

Holonomy saddles and supersymmetry

Chiung Hwang,^{*} Sungjay Lee,[†] and Piljin Yi[‡]*School of Physics, Korea Institute for Advanced Study, Seoul 02455, Korea*

(Received 2 May 2018; published 14 June 2018)

In gauge theories on a spacetime equipped with a circle, the holonomy variables, living in the Cartan torus, play special roles. With their periodic nature properly taken into account, we find that a supersymmetric gauge theory in d dimensions tends to reduce in the small radius limit to a disjoint sum of multiple $(d - 1)$ dimensional theories at distinct holonomies, called H -saddles. The phenomenon occurs regardless of the spacetime dimensions, and here we explore such H -saddles for $d = 4$ $\mathcal{N} = 1$ theories on T^2 fibered over Σ_g , in the limits of elongated T^2 . This naturally generates novel relationships between 4d and 3d partition functions, including ones between 4d and 3d Witten indices, and also leads us to reexamine recent studies of the Cardy exponents and the Casimir energies and of their purported connections to the 4d anomalies.

DOI: [10.1103/PhysRevD.97.125013](https://doi.org/10.1103/PhysRevD.97.125013)

I. GLUING GAUGE THEORIES ACROSS DIMENSIONS

Gauge theories in a spacetime with a circle admit holonomy variables as special degrees of freedom. With the spacetime sufficiently noncompact, the infrared properties of the theory are often characterized by the vacuum expectation values (vev) of the Wilson line operator [1], or the traced holonomy along the circle.

In many theories, the holonomy variables are not exactly flat at the quantum level and the Wilson line often serves as an order parameter. For example, 4d $\mathcal{N} = 1$ pure $SU(N)$ Yang-Mills on a large circle, or on a circle with supersymmetric boundary condition, are known to admit N distinct vacua, whose confining nature is dictated by equally spaced eigenvalues of the holonomy, hence a vanishing Wilson line expectation value. If we replace the circle by a sufficiently small thermal circle, with the antiperiodic boundary condition on gauginos, the eigenvalues become clustered at the origin, signaling a deconfined phase at high temperature as evidenced by a nonvanishing Wilson line vev.

If supersymmetry is extended enough to ensure that these variables correspond to genuine flat directions at quantum level, compactification on the circle generates an infinite number of superselection sectors, labeled by the

holonomy. A more typical situation with minimal supersymmetry is, on the other hand, that at generic vev the supersymmetry is spontaneously broken; one finds some discrete choices of the holonomy vev with the supersymmetry intact. In either case, the process of the dimensional reduction, as the circle size is taken to zero, is typically ambiguous until we specify at which holonomy vev this is done. When the holonomy is nontrivial, the net effect is that of the Wilson line symmetry breaking.

When the space is compact or, more precisely, has no more than two extended directions, on the other hand, the special nature of the holonomy variables manifest somewhat differently, as they must be integrated over for the path integral. For example, the localization for the twisted partition functions produces integration over gauge holonomy variables at the end of the procedure. This means that one must be rather careful in taking a small radius limit. If one naively replaces this integration over the holonomy, living in the Cartan torus, by one over \mathbb{R}^{rank} , the Cartan subalgebra, one ends up computing partition function of a dimensionally reduced theory in one fewer dimension, with the vev of the holonomy variable naively frozen at the identity.

As we commented already, however, dimensional reduction of a single supersymmetric gauge theory on a circle may produce distinct gauge theories in one fewer dimension, depending on what holonomy vev's are available and chosen. For partition function computations on a compact spacetime with a circle, then, this ambiguity of the dimensional reduction must also manifest. How does this happen? Since the original integration range is over the Cartan torus rather than the Cartan subalgebra and since the periodic nature of the holonomy variables is not to be ignored so easily, the answer is quite clear: As we scan the holonomy

^{*}chwang@kias.re.kr[†]sjlee@kias.re.kr[‡]piljin@kias.re.kr

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). 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.

along the Cartan torus, we often find special places where the Wilson line symmetry breaking leads to supersymmetric gauge theories in one fewer dimension.

This translates to the supersymmetric partition function Ω_d^G of theory G in d -dimensions reducing, in an appropriate scaling limit, to a discrete sum of $(d-1)$ -dimensional partition functions Z_{d-1}^H of theories H 's sitting at special holonomies u_H , modulo some prefactors, as

$$\Omega_d^G \rightarrow \sum_{u_H} \sim Z_{d-1}^H, \quad (1.1)$$

where these u_H 's are distributed discretely along the periodic Cartan torus. In the vanishing radius limit, distinct u_H 's are infinitely far from one another, so that taking the naive limit of replacing the holonomies by scalars amounts to concentrating on a small neighborhood near a single u_H . Since $u_H \neq 0$ would be infinitely far away from $u_H = 0$ from the perspective of dimensionally reduced theories, one is often misled to consider the theory at $u_H = 0$, tantamount to replacing the Cartan torus by the Cartan subalgebra, and ends up computing a wrong scaling limit of Ω_d^G .

We will call these special holonomy values u_H 's (and the supersymmetric theories sitting there) the holonomy saddles, or H -saddles. The authors of Ref. [2] had introduced this concept and thereby resolved a 15-year-old puzzle [3–5] on Witten indices of 1d pure Yang-Mills theories [6–11]; in retrospect, the puzzle had originated from a simple misconception that only the naive $u_H = 0$ saddle (and its images under the shift by the center) contributes to the right-hand side. Since the holonomy moduli space is present universally for spacetimes with a circle, at least classically, and since the holonomy must be integrated over for compact enough space, it is clear that this H -saddle phenomenon will occur for twisted partition functions regardless of spacetime dimensions.

For field theory Witten indices [12], for example, H -saddles dictate how the Witten indices of gauge theories in adjacent dimensions could be related. Witten indices can easily differ in different dimensions despite the standard rhetoric that compactification on torus does not change the number of vacua. A well-known modern example of such disparities is how the 1d wall-crossing phenomena does not manifest in 2d elliptic genera. H -saddles now give us a rather concrete way to relate such topologically protected quantities across dimensions, in a very definite manner.

The importance of the holonomy in relating supersymmetric theories between different dimensions has been noted elsewhere, if somewhat sporadically. Notable examples are due to Aharony and collaborators [13,14] who observed how a Seiberg-dual pair of 4d/3d theories may translate to multiple such in 3d/2d as well as an even earlier work in Ref. [15] where, again, a 2d limit of a 3d mirror symmetry is explored. Our study can be viewed as an effort to explore such phenomena much more systematically and

concretely, now armed with varieties of exact partition functions, and to consider other ramifications. Also related are Refs. [16,17] which found exceptions to the purported universal connection between the Cardy exponents and the anomaly coefficients [18]. What we find here is that such a universal expression is often an artifact of ignoring H -saddles other than the naive one at $u_H = 0$ and that when the theory comes with matter fields in gauge representations bigger than the defining ones, this “exception” tends to occur generically for all acceptable spacetimes, including $S^1 \times S^3$. Furthermore, we will find similar failures for the Casimir limit in general, although this side proves to be more subtle.

We wish to emphasize that this phenomenon is inherent to the supersymmetric gauge theories themselves, rather than merely a property of the partition functions thereof. Note that the latter quantities need compact spacetime for their definition. When the spacetime has at least three noncompact directions, these special values of the holonomy give various superselection sectors where the theories in one less dimension are equipped with supersymmetry intact at quantum level. Nevertheless, the partition functions in general and the Witten indices in particular offer handy tools for classifying these special holonomies, which is why we concentrate on computation of these quantities in this paper.

This paper is organized as follows. In the rest of this introductory section, which also serves as a rough summary, we will overview supersymmetric twisted partition functions and give a broad characterization of H -saddle phenomena. This phenomenon of H -saddles and their consequences will be studied in the subsequent sections for a large class of 4d $\mathcal{N} = 1$ theories defined on compact spacetimes which are T^2 fiber bundles over smooth Riemannian surfaces.

Section II will review a recent construction of A-twisted partition functions in such backgrounds, and recall the detailed computational procedure. This is then extended to the so-called “physical” backgrounds, one special case of which is the superconformal index (SCI). Section III will classify the Bethe vacua in the small and the large τ limits. The Bethe vacua are easily seen to be clustered into subfamilies, each of which can be regarded as the Bethe vacua of some 3d theories sitting at special value of the holonomy. Although the latter viewpoint is physically better motivated in the small τ limit, which we can really view as a compactification to 3d, the other limit of large τ follows the same pattern thanks to $SL(2, \mathbb{Z})$ property of the fiber T^2 . Even when the $SL(2, \mathbb{Z})$ is not available, such as in SCI's, such clustering of Bethe vacua do occur as well, although, as we will see in Sec. IV.

These limiting behaviors of Bethe vacua imply that a 4d gauge theory typically decomposes into a disjoint sum of several, potentially distinct 3d theories: The 3d limit of 4d supersymmetric partition functions becomes a sum of

partition functions of these 3d theories, albeit with extra exponential factors. A special case of this is the 4d Witten index, reexpressed as a sum of Witten indices of the associated 3d theories at H -saddles, clearly without the extra exponential prefactors. In Sec. IV, we explore such limits for various background geometries and spacetime. One noteworthy corollary here is that the Cardy exponents, and even the Casimir energies, to a lesser degree, would generally deviate from the existing proposals [18–20], connected to various 4d anomalies. As we will see these proposals are often tied to the naive $u_H = 0$ saddle which may or may not be the dominant saddle. We should note, however, that the Casimir limit of SCI's is somewhat special in that the microscopic derivations in Refs. [21,22] and the anomaly connection thereof proved to be robust, despite the presence of nontrivial H -saddles. We comment on this toward the very end of this paper.

A. Twisted partition functions and the Euclidean time

Twisted partition functions, to be denoted by Ω throughout this paper, are obtained by computing the partition function with an insertion of the chirality operator $(-1)^{\mathcal{F}}$. A requisite for $(-1)^{\mathcal{F}}$ is that there is a notion of natural Euclidean time coordinate, forming a circle S^1 . With the natural \mathbb{Z}_2 action of supersymmetry, say, \mathcal{Q} , which anti-commutes with the chirality operator, this insertion allows generic bosonic states to cancel against fermionic states, and leaves behind a special subset of the Hilbert space.

When the theory is suitably gapped and the space is taken to be T^{d-1} , this quantity would compute the Witten index [12], integral and enumerative of supersymmetric ground states. In recent years, many types of supersymmetric partition functions have been proposed with the accompanying computational tricks under the banner of the localization. The superconformal index [23,24] is one such class of well-known and much-computed objects, while the elliptic genera in 2d [25,26] and the refined Witten indices in 1d have been developed to a very sophisticated level [27,28].

The length of Euclidean time circle, β , may be interpreted as the inverse temperature. For the twisted version, however, this parameter is often argued to disappear from the end result, since supercharges \mathcal{Q} act as a one-to-one map for positive energy bosonic and fermionic states. This disappearance is, of course, a desired feature of the index, since the latter was designed, to begin with, to count Bose-Fermi asymmetry of the ground state sector. The twisted partition functions

$$\mathrm{Tr}(-1)^{\mathcal{F}} e^{-\beta\mathcal{Q}^2} \quad (1.2)$$

are thus argued to be projected to the ground state sector

$$\mathrm{Tr}_{\mathrm{kernel}(\mathcal{Q})}(-1)^{\mathcal{F}}, \quad (1.3)$$

which is necessarily integral and enumerative.

This is, however, not quite true in general. If the theory admits a continuum spectrum whose energies are bounded below by $E_{\mathrm{gap}} > 0$, the trace (1.2) actually produces

$$\mathrm{Tr}(-1)^{\mathcal{F}} e^{-\beta\mathcal{Q}^2} = \mathrm{Tr}_{\mathrm{kernel}(\mathcal{Q})}(-1)^{\mathcal{F}} + O(e^{-\beta E_{\mathrm{gap}}}). \quad (1.4)$$

This subtlety is relatively easy to handle since one may be able to scale $E_{\mathrm{gap}} \rightarrow +\infty$ first, without affecting the ground state counting. When $E_{\mathrm{gap}} = 0$, on the other hand, separating out the continuum contributions becomes something of an art. One popular scheme in the face of such gapless asymptotic directions is to insert the chemical potentials ν 's for global symmetries F 's,

$$\mathrm{Tr}(-1)^{\mathcal{F}} e^{\nu F} e^{-\beta\mathcal{Q}^2}, \quad (1.5)$$

where $[F, \mathcal{Q}] = 0$ is needed for this quantity to remain controllable. We may even have F involving an R -charge, as long as we choose one particular supercharge \mathcal{Q} carefully so that the two mutually commute. In many practical examples, coming out of string theory, this option is available and exploited.

Although such an insertion of chemical potentials may appear an innocent device to keep track of global charges of states, this is true only for theories suitably gapped to begin with. With gapless theories, this chemical potential modifies the Lagrangian in such a way that asymptotic directions that transform under F become massive. Since one gaps the asymptotic dynamics artificially, one should not expect the twisted partition function to behave nicely in the $\nu = 0$ limit.

Recovering information about the original theory prior to turning on ν is hardly straightforward although well-established routines exist for a handful of classes of theories. The Atiyah-Patodi-Singer index theorem, applicable to non-linear sigma models onto manifolds with boundary, is one such classic example while a more recent one is the 1d gauged quiver quantum mechanics as explained in Ref. [10]. Beyond these few, however, no general prescription is known. Despite such difficulties, the twisted partition functions of such mass-deformed theories proved to be very useful for some tasks, e.g., most notably, checking Strong-Weak dualities [19,29–33].

What do we do to actually evaluate such objects? The popular trick of the localization naturally enters the story when chemical potentials are turned on. The chemical potentials ν tend to push the dynamics to a small subset of the configuration space or even to a small part of the spacetime; the localization method is then invoked to amplify this effect maximally, whereby the path integral is reduced to that of a Gaussian path integral followed by a finite number of leftover zero-mode integrals from vector multiplets.

An interesting fact about the localization routine is that, in the final expression, β as in $e^{-\beta\mathcal{Q}^2}$ automatically drops out. This may happen because the system is fully gapped by

ν so that the naive Bose-Fermi cancellation works perfectly. In fact, this is the case for main examples of this paper, namely 4d $\mathcal{N} = 1$ theories which are maximally mass deformed by ν 's. As such, we will work with

$$\Omega(\nu) \equiv \text{Tr}(-1)^{\mathcal{F}} e^{\nu F} e^{-\beta \mathcal{Q}^2} \Big|_{\text{localization}}. \quad (1.6)$$

Sometimes this lift of the asymptotic flat direction by ν is incomplete, which tends to happen in odd spacetime dimensions. In such cases, the localization still removes β by computing, implicitly, a limit of $\beta \rightarrow 0$,

$$\text{Tr}(-1)^{\mathcal{F}} e^{\nu F} e^{-\beta \mathcal{Q}^2} \Big|_{\text{localization}} = \lim_{\beta \rightarrow 0} \text{Tr}(-1)^{\mathcal{F}} e^{\nu F} e^{-\beta \mathcal{Q}^2}. \quad (1.7)$$

This has been first noted for 1d systems [10] and further checked in Ref. [2].

Although we have described how the Euclidean time span β naturally drops out in the localization computation, the resulting twisted partition function Ω can actually retain β indirectly, via the chemical potential ν understood as a holonomy associated with an external flavor gauge field,

$$i\partial_t \rightarrow i\partial_t + \frac{\nu F}{\beta}. \quad (1.8)$$

In the small β limit, one has an option of keeping ν finite or keeping the alternate variable $\tilde{\nu}$ finite with

$$\nu = \beta \tilde{\nu}. \quad (1.9)$$

The so-called ‘‘real’’ masses in 3d are, for example, nothing but such finite $\tilde{\nu}$.

We can think about something similar for the gauge holonomy variables, u , which enter the localization formulas for Ω as

$$\Omega^G(\nu) = \int d^{\text{rank}} u g_G(u; \nu). \quad (1.10)$$

We introduced the label G to denote the theory and $g_G(u, \nu)$ is from the Gaussian integrals over nonzero modes. If we introduce the similarly rescaled variables $\tilde{u} = u/\beta$ in the small β limit, the following object where the integral is taken over \tilde{u} instead of u ,

$$\sim \int d^{\text{rank}} \tilde{u} \lim_{\beta \rightarrow 0} \beta^{\text{rank}} g_G(\beta \tilde{u}; \beta \tilde{\nu}) \quad (1.11)$$

appears naturally. Since β is taken to be arbitrarily small, the periodic nature of u is now lost. What would such an integral compute?

To be precise, let us consider a spacetime of type $S^1 \times M_{d-1}$. In the small radius limit, the dimensional reduction on S^1 produces a theory on M_{d-1} with the same

field content as the original theory. We will label this theory on M_{d-1} by the same label G , whose partition function would also produce a localized path integral as

$$\mathcal{Z}^G(\tilde{\nu}) = \int d^{\text{rank}} \tilde{u} f_G(\tilde{u}; \tilde{\nu}). \quad (1.12)$$

Past experiences with such objects tell us that the limit is often equipped with an extra exponential factor,

$$\beta^{\text{rank}} g_G(\beta \tilde{u}; \beta \tilde{\nu}) \rightarrow e^{S_G^{\text{Cardy}}/\beta + \text{subleading terms}} f_G(\tilde{u}; \tilde{\nu}) \quad \text{as } \beta \rightarrow 0, \quad (1.13)$$

where S_G^{Cardy} is the Cardy exponent [34]. Then, the naive expectation is

$$\Omega^G(\beta \tilde{\nu}) \rightarrow e^{S_G^{\text{Cardy}}/\beta + \text{subleading terms}} \times \mathcal{Z}^G(\tilde{\nu}) \quad \text{as } \beta \rightarrow 0, \quad (1.14)$$

where the two partition functions were computed for the one and the same gauge theory, G , only in two different dimensions. The exponent S_G^{Cardy} would dictate ‘‘high temperature behavior’’ of the twisted partition function.

