[43] Ilija Burić, Ivan Gusev, and Andrei Parnachev, Holographic correlators from thermal bootstrap, arXiv:2508.08373.
[44] David Poland, Valentina Prilepina, and Petar Tadić, The five-point bootstrap, J. High Energy Phys. 10 (2023) 153.
[45] David Poland, Valentina Prilepina, and Petar Tadić, Improving the five-point bootstrap, J. High Energy Phys. 05 (2024) 299.
[46] David Poland, Valentina Prilepina, and Petar Tadić, Mixed five-point correlators in the 3D Ising model, J. High Energy Phys. 10 (2025) 237.
[47] Alexandros Stratoudakis, Dispersion relations for conformal field theories at finite temperature, University of Crete Institutional Repository (2024), p. 35, https://elocus.lib.uoc.gr/dlib/d/e/0/metadata-dlib-1719839923-431949-31018.tkl.
[48] Agnese Bissi, Parijat Dey, and Tobias Hansen, Dispersion relation for CFT four- point functions, J. High Energy Phys. 04 (2020) 092.
[49] Simon Caron-Huot, Analyticity in spin in conformal theories, J. High Energy Phys. 09 (2017) 078.
[50] A. Liam Fitzpatrick, Kuo-Wei Huang, David Meltzer, Eric Perlmutter, and David Simmons-Duffin, Model-dependence of minimal-twist OPEs in $d > 2$ holographic CFTs, J. High Energy Phys. 11 (2020) 060.
[51] Kuo-Wei Huang, Toward null-state equations in $d > 2$, J. High Energy Phys. 11 (2023) 203.
[52] Kuo-Wei Huang, Resummation of multistress tensors in higher dimensions, Phys. Rev. D **111**, 046016 (2025).