Referee Report

arXiv:2607.24913

"Neural Spectral Bias and Conformal Correlators II: Modular and Annulus Bootstrap"

Summary of the paper

The manuscript extends the authors' anchored neural-network approach to two-dimensional torus and annulus partition functions. Section 2 reviews two established geometric reformulations. First, a torus is represented as a double cover of the sphere, and modular S-invariance becomes the crossing equation of four $\mathbb{Z}_2$ twist fields. On the diagonal slice $z = \bar{z}$, or equivalently $\tau = it$, this gives equation (2.12). Second, annulus open/closed duality is represented as the crossing equation (2.20) for a mixed four-point function of defect-changing endpoint operators.

Section 3 supplies the numerical prescription. For the torus, the authors multiply the line correlator by $B_{\mathrm{t}}(z) = \sqrt{t(z)}|\eta(\tau(z))|^2$, separate a prescribed leading endpoint term from a gap-weighted neural correction in equations (3.5)–(3.6), and train a two-hidden-layer width-64 GELU network using a crossing residual and one exact anchor value. For the annulus, two networks represent the open and closed channels. Cardy and Verlinde data determine the leading weights, next weights, multiplicities, degeneracies, and endpoint powers in equations (3.17)–(3.21), while equation (3.25) anchors both channels at one value of z.

Sections 4 and 5 test the method on known compact CFT partition functions. The torus examples include unitary A-, D-, and E-series minimal models, the non-unitary Lee–Yang theory, and Wess–Zumino–Witten models. The annulus examples cover a substantial set of Cardy boundary pairs in the Ising and tricritical Ising theories and in several WZW models. Section 6 considers non-compact targets: the free boson, the Liouville torus amplitude, and Liouville annuli with ZZ and FZZ boundary conditions. The central new claim is empirical rather than a uniqueness theorem: despite the large family of functions satisfying the imposed crossing and anchor constraints, the optimization bias of the network is said to select a function close to the known physical answer.

Assessment of correctness and clarity

The paper is ambitious, clearly organized, and based on an interesting idea. The twist-field and defect-correlator reformulations provide a natural bridge between modular consistency and the authors' earlier neural crossing framework. The range of benchmark theories is a genuine strength, especially the simultaneous treatment of the two annulus channels and the inclusion of non-unitary and non-compact examples. Many of the central regions of the reported curves are indeed close to their exact targets.

I do not, however, think that the present numerical evidence supports the accuracy, universality, or mechanism claimed in the abstract and conclusion. Most importantly, equation (3.26) is not a relative error and can label an order-one fractional discrepancy as a tiny percentage when the exact correlator is small. The numerical data contain several concrete examples of this problem. In addition, equations (3.4)–(3.6) do not have the exact torus endpoint expansion asserted for them: the correction omits a universal kinematic term and a logarithmic factor. There is also no held-out

collocation test for most compact examples, the two WZW cases require severe post-selection, and no control experiment establishes that the successful cases are caused specifically by neural spectral bias.

These issues do not make the overall programme unpromising, but they affect the paper's main conclusions rather than only its presentation. A substantial new numerical analysis and a more precise statement of scope are needed.

Novelty and significance

The most distinctive contribution is the broad empirical application of the anchored function-space method to modular and boundary consistency, particularly the two-channel annulus construction. This is a useful extension of the authors' preceding papers. At the same time, the network architecture, optimizer, loss strategy, and claimed implicit bias are largely inherited from those works, while the torus/twist map, annulus defect map, Cardy states, Verlinde multiplicities, and rational-CFT characters are established ingredients. The advance is therefore a synthesis and benchmark study, not a new result about modular invariance or boundary CFT by itself.

The function-space perspective is complementary to methods that optimize spectral data directly. It has the advantage of not requiring positivity and can therefore be applied to Lee–Yang and Liouville theory. Conversely, it does not by itself enforce integrality, positivity, a physical q-expansion, or the remaining sewing constraints. This tradeoff and the relation to the authors' earlier work and to the complementary spectrum-space modular approach cited in the Introduction should be developed much more explicitly. The work could be significant as a numerical proof of concept if the empirical claims are put on a controlled and reproducible footing.