B. Holonomy saddles: An overview

However, comparing (1.10) and (1.12), one easily realizes that this is too rash. A limiting formula like (1.14) would hold if and only if the toroidal du integration in (1.10) can be opened up to a planar integration in (1.12); since this is a discontinuous process, this may be justified only if, in the small β limit, the infinitesimal region around $u = 0$ contributes dominantly to the integral.

As a simple counterexample, which may look trivial but is illuminating nevertheless, consider an $SU(2)$ gauge theory with matter multiplets with integral isospins only. Suppose that we choose the range of u suitable for the odd isospins, say $[0, 1)$ in our convention where weight vectors are normalized to be integral and the holonomies are divided by 2π . The integrand $g(u; z)$ would then be invariant under the shift related by the center, $u \rightarrow u + 1/2$; the expansion of g around $u = 1/2$ will look exactly the same as that around $u = 0$, so that the integral near $u = 0$ and that near $u = 1/2$ contribute exactly the same amount. Although this particular problem is easily countermanded by an overall factor 2, it does warn us of a generic danger in confining ourselves to small regions near $u = 0$ when u is a periodic variable.

What happens generically is that the small β limit of $\Omega(\beta \tilde{\nu})$ is actually a sum of \mathcal{Z} 's for several disjoint theories on M_{d-1} , such that the limit has the form

$$\Omega^G(\beta \tilde{\nu}) \rightarrow \sum_{u_H} e^{S_H^{\text{Cardy}}/\beta + \text{subleading terms}} \times \mathcal{Z}^H(\tilde{\nu}) \quad \text{as } \beta \rightarrow 0, \quad (1.15)$$

instead of (1.14). The summand is labeled by special holonomies values u_H around which the dimensional

reduction gives a theory H , with potentially smaller field content than the naive dimensional reduction of the original theory G . The integration over the toroidal u 's reduces to patches of planar integrations near such u_H 's while contributions from the rest become suppressed by $e^{-1/\beta}$. The accompanying limit in the localization formulas should be similarly

$$\begin{aligned} & \beta^{\text{rank}} g_G(u_H + \beta \tilde{u}; \beta \tilde{v}) \\ & \rightarrow e^{S_H^{\text{candy}}/\beta + \text{subleading terms}} \times f_H(\tilde{u}; \tilde{v}) \quad \text{as } \beta \rightarrow 0 \\ \mathcal{Z}^H(\tilde{v}) &= \int d^{\text{rank}} \tilde{u} f_H(\tilde{u}; \tilde{v}). \end{aligned} \quad (1.16)$$

The discrete locations u_H are infinitely separated from one another, in the limit of $\beta \rightarrow 0$, and thus cannot be captured by the \tilde{u} integration near the origin alone.

With at least one nontrivial u_H , one must ask which of these saddles is dominant in the small β limit; one might have expected that the theory G at the naive saddle at $u_H = 0$ is the dominant one, given its largest light field content, but it turns out this is generally false. In particular, when M_{d-1} is T^{d-1} whereby Ω^G and \mathcal{Z}^H would both compute the Witten indices, each of admissible u_H generically contributes on equal footing. In other words, (1.15) would reduce to

$$\mathcal{I}_d^G(\beta \tilde{v}) \rightarrow \sum_{u_H} \mathcal{I}_{d-1}^H(\tilde{v}) \quad \text{as } \beta \rightarrow 0. \quad (1.17)$$

This means that u_H must be such that the theory H there must have supersymmetric vacua.¹

Let us take $d = 4$ $\mathcal{N} = 1$ theories, which will be our main examples. With generic holonomies, u , the 3d theory would be a product of free $U(1)$'s whose vacuum manifolds are generically lifted by a combination of induced FI constants or Chern-Simon levels. Light charged multiplets, say with the charge λ with respect to the Cartan $U(1)$'s, would be needed for vacua with unbroken supersymmetry. This constrains the position u to quantized values, u_H ,

$$\lambda \cdot u_H \in \mathbb{Z} \quad (1.18)$$

for each such λ . In this manner, a contributing H -saddle is equipped with a set of unbroken charges λ 's, which in turn defines the 3d theory H at u_H , modulo UV couplings in the 3d sense inherited and computable from the original 4d theory G . What we described here is a little simplified; it turns out that when the matter content is not symmetric under charge conjugation, one can actually have an

¹When the theory possesses a gapless asymptotic sector, so that the twisted partitions do not compute the true index, this condition should be relaxed since the twisted partition functions capture the so-called bulk part of the true index.

H -saddle with decoupled $U(1)$'s or pure Yang-Mills sectors, as long as appropriate Chern-Simons coefficients are generated from integrating out heavy modes. What remains unchanged, though, is that contributing u_H 's occur discretely. See Sec. III for a complete characterization of H -saddles.

What we described above is a generic feature of gauge theories, due to the special roles played by the holonomy variables: The toroidal nature of the holonomy variables appears lost in the small radius limit, yet the periodic nature should not be ignored. Integrating over such holonomy variables, such as for gauge theories on compact spacetime with a circle, we must remember to keep careful track of these holonomies. For supersymmetric partition functions, it so happens that there are multiple saddles which contribute to the total expression, each of which can be understood as a partition function of some other theories in one fewer dimension.

In this paper, we will consider implications of H -saddles in the context of 4d $\mathcal{N} = 1$ theories on T^2 fibered over Riemannian surfaces of arbitrary genus Σ_g . General partition functions of this class were given very recently via the so-called Bethe ansatz equation (BAE) [19]. In this approach one first consider compactification on T^2 reducing the system to 2d, and vacua and partition functions are found via the effective 2d twisted superpotential of Coulombic variables. The vacua thus found is called Bethe vacua [35]. As such, this construction works for a restricted class of 4d theories, where, given the matter content, the superpotential is appropriately suppressed to allow maximal flavor symmetry. On the other hand, the construction is ideal for the investigation of the H -saddle phenomena since the latter turns out to be quite manifest in the classification of BAE vacuum solution. T^2 fiber has two circles, whose relative size is encoded in the complex structure τ . In the large and the small τ limit, one of the two circles becomes very small relative to the other, and as such the phenomenon of H -saddles emerges on the smaller of the two circle directions.

The 2d twisted superpotential \mathcal{W} is naturally a function of the pair of the holonomies along T^2 , packaged into rank-many complex coordinates u . Bethe vacua are particular holonomy values, u_* , where $e^{2\pi i \partial \mathcal{W}} = 1$. This vacuum condition is periodic under integral shifts, $u \rightarrow u + n + m\tau$ where τ is the complex structure of T^2 , which is a gauge equivalence. What we will discover is that these Bethe vacua appear in clusters, scattered at discrete places in the unit cell,

$$u_*/\tau \simeq \tilde{u}_H + \tilde{\sigma}_*, \quad u_* \simeq u_H + \sigma_*, \quad (1.19)$$

where $\tilde{u}_H \sim 1/\tau$ and $u_H \sim \tau$ with coefficients between 0 and 1 for $\tau \rightarrow i0^+$, $i\infty$ limits, respectively. Each such H -saddle would come with multiple and finite $\tilde{\sigma}_*$'s and σ_* 's, which represents supersymmetric vacua in the reduced 3d theory at such H -saddles. Depending on which circle is called the

Euclidean time, the limit will also compute the Cardy exponents or the Casimir energies at each of such H -saddles. The question of which saddle dominates becomes a non-trivial issue, generically compromising folklore on universality of such exponents. We will revisit the asymptotics of the partition functions in Sec. IV.

II. 4D $\mathcal{N}=1$ PARTITION FUNCTIONS AND BAE

Recently the supersymmetric partition function of a four-dimensional $\mathcal{N}=1$ theory on \mathcal{M}_4 was discussed [19] where \mathcal{M}_4 is a torus bundle over a Riemannian surface Σ_g :

$$T^2 \rightarrow \mathcal{M}_4 \rightarrow \Sigma_g. \quad (2.1)$$

The partition function is obtained by considering an A-twisted theory on Σ_g via nontrivial background flux $\mathfrak{n}_R = g - 1$ for the $U(1)_R$ symmetry group. Here we give a quick summary of results in Ref. [19]. We will use G to denote the $\mathcal{N}=1$ theory itself, while \mathcal{G} and \mathfrak{G} are the gauge group and the associated Lie algebra, respectively.

Before we plunge into details, a cautionary remark is in order. Much of what follows will be phrased in terms of gauge holonomies on T^2 , valued in two copies of the Cartan torus. As was emphasized by Witten and others [36–40], however, the space of connections on T^2 and higher dimensional torii can in general admit disconnected components, even when the gauge group is connected. Such possibilities are not taken into account in the computations outlined below, so we will confine ourselves, in this paper, to theories with $\mathcal{G} = SU(r+1)$, $Sp(r)$.

A. A-twisted background

Compactifying on T^2 , one has a two-dimensional $\mathcal{N}=(2,2)$ supersymmetric theory with infinite Kaluza-Klein modes. Performing the path integral via localization, summing over the magnetic flux sector, and then evaluating the resulting residue formulas, the partition function is written universally as a sum over the so-called Bethe vacua,

$$\begin{aligned} \Omega &= \sum_{u_* \in \mathcal{S}_{\text{BE}}} \mathcal{F}_1(u_*, \nu; \tau)^{p_1} \mathcal{F}_2(u_*, \nu; \tau)^{p_2} \mathcal{H}(u_*, \nu; \tau)^{g-1} \\ &\times \prod_{\alpha} \Pi_A(u_*, \nu; \tau)^{n_{\alpha}}. \end{aligned} \quad (2.2)$$

where $p_1, p_2 \in \mathbb{Z}$ are the two Chern numbers of the circle bundles:

$$p_1 = \frac{1}{2\pi} \int_{\Sigma_g} dA_{KK_1}, \quad p_2 = \frac{1}{2\pi} \int_{\Sigma_g} dA_{KK_2}. \quad (2.3)$$

$\tau = \tau_1 + i\tau_2$ is the modular parameter of the torus with $\tau_2 = \frac{\beta_2}{\beta_1}$, where β_1 and β_2 are two radii of the torus. Note that one can perform a modular transformation of T^2 such that $(p_1, p_2) = (p, 0)$.

Now let us see how field theory data enter this universal formula. The basic variables are holonomies along T^2 . The gauge holonomies u_1, u_2 , along these two circles

$$a_1 = \frac{1}{2\pi} \int_{S^1_{\beta_1}} A, \quad a_2 = \frac{1}{2\pi} \int_{S^1_{\beta_2}} A, \quad (2.4)$$

of 2π period each, are combined to

$$u = a_1 \tau - a_2 \quad (2.5)$$

and, similarly for the complexified flavor holonomy

$$\nu = a_1^{(F)} \tau - a_2^{(F)}. \quad (2.6)$$

All of these holonomy variables obey

$$u_a \sim u_a + 1 \sim u_a + \tau, \quad \nu_A \sim \nu_A + 1 \sim \nu_A + \tau, \quad (2.7)$$

under the respective large gauge transformations.

The effective action of the theory is fully governed by two holomorphic functions \mathcal{W} and Ω , which are called the effective twisted superpotential and the effective dilaton. Each component in (2.2) and (2.30) is then obtained from those two quantities,

$$\mathcal{F}_1(u, \nu; \tau) = \exp \left[2\pi i \frac{\partial \mathcal{W}}{\partial \tau} \right], \quad (2.8)$$

$$\mathcal{F}_2(u, \nu; \tau) = \exp \left[2\pi i \left(W - u_a \frac{\partial \mathcal{W}}{\partial u_a} - \nu_A \frac{\partial \mathcal{W}}{\partial \nu_A} - \tau \frac{\partial \mathcal{W}}{\partial \tau} \right) \right], \quad (2.9)$$

$$\mathcal{H}(u, \nu; \tau) = e^{2\pi i \Omega(u, \nu; \tau)} \left(\det_{ab} \frac{\partial^2 \mathcal{W}(u, \nu; \tau)}{\partial u_a \partial u_b} \right), \quad (2.10)$$

$$\begin{aligned} \Phi_a(u, \nu; \tau) &= \exp \left[2\pi i \frac{\partial \mathcal{W}}{\partial u_a} \right], \\ \Pi_A(u, \nu; \tau) &= \exp \left[2\pi i \frac{\partial \mathcal{W}}{\partial \nu_A} \right]. \end{aligned} \quad (2.11)$$

For semisimple \mathcal{G} , the W-bosons and their superpartners do not contribute to the effective twisted superpotential, and only charged chiral multiplets contribute. The contribution of a single chiral multiplet is given by

$$\begin{aligned} \mathcal{W}_{\Phi} &= -\frac{u^3}{6\tau} + \frac{u^2}{4} - \frac{u\tau}{12} + \frac{1}{24} \\ &+ \frac{1}{(2\pi i)^2} \sum_{k=0}^{\infty} (\text{Li}_2(xq^k) - \text{Li}_2(x^{-1}q^{k+1})), \end{aligned} \quad (2.12)$$

where Li_2 is a polylogarithm function. We have defined $x = e^{2\pi i u}$ and $q = e^{2\pi i \tau}$.

Classification of the Bethe vacua \mathcal{S}_{BE} starts with solving

$$1 = \Phi_a, \quad (2.13)$$

which can be expressed more explicitly as

$$\begin{aligned} \Phi_a(u, \nu; \tau) &= \prod_i \prod_{\rho_i} \Psi(\rho_i \cdot u + \nu_i; \tau)^{\rho_i^a}, \\ \Psi(w; \tau) &\equiv e^{-\pi i w^2 / \tau} \theta(w, \tau)^{-1}, \\ \theta(w, \tau) &= q^{1/12} t^{-1/2} \prod_{k \geq 0} (1 - tq^k)(1 - t^{-1}q^{k+1}), \end{aligned} \quad (2.14)$$

with $t = e^{2\pi i w}$, where $\nu_i \equiv \nu \cdot F_i$ is the net sum of flavor chemical potentials for the i th multiplet with flavor charges F_i . The product is over chiral multiplets, labeled by i for each gauge multiplet and the charges ρ_i thereof with respect to the Cartan. Then, \mathcal{S}_{BE} , which is nothing but the set of the supersymmetric vacua for the 2d twisted superpotential, may be defined as

$$\begin{aligned} \mathcal{S}_{\text{BE}} &= \{u_* | \Phi_a(u_*, \nu; \tau) = 1, \forall a, w \cdot u_* \neq u_*, \\ &\quad \forall w \in W_G\} / W_G. \end{aligned} \quad (2.15)$$

The additional constraint, that vacua invariant under any part of the Weyl group W_G are to be ignored, has been noted in the past literature, most notably Refs. [41,42].

With the explicit form of \mathcal{W} in (2.12), one can similarly compute the rest. The effective dilaton contribution is given by

$$\begin{aligned} e^{2\pi i \Omega} &= \left(\prod_i \prod_{\rho_i \in \mathfrak{R}_i} \Psi(\rho_i \cdot u + \nu_i; \tau)^{r_i - 1} \right) \\ &\quad \times \left(\eta(\tau)^{-2\text{rank}(G)} \prod_{\alpha \in \mathfrak{g}} \Psi(\alpha \cdot u; \tau) \right), \end{aligned} \quad (2.16)$$

where the product in the second parentheses is taken over the roots of $\mathfrak{g} = \text{Lie}(G)$. From the effective dilaton (2.16) and the effective twisted superpotential (2.12), one can obtain the handle-gluing operator \mathcal{H} as well. The explicit forms of the fibering operators \mathcal{F} and the flavor flux operators Π_α are

$$\mathcal{F}_{1,2}(u, \nu; \tau) = \prod_i \prod_{\rho_i \in \mathfrak{R}_i} \Xi_{1,2}(\rho_i \cdot u + \nu_i; \tau), \quad (2.17)$$

$$\Pi_A(u, \nu; \tau) = \prod_i \prod_{\rho_i \in \mathfrak{R}_i} \Psi(\rho_i \cdot u + \nu_i; \tau)^{\omega_i^a}, \quad (2.18)$$

where we have defined

$$\Xi_1(w; \tau) = e^{\frac{\pi i}{3\tau^2} w^3 - \frac{\pi i}{6} w} \Gamma_0(u; \tau), \quad (2.19)$$

$$\Xi_2(w; \tau) = e^{2\pi i \left(\frac{w^3}{6\tau} - \frac{w^2}{4} + \frac{w\tau}{12} + \frac{1}{24} \right)} \prod_{k=0}^{\infty} \frac{f(w+k\tau)}{f(-w+(k+1)\tau)}, \quad (2.20)$$

with

$$\Gamma_0(w; \tau) = \prod_{n=0}^{\infty} \left(\frac{1 - t^{-1}q^{n+1}}{1 - tq^{n+1}} \right)^{n+1}, \quad (2.21)$$

$$f(w) = \exp \left[\frac{1}{2\pi i} \text{Li}_2(e^{2\pi i w}) + w \log(1 - e^{2\pi i w}) \right]. \quad (2.22)$$

For the flux operator, the products are taken over every chiral multiplet and the weights of its representation. ρ_i^a is the a th Cartan charge of the gauge weight ρ_i while ω_i^a is the a th Cartan charge of the flavor weight ω_i . With nontrivial background flux \mathfrak{n}_α for the flavor group, the flavor flux operator Π_A contributes to the partition function as in (2.2).

Before proceeding, however, we should mention a few caveats. The most obvious is the presence of nonanomalous $U(1)_R$ symmetry. Since this symmetry is used for the topological A-twisting [43], one must further require the $U(1)_R$ charges of chiral multiplets r_i be integral. As such, neither for pure $\mathcal{N} = 1$ Yang-Mills theories nor for typical $\mathcal{N} = 1$ superconformal theories would this methodology be applicable. Later, however, we will go to a slightly different class of spacetime, with the same geometry but different fluxes, so that the integrality of the $U(1)_R$ charges can be relaxed. There, the partition function formula should be applicable to $\mathcal{N} = 1$ superconformal theories with a -maximized R -charges.

A less obvious caveat, although quite rampant in the exact partition function computations, comes from the flavor chemical potentials. As mentioned earlier, the latter means that the theory is artificially mass deformed, and that we may not be able to recover physics of the original undeformed theory easily. This danger is present in all exact twisted partition function computations via the localization, but perhaps a little more so in this class since this computation turns on a chemical potential for each and every chiral multiplet: One must always take extreme care in interpreting the results.

With $SU(N)$ theories with N_f fundamental flavors, for example, this current computation would give a simple numerical Witten index for all N_f if we take $\mathcal{M}_4 = T^4$. However, such theories are often equipped with a manifold of the vacuum moduli, which, since the number of spacetime dimensions is larger than two, should have made the notion of the Witten index ill defined. One must really regard these partition functions as probing theories that are compactified on T^2 with flavor holonomies necessarily turned on along the two circles.

B. Alternate backgrounds and superconformal index

So far we have considered an A-twisted theory, whose supersymmetric background includes the nontrivial $U(1)_R$ gauge field of

$$\nu_R = 0, \quad \mathfrak{n}_R = g - 1. \quad (2.23)$$

On the other hand, there is another class of supersymmetric backgrounds without the $U(1)_R$ flux, which is called “physical gauge” in Ref. [19]:

$$\nu_R = \frac{1-g}{p}\tau, \quad \mathfrak{n}_R = 0, \quad (2.24)$$

with $g-1 \equiv 0 \pmod{p}$. This is achieved by starting with $p_1 = p$ and $p_2 = 0$, and taking a large gauge transformation on R -symmetry that removes the R -flux in favor of the R -chemical potential.

$$\mathcal{F}^{\text{phys}}(u, \nu, \nu_R; \tau) = \left(\prod_i \prod_{\rho_i \in \mathfrak{R}_i} \Xi_1(\rho_i \cdot u + \nu_i + \nu_R(r_i - 1); \tau) \right) \left((-1)^{\frac{l_R(l_R+1)}{2} \text{rank}(\mathcal{G})} \eta(\tau)^{2l_R \text{rank}(\mathcal{G})} \prod_{\alpha \in \mathfrak{G}} \Xi_1(\alpha \cdot u + \nu_R; \tau) \right), \quad (2.26)$$

where

$$\nu_R = l_R \tau, \quad l_R = \frac{1-g}{p} \in \mathbb{Z}. \quad (2.27)$$

In addition, in the physical gauge, there is no effective dilaton contribution because $\mathfrak{n}_R = 0$. Thus, the handle-gluing operator is simply the Hessian determinant,

$$H^{\text{phys}}(u, \nu, \nu_R; \tau) = \det_{ab} \left(\frac{1}{2\pi i} \frac{\partial \log \Phi_a^{\text{phys}}}{\partial u_b} \right). \quad (2.28)$$

As a result, the partition function in the physical gauge is given by

$$\Omega^{\text{phys}} = \sum_{u_* \in \mathcal{S}_{\text{BE}}^{\text{phys}}} \mathcal{F}_1^{\text{phys}}(u_*, \nu, \nu_R; \tau) H(u_*, \nu, \nu_R; \tau)^{g-1} \times \prod_{\alpha} \Pi_A^{\text{phys}}(u_*, \nu, \nu_R; \tau)^{n_{\alpha}}, \quad (2.29)$$

where we restricted ourselves to the case $p_1 = p$ and $p_2 = 0$, with the Bethe vacua

$$\mathcal{S}_{\text{BE}}^{\text{phys}} = \{u_* | \Phi_a^{\text{phys}}(u_*, \nu, \nu_R; \tau) = 1, \forall a, w \cdot u_* \neq u_*, \forall w \in W_{\mathcal{G}}\} / W_{\mathcal{G}}. \quad (2.30)$$

Furthermore, Closset *et al.* advocated that once we arrive at this so-called “physical gauge,” the integral restriction on r_i can be lifted.

As such, this partition function is supposed to compute a limit of the superconformal index [23,24] if we take $g = 0$ and $p = 1$,

$$\Omega_{S^1 \times S^3}(q; x) \equiv \text{Tr}_{S^3} [(-1)^{\mathcal{F}} q^{2J+R} x^{G_F} e^{-\beta_2 H}], \quad (2.31)$$

The authors of Ref. [19] proposed that the partition function in this background can be written in a similar manner but with different operators \mathcal{H} , \mathcal{F} , etc., which will be our working assumption, below. The flux operators are

$$\Phi_a^{\text{phys}}(u, \nu, \nu_R; \tau) = \prod_i \prod_{\rho_i \in \mathfrak{R}_i} \Psi(\rho_i \cdot u + \nu_i + \nu_R(r_i - 1); \tau)^{\rho_i^a},$$

$$\Pi_A^{\text{phys}}(u, \nu, \nu_R; \tau) = \prod_i \prod_{\rho_i \in \mathfrak{R}_i} \Psi(\rho_i \cdot u + \nu_i + \nu_R(r_i - 1); \tau)^{\omega_i^a}, \quad (2.25)$$

while the fibering operator for the circle 1 is

where the pair of rotational chemical potentials that enter the usual superconformal index are identified. For the superconformal index, the large and the small radius limits are already discussed in literature quite extensively [16–18,20–22]. We reexamine these limits of the partition function using the Bethe formalism. The subtlety with the holonomy should be again present, and many results of the previous section carry over to the new background verbatim. Reference [19] initially motivated this construction via a large $U(1)_R$ transformation, as outlined above, from the A-twisted cases with $p \neq 0$. This cannot be really considered a derivation since the nonintegral values of r_i 's, inevitable for $\mathcal{N} = 1$ superconformal field theories, would be detrimental to such a process. On the other hand, an alternate justification was also given by the same authors, where these BAE expressions for $g = 0$, $p = 1$ are transformed to the conventional form of the superconformal index via contour manipulations. See Eq. (4.52).

III. BETHE VACUA FOR ELONGATED T^2

In this paper, for the sake of convenience, we will regard circle 2 the Euclidean time. Then the large and the small Euclidean times correspond to, respectively,

$$\tau = \dots + i \frac{\beta_2}{\beta_1} \rightarrow i\infty \quad \text{or} \quad i0^+. \quad (3.1)$$

The small τ limit can be viewed as compactification along circle 2, while the large τ limit would be viewed as the compactification along circle 1. This interpretation is possible as the size of the base Σ_g appears nowhere in the partition functions, and also because only the ratio of the two radii appears. For either compactification, the Kaluza-Klein towers will acquire a large mass shift at typical values of holonomy. This means that the low energy effective theory in the remaining three dimensions would be rank-many free

$U(1)$ theories. At such a generic point we will find that the 3d theory has supersymmetry spontaneously broken and thus cannot contribute solutions to the BAE.

What we will find is that solving 4d BAE will produce vacua clustered at some discrete and special choices of the holonomy. As $\tau \rightarrow i\infty$, these special holonomies u_H line up along circle 1, while for the other limit $\tau \rightarrow i0^+$ they line up along circle 2. At such special places u_H , i.e., at H -saddles, with the holonomy, the gauge charge set of the chiral multiplet will split,

$$\{\rho\} = \{\lambda\} \cup \{\hat{\rho}\}, \quad (3.2)$$

where those chirals associated with λ 's will produce light 3d chiral multiplets at the bottom of the KK tower, while those associated with $\hat{\rho}$ will produce KK towers with no such light 3d field. The vector multiplets in the adjoint representation would be also similarly decomposed, and a spontaneous symmetry breaking by a Wilson line will occur

$$\mathcal{G} \rightarrow \mathcal{H}, \quad (3.3)$$

where the unbroken, 3d gauge group \mathcal{H} has the same rank as the 4d gauge group \mathcal{G} . The new 3d gauge theory H , typically with smaller light field content, both vectors and chirals, than the naive dimensional reduction of the theory G , appears. The holonomy u_H 's and the new theories there, we will collectively call H -saddles [2].

This classification of H -saddles, to be explained in detail below, could include some special cases. The case with $\{\lambda\} = \{\rho\}$, for example at $u_H = 0$, would produce the 3d theory with the same field content as the naive dimensional reduction. Most of the related literature have assumed, effectively, that this type of H -saddle is either the only kind or the dominant one. The other extreme $\{\lambda\} = \emptyset$, or more generally the cases where λ do not span the charge vector space, would produce 3d theory with a pure gauge sector. Also, H -saddles that include an Abelian subgroup in \mathcal{H} require more care, since a large 3d Fayet-Iliopoulos constant can be generated even though the 4d theory had no such Abelian factor. For the latter types of H -saddles, a little more care must be given, which we will go through in Secs. III C. and III D.

While most of this section is devoted to the A-twisted case, the ‘‘physical’’ version is really no different. The BAE equations remain identical to those of the A-twisted case, except the additional $U(1)_R$ chemical potential $(r_i - 1)l_R\tau$ for the chiral fields. As such, the large τ limit of the ‘‘physical’’ version requires extra care, which will be addressed in Sec. III E.

A comment on a notation is in order, to avoid confusion. For a holonomy variable w with the natural periods, τ and 1, we will define its ‘‘fractional’’ part, $\{w/\tau\}$, as

$$\{w/\tau\} \equiv w/\tau - m, \quad (3.4)$$

where

$$m \equiv \lfloor w/\tau \rfloor \quad (3.5)$$

is an integer such that the real part of $w/\tau - m$ lies in $[0, 1)$. It follows that, for example,

$$\{\lambda \cdot u_H/\tau\} = 0, \quad \{\hat{\rho} \cdot u_H/\tau\} \neq 0, \quad (3.6)$$

in the large radius limit of A-twist cases (3.2), and

$$\{\lambda \cdot \tilde{u}_H/\tilde{\tau}\} = 0, \quad \{\hat{\rho} \cdot \tilde{u}_H/\tilde{\tau}\} \neq 0, \quad (3.7)$$

with $\tilde{u} \equiv u/\tau$ and $\tilde{\tau} = -1/\tau$, in the small radius limit of A-twisted cases. Note that we also use the same curly bracket $\{\cdot\cdot\cdot\}$ as a symbol for sets, as is customary. Hopefully, the distinction between these two is self-evident.

A. H -saddles in the small $\tilde{\tau} = -1/\tau$ limit

We will start with $\tau \rightarrow i\infty$, or $q \rightarrow 0$, although the discussion below may appear more natural in the other limit of $\tau \rightarrow i0^+$. In the small τ limit, the role of the two circles will be exchanged, so that our finding here will carry over almost verbatim, via an $SL(2, \mathbb{Z})$ action.

We start by noting that there are two different types of solutions to the 4d BAE equation in such a limit. The first type comes from assuming u remains finite under the scaling, and as such we find

$$\begin{aligned} \Psi(w; \tau) &\rightarrow \frac{q^{-1/12}}{t^{-1/2} - t^{1/2}} \\ \Phi_a(u; \tau) &\rightarrow \Phi_a^{3d} \equiv \Lambda_G^a \prod_i \prod_{\rho_i} [x^{-\rho_i/2} y_i^{-1/2} - x^{\rho_i/2} y_i^{1/2}]^{-\rho_i^a} \end{aligned} \quad (3.8)$$

with

$$\Lambda_G^a = q^{-\sum_i \rho_i^a / 12}, \quad (3.9)$$

where $y_i \equiv e^{2\pi i v_i}$ and $x^{\rho_i} \equiv \prod_a x_a^{\rho_i^a}$. The resulting BAE equations, $\Phi_a^{3d} = 1$, look exactly the same as the 3d BAE equations [30–33] of a dimensionally reduced theory with the same field content. The only unexpected feature is that $2\pi i\tau$ now plays the role of a UV FI constant for the trace-part $U(1)$ gauge field when the latter is present.

If $\sum \rho_i^a \neq 0$, the locations of the solutions, x_a , could be too far away and conflict with the above truncation. Thankfully, however, this never really happens for 4d theories with no $U(1)$ factor, as is necessary for the asymptotic freedom. With the gauge group \mathcal{G} being at most a product of semisimple Lie groups, we find

$$\Lambda_G^a = 1 \quad (3.10)$$

generally.²

On the other hand, a very different kind of solutions also exist. Suppose we consider a regime where x scales to zero, along with $q \rightarrow 0$. For example let us look for solutions near $x_a^{n_a} \sim q$. At such points $q \rightarrow 0$ might leave a factor $(1 - q/x^{\rho_i} y_i)$ as well. Note that we can still make use of the infinite product formulas only if $q/x^{\rho_i} \rightarrow 0$ as well, so one must first shift the argument $\rho_i \cdot u$ by an integral multiple of τ , using the identity

$$\theta(w; \tau) = (-1)^m t^m q^{m^2/2} \theta(w + m\tau; w), \quad (3.14)$$

bringing

$$\begin{aligned} \Psi(w, \tau) &= (-1)^m \Psi(\{w/\tau\}\tau; \tau) \\ &= (-1)^m e^{-\pi i \tau \{w/\tau\}^2 + \pi i \tau \{w/\tau\}} q^{-1/12} \\ &\quad \times \frac{1}{\prod_{k \geq 0} (1 - \{t\} q^k) (1 - q^{k+1}/\{t\})}, \end{aligned} \quad (3.15)$$

where $\{t\} \equiv e^{2\pi i \tau \{w/\tau\}}$ with $\{w/\tau\}$ obtained by an additive shift of w by $-m\tau$ for an integer m ; the real part of $\{w/\tau\}$ lies between 0 and 1. The aim is to make the infinite product well defined in the limit of $q \rightarrow 0$. At generic values of w , each of the infinite products reduces to 1, which shows that there is no solution to 4d BAE, $\Phi_a = 1$.

The new type of solutions will have to cluster around special places, called H -saddles, where the real part of $(\rho_i \cdot u + \nu_i)/\tau$ for ρ_i 's are integral and not necessarily zero. In the large τ limit, with ν_i kept finite, such H -saddles, say u_H , are located at

$$\lambda_i \cdot u_H \simeq m_{\lambda_i} \tau \quad (3.16)$$

²This can be seen from the Weyl character formulas, with the Cartan generators \mathcal{C} 's in an irreducible representation \mathfrak{R} of a semisimple gauge group \mathcal{G} ,

$$\chi_{\mathcal{R}}(e^u) \equiv \text{tr}_{\mathcal{R}} e^{u \cdot \mathcal{C}} = \frac{\sum_w (-1)^{|w|} e^{u \cdot w (\lambda_{\mathcal{R}} + \rho_w)}}{\sum_w (-1)^{|w|} e^{u \cdot w (\rho_w)}}, \quad (3.11)$$

where $\lambda_{\mathcal{R}}$ is the highest weight, $\rho_w \equiv \sum_{\alpha \in \Delta_+} \alpha/2$ is the Weyl vector, and the sums on the right-hand side are over the Weyl group. It follows that $\chi_{\mathcal{R}}(e^u) = \chi_{\mathcal{R}}(e^{w(u)})$ for any Weyl reflection w and thus

$$\sum_{\rho \in \mathfrak{R}} \rho \cdot w(u) = \sum_{\rho \in \mathfrak{R}} \rho \cdot u, \quad (3.12)$$

Since u is arbitrary and since this holds for any Weyl reflection w , it follows

$$\sum_{\rho \in \mathfrak{R}} \rho^a = 0 \quad (3.13)$$

for each irreducible representation \mathfrak{R} of \mathcal{G} .

for some subset λ_i 's of charge vectors ρ_i 's and accompanying integers m 's. We can then invoke the language of Wilson-line symmetry breaking and consider u_H as the point where λ_i is nearly massless while the others are heavily gapped. The holonomy in question is along the direction 1, which was the fiber circle.

In the neighborhood of u_H , we write for the nearly massless ones

$$\lambda_i \cdot u + \nu_i = m_{\lambda_i} \tau + \lambda_i \cdot \sigma + \nu_i, \quad (3.17)$$

with σ and ν_i understood to remain finite while $\tau \rightarrow i\infty$. In such a neighborhood of u_H , then, we introduce new shifted Cartan variables $\sigma_a \equiv (u - u_H)_a$ such that

$$x_a^{\lambda_i} = q^m z_a^{\lambda_i}, \quad z_a \equiv e^{2\pi i \sigma_a}. \quad (3.18)$$

The shift of u_a to σ_a is integral in τ , so can at most change the sign of Ψ and Φ_a 's, and for these light fields, we can proceed exactly the same way as the naive limit as in (3.8). For heavy fields, from $\hat{\rho}$'s, on the other hand, the infinite product for the relevant θ reduces to 1, and leaves behind a prefactor only. Then, near u_H , contributions to Φ_a from light and heavy modes accumulate to

$$\begin{aligned} \Phi_a(u; \tau) &\rightarrow \Phi_a^{3d;H} \\ &\equiv (-1)^M \Lambda_H \prod_i \prod_{\lambda_i} [z^{-\lambda_i/2} y_i^{-1/2} - z^{\lambda_i/2} y_i^{1/2}]^{-\lambda_i^a} \end{aligned} \quad (3.19)$$

for some integer M , with

$$\Lambda_H^a \equiv \prod_i \prod_{\hat{\rho}_i} e^{-\pi i \tau \hat{\rho}_i^a (\{(\hat{\rho}_i \cdot u_H + \hat{\rho}_i \cdot \sigma + \nu_i)/\tau\}^2 - \{(\hat{\rho}_i \cdot u_H + \hat{\rho}_i \cdot \sigma + \nu_i)/\tau\})}, \quad (3.20)$$

where we used $\Lambda_G = 1$.

As such, the solutions to the 4d BAE are neatly decomposed into union of the 3d BAE solutions at these distinct H -saddles,

$$\{u_* : 1 = \Phi_a(u_*)\}_{\tau \rightarrow i\infty} = \cup_{u_H} \{u_H + \sigma_* : 1 = \Phi_a^{3d;H}(\sigma_*)\}, \quad (3.21)$$

where $\lambda_i \cdot u_H/\tau$ are real and integral for some subset λ_i 's of ρ_i 's. For actual admissible BAE vacua, we must exclude those solutions that are fixed under some Weyl reflection but it is clear that this does not interfere with this classification into clusters around H -saddles. There are two special subclasses of H -saddles that are noteworthy. One is when $\{\lambda\} = \{\rho\}$, which occurs when the gauge holonomy is trivial. This of course corresponds to the naive dimensional reduction, where the 4d BAE reduces to $1 = \Phi_a^{3d}$ above. That is, we included the first type of solutions on equal footing as well, by assigning them to $u_H = 0$. The other

classes are when the unbroken matter charges λ_i 's do not span the entire weight vector space. This means that the reduced 3d theory includes pure gauge sectors with no chiral fields coupled. The latter deserves a more in-depth discussion which is postponed to the end of this section, as this requires a more explicit evaluation of Λ_H 's.

B. H -saddles in the small τ limit

We start with the identity

$$\theta(w, \tau) = i e^{-\pi i w^2 / \tau} \theta(w/\tau, -1/\tau), \quad (3.22)$$

which implies that

$$\Psi(w, \tau) = \frac{i}{\theta(w/\tau, -1/\tau)} = \frac{i}{\theta(\tilde{w}, \tilde{\tau})},$$

$$\tilde{w} \equiv \frac{w}{\tau}, \quad \tilde{\tau} \equiv -\frac{1}{\tau}. \quad (3.23)$$

As such, the BAE in the small τ limit can be cast as

$$1 = \Phi_a(u, \nu; \tau) \equiv \prod_i \prod_{\rho_i} \tilde{\Psi}(\rho_i \cdot \tilde{u} + \tilde{\nu}_i; \tilde{\tau})^{\rho_i^a}, \quad (3.24)$$

with

$$\tilde{\Psi}(w, \tau) = i e^{-\pi i w^2 / \tau} \frac{1}{\theta(\tilde{w}, \tilde{\tau})}, \quad (3.25)$$

since the quadratic exponents in the prefactor will cancel out for Φ_a thanks to gauge and axial anomaly cancellation. This way, the small τ analysis will follow the large τ analysis almost verbatim.

Again, there are two types of solutions. The first class comes with finite \tilde{u} 's, or equivalently,

$$u \sim \tau, \quad \text{as } \tau \rightarrow i0^+, \quad (3.26)$$

which corresponds to a trivial holonomy along the time circle. In this obvious saddle, all chiral fields acquire finite mass $\sim \nu_i$ and thus are in equal footing. The 4d BAE reduces to, as $\tilde{q} \rightarrow 0$,

$$1 = \tilde{\Phi}_a^{3d} \equiv i^{\#} \prod_i \prod_{\rho_i} [\tilde{x}^{-\rho_i/2} \tilde{y}_i^{-1/2} - \tilde{x}^{\rho_i/2} \tilde{y}_i^{1/2}]^{-\rho_i^a}, \quad (3.27)$$

which is essentially the same equation as (3.8) of the large radius limit, once we replace x , y , and q by \tilde{x} , \tilde{y} , and \tilde{q} .

Similarly to the large τ limit, more solutions appear as we allow \tilde{u} to scale with the large $\tilde{\tau} = -1/\tau$, such that,

$$\lambda_i \cdot \tilde{u} = m_{\lambda_i} \tilde{\tau} + \lambda_i \cdot \tilde{\sigma}, \quad (3.28)$$

again exactly as before, for some proper subset λ_i 's of ρ_i 's, or equivalently

$$\lambda_i \cdot u \simeq -m_{\lambda_i} + \tau \lambda_i \cdot \tilde{\sigma}. \quad (3.29)$$

As in the previous large τ case, the subset λ_i 's represents light degrees of freedom among the matter fields. Shifting the tilded variables similarly, vacua around such an H -saddle solve

$$1 = \tilde{\Phi}_a^{3d;H} \equiv (-1)^{M'/2} \tilde{\Lambda}_H \prod_i \prod_{\lambda_i} [\tilde{z}^{-\lambda_i/2} \tilde{y}_i^{-1/2} - \tilde{z}^{\lambda_i/2} \tilde{y}_i^{1/2}]^{-\lambda_i^a} \quad (3.30)$$

for some integer M' , with

$$\tilde{\Lambda}_H^a \equiv \prod_i \prod_{\hat{\rho}_i} e^{-\pi i \tilde{\tau} \hat{\rho}_i^a (\{(\hat{\rho}_i \cdot \tilde{u}_H + \hat{\rho}_i \cdot \tilde{\sigma} + \tilde{\nu}_i) / \tilde{\tau}\}^2 - \{(\hat{\rho}_i \cdot \tilde{u}_H + \hat{\rho}_i \cdot \tilde{\sigma} + \tilde{\nu}_i) / \tilde{\tau}\})}. \quad (3.31)$$

As before, the latter ignores overall the phase factor.

The solutions to the 4d BAE can be again grouped into

$$\{u_* : 1 = \Phi_a(u_*)\}_{\tau \rightarrow i0^+} = \cup_{\tilde{u}_H} \{\tilde{u}_H + \tilde{\sigma}_* : 1 = \tilde{\Phi}_a^{3d;H}(\tilde{\sigma}_*)\}, \quad (3.32)$$

where, again, $\lambda_i \cdot \tilde{u}_H / \tilde{\tau}$ are real and integral for some subset λ_i 's of ρ_i 's. As before, the necessary exclusion of those solutions fixed under some Weyl reflection should be performed, which does not interfere with this H -saddle classification of solutions. The sum includes the special case of $\{\lambda_i\} = \{\rho_i\}$, where the 3d BAE is nothing but (3.27). Note that the locations of H -saddles are along direction 2 in this small τ limit while they were along direction 1 in the large τ limit. It is reasonably clear that the solutions to the BAE can be matched, between the large radius limit and the small radius limit, one on one and saddle by saddle. As before, the naive H -saddle at $u_H = 0$ as well as those that involve a pure gauge sector should be included. See Sec. III D.

C. 3d UV couplings

In both limits, the naive reduction on either circle gives 3d BAE, $1 = \Phi^{3d}$ or $1 = \tilde{\Phi}^{3d}$, whereby one recovers vacua with negligible holonomies. This by itself does not count all 4d BAE vacua, however. In order to account for all vacua, one must consider the possibility of turning on some nontrivial holonomies, leading to $1 = \Phi^{3d;H}$ near u_H in the large radius limit, or $1 = \tilde{\Phi}^{3d;H}$ near \tilde{u}_H in the small radius limit. The holonomy in question is along circle 1 in the large radius limit, hence along the fiber circle, and along circle 2 in the small radius limit, hence along the time circle, respectively. Either way, the effective 3d theory at a given H -saddle comes with reduced field content: only those associated with weights λ_i 's remain ‘‘light,’’ while the

rest acquire large masses of order β_2/β_1 or β_1/β_2 , respectively.

When we consider a particular H -saddle and 3d effective theory sitting there, the effect of the heavy modes can manifest via induced couplings.³ In 3d theories, the auxiliary D -term shows up as

$$(\zeta + \sigma \cdot \kappa + \mu \cdot \kappa^F) \cdot D \quad (3.33)$$

with the FI constant ζ , the gauge Chern-Simons level κ , and the gauge-flavor-mixed Chern-Simons level κ^F . σ is the real scalar in the Cartan part of the 3d vector multiplet while μ is the real masses associated with the flavor symmetries.

This means that one-loop of chiral multiplet of charge Q and an effective mass $M(\sigma) = Q \cdot \sigma + M_0$ will induce a shift [29],

$$\Delta(\zeta + \sigma \cdot \kappa + \mu \cdot \kappa^F) = \frac{1}{2} Q |Q \cdot \sigma + M_0|. \quad (3.34)$$

In the $\tau = \tau_1 + i\tau_2 \rightarrow i\infty$ limit, the heavy modes have masses M_0 of order $|\tau_2| \gg |\sigma|$, and the leading terms will induce $\zeta \sim |\tau|$ as

$$\begin{aligned} \zeta + \sigma \cdot \kappa + \mu \cdot \kappa^F &= \frac{1}{2} \sum_i \sum_{\hat{\rho}_i} \sum_{n \in \mathbb{Z}} |\text{Im}(n\tau + \hat{\rho}_i \cdot u_H + \hat{\rho}_i \cdot \sigma + \nu_i)| \\ &+ \frac{1}{2} \sum_i \sum_{\lambda_i} \sum_{n \in \mathbb{Z}} |\text{Im}(n\tau + \lambda_i \cdot \sigma + \nu_i)|. \end{aligned} \quad (3.35)$$

Regularizing the sum, we find, as $\tau \rightarrow i\infty$,

$$\begin{aligned} \zeta^a + (\sigma \cdot \kappa + \mu \cdot \kappa^F)^a &\simeq \frac{\tau_2}{2} \sum_{i, \rho_i} \rho_i^a (\bar{\epsilon}_{\rho_i} + \text{Im}(\rho_i \cdot \sigma + \nu_i)/\tau_2) \\ &- (\bar{\epsilon}_{\rho_i} + \text{Im}(\rho_i \cdot \sigma + \nu_i)/\tau_2)^2 \\ &\simeq \frac{\tau_2}{2} \sum_{i, \rho_i} (\bar{\epsilon}_{\rho_i} - \bar{\epsilon}_{\rho_i}^2) \rho_i^a \\ &+ \frac{1}{2} \sum_{i, \rho_i} (1 - 2\bar{\epsilon}_{\rho_i}) \rho_i^a \text{Im}(\rho_i \cdot \sigma + F_i \cdot \nu), \end{aligned} \quad (3.36)$$

where $\bar{\epsilon}_{\rho_i} \equiv \{\rho_i \cdot u_H/\tau\}$ so that $\bar{\epsilon}_{\lambda_i} = 0$. Repeating the exercise for the small τ limit, we find

³Although we refer to induced 3d couplings at H -saddles here, the computation is straightforwardly extended to arbitrary holonomy values. At H -saddles, one typically considers split ρ 's to λ 's for light modes and $\hat{\rho}$'s for heavy modes, and the UV contribution comes from the latter. At generic holonomy, however, $\{\lambda\} = \emptyset$ and the contribution comes from all charged chiral multiplets.

$$\begin{aligned} &\zeta^a + (\bar{\sigma} \cdot \bar{\kappa} + \bar{\mu} \cdot \bar{\kappa}^F)^a \\ &\simeq \frac{\tilde{\tau}_2}{2} \sum_{i, \rho_i} (\bar{\epsilon}_{\rho_i} - (\bar{\epsilon}_{\rho_i})^2) \rho_i^a \\ &+ \frac{1}{2} \sum_{i, \rho_i} (1 - 2\bar{\epsilon}_{\rho_i}) \rho_i^a \text{Im}(\rho_i \cdot \bar{\sigma} + F_i \cdot \bar{\nu}), \end{aligned} \quad (3.37)$$

while $\bar{\nu} = \nu/\tau$ as with others and $\bar{\epsilon}_{\rho_i} \equiv \{\rho_i \cdot \tilde{u}_H/\tilde{\tau}\}$. We use the common symbol $\bar{\epsilon}_\rho$ on the large and the small radius limits since these two sets of numbers are really identical.

In either expressions, we can infer the UV contributions to these couplings in 3d sense, by expanding in $1/\tau_2$ ($1/\tilde{\tau}_2$) and dropping λ_i contributions in the second sums for the Chern-Simons level, e.g.,

$$\begin{aligned} \zeta_{\text{UV}}^a + (\sigma \cdot \kappa_{\text{UV}} + \mu \cdot \kappa_{\text{UV}}^F)^a &\simeq \frac{\tau_2}{2} \sum_{i, \hat{\rho}_i} (\bar{\epsilon}_{\hat{\rho}_i} - \bar{\epsilon}_{\hat{\rho}_i}^2) \hat{\rho}_i^a + \frac{1}{2} \sum_{i, \hat{\rho}_i} (1 - 2\bar{\epsilon}_{\hat{\rho}_i}) \hat{\rho}_i^a \text{Im}(\hat{\rho}_i \cdot \sigma + F_i \cdot \mu) \end{aligned} \quad (3.38)$$

with $\mu = \nu$. For the small τ limit, we take $\mu = \tilde{\nu}$ and replace σ in favor of $\bar{\sigma}$. Although the KK mode sums associated with the charge λ could have contributed to ζ_{UV} , they cancel against the same contributions from $\hat{\rho}$'s, thanks to the observation we made earlier, $\sum \rho = 0$, for each irreducible representation for any semisimple group. With this, it is clear that these are precisely the couplings responsible for the prefactor Λ_H and $\tilde{\Lambda}_H$,

$$\Lambda_H^a \simeq e^{-2\pi(\zeta_{\text{UV}} + \sigma \cdot \kappa_{\text{UV}})^a}, \quad \tilde{\Lambda}_H^a \simeq e^{-2\pi(\zeta_{\text{UV}} + \bar{\sigma} \cdot \bar{\kappa}_{\text{UV}})^a}, \quad (3.39)$$

provided that the left-hand sides are appropriately expanded in $1/\tau_2$ ($1/\tilde{\tau}_2$) and truncated to the leading order.

The FI constant ζ_{UV} must be present only for the Abelian part of the subgroup \mathcal{H} left unbroken by the holonomy u_H , which we wish to confirm as a consistency check. Clearly this would hold if $\alpha \cdot \zeta_{\text{UV}} = 0$ for any root α that belongs to the unbroken groups \mathcal{H} . Since the symmetry breaking to \mathcal{H} is due to the holonomy, this means that the irreducible representation \mathfrak{R}_i of \mathcal{G} will decompose into various spin s representations under $SU(2)_\alpha \subset \mathcal{H} \subset \mathcal{G}$, and that

$$\{\bar{\epsilon}_\rho | \rho \in \mathfrak{R}_i\} \rightarrow \{\bar{\epsilon}_l | \mathfrak{R}_i = \bigoplus_l [s_l]\}. \quad (3.40)$$

At the holonomy such that $\bar{\epsilon}_\alpha = 0$, therefore, we have

$$\alpha \cdot \sum_{\rho \in \mathfrak{R}_i} \rho (\bar{\epsilon}_\rho - \bar{\epsilon}_\rho^2) = \sum_{s_l} (\bar{\epsilon}_l - \bar{\epsilon}_l^2) \sum_{\mu \in [s_l]} \alpha \cdot (\mu + \dots), \quad (3.41)$$

where $\bar{\epsilon}_l$ denotes the common value of those $\bar{\epsilon}_\rho$'s that fall into the l th irreducible representation, say with spin s_l , of $SU(2)_\alpha$. μ 's are the weights of spin $[s_l]$ representation, embedded into those of \mathfrak{R}_i , and the ellipsis denotes the part invariant under $SU(2)_\alpha$. We thus find

$$\alpha \cdot \zeta_{\text{UV}} = \alpha \cdot \left(\sum_l (\bar{\epsilon}_l - \bar{\epsilon}_l^2) \sum_{\mu \in [s_l]} \mu \right) = 0, \quad (3.42)$$

as expected, where in the last step we again used $\sum_{\mu \in \mathfrak{R}} \mu = 0$ for any irreducible representation \mathfrak{R} of a (semi)simple Lie group. Generalization to the entire set of α with $\bar{\epsilon}_\alpha = 0$ is immediate.

Also, the Chern-Simons coefficients should be appropriately quantized. Indeed, we find the gauge Chern-Simons levels in the UV

$$\begin{aligned} \kappa_{\text{UV}}^{ab} &= \frac{1}{2} \sum_{i, \hat{\rho}_i} \hat{\rho}_i^a \hat{\rho}_i^b (1 - 2\bar{\epsilon}_{\hat{\rho}_i}) \\ &= \frac{1}{2} \sum_{i, \rho_i} \rho_i^a \rho_i^b (1 - 2\bar{\epsilon}_{\rho_i}) - \frac{1}{2} \sum_{i, \lambda_i} \lambda_i^a \lambda_i^b \\ &= \frac{1}{2} \sum_{i, \rho_i} \rho_i^a \rho_i^b (1 + 2[\rho_i \cdot u_H / \tau]) - \frac{1}{2} \sum_{i, \lambda_i} \lambda_i^a \lambda_i^b \end{aligned} \quad (3.43)$$

holds since $\bar{\epsilon}_\lambda = 0$ and since the gauge anomaly cancellation demands $\sum_{i, \rho_i} \rho_i^a \rho_i^b \rho_i^c = 0$. Recall that $[\dots]$ means the real integral part, as in (3.5). All quantities in the sums are manifestly integral, so the induced UV Chern-Simons coefficients are integral up to the overall factor 1/2. Exactly the same applies to $\tilde{\kappa}_{\text{UV}}$.

The factor 1/2, which some may find troublesome, is not a problem at all. For many theories, such as the SQCD type where the fundamental chirals has to appear in pairs, we expect that this is countermanded by the 4d spectrum. More to the point, the half-quantized Chern-Simons coefficient is usually an indication that we must be more careful about the effective action coming from integration out of massive fermions. The usual statement that this leads to Chern-Simons action is known to miss the global structure of the effective action; whenever the Chern-Simons coefficient generated is half-integral, and thus apparently variant under large gauge transformation, the effective action is actually an eta-invariant with full gauge invariance [44,45].

D. Locating H -saddles with pure gauge sectors

So far we have pretended that the H -saddle would come with charge vectors $\{\lambda | \bar{\epsilon}_\lambda = 0\}$, enough of them to span the entire charge vector space. However, this needs not be the case in general, as one can have 3d theories with unbroken supersymmetry when the Chern-Simons is nontrivial. Also a further issue arises when the unbroken gauge group \mathcal{H} contains a $U(1)$ factor with a UV FI constant generated. In the latter cases, some topological vacua may appear shifted far away from σ , $\tilde{\sigma} \sim O(1)$, potentially muddying the classification of the H -saddles. Here we wish to address issues related to such H -saddles.

Consider the case where \mathcal{H} contains no Abelian sector. Let us write

$$\mathcal{H} = \dots \oplus \mathcal{K} \oplus \dots, \quad (3.44)$$

where \mathcal{K} is a semisimple Lie group with no light 3d chiral multiplet coupled. Where would such an unbroken group be found? Recall that the location of the H -saddle was determined, so far, by the condition

$$\bar{\epsilon}_\lambda = 0 \quad (3.45)$$

for some subset of matter charges, λ 's. In this current case, no such λ knows about \mathcal{K} . Instead the location of the H -saddle is determined by the spontaneous symmetry breaking as

$$\bar{\epsilon}_\alpha = 0, \quad \alpha \in \mathfrak{K}, \quad (3.46)$$

where \mathfrak{K} is the Lie algebra of \mathcal{K} . When \mathcal{H} contains no Abelian subgroup, this combination of $\{\alpha\} \cup \{\lambda\}$ should span the entire weight space of \mathfrak{G} , the Lie algebra of \mathcal{G} , and again determine the acceptable positions u_H discretely.

In the absence of light matter fields coupled to \mathcal{K} , and since no UV FI constant would exist for such non-Abelian group, the 3d supersymmetric vacua in question are all “topological” types [29]. For $\mathcal{K} = Sp(r)$, for example, one can take a simple basis for the Cartan $U(1)^r$ such that $\sigma = \sum_1^r \sigma_s \mathcal{C}_s$ with chiral fields in the defining representations have unit charges with respect to one and only one \mathcal{C}_s . As such, the reduced BAE will take the simple form, after some rescaling

$$1 = (z_s)^{2\kappa_{\text{UV}}^{Sp(r)}}, \quad s = 1, \dots, r. \quad (3.47)$$

We remove solutions with $z_s = \pm 1$ for some s or those with $z_s = z_t$ for some $s \neq t$, and identify those related by Weyl transformations, $W = S_r \times (\mathbb{Z}_2)^r$. Then the vacua are labeled by unordered distinct r phases, $e^{\pi i n / \kappa_{\text{UV}}^{Sp(r)}}$, with $1 \leq n < \kappa_{\text{UV}}^{Sp(r)}$

$$\binom{|\kappa_{\text{UV}}^{Sp(r)}| - 1}{r}. \quad (3.48)$$

Similarly, for $\mathcal{K} = SU(r+1)$, a simple choice is $\sigma = \sum_1^r \sigma_s \mathcal{C}_s - (\sum_s \sigma_s) \mathcal{C}_{r+1}$ in the redundant basis, whereby the reduced BAE equation becomes

$$1 = (z_1/z_{r+1})^{\kappa_{\text{UV}}^{SU(r+1)}} = \dots = (z_r/z_{r+1})^{\kappa_{\text{UV}}^{SU(r+1)}} \quad (3.49)$$

with $z_{r+1} \equiv [\prod_1^r z_s]^{-1}$ understood. Each equation yields $|\kappa_{\text{UV}}^{SU(r+1)}| - 1$ acceptable solutions, $z_s/z_{r+1} \neq 1$, upon which we further impose $z_s/z_{r+1} \neq z_t/z_{r+1}$ for all pairs $s < r$ as well. In the end, the number of acceptable Weyl-inequivalent solutions is, again

$$\binom{|\kappa_{UV}^{SU(r+1)}| - 1}{r}. \quad (3.50)$$

These dovetail precisely with the 3d index computation by Witten [46] once we extend the latter to $\mathcal{N} = 2$; the only new ingredient for $\mathcal{N} = 2$ is to take $\kappa' = \kappa_{UV} - h$, where h is the dual Coxeter number, instead of $\kappa' = \kappa_{UV} - h/2$, for the bosonic theory in the end, since the adjoint fermion content is doubled between the two. In both cases, therefore, a necessary condition for the existence of an H -saddle involving a pure $Sp(r)$ sector or a pure $SU(r+1)$ sector is $\kappa_{UV} \geq r + 1$.

Now let us allow \mathcal{K} to include an Abelian factor. Unbroken $U(1)$'s can come with large 3d FI constants as we saw in the previous section, which will interfere with reduction of 4d BAE to 3d BAE. If one started with 4d chiral multiplets in real or pseudoreal representations, such FI constants would cancel out exactly, but of course this need not be the case. If one finds large FI constants, say at some u_H , that scale as $\text{Im}\tau$ (or $\text{Im}\tilde{\tau}$), the reduced 3d BAE of such $U(1) \subset \mathcal{K}$ cannot be solved for σ (or $\tilde{\sigma}$) kept finite. What this really means is that H -saddles must be looked for, with the condition of $\zeta_{UV} = 0$ imposed simultaneously, i.e.,

$$\zeta^s \propto \sum_{i,\hat{\rho}_i} \hat{\rho}_i^s (\bar{e}_{\hat{\rho}_i} - \bar{e}_{\hat{\rho}_i}^2) = \sum_{i,\rho_i} \rho_i^s (\bar{e}_{\rho_i} - \bar{e}_{\rho_i}^2) \rightarrow 0. \quad (3.51)$$

Recall that for this case, neither a matter charge λ nor a root α can be invoked to fix the location of u^a . Instead, we have this $\zeta^s = 0$ condition, again fixing the holonomy u_H to discrete possibilities.

Once this necessary condition is met and the candidate location for the H -saddle is found, we must decide whether such an H -saddle will actually contribute, i.e., whether the reduced 3d BAE admits nontrivial vacua nearby. Because we are dealing with pure $\mathcal{N} = 2$ gauge theory in three dimensions, the latter is possible only if $\kappa_{UV} \neq 0$. At a saddle with decoupled $U(1)_a$ unbroken group the actual supersymmetric vacua are determined by the 3d BAE,

$$C_s = (z_s)^{\kappa_{UV}^s} \quad (3.52)$$

for some finite constant $C_{U(1)}$, so the number of them is

$$|\kappa_{UV}^s|. \quad (3.53)$$

Therefore, we have found that an H -saddle involving a decoupled $U(1)_s$ gauge sector is possible provided that $\zeta_{UV}^s = 0$ and $\kappa_{UV}^s \neq 0$.

The question of H -saddles with $\mathcal{K} = U(1) \subset \mathcal{H}$ but now with light 3d charged matter field, a general case of (2) above is a little more involved. Suppose we located a candidate H -saddle using a condition of type $\bar{e}_\lambda = 0$ for some $U(1)$ -coupled charge vector λ . A schematic form of the rank 1 3d BAE at such an H -saddle is

$$1 \simeq q^\xi z^{\kappa - \sum Q^2/2 + \sum (Q')^2/2} \frac{\prod_Q (yz^Q - 1)^Q}{\prod_{Q'} (z^{Q'} - y)^{Q'}}, \quad (3.54)$$

or

$$\prod_{Q'} (z^{Q'} - y)^{Q'} \simeq q^\xi z^{\kappa - \sum Q^2/2 + \sum (Q')^2/2} \prod_Q (yz^Q - 1)^Q, \quad (3.55)$$

where $-Q$ and Q' denote, collectively, the light charges of negative and positive signs respectively, and $\xi \equiv \zeta_{UV}/\text{Im}\tau$ and $\kappa = \kappa_{UV}$ or $\xi \equiv \tilde{\zeta}_{UV}/\text{Im}\tilde{\tau}$ and $\kappa = \tilde{\kappa}_{UV}$, in the large or in the small τ limits, respectively. Here, we will consider a large τ limit, or $q \rightarrow 0$, without loss of generality.

Suppose that $\xi > 0$. For $\xi < 0$, we get the same result after flipping $Q \leftrightarrow Q'$ and $\kappa \leftrightarrow -\kappa$. If $\xi = 0$, there is no issue, to begin with, as all solutions would be $O(1)$ and do not scale with q . Setting $q = 0$ for the moment, we find Q' finite solutions $z \sim y^{1/Q'}$ to (3.55), each of which are Q' times degenerate. As we turn back on small q , these would split but remain finite. To enumerate the other worrisome solutions that scale with q or $1/q$, it is useful to define $k \equiv \kappa - \sum Q^2/2 + \sum (Q')^2/2$, $l \equiv \kappa + \sum Q^2/2 - \sum (Q')^2/2$. We then find,

- (i) $k \geq 0, l > 0$,
 l large solutions $z \sim q^{-\xi/l}$;
 the total number of solutions are $\sum (Q')^2 + l = \kappa + \sum Q^2/2 + \sum (Q')^2/2$;
- (ii) $k \geq 0, l \leq 0$,
 no new solutions;
 the total number is $\sum (Q')^2$;
- (iii) $k < 0, l \leq 0$,
 $-k$ small solutions $z \sim q^{\xi/|k|}$;
 the total number is $\sum (Q')^2 - k = -\kappa + \sum Q^2/2 + \sum (Q')^2/2$;
- (iv) $k < 0, l > 0$,
 l large solutions $z \sim q^{-\xi/l}$ and $-k$ small solutions $z \sim q^{\xi/|k|}$;
 the total number is $\sum (Q')^2 + l - k = \sum Q^2$.

Among these vacua, the large and the small ones $z \sim q^\#$ should be taken only as a qualitative indication that somewhere far away there exist supersymmetric and topological vacua of free $U(1)$ Chern-Simons theory. The truncation to (3.54) is justified only at finite values of σ , and thus precise locations of these additional vacua should be worked out by going back to the 4d BAE. Why is the reduction to 3d theory, which has worked flawlessly so far, compromised? Simply because the dimensional reduction ends up with 3d FI constant which still remembers the large value of τ and thus the fact that the purported 3d theory came from 4d theory with the extremely elongated T^2 . The number of vacua found for such $U(1)$ factor in the preceding analysis should still hold, but the precise

locations of those at $z \sim q^\#$ are not to be trusted. Rather, one must really view this situation as a sum of distinct H -saddles consisting of two types. One is $U(1)$ theory with charge matters, but with its topological vacua due to very large FI constant excised. The others are free $U(1)$ theory elsewhere in the u -space, with no light matter field coupled and supersymmetric vacua, due to Chern-Simons level k or l , as in (3.52).

To summarize, locating H -saddles involves three sets of data,

$$\bar{\epsilon}_\lambda = 0, \quad \bar{\epsilon}_\alpha = 0, \quad \zeta_{\text{UV}} = 0, \quad (3.56)$$

and one proceeds by collecting at least rank-many conditions to fix discrete locations in the Cartan torus spanned by u 's. In particular, when we end up a $U(1)$ factor coupled with charged 3d matter and large UV FI constant $\zeta \sim \text{Im}\tau$, $\text{Im}\bar{\tau}$, we must take care to discard the far-away topological vacua $\sigma \sim \zeta$ from such an H -saddle and instead look for a nearby saddle with a free $U(1)$ factor at vanishing FI constant and nonvanishing Chern-Simons level. With finite chemical potentials and a matter content symmetric under the charge conjugation, this latter complication never appears. On the other hand, such $U(1)$ cases will be more typical in the large radius limit for the so-called ‘‘physical’’ version, regardless of matter content, because of the large $U(1)_R$ chemical potential. Next, we now move to this last type of H -saddles.

E. H -saddles with large chemical potentials

As we hinted already, the large radius limit of the ‘‘physical’’ $S^1 \times S^3$ partition function deviates a little from the main story of this paper. Apart from why this has to be so from the viewpoint of how these objects are constructed, we can also trace the difference at a mathematical level to the large $U(1)_R$ chemical potential $(r-1)\tau$. The latter shifts the argument of various operators by a large amount in the large τ limit, common for each chiral multiplets in a single irreducible gauge representation.

With nonintegral r 's, in particular, necessary at the superconformal point, one immediate consequence is that, even if the 4d theory came with charge-conjugation symmetric gauge representation, the light 3d field content, if any, would not be generically so; the positively charged matter and the negatively charged matter would become light at different holonomies. At candidate H -saddles, one will typically encounter unbroken $U(1)$ gauge theories with large uncanceled ζ_{UV} , which shifts the location of the saddle to far away, and makes the search for genuine H -saddle qualitatively different from the other cases. For this reason, we will denote these rather distinct H -saddle values of u 's by introducing the notation \hat{u}_H .

One obvious place to look for a saddle, independent of details, is $\hat{u}_H = 0$. Here the 4d gauge group \mathcal{G} will descend to 3d intact, while chiral multiplets with typical values of r_i

will become all massive. As such no FI constant would be generated, as the 3d gauge group $\mathcal{H} = \mathcal{G}$ would have no $U(1)$ factor. For a qualitative understanding, we will confine our attention to theories with a single classical Lie group \mathcal{G} as the gauge group and further assume that $0 < r_i < 1$ for all matter multiplets. Is there an H -saddle located at the naive choice $\hat{u}_H = 0$?

The pure \mathcal{G} Yang-Mills-Chern-Simons theory there would have no supersymmetric vacua unless one finds a sufficiently large UV Chern-Simons level, which can be easily computed as

$$\kappa_{\text{UV}}^{ab} = \delta^{ab} \gamma_{\mathcal{G}} T_{\text{def}}^{(2)} \kappa_{\text{UV}}^{\mathcal{G}}, \quad (3.57)$$

where $T_{\text{def}}^{(2)}$ is that of the defining representation and $\gamma_{SO} = 1/2$ and $\gamma_{SU} = \gamma_{Sp} = 1$.⁴ Then, we find

$$\kappa_{\text{UV}}^{\mathcal{G}} = \frac{1}{2\gamma_{\mathcal{G}} T_{\text{def}}^{(2)}} \sum_i T_i^{(2)} (1 - 2r_i). \quad (3.58)$$

On the other hand, the Adler-Bell-Jackiw anomaly cancellation requires

$$T_{\text{adj}}^{(2)} = \sum_i T_i^{(2)} (1 - r_i), \quad (3.59)$$

so that

$$\begin{aligned} \kappa_{\text{UV}}^{\mathcal{G}} &= \frac{T_{\text{adj}}^{(2)}}{\gamma_{\mathcal{G}} T_{\text{def}}^{(2)}} \times \sum_i \frac{T_i^{(2)}}{T_{\text{adj}}^{(2)}} (1/2 - r_i) \\ &= \frac{T_{\text{adj}}^{(2)}}{\gamma_{\mathcal{G}} T_{\text{def}}^{(2)}} \cdot \left(1 - \frac{3}{2} \cdot \frac{\sum_i T_i^{(2)}}{3T_{\text{adj}}^{(2)}} \right). \end{aligned} \quad (3.60)$$

Note that the second term inside the parentheses cannot exceed $3/2$, once we demand the asymptotic freedom. If the theory contains a single type of chiral multiplets, the asymptotic freedom combined with $0 < r$ implies

$$|\kappa_{\text{UV}}^{\mathcal{G}}| < \frac{T_{\text{adj}}^{(2)}}{2\gamma_{\mathcal{G}} T_{\text{def}}^{(2)}} = h_{\mathcal{G}}, \quad (3.61)$$

with the dual Coxeter number $h_{\mathcal{G}}$. Recall that the counting of 3d vacua for pure YMCS theories is dictated by the difference between $|\kappa|$ and h . This leads us to suspect that, for all asymptotically free theories that flow to CFT, the naive $\hat{u}_H = 0$ saddle is absent.

For $SU(N_c)$ theory with N_f fundamental and N_f antifundamental chirals, e.g., (3.58) gives, since $r = 1 - N_c/N_f$ by the ABJ anomaly cancellation,

⁴For actual vacuum counting via BAE, however, we restrict ourselves at most to SU and Sp cases: See the top of Sec. II.

$$\kappa_{\text{UV}} = 2N_c - N_f \quad (3.62)$$

for 4d conformal field theories, $3N_c/2 \leq N_f < 3N_c$. With

$$|\kappa_{\text{UV}}| < h_{SU(N_c)} = N_c, \quad (3.63)$$

(3.50) tells us that the naive saddle at the origin, $\hat{u}_H = 0$, has no supersymmetric vacua and thus is not an H -saddle. For SQCD theories, the saddle at origin is actually absent. Similar considerations for $Sp(r)$ theory with $2N_f$ fundamental flavors show that, again there is no H -saddle at $\hat{u}_H = 0$ for asymptotically free theories $N_f \leq 3r + 2$; for $N_f = 3r + 2$, one finds $\kappa = -r$.

With no H -saddle at the origin, the next places are those holonomies with vanishing UV FI constants,

$$\sum \rho^a (\hat{\epsilon}_\rho - \hat{\epsilon}_\rho^2) = 0. \quad (3.64)$$

Suppose that the matter content is symmetric under charge conjugation, such that charge vectors always come in pairs $(\rho, -\rho)$. Then, places where this happens generically are

$$\rho \cdot \hat{u}_H \in \tau\mathbb{Z}/2, \quad (3.65)$$

which allows $\epsilon_{\hat{\rho}} = \epsilon_{-\hat{\rho}}$ and thus pairwise cancellations in the sum (3.64). Assuming $r_i \neq 1/2$, the theory reduces to pure Yang-Mills type and the Chern-Simons level is

$$\kappa_{\text{UV}}^{ab} = \delta^{ab} \gamma_G T_{\text{def}}^{(2)} \kappa_{\text{UV}}^G + \sum_{i, \rho_i} \rho_i^a \rho_i^b [\rho_i \cdot \hat{u}_H / \tau + r_i], \quad (3.66)$$

where κ_{UV}^G is the Chern-Simons level at $\hat{u}_H = 0$ as in (3.60). Coming back to $SU(2)$ theory with $2N_f$ fundamental flavors, we find that reduced theory is a pure $SU(2)$ YMCS with

$$\kappa_{\text{UV}} = -2, 4, \quad \text{for } N_f = 3, 5, \quad (3.67)$$

implying 1 and 3 BAE vacua, respectively, which are consistent with the Witten index of the original 4d theories.

$N_f = 4$ with $r = 1/2$ at SCFT also admit $\hat{u}_H = \tau/2$ as an H -saddle; the reduced theory is an $SU(2)$ theory with $2N_f = 8$ fundamental chirals and vanishing UV Chern-Simons level. The number of vacua for this 3d theory is usually expected to be three. However, the actual theory at this H -saddle, being a reduction from 4d where the baryonic $U(1)$ is anomalous, and, because this theory cannot have UV FI constant, one of these potential vacua is pushed to the Coulombic infinity. The number of vacua at the $\hat{u}_H = \tau/2$ saddle is actually 2 which is again consistent with the 4d Witten index. These suggest that for SQCD theories, $\hat{u}_H = \tau/2$ is the only H -saddle in the Casimir limit.

IV. 4D THEORY AS A DISJOINT SUM OF 3D THEORIES

These discussions lead us to a clear definition of the H -saddle for general supersymmetric gauge theories on a compact spacetime with a small circle or a small circle bundle. In the small radius limit, a d dimensional partition function Ω_d will reduce to a sum of $(d-1)$ dimensional partition functions, \mathcal{Z}_{d-1}^H ,

$$\Omega_d \rightarrow \sum_{u_H} c_H \mathcal{Z}_{d-1}^H \quad (4.1)$$

labeled by some discrete choices of the holonomy including the trivial one. The prefactors c_H capture contributions from the Kaluza-Klein towers as well as massive multiplets. For the partition functions we have been studying,

$$\Omega_4^{g; p_1, p_2} = \sum_{u_*} \mathcal{H}^{g-1} \mathcal{F}_1^{p_1} \mathcal{F}_2^{p_2} \quad (4.2)$$

with the two circles (p_1, p_2) -fibered over a genus g surface, we find that this decomposes, both in the large and in the small τ limits,

$$\Omega_4^{g; p_1, p_2} \rightarrow \sum_{u_H} \left(\sum_{\sigma_*} \mathcal{H}^{g-1} \mathcal{F}_1^{p_1} \mathcal{F}_2^{p_2} \right), \quad (4.3)$$

where the 4d BAE vacua are reorganized into sets of 3d BAE vacua for mutually disjoint 3d theories at various H -saddles. As we already emphasized, we find such decomposition even in the large τ limit, because, in effect, this is equivalent to a small radius limit of circle 1.

For $g \neq 1$, we can reorganize the sum over 3d vacua σ_* at each u_H , in terms of the 3d BAE partition function, \mathcal{Z}_3^H and a multiplicative factor c_H . The latter will generically have an exponential behavior, interpreted as the Cardy exponent in the small τ limit and as the Casimir energy in the large τ limit. Previous estimates of such leading exponents have effectively considered only the naive saddle at $u_H = 0$ [20,22]. The results from such computations, interpreted as being related to 4d anomaly polynomials, must be therefore questioned.

A. 4d Witten index is a sum of 3d Witten indices

Before we plunge into quantitative studies, it is worthwhile to consider the simplest case of $p_{1,2} = 0$ and $g = 1$. For T^4 , the BAE computes the numerical Witten index, with the summand at each u_* equal to 1. This means that we have an intuitive relation

$$\mathcal{I}_4^G = \sum_{u_H} \mathcal{I}_3^H, \quad (4.4)$$

whereby the 4d Witten is reconstructed from those of several 3d theories sitting at distinct holonomies. Regardless of the details of computations to follow for

different g 's and p 's, this by itself tells us that the small radius limit of 4d theory cannot be regarded as a single 3d theory. If we are considering supersymmetric theories in compact spacetime, therefore, the 4d theory in the small radius limit should be considered as a disjoint sum of 3d theories.

Since the same set of u_H 's enters such decompositions for all $\mathcal{M}_4^{g;p_1,p_2}$'s, this also means that an H -saddle will occur if and only if the reduced 3d theory there has a nontrivial Witten index. The latter condition can be considered as the most important single property of H -saddles. The class of theories we are considering in this paper are maximally mass deformed by flavor holonomies so that the partition functions and Witten indices are all integral. As such, an H -saddle would occur if and only if the Witten index of the reduced 3d theory at the candidate holonomy is nonvanishing.

On the other hand, the notion of an H -saddle clearly goes beyond the particular class of theories or background geometries we are considering in this paper. More generally twisted partition functions are often not enumerative, resulting in nonintegral twisted partition functions. As we recalled in the Introduction, a twisted partition function would generally compute the analog of the ‘‘bulk index.’’ In such cases, the defining property of the H -saddle should be extended to allow a nonvanishing supersymmetric partition function of the reduced theory at the candidate holonomy.

If we were considering the 4d theory on $S^1 \times \mathbb{R}^3$, the holonomies would label superselection sectors; the dimensional reduction process is ambiguous until we specify the holonomy or compute the vacuum expectation value of the holonomy. The above relation tells us that there are discrete choices of u_H whereby the dimensional reduction produces distinct 3d theories whose 3d supersymmetry is not spontaneously broken, and that the 4d Witten index is reproduced only after we sum over the Witten indices of these 3d theories at distinct u_H 's.

Such a behavior of 4d theory on a circle, producing multiple 3d theories in the small radius limit, has been noted previously by Seiberg and collaborators while studying how 4d dualities reduce to 3d dualities [13]. As the above relation shows, a dual pair of 4d theories would produce, each, several 3d theories which must be collectively dual to each other. Whether or not this implies individual 3d dualities, say, in our language at H -saddle pairs, is in principle another matter. For 4d theory as a starting point, however, the interpretation of u_H as the superselection sector label does suggest that the duality will hold for 3d theories pairwise, or in our terminology, H -saddle by H -saddle.

The robust nature of the Witten index under small deformations is often invoked to simplify index computations. One such would be the insensitivity to the size of the circles in T^4 , but this, if used improperly, seems to imply that Witten indices agree between theories in different

dimensions if one is a dimensional reduction of the other. However, we already know, via many examples, that this is not quite correct. For instance, Witten pointed out how the index of 4d $\mathcal{N} = 1$ pure Yang-Mills is sensitive to disconnected sectors of mutually commuting holonomies along T^3 [40]; such sectors would be dropped if one or more radii of T^3 is taken to zero literally.

Our relation is yet another reminder that such topological invariance argument should not be taken too far. The problem with the zero radius limit of a spacetime circle is that the compact space of holonomies becomes non-compact in a zero radius limit, and cannot be considered a small deformation. The above formula, which is far more general than the particular class of theories here and the partition functions thereof, gives a neat way to relate Witten indices of gauge theories in adjacent dimensions.

We close with two simple examples. The first is the canonical SQCD, namely $SU(N)$ theories with N_f fundamental and N_f antifundamental chirals. For these, it is clear that the only H -saddle is the one at $u_H = 0$, hence we have

$$\mathcal{I}_4^G = \mathcal{I}_3^H|_{u_H/\tau=0}, \quad (4.5)$$

where the reduced H theory at $u_H/\tau = 0$ has the same gauge group and the same chiral multiplet content as its 4d cousin G . Indeed, Closset *et al.* [19] found

$$\mathcal{I}_4^G = \mathcal{I}_3^H|_{u_H/\tau=0} = \binom{N_f - 2}{N - 1}. \quad (4.6)$$

Since the matter representation is real collectively, neither the FI constant nor the Chern-Simon level arise at UV.

The other, less trivial, example is an $SU(2)$ theory with two fundamental chirals and two adjoint chirals. For this, a nontrivial H -saddle at $u_H/\tau = 1/2$ is present as well as the naive one at $u_H = 0$. While we are formulating things via the large radius limit, the small radius limit is found, verbatim, by replacing the variables to the tilded ones with the identical result. As such we have

$$\mathcal{I}_4^G = \sum_{u_H/\tau=0,1/2} \mathcal{I}_3^H \quad (4.7)$$

where the 3d theory at $u_H/\tau = 1/2$ has the two adjoint chirals only. Again no UV 3d coupling is generated at either saddle, and 3d BAE vacua can be counted straightforwardly. We find

$$\mathcal{I}_4^G = \mathcal{I}_3^H|_{u_H/\tau=0} + \mathcal{I}_3^H|_{u_H/\tau=1/2} = 8 + 6 = 14. \quad (4.8)$$

The main feature of the latter example, relative to the first, is a chiral multiplet with gauge representation beyond the fundamental one. In fact, the existence of a chiral multiplet in a gauge representation larger than the defining one, for

classical ones at least, is one obvious criterion for a nontrivial H -saddle to exist.

Note that, of these, the first example is not compatible with a general A-twist background, since the anomaly-free $U(1)_R$ charge is not integral. For $\Sigma_g = T^2$, however, the A-twist is null, so we do not need to restrict the $U(1)_R$ charge to be integral. And as long as one can find non-anomalous $U(1)_R$, its chemical potential can be turned on. For this reason, this recursive computation of the Witten index can be used for more examples of theories than generic geometries of this class would allow. Of course, this is up to the major caveat that theories being considered are all equipped with real masses in the 3d sense, as is a common downside of the BAE formulation. One must take care, in general, not to confuse the Witten index computed this way with those of the vanilla 4d $\mathcal{N} = 1$ theories on R^4 .

We close this subsection with a caveat. In relating 4d theories to one or more 3d theories, obtained by dimensional reduction at such saddles, we are always speaking of the small radius limit. This means that constraints from the 4d anomaly, for example, should be considered valid even in the said 3d limit. One example is the $SU(N_c)$ SQCD, whose strict 3d form allows an extra $U(1)$ flavor symmetry which would be anomalous in 4d. Our 3d theories at H -saddles are the ones without such a global symmetry; this affects the allowed superpotential and hence the 3d Witten index as well.

B. Asymptotics

Now let us turn to other, more involved partition functions. In the literature, some 4d partition functions have been discussed with a particular interest on their asymptotic behavior [16–18,20–22]. Especially, Ardehali first observed the influence of the holonomy on the Cardy limit of the superconformal index [16], i.e., the partition function on $S^3 \times S^1$, which is later extended to more general manifolds by Di Pietro and Honda [17]. The latter discussed the Cardy limit of the $\mathcal{M}_3 \times S^1$ partition function, with explicit examples $\mathcal{M}_3 = L(n, 1)$, $\Sigma_g \times S^1$.

On the other hand, for the Casimir limit, the role of the holonomy is rarely discussed as far as we are aware. In this section, we provide a unified way of examining both the Cardy limit and the Casimir limit of the partition function, which manifests itself in the H -saddle approach.

1. Asymptotics of \mathcal{H}

The handle-gluing operator is, with r_i being the $U(1)_R$ charge of the i th chiral multiplet,

$$\mathcal{H} \equiv \eta(\tau)^{-2\text{rank}(\mathcal{G})} \prod_{\alpha} \Psi(\alpha \cdot u; \tau) \prod_i \prod_{\rho_i} \Psi(\rho_i \cdot u + \nu_i; \tau)^{r_i-1} \times \det \left[\frac{\partial_a \log \Phi_b}{2\pi i} \right], \quad (4.9)$$

of which the last piece can, at most, contribute logarithmic corrections in the exponent. The large and the small τ limit of Ψ 's were already explored. These may be combined to, for the gauge multiplet contributions, at each H -saddle,

$$\eta(\tau)^{-2\text{rank}(\mathcal{G})} \prod_{\alpha} \Psi(\alpha \cdot u; \tau) \Big|_{\tau \rightarrow i\infty} \sim q^{-\dim(\mathcal{G})/12} e^{\pi i \tau \sum_{\alpha} \epsilon_{\alpha} (1 - \epsilon_{\alpha})}, \quad (4.10)$$

where $\epsilon_{\alpha} = \{\alpha \cdot (u_H + \sigma) / \tau\}$ are real numbers between 0 and 1,

$$\eta(\tau)^{-2\text{rank}(\mathcal{G})} \prod_{\alpha} \Psi(\alpha \cdot u; \tau) \Big|_{\tau \rightarrow i0^+} \sim \tilde{q}^{-\dim(\mathcal{G})/12} e^{\pi i \tilde{\tau} \sum_{\alpha} \tilde{\epsilon}_{\alpha} (1 - \tilde{\epsilon}_{\alpha})}, \quad (4.11)$$

where $\tilde{\epsilon}_{\alpha} = \{\alpha \cdot (\tilde{u}_H + \tilde{\sigma}) / \tilde{\tau}\}$ are real numbers between 0 and 1. The chiral multiplet contributions can be written similarly as

$$\prod_i \prod_{\rho_i} \Psi(\rho_i \cdot u + \nu_i; \tau)^{r_i-1} \Big|_{\tau \rightarrow i\infty} \sim q^{-\sum_i \sum_{\rho_i} (r_i-1)/12} e^{\pi i \tau \sum_i (r_i-1) \sum_{\rho_i} \epsilon_{\rho_i} (1 - \epsilon_{\rho_i})}, \quad (4.12)$$

and

$$\prod_i \prod_{\rho_i} \Psi(\rho_i \cdot u + \nu_i; \tau)^{r_i-1} \Big|_{\tau \rightarrow i0^+} \sim \tilde{q}^{-\sum_i \sum_{\rho_i} (r_i-1)/12} e^{\pi i \tilde{\tau} \sum_i (r_i-1) \sum_{\rho_i} \tilde{\epsilon}_{\rho_i} (1 - \tilde{\epsilon}_{\rho_i})}, \quad (4.13)$$

where $\epsilon_{\rho_i} = \{\rho_i \cdot (u_H + \sigma) + \nu_i / \tau\}$ and $\tilde{\epsilon}_{\rho_i} = \{\rho_i \cdot (\tilde{u}_H + \tilde{\sigma}) + \tilde{\nu}_i / \tilde{\tau}\}$ are also real numbers between 0 and 1.

2. Asymptotics of \mathcal{F}_1

The first fibering operator is given by

$$\mathcal{F}_1 = \prod_i \prod_{\rho_i} \Xi_1(\rho_i \cdot u + \nu_i; \tau) = \prod_i \prod_{\rho_i} e^{2\pi i \left(\frac{(\rho_i \cdot u + \nu_i)^3}{6\tau^2} - \frac{\rho_i \cdot u + \nu_i}{12} \right)} \Gamma_0(\rho_i \cdot u + \nu_i; \tau), \quad (4.14)$$

where

$$\Gamma_0(u; \tau) = \prod_{n=0}^{\infty} \left(\frac{1 - x^{-1} q^{n+1}}{1 - x q^{n+1}} \right)^{n+1}. \quad (4.15)$$

To find the large radius limit of \mathcal{F}_1 , again we decompose $\rho_i \cdot u + \nu_i$ into $(\epsilon_{\rho_i} + m_{\rho_i})\tau$ where ϵ_{ρ_i} belongs in the range $0 \leq \epsilon_{\rho_i} < 1$ and m_{ρ_i} is the integer part. Using

$$\Xi_1(u + m\tau; \tau) = e^{-\frac{m}{2}(m^2+m)} \Psi(u; \tau)^{-m} \Xi_1(u; \tau) \quad (4.16)$$

for an integer m , one can find the large radius limit of \mathcal{F}_1 as follows:

$$\mathcal{F}_1|_{\tau \rightarrow i\infty} \sim \prod_i \prod_{\rho_i} e^{\pi i \tau (\frac{\epsilon_i^3}{3} + \epsilon_i^2 m_{\rho_i} - \epsilon_i m_{\rho_i} - \frac{\epsilon_i \rho_i}{6} + \frac{m_{\rho_i}}{6})}, \quad (4.17)$$

with $\rho_i \cdot u + \nu_i = (\epsilon_{\rho_i} + m_{\rho_i})\tau$.

The identity (4.16) also resolves an apparent puzzle with this asymptotic formula. Note that under a large gauge transformation u_a can be shifted to $u_a + \tau$. This will induce shift of both ϵ_ρ 's and m_ρ 's, under which the exponent of (4.17) does not look particularly invariant. Let us first look at how \mathcal{F}_1 transforms. Since $\rho \cdot u$ will shift by $\rho_a \tau$, the transformation is

$$\mathcal{F}_1 \rightarrow \left[\prod_i \prod_{\rho_i} (-1)^{\rho_i^a} \Psi^{-\rho_i^a} \right] \times \mathcal{F}_1 = \Phi_a^{-1} \times \mathcal{F}_1, \quad (4.18)$$

where we used $\sum_\rho \rho^a = 0$ for each irreducible representations, as was shown in Sec. III.2. What this formula tells us is that although the flux operator \mathcal{F}_1 is not invariant as a function of u under such large gauge transformations, its values at supersymmetric vacua, where $1 = \Phi_a$, are invariant. Therefore, although the leading exponent in (4.17) may look odd, its values at H -saddles are really invariant under $u_a \rightarrow u_a + \tau$. The same kind of invariance will work for \mathcal{F}_2 under $u_a \rightarrow u_a + 1$, for the small τ limit, as the two are related by S -transformation.

On the other hand, the small radius limit of \mathcal{F}_1 can be obtained using the S -transformation. First note that Ξ_1 satisfies an identity [19]

$$\Xi_1(u; \tau) = e^{\frac{\pi i u^3}{\tau^3}} \Xi_2\left(\frac{u}{\tau}; -\frac{1}{\tau}\right), \quad (4.19)$$

where Ξ_2 is defined by

$$\Xi_2(u; \tau) = e^{2\pi i (\frac{u^3}{6\tau} - \frac{u^2}{4} + \frac{u}{12} + \frac{1}{24})} \prod_{k=0}^{\infty} \frac{f(u + k\tau)}{f(-u + (k+1)\tau)}, \quad (4.20)$$

and

$$f(u) = \exp\left[\frac{1}{2\pi i} \text{Li}_2(e^{2\pi i u}) + u \log(1 - e^{2\pi i u})\right]. \quad (4.21)$$

Ξ_2 satisfies

$$\Xi_2(u + m\tau; \tau) = e^{\frac{\pi i m}{6}} \Xi_2(u; \tau), \quad (4.22)$$

and, as a result, Ξ_1 can be written as

$$\Xi_1(\rho_i \cdot u + \nu_i; \tau) = e^{-\pi i \tilde{\tau}^2 (\frac{\epsilon_{\rho_i} + \tilde{m}_{\rho_i}}{3} + \frac{\pi i (\tilde{m}_{\rho_i})}{6})} \Xi_2(\tilde{\epsilon}_{\rho_i} \tilde{\tau}; \tilde{\tau}), \quad (4.23)$$

where $\rho_i \cdot \tilde{u} + \tilde{\nu}_i$ is decomposed into $(\tilde{m}_{\rho_i} + \tilde{\epsilon}_{\rho_i})\tilde{\tau}$ with $0 \leq \tilde{\epsilon}_{\rho_i} < 1$ and an integer \tilde{m}_{ρ_i} . The cubic phase term will

vanish after summed over all the multiplets due to the anomaly-free condition. $f(u)$ comes from the 1-loop determinant of a chiral multiplet on S^3 and converges to 1 for large u ,

$$f(u)|_{u \rightarrow i\infty} = 1. \quad (4.24)$$

Thus, \mathcal{F}_1 has the following asymptotic behavior for $\tau \rightarrow i0^+$:

$$\mathcal{F}_1|_{\tau \rightarrow i0^+} \sim \prod_i \prod_{\rho_i} e^{\pi i \tau^2 (\frac{\epsilon_i^3}{3} - \frac{\epsilon_i^2}{2} + \frac{\epsilon_i \rho_i}{6})}. \quad (4.25)$$

3. Asymptotics of \mathcal{F}_2

The second fibering operator \mathcal{F}_2 is given by

$$\mathcal{F}_2 = \prod_i \prod_{\rho_i} \Xi_2(\rho_i \cdot u + \nu_i) \quad (4.26)$$

$$= \prod_i \prod_{\rho_i} e^{2\pi i (\frac{\rho_i \cdot u + \nu_i}{6\tau} - \frac{(\rho_i \cdot u + \nu_i)^2}{4} + \frac{(\rho_i \cdot u + \nu_i)\tau}{12} + \frac{1}{24})} \times \prod_{k=0}^{\infty} \frac{f(\rho_i \cdot u + \nu_i + k\tau)}{f(-\rho_i \cdot u - \nu_i + (k+1)\tau)}. \quad (4.27)$$

In the large radius limit, we find the following limit of \mathcal{F}_2 :

$$\mathcal{F}_2|_{\tau \rightarrow i\infty} = \prod_i \prod_{\rho_i} e^{\pi i \tau^2 (\frac{\epsilon_i^3}{3} - \frac{\epsilon_i^2}{2} + \frac{\epsilon_i \rho_i}{6})}, \quad (4.28)$$

with $\rho_i \cdot u + \nu_i = (\epsilon_{\rho_i} + m_{\rho_i})\tau$. Similarly one can also find the small radius limit of \mathcal{F}_2 using the S -transformation (4.19):

$$\mathcal{F}_2|_{\tau \rightarrow i0^+} = \prod_i \prod_{\rho_i} e^{-\pi i \tilde{\tau}^2 (\frac{\tilde{\epsilon}_i^3}{3} + \tilde{\epsilon}_i^2 m_{\rho_i} - \tilde{\epsilon}_i \tilde{m}_{\rho_i} - \frac{\tilde{\epsilon}_i \rho_i}{6} + \frac{m_{\rho_i}}{6})}, \quad (4.29)$$

where $\rho_i \cdot \tilde{u} + \tilde{\nu}_i = (\tilde{m}_{\rho_i} + \tilde{\epsilon}_{\rho_i})\tilde{\tau}$ with $\tilde{\epsilon}_{\rho_i} \in [0, 1)$. Note that the small and the large τ limits of \mathcal{F}_2 mirror, under $\tau \rightarrow -1/\tau$, the large and the small τ limits of \mathcal{F}_1 , faithfully and respectively.

4. Asymptotics of $\mathcal{F}^{\text{phys}}$

Unlike the A-twist gauge, the physical handle-gluing operator only contains Jacobian factor H^{phys} , which does not contribute to the leading term of the partition function. The leading contribution then only comes from $\mathcal{F}^{\text{phys}}$. From (4.25) one can see that each component of $\mathcal{F}^{\text{phys}}$ has the following asymptotic behavior.

In the small τ limit, with

$$\tilde{\epsilon}'_{\rho_i} = \{(\rho_i \cdot (u_H + \tilde{\sigma}) + \tilde{\nu}_i)/\tilde{\tau} + l_R(r_i - 1)/\tilde{\tau}\}, \quad (4.30)$$

$$\tilde{\epsilon}'_\alpha = \{\alpha \cdot (u_H + \tilde{\sigma})/\tilde{\tau} + l_R/\tilde{\tau}\}, \quad (4.31)$$

we find similarly

$$\begin{aligned} & \prod_i \prod_{\rho_i \in \mathfrak{R}_i} \Xi_1(\rho_i \cdot u + \nu_i + l_R \tau(r_i - 1); \tau) \Big|_{\tau \rightarrow i0^+} \\ & \sim \prod_i \prod_{\rho_i} e^{\pi i \tilde{\tau}^2 \left(\frac{(\tilde{\epsilon}'_{\rho_i})^3}{3} - \frac{(\tilde{\epsilon}'_{\rho_i})^2}{2} + \frac{\tilde{\epsilon}'_{\rho_i}}{6} \right)}, \end{aligned} \quad (4.32)$$

and

$$\begin{aligned} & (-1)^{\frac{l_R(l_R+1)}{2} \text{rank}(\mathcal{G})} \eta(\tau)^{2l_R \text{rank}(\mathcal{G})} \prod_\alpha \Xi_1(\alpha \cdot u + l_R \tau; \tau) \Big|_{\tau \rightarrow i0^+} \\ & \sim \prod_{\mathfrak{G}} e^{\pi i \tilde{\tau}^2 \left(\frac{(\tilde{\epsilon}'_\alpha)^3}{3} - \frac{(\tilde{\epsilon}'_\alpha)^2}{2} + \frac{\tilde{\epsilon}'_\alpha}{6} \right)}, \end{aligned} \quad (4.33)$$

where the last product is taken over all the gauge generators, again. It is important to note here that the set of H -saddles and the subsequent values of $\tilde{\epsilon}$'s to be used in the subsequent expansion of the exponents are no different from the preceding discussion of the small τ limit of A -twisted cases. This happens because the shift due to ν_R is negligible as $\tau \rightarrow i0^+$, as far as the values of u_H are concerned. Clearly this is not the case for the other limit $\tau \rightarrow i\infty$.

As we noted already in Sec. III, the large radius limit $\tau \rightarrow i\infty$ for “physical” cases follows a different pattern due to the large shift $\nu_R = l_R \tau = \frac{1-g}{p} \tau$. With

$$\epsilon'_{\rho_i} = \{(\rho_i \cdot (\hat{u}_H + \sigma) + \nu_i)/\tau + l_R(r_i - 1)\}, \quad (4.34)$$

$$\epsilon'_\alpha = \{\alpha \cdot (\hat{u}_H + \sigma)/\tau + l_R\}, \quad (4.35)$$

where m'_{ρ_i}, m'_α are the remaining integer parts, the exponential behavior goes as

$$\begin{aligned} & \prod_i \prod_{\rho_i \in \mathfrak{R}_i} \Xi_1(\rho_i \cdot u + \nu_i + l_R \tau(r_i - 1); \tau) \Big|_{\tau \rightarrow i\infty} \\ & \sim \prod_i \prod_{\rho_i} e^{\pi i \tau \left(\frac{(\epsilon'_{\rho_i})^3}{3} + (\epsilon'_{\rho_i})^2 m'_{\rho_i} - \epsilon'_{\rho_i} m'_{\rho_i} - \frac{\epsilon'_{\rho_i}}{6} + \frac{m'_{\rho_i}}{6} \right)}, \end{aligned} \quad (4.36)$$

and

$$\begin{aligned} & (-1)^{\frac{l_R(l_R+1)}{2} \text{rank}(\mathcal{G})} \eta(\tau)^{2l_R \text{rank}(\mathcal{G})} \prod_\alpha \Xi_1(\alpha \cdot u + l_R \tau; \tau) \Big|_{\tau \rightarrow i\infty} \\ & \sim \prod_{\mathfrak{G}} e^{\pi i \tau \left(\frac{(\epsilon'_\alpha)^3}{3} + (\epsilon'_\alpha)^2 m'_\alpha - \epsilon'_\alpha m'_\alpha - \frac{\epsilon'_\alpha}{6} + \frac{m'_\alpha}{6} \right)}, \end{aligned} \quad (4.37)$$

where the last product is taken over all generators of the gauge group with $\epsilon = 0$ understood for the Cartan generators.

5. H -saddles from the small radius limit of a fibered circle

We close with a minor consistency check on the notion of the H -saddle by considering the collapsing circle with nontrivial winding number p over the base. The Cardy limit $\beta_1 \rightarrow 0$ with $p_1 \neq 0$ would be the canonical example, while the Casimir limit $\beta_1 \rightarrow \infty$ with $p_2 \neq 0$ shares the same issue since as far as our partition functions goes this is equivalent to the other Cardy limit $\beta_2 \rightarrow 0$. The winding number p of the collapsing circle is not part of 3d spacetime data, so should not enter the 3d partition functions \mathcal{Z}_3^H 's, since the notion of the H -saddle relies on the existence of 3d theories that makes sense without referring to its 4d origin. It would be allowed to enter the coefficients c_H 's which serve as the glue between the 3d theories at H -saddles and the original 4d theory.

In view of the lengthy discussions in Sec. III, it should be relatively clear that the part of $(\mathcal{F}_1)^{p_1}$ that could have contributed to \mathcal{Z}_3^H in the large τ limit resides entirely in Γ_0 of Eq. (2.21). However the latter function reduces to 1 universally as $q \rightarrow 0$, regardless of the field content. The winding number p_1 therefore contributes at most to c_H 's, in this limit, via the surviving exponential prefactors in front of Γ_0 's, and does not interfere with the 3d theory at the H -saddles. Then, $SL(2, \mathbb{Z})$ automatically implies that the same happens for the $\beta_2 \rightarrow 0$ limit with $p_2 \neq 0$, as \mathcal{F}_2 in the small τ limit is nothing but \mathcal{F}_1 in the large τ limit modulo exponential prefactors, which are again harmless for the issue here.

For physical cases, one should look at how $\mathcal{F}^{\text{phys}}$ behaves in the $\tau \rightarrow i\infty$ limit. The relevant part of the latter fibering operator is made up of Ξ_1 's, or Γ_0 's therein, so again, $\mathcal{F}^{\text{phys}}$ reduces to a product of exponential functions: The winding number p can contribute to c_H 's at most, again as promised.

C. Cardy and Casimir

In the small and the large τ limits, we found exponential behaviors of the partition function which differ between different H -saddles. The partition function on A -twisted geometries, for example, always has an H -saddle at $u_H = 0$, and the exponential behavior there follows a universal form,

$$\mathcal{H}^{g-1} \Big|_{u_H=0} \sim [e^{2\pi i \tau \cdot (-\text{tr}_f R)/12}]^{g-1} \quad \text{or} \quad [e^{2\pi i \tilde{\tau} \cdot (-\text{tr}_f R)/12}]^{g-1}, \quad (4.38)$$

in the respective limits of large τ or $\tilde{\tau} = -1/\tau$. $\text{tr}_f R$ means the trace of $U(1)_R$ charge over all 4d fermions. The exponents inferred from this universal part have been identified in the past and given interpretation of the Casimir energy and the Cardy exponent with respect to the large and the small radius limit of β_2 [19]. The same expression also appeared for the Cardy limit of SCI's, which was then related to conformal anomaly coefficients [18].

Existence of H -saddles at $u_H \neq 0$ and the different leading exponents at such places, however, tell us that the Cardy exponent and the Casimir energy may be rather different in general. In this last section, we will explore this issue. For the sake of simplicity we will confine our attention to pure imaginary $\tau = i\beta_2/\beta_1$ and consider only those 4d theories whose chiral field content is invariant under the charge conjugation symmetry, $\rho \rightarrow -\rho$.

I. A-twist

We find, at each H -saddle at u_H , the leading exponents of \mathcal{H}^{g-1} in the large τ limit is

$$(g-1) \times \left[-\frac{1}{12}(\text{tr}_f R) + \frac{1}{2} \sum_{\alpha} \epsilon_{\alpha} (1 - \epsilon_{\alpha}) + \frac{1}{2} \sum_i (r_i - 1) \sum_{\rho_i} \epsilon_{\rho_i} (1 - \epsilon_{\rho_i}) \right] \quad (4.39)$$

multiplied by $2\pi i\tau$, instead of the universal form

$$(g-1) \times \left[-\frac{1}{12}(\text{tr}_f R) \right] \quad (4.40)$$

at $u_H = 0$. Clearly the H -saddle with the dominant contribution and the exponent thereof may be identified only after comparing this expression at different H -saddles. Furthermore, the actual exponent is given by this multiplied by $(g-1)$, so the dominant contributions for $g=0$ and the dominant contributions for $g>1$ will generically come from different H -saddles. For the small τ limit, the same formulas work with ϵ 's and τ replaced by $\tilde{\epsilon}$'s and $\tilde{\tau}$.

Note that, once we begin to identify 3d BAE vacua and evaluate the sum, ϵ 's at a given H -saddle would be really

$$\epsilon_{\rho} = \bar{\epsilon}_{\rho} + \frac{\rho_i \cdot \sigma_* + \nu_i}{\tau} \quad \text{or} \quad \tilde{\epsilon}_{\rho} = \bar{\epsilon}_{\rho} + \frac{\rho_i \cdot \tilde{\sigma}_* + \tilde{\nu}_i}{\tilde{\tau}} \quad (4.41)$$

etc., for multiple σ_* 's found by solving the 3d BAE at u_H . Expanding (4.39), the leading term

$$(g-1) \times \left[-\frac{1}{12}(\text{tr}_f R) + \frac{1}{2} \sum_{\alpha} \bar{\epsilon}_{\alpha} (1 - \bar{\epsilon}_{\alpha}) + \frac{1}{2} \sum_i (r_i - 1) \sum_{\rho_i} \bar{\epsilon}_{\rho_i} (1 - \bar{\epsilon}_{\rho_i}) \right] \quad (4.42)$$

must be augmented by the subleading pieces

$$\begin{aligned} & \frac{(g-1)}{\tau} \times \left[\sum_{\alpha} \bar{\epsilon}_{\alpha} \alpha + \sum_i (r_i - 1) \sum_{\rho_i} \bar{\epsilon}_{\rho_i} \rho_i \right] \cdot \sigma_*, \\ & - \frac{(g-1)}{\tilde{\tau}} \times \left[\sum_{\alpha} \bar{\epsilon}_{\alpha} \alpha + \sum_i (r_i - 1) \sum_{\rho_i} \bar{\epsilon}_{\rho_i} \rho_i \right] \cdot \tilde{\sigma}_*, \end{aligned} \quad (4.43)$$

which, combined with the overall factor $2\pi i\tau$ ($2\pi i\tilde{\tau}$), supply finite and σ_* ($\tilde{\sigma}_*$) dependent phases. Thus, cancellations between 3d BAE vacua in favor of smaller exponents at a given H -saddle cannot be ruled out in general. Although such cancellations do not appear to be commonplace, we will identify a few examples of this kind later.

For $p_{1,2} \neq 0$, there is a further exponential contribution of the form, via $\mathcal{F}_1^{p_1} \mathcal{F}_2^{p_2}$ in the sum. In the large τ limit, the additional terms, to be added to (4.39), are

$$\begin{aligned} p_1 \times & \left[\sum_i \sum_{\rho_i} \frac{1}{2} \left(\frac{\epsilon_{\rho_i}^3}{3} + \epsilon_{\rho_i}^2 m_{\rho_i} - \epsilon_{\rho_i} m_{\rho_i} - \frac{\epsilon_{\rho_i}}{6} + \frac{m_{\rho_i}}{6} \right) \right] \\ & + p_2 \times \left[\sum_i \sum_{\rho_i} \frac{\tau}{2} \left(\frac{\epsilon_{\rho_i}^3}{3} - \frac{\epsilon_{\rho_i}^2}{2} + \frac{\epsilon_{\rho_i}}{6} \right) \right], \end{aligned} \quad (4.44)$$

again modulo the large multiplicative factor $2\pi i\tau$. For the small τ limit, we merely need to exchange the asymptotic forms of \mathcal{F}_1 and of \mathcal{F}_2 and replace $\epsilon_{\rho_i} \rightarrow \tilde{\epsilon}_{\rho_i}$, $m_{\rho_i} \rightarrow \tilde{m}_{\rho_i}$ and $\tau \rightarrow \tilde{\tau}$.

In particular, with the restriction of the matter content to be symmetric under the charge conjugation, all terms that involve $\rho \cdot u$ cancel away leaving behind those involving powers of ν_i 's. The above then reduces to, e.g., for the large τ limit,

$$\begin{aligned} p_1 \times & \left[\sum_i \frac{\nu_i}{\tau} \sum_{\rho_i} \frac{1}{2} \left(\bar{\epsilon}_{\rho_i}^2 + 2\bar{\epsilon}_{\rho_i} m_{\rho_i} - m_{\rho_i} - \frac{1}{6} \right) \right] \\ & + p_2 \times \left[\sum_i \nu_i \sum_{\rho_i} \frac{1}{2} \left(\bar{\epsilon}_{\rho_i}^2 - \bar{\epsilon}_{\rho_i} + \frac{1}{6} \right) \right], \end{aligned} \quad (4.45)$$

the latter of which contributes ν -dependent pieces to the Casimir energy, while the former $1/\tau$ term contributes a finite imaginary piece to the exponent.

2. Physical

For the ‘‘physical’’ case, the H -saddle behavior is different between the large radius limit and the small radius limit, as we saw in the previous subsection. The small radius limit itself is on par with that of an A-twisted case, except that only $\mathcal{F}^{\text{phys}}$ contributes the leading exponential

$$\begin{aligned} p \times & \sum_i \sum_{\rho_i \in \mathfrak{R}_i} \left[\frac{\tilde{\tau}}{2} \left(\frac{\bar{\epsilon}_{\rho_i}^3}{3} - \frac{\bar{\epsilon}_{\rho_i}^2}{2} + \frac{\bar{\epsilon}_{\rho_i}}{6} \right) + \frac{1}{2} [\rho_i \cdot \tilde{\sigma} + \tilde{\nu}_i + l_R(r_i - 1)] \right. \\ & \left. \times \left(\bar{\epsilon}_{\rho_i}^2 - \bar{\epsilon}_{\rho_i} + \frac{1}{6} \right) \right], \end{aligned} \quad (4.46)$$

from matters, which reduces to

$$\sum_i \sum_{\rho_i \in \mathfrak{R}_i} \left[p \times \frac{\tilde{v}_i}{2} \left(\bar{e}_{\rho_i}^2 - \bar{e}_{\rho_i} + \frac{1}{6} \right) + (1-g) \times \frac{r_i - 1}{2} \left(\bar{e}_{\rho_i}^2 - \bar{e}_{\rho_i} + \frac{1}{6} \right) \right], \quad (4.47)$$

on theories with matter content which is symmetric under charge conjugation. The contribution from the vector is

$$(1-g) \times \sum_{\mathfrak{G}} \left[\frac{1}{2} \left(\bar{e}_\alpha^2 - \bar{e}_\alpha + \frac{1}{6} \right) \right]. \quad (4.48)$$

Both appear in the exponent with $2\pi i \tilde{\tau}$ multiplied.

The expression (4.47) plus (4.48), with \tilde{v}_i set to zero, has been isolated for the high-temperature limit of the 4d superconformal index [16], i.e., $p = 1$ and $g = 0$, and govern the asymptotic behavior of the integrand prior to the holonomy integration, called V_{eff} as in Ref. [16] modulo a constant shift. One subtlety is that the expressions we found via the BAE are meant to be evaluated and used at discrete places, $u = u_H$'s, so agreement with Ref. [16] requires that the maximum of V_{eff} necessarily occurs at an H -saddle. In fact, this is very likely since V_{eff} is a piece-wise linear function, as a consequence of ABJ anomaly cancellation, and the derivative changes only at points where one or more charged fields become massless. Thus the local maximum and minimum can only occur at places where $q \cdot u \in \mathbb{Z}$ for some charge q , and for a full agreement we only need to exclude places, u_0 where a vector multiplet of charge α becomes massless and no chiral multiplets are.

Since the contributions of the chiral multiplets to V_{eff} , after using the anomaly condition, cannot change abruptly there and since contribution from the α -charged vector will make a sharp turn, it suffices to consider how the derivative of

$$V_\alpha = -e_\alpha^2 + \epsilon_\alpha - \frac{1}{6} \quad (4.49)$$

behaves at $\epsilon_\alpha = 0$, i.e., where $\alpha \cdot u$ becomes an integer. Let us consider a small neighborhood around u_0 parametrized by $-1 < t < 1$ as $u = u_0 + tv$ where v is an arbitrary direction. Depending on the sign of $\alpha \cdot v$, the integer part of $\alpha \cdot u = n_\alpha + t\alpha \cdot v$ changes from $n_\alpha - 1$ to n_α or n_α to $n_\alpha - 1$ as t crosses $t = 0$. Thus, the vector multiplet contribution turns sharply at $t = 0$. Computing the derivatives before and after, it is easy to see that the turn

$$\left. \frac{dV_\alpha}{dt} \right|_{t=0^+} - \left. \frac{dV_\alpha}{dt} \right|_{t=0^-} = 2|\alpha \cdot v| \quad (4.50)$$

is positive for arbitrary v . The point u_0 cannot be a local maximum, which means that the maximum of V_{eff} cannot occur at such a point. It would occur at one of the H -saddles, therefore, which gives a full agreement on

the Cardy exponent between the previous approach and the BAE.

One can also deduce the large radius limit of the ‘‘physical’’ case. Recall, for each BAE solution, the leading exponent from $\mathcal{F}^{\text{phys}}$ is given by

$$p \times \left[\sum_i \sum_{\rho_i \in \mathfrak{R}_i} \frac{1}{2} \left(\frac{(\epsilon'_{\rho_i})^3}{3} + (\epsilon'_{\rho_i})^2 m'_{\rho_i} - \epsilon'_{\rho_i} m'_{\rho_i} - \frac{\epsilon'_{\rho_i}}{6} + \frac{m'_{\rho_i}}{6} \right) + \sum_{\mathfrak{G}} \frac{1}{2} \left(\frac{(\epsilon'_\alpha)^3}{3} + (\epsilon'_\alpha)^2 m'_\alpha - \epsilon'_\alpha m'_\alpha - \frac{\epsilon'_\alpha}{6} + \frac{m'_\alpha}{6} \right) \right], \quad (4.51)$$

with $2\pi i \tau$ multiplied. Since the partition function in total is obtained by summing up the contributions with those leading exponentials, the simplest guess would be that the Casimir energy equals the smallest exponent among the values of (4.51) evaluated at the BAE solutions.

3. Cancellations in the Casimir limit

However, we also encounter a large class of examples where the Casimir energies do not equal the smallest exponents computed above. The primary examples are found in the superconformal indices, i.e., the partition function in ‘‘physical’’ background with $p = 1$, $g = 0$. The Casimir energy of the resulting SCI's turns out to be equal to the value of (4.51) at $\hat{u}_H = 0$, despite the presence of nontrivial H -saddles. This holds, in many cases for SCI, even with the naive $\hat{u}_H = 0$ saddle absent.

This surprising fact can be demonstrated by rewriting the partition function as a unit circle contour integral,

$$\sum_{u_* \in S_{\text{BE}}} \mathcal{F}^{\text{phys}}(u_*, \nu; \tau) H(u_*, \nu; \tau)^{-1} = \frac{1}{|W_G|} \int_{|x|=1} \frac{dx}{2\pi i x} \mathcal{F}^{\text{phys}}(x, y; q), \quad (4.52)$$

the leading factor of $\mathcal{F}^{\text{phys}}$, which would have generated the holonomy-dependent Casimir energy,

$$\left(\prod_i \prod_{\rho_i \in \mathfrak{R}_i} e^{\frac{\pi i}{3\tau^2} [\rho_i \cdot u + \nu_i + \tau(r_i - 1)]^3 - \frac{\pi i}{6} [\rho_i \cdot u + \nu_i + \tau(r_i - 1)]} \right) \times \left(\prod_{\mathfrak{G}} e^{\frac{\pi i}{3\tau^2} (\alpha \cdot u + \tau)^3 - \frac{\pi i}{6} (\alpha \cdot u + \tau)} \right) = e^{\sum_i \dim(\mathfrak{R}_i) \left[\frac{\pi i}{3\tau^2} (\nu_i + \tau(r_i - 1))^3 - \frac{\pi i}{6} (\nu_i + \tau(r_i - 1)) \right] + \frac{\pi i}{6} \tau \dim(\mathfrak{G})} \quad (4.53)$$

becomes independent of u due to anomaly conditions. Thus, it comes out of the integral, and the leading exponent of q is fixed by the value of (4.51) at $\hat{u}_H = 0$ [20]. When we come back to BAE form, this happens via numerous cancellations between BAE vacua and sometimes even between H -saddles.

This cancellation is possible in part because the position of H -saddles is aligned along the real axis of $2\pi i u$ in this case. This should be contrasted to the Cardy limit, where the H -saddles are located along the unit circle $|x|=1$ so that H -saddle phenomena manifests even in this alternate integral formula. When the H -saddle occurs along the unit circle $|x|=1$, the cancellations due to the anomaly cancellation condition no longer works because the infinite product formula must be rewritten in new shifted variables whenever one crosses such an H -saddle; this was at the heart of the

H -saddle computation. The previous observation by Ardehali on Cardy exponents [16] has effectively captured this H -phenomenon on such a unit circle version of the superconformal indices. In contrast, such a cancellation does not happen in the Cardy limit.

Something similar happens for the Casimir limit of the A-twist case when the fibration is nontrivial. To see this, we should keep the finite part of the leading terms of \mathcal{F} and \mathcal{H} . For simplicity, we focus on the rank-1 case with $p_1 = p$, $p_2 = 0$. For massive matter fields at a given H -saddle, u_H , the contribution from \mathcal{F} is

$$\prod_i \prod_{\rho_i} e^{\pi i \tau \left(\frac{\bar{\epsilon}_{\rho_i}^3}{3} + \bar{\epsilon}_{\rho_i}^2 m_{\rho_i} - \bar{\epsilon}_{\rho_i} m_{\rho_i} - \frac{\bar{\epsilon}_{\rho_i} + m_{\rho_i}}{6} \right) + \pi i \left(\bar{\epsilon}_{\rho_i}^2 + 2\bar{\epsilon}_{\rho_i} m_{\rho_i} - m_{\rho_i} - \frac{1}{6} \right) (\rho_i \sigma + \nu_i) - \frac{\pi i}{2} (m_{\rho_i}^2 + m_{\rho_i})}. \quad (4.54)$$

For massless matter fields, i.e., for $\rho_i = \lambda_i$ such that $\epsilon_{\lambda_i} = 0$, we have an additional factor

$$\times (1 - z^{\lambda_i} y_i)^{m_{\lambda_i}} \quad (4.55)$$

with $z = e^{2\pi i \sigma}$. Using $m_{\lambda_i} = \lambda_i u_H / \tau$, the contribution from massless fields can be written as follows:

$$\begin{aligned} & \prod_i \prod_{\lambda_i} e^{-\frac{\pi i}{2} (\lambda_i^2 u_H^2 / \tau^2 + \lambda_i u_H / \tau)} q^{\lambda_i u_H / \tau / 12} z^{-\lambda_i / 12} y_i^{-1/12} [z^{-\lambda_i / 2} y_i^{-1/2} - z^{\lambda_i / 2} y_i^{1/2}]^{\lambda_i u_H / \tau} \\ & = \prod_i \prod_{\lambda_i} e^{-\frac{\pi i}{2} (\lambda_i^2 u_H^2 / \tau^2 + \lambda_i u_H / \tau - 2M u_H / \tau)} q^{\lambda_i u_H / \tau / 12} z^{-\lambda_i / 12} y_i^{-1/12} \Lambda_H^{u_H / \tau} (\Phi_a^{3d; H})^{-u_H / \tau}, \end{aligned} \quad (4.56)$$

where $M = \sum_i \sum_{\rho_i} m_{\rho_i} \rho_i$ and Λ_H is defined in (3.20). Note that u_H / τ is a rational number in $[0, 1)$. Thus, at $\sigma = \sigma_*$, the last factor becomes a root of unity,

$$(\Phi_a^{3d; H})^{-u_H / \tau} |_{\sigma=\sigma_*} = e^{-2\pi i k u_H / \tau}. \quad (4.57)$$

\mathcal{H} consists of two parts: $e^{2\pi i \Omega}$ and H . For massive matter fields, the leading term of $e^{2\pi i \Omega}$ is given by

$$\prod_i \prod_{\rho_i} e^{\pi i \tau (r_i - 1) \left(-\bar{\epsilon}_{\rho_i}^2 + \bar{\epsilon}_{\rho_i} - \frac{1}{6} \right) + \pi i (r_i - 1) \left(-2\bar{\epsilon}_{\rho_i} + 1 \right) (\rho_i \sigma + \nu_i) + \pi i (r_i - 1) m_{\rho_i}}, \quad (4.58)$$

while for massless matter fields, we have an additional factor

$$\times (1 - z^{\lambda_i} y_i)^{-(r_i - 1)}. \quad (4.59)$$

The same expansion can be made for vector fields, by replacing $\rho_i \rightarrow \alpha$ and $r_i \rightarrow 2$. Moreover, for $SU(2)$, H is explicitly written as

$$\sum_i \sum_{\rho_i} |\rho_i|^2 \left[\frac{1}{2} - \{(\rho_i (\hat{u}_H + \sigma) + \nu_i) / \tau\} + \sum_{k=0}^{\infty} \frac{\{x_H^{\rho_i} z^{\rho_i} y_i\} q^k}{1 - \{x_H^{\rho_i} z^{\rho_i} y_i\} q^k} - \sum_{k=0}^{\infty} \frac{\{x_H^{-\rho_i} z^{-\rho_i} y_i^{-1}\} q^{k+1}}{1 - \{x_H^{-\rho_i} z^{-\rho_i} y_i^{-1}\} q^{k+1}} \right]. \quad (4.60)$$

For massive fields, the first line is the leading contribution of order q^0 while for massless fields, there is an extra $\mathcal{O}(q^0)$ contribution $+\frac{z^{\lambda_i} y_i}{1 - z^{\lambda_i} y_i}$. A similar expansion is made for physical gauge as well by replacing $\nu_i \rightarrow \nu_i + \nu_R (r_i - 1)$.

4. Explicit examples with $\mathcal{G} = SU(2)$: The Casimir limit

We now explore some explicit examples for the Casimir limit; recall that this side is prone to further subtleties beyond H -saddles. Let us discuss the A-twist case first.

Considering asymptotically free theories of $SU(2)$, allowed representations are those with isospin $\frac{1}{2} \leq s \leq \frac{3}{2}$. For a model with few number of matters, BAE tends to be trivial due to a lack of enough flavor symmetry and cannot be discussed using the A-twist formalism. The Intriligator-Seiberg-Shenker model is such an example [47]. It has no anomaly-free flavor symmetry and, as a result, has the fixed anomaly-free R -charge $R = 3/5$, which allows A-twist only on a manifold of genus $g \in 5\mathbb{Z}$ due to the Dirac quantization condition for R -charges. Thus, we relegate the discussion of this model to the physical gauge case, and here consider $SU(2)$ with fundamentals and adjoints.

The RGG anomaly condition restricts R -charges of fundamentals and adjoints such that

$$\sum_{i=1}^{N_f} (r_i - 1) + 4 \sum_{j=1}^{N_a} (\tilde{r}_j - 1) + 4 = 0, \quad (4.61)$$

where r_i and \tilde{r}_j are the R -charges of fundamentals and adjoints respectively. For simplicity, we take

$$r_i = 1 + \frac{4(N_a - 1)}{N_f}, \quad \tilde{r}_j = 0. \quad (4.62)$$

With one adjoint, the numbers of flavors allowed by the asymptotically free condition are $N_f = 2, 4, 6$. In those cases, however, the exponent (4.39) is independent of u , so not very interesting in our discussion. Instead, we discuss the $SU(2)$ model with two fundamentals and two adjoints. Because of the adjoints, the H -saddles for this model are located at $u_H/\tau = 0$ and $u_H/\tau = 1/2$.

Take the H -saddle at $u_H/\tau = 1/2$. At this H -saddle only the vector field and the adjoint matter fields are massless while the fundamental matter fields become massive. Thus, the reduced BAE is given by

$$\frac{\prod_{j=1}^2 (z^2 - w_j)^2}{\prod_{j=1}^2 (1 - z^2 w_j)^2} = 1. \quad (4.63)$$

The equation has eight solutions, which are classified into two classes S_{\pm} satisfying

$$\left. \frac{\prod_{j=1}^2 (z^2 - w_j)}{\prod_{j=1}^2 (1 - z^2 w_j)} \right|_{z=z_*} = \pm 1, \quad z_* \in S_{\pm}. \quad (4.64)$$

For positive sign, the equation reduces to

$$z^4 = 1, \quad (4.65)$$

which has solutions $z = \pm 1, \pm i$. Among them, since $z = \pm 1$ are Weyl invariant, only $z = \pm i$ are relevant solutions. For negative sign, on the other hand, the equation can be reorganized into

$$(1 + w_1)(1 + w_2) = \frac{z^4(w_1 + w_2) - 2z^2(1 + w_1 w_2) + w_1 + w_2}{(1 - z^2)^2}. \quad (4.66)$$

The fibering operator and the handle-gluing operator at $u_H/\tau = 1/2$ are expanded as follows:

$$\begin{aligned} \mathcal{F} &= \frac{e^{\frac{2i}{\tau} f(\nu_i, \mu_i)} \prod_{j=1}^2 (1 - z^2 w_j)}{w_1^{7/12} w_2^{7/12} \prod_{j=1}^2 (z^2 - w_j)} + \mathcal{O}(q^{\frac{1}{2}}), \\ \mathcal{H} &= q^{\frac{7}{12}} \times \frac{4e^{\frac{2i}{\tau} h(\nu_