Recommendation

Major revision. I would be willing to reconsider the manuscript after the quantitative accuracy claims have been recomputed with appropriate metrics and held-out tests, the endpoint ansatz has been corrected, the WZW post-selection has been treated prospectively, and the claim of a specifically neural spectral-bias mechanism has been supported by controls. In its present form, the manuscript is not yet suitable for publication in JHEP or PRD.

Major comments

1. **Equation (3.26) materially overstates accuracy for small correlators.**

   The quantity called the "prediction relative error" is

   $$\frac{\widetilde{G}^{\mathrm{pred}}(z) - \widetilde{G}^{\mathrm{exact}}(z)}{1 + |\widetilde{G}^{\mathrm{exact}}(z)|}.$$

   For $|\widetilde{G}^{\mathrm{exact}}| \ll 1$, this is essentially a signed absolute error, not a fractional error. The distinction changes the interpretation of several headline results. For the ZZ–FZZ example in Figure 15, the open value at $z = 0.5$ is 2.63×10^{-4}, whereas the ensemble mean is 3.08×10^{-4},

a 17.1% fractional bias; the closed discrepancy is approximately 7.2%. At $z = 0.05$, the open exact and mean values are approximately 2.5313×10^{-7} and 5.9022×10^{-7}, respectively, a 133% discrepancy.

The problem is not confined to this deliberately difficult example. In Figure 23, for the closed channel of the Ising $(\mathbf{1}, \boldsymbol{\varepsilon})$ annulus at $z = 0.9$, the exact and mean values are approximately 0.0062849 and 0.0115758, an 84.2% fractional discrepancy. Equation (3.26) assigns only about 0.526% to the same difference, consistent with the apparently small number in Table 1. Similarly, the data underlying Figure 16 have fractional discrepancies of about 62.7% in the closed channel at $z = 0.01$ and 10.5% in the open channel at $z = 0.9$, despite the statement in Section 6 that the reconstruction is accurate to a few parts in 10^4 across the full range.

The authors should retain equation (3.26), if useful, only under a name such as "regularized absolute error." Every table and figure should also report a genuine fractional error wherever the target is nonzero, an absolute error, and integrated and supremum norms with a clearly stated normalization. For positive functions spanning many orders of magnitude, an error in $\log G$ is also informative. It must be stated whether the reported "maximum" takes an absolute value, since equation (3.26) itself is signed. The abstract, Section 6, and the conclusion should be revised after the results are recomputed with these metrics.

2. **The torus ansatz has an incorrect endpoint expansion.**

   Equation (3.4) gives only the leading asymptotic term of B_{t}, and equation (3.6) treats the first correction as if it were determined only by the primary gap. Using the standard theta-function identity for η^2 and the small-z expansions of the elliptic integrals gives

$$B_{\mathrm{t}}(z) = c_0 z^{1/6} \sqrt{\log(16/z)} \left[1 + z \left(\frac{1}{12} - \frac{1}{4 \log(16/z)} \right) + \cdots \right].$$

   Consequently, even before the first non-vacuum primary is included, $\widetilde{G} - L$ contains a universal kinematic correction beginning at power $z^{7/6}$, with logarithmic factors. By contrast, equation (3.6) forces the neural correction to begin as $z^{1/6+2\Delta_{\mathrm{gap}}} \mathrm{NN}(z)$ and omits the $\sqrt{\log(16/z)}$ multiplying the physical $z^{1/6+2\Delta_{\mathrm{gap}}}$ term. A finite GELU MLP is regular at $z = 0$, so it cannot repair the claimed asymptotic form there.

   This is especially clear when $\Delta_{\mathrm{gap}} > 1/2$. For $\widehat{\mathfrak{su}}(3)_1$, where $\Delta_{\mathrm{gap}} = 2/3$, and for $\widehat{\mathfrak{su}}(4)_1$, where $\Delta_{\mathrm{gap}} = 3/4$, the universal $z^{7/6}$ correction occurs before the power permitted by equation (3.6). The finite training cutoff $z \geq 0.01$ can mask this structural mismatch but does not remove it. The footnote after equation (3.6), which says that a logarithm can be added without changing the displayed numerics, does not establish the endpoint claim.

   A clean correction would be to use the exact $B_{\mathrm{t}}(z)$, rather than only its leading asymptotic, as the vacuum piece and to factor $B_{\mathrm{t}}(z) z^{2\Delta_{\mathrm{gap}}}$ from the primary correction. Another valid parametrization would have to include the universal $\min(1, 2\Delta_{\mathrm{gap}})$ ordering and all required logarithms. The authors should derive the endpoint expansion explicitly and rerun the benchmarks with an asymptotically consistent ansatz.

3. **Most compact-CFT accuracy plots are in-sample tests.**

   Equations (3.8) and (3.23) state that the crossing loss is trained on 90 uniform points. The compact-model ensemble and error curves are reported on the same 90 points, with the same endpoints and spacing. Thus the reported crossing residuals and target errors do not test behavior between collocation points. This is important because the constraints are imposed

only discretely and the problem is underdetermined; a smooth-looking plot does not provide a quantitative generalization test.

Each benchmark should be evaluated on a dense, disjoint grid that is never used for training, preferably with both uniform and logarithmic sampling near the endpoints. The paper should report the crossing residual and the error against the exact target on that grid. Convergence under changes in the number and placement of training points should also be shown. Random or resampled collocation points during training would provide a useful check. Claims about the "full interval" must either be supported by tests much closer to $z = 0, 1$, using the corrected asymptotics, or be restricted to the finite numerical interval actually studied.

4. **The manuscript does not isolate neural spectral bias as the selection mechanism.**

   Equation (2.12) leaves infinite functional freedom: one may choose a smooth function on one side of $z = 1/2$ and define the other side by crossing. The annulus equation (2.20), with two channel functions, has analogous freedom. Endpoint powers and anchors restrict this space but do not select a unique answer. The observation that a particular optimization often returns a nearby physical benchmark is interesting, but it does not by itself show that "neural spectral bias in the lazy regime" is the cause.

   The paper contains no measurement of parameter displacement or neural tangent kernel stability demonstrating lazy training, and it does not analyze the frequency content of the learned correction. It also lacks non-neural controls. The authors should compare the MLP with, at minimum, a Chebyshev or spline representation with an explicit smoothness penalty, a random-feature or kernel regression model, and the linearized neural-tangent-kernel limit of their own architecture. Ablations in width, depth, activation, weight decay, anchor weight, and optimizer are needed. If several minimum-complexity methods select the same answer, the result may be a useful implicit-regularization phenomenon but should not be presented as uniquely neural. If the MLP is special, these controls should make that distinction visible.

5. **The WZW post-selection is substantial counter-evidence to the universality claim.**

   Section 4.4 reports a bimodal ensemble for $\widehat{\mathfrak{su}}(2)_2$. Only 92 of 1000 seeds are retained in Figure 8 because they trigger early stopping and land near the exact answer. Appendix A.1 applies the same filter to $\widehat{\mathfrak{su}}(2)_1$, retaining 184 of 1000 seeds. Thus the large majority of initializations select a spurious branch even though they are trained with the same formal constraints. Reporting only the minority peak gives an overly favorable impression of reliability.

   The complete distributions, losses, endpoint behavior, and true errors of both peaks should be shown. The authors should explain whether the early-stopping criterion and its use as a physical-solution classifier were fixed before these exact answers were inspected. An exact-answer-independent rule must be specified and tested prospectively on separate models. The overall success probability and computational cost should be reported. If no blind criterion distinguishes the two branches, these examples should be presented as a limitation of the method rather than evidence that early stopping identifies the physical correlator.

6. **The scope is the imaginary-modulus line, not a full modular partition function.**

   Section 2.1 explicitly restricts to $z = \bar{z}$, corresponding to $\tau = it$, and imposes only the S-relation (2.12). The calculation does not learn the dependence on a general complex τ, test T-invariance, recover spin dependence, or enforce the remaining sewing constraints. A one-dimensional function close to a known thermal-line partition function need not extend to a physical modular invariant with the required $q, \bar{q}$ coefficients.

The phrases "full partition functions" and "modular-invariant partition functions" should be replaced by "thermal-line profiles" or similarly precise language unless the analysis is extended off the line and T-invariance is imposed. A useful intermediate physicality test would be to extract withheld low-lying q-series coefficients and examine their expected integrality or sign properties where applicable. The annulus results should likewise distinguish satisfaction of the imposed open/closed functional equation from reconstruction of a complete consistent boundary CFT.

7. **The information supplied to the network and its practical robustness need an honest accounting.**

   The torus input is not only "a gap and an anchor." It includes the central charge, vacuum normalization, exact endpoint structure, a specific gap-dependent ansatz, and an exact target value in equation (3.10). The annulus setup additionally uses channel-specific leading and next weights, leading multiplicities and degeneracies derived from Cardy/Verlinde data, endpoint exponents, and two exact channel values in equation (3.25). Section 6 then introduces target-specific exponentiation and rescalings. These are legitimate inputs for a benchmark, but they should be itemized in the abstract and conclusions, and the case-specific changes weaken the claim of one universal prescription.

   The authors should study sensitivity to the anchor location and its uncertainty, perturbations of the assumed gap and next weight, errors in the leading coefficient, and the number of anchors. Such tests are essential for an application in which the exact answer is not already known. At present every target is exactly known and trained separately; the work demonstrates interpolation of known functions rather than a new physical prediction. Its significance would increase considerably if it predicted withheld spectral coefficients, reconstructed a genuinely unknown observable, or used independently determined noisy data without tuning to the final target.

8. **The annulus derivation and numerical record should be made fully auditable.**

   The annulus part is the most novel component, but several inputs are stated rather than derived. The manuscript should show the normalization leading to the prefactor in equation (2.18), explain carefully why the endpoint operators in equation (2.19) have the stated universal dimension for the defect configuration used, and derive the channel exponents and coefficients in equations (3.18)–(3.21). A table listing h^X_{lead}, h^X_{next}, n^X_0, and every effective exponent for each displayed boundary pair would make the physical input transparent. The clipping prescription (3.21) also removes an explicitly vanishing endpoint power whenever the raw exponent is positive; the authors should demonstrate that the correct endpoint behavior is still obtained rather than merely fitted on the truncated grid.

   For a numerical paper, every result in Tables 1–4 and Figures 4–29 should be reproducible from a frozen archive. The archive should contain the annulus, WZW, and Liouville training scripts, exact-target generation, configuration files or command lines, software versions, seed lists, unfiltered per-seed outputs, and the post-selection rule. The manuscript currently states that the companion repository contains the explicit code for all runs; the cited immutable record should be checked to ensure that this is in fact true for every reported example.

Minor comments

1. Section 2.1 says that, after complexifying τ and $\bar{\tau}$, the partition function is holomorphic "on the complex plane." The appropriate statement is separate holomorphy on the relevant product of upper and lower half-planes, with the Euclidean partition function obtained on the conjugate slice.

2. The exact universal prefactor relating the torus partition function to the twist correlator should be written explicitly, as is done for the annulus in equation (2.18). This would also make the conventions used in Section 3 easier to verify.

3. Equation (4.19) is introduced after a discussion of arbitrary simple $\mathfrak{g}$, but its S_N Weyl group, $k + N$, and lattice notation are specialized to $\mathfrak{su}(N)$. The scope should be narrowed or a genuinely general Weyl–Kac formula should be given.

4. The manuscript should distinguish standard deviation across seeds, standard error of the ensemble mean, and uncertainty in the exact target. Several expressions such as $(9.50 \pm 11.8) \times 10^{-6}$ also need consistent significant figures and an explanation of the highly skewed loss distribution they imply.

5. Histograms only at $z = 0.5$ can conceal the large endpoint errors discussed above. Distributions of per-seed functional norms and endpoint errors would be more informative than a single-point histogram.

6. The statement that the same torus partition function applies to time-like and space-like Liouville theory for any real central charge requires qualification. The integration contour, zero-mode normalization, and interpretation of time-like Liouville are subtle and should not be passed over as an automatic continuation.

7. The comparison with the spectrum-optimization approach cited in the Introduction should explain what physical constraints each method enforces, what information each requires, and what output can be certified. A one-sentence comparison is insufficient given the incremental nature of the present extension.

8. The paper should cite general literature on spectral bias, lazy training, neural tangent kernels, and implicit regularization, rather than using only the authors' preceding applications to support those concepts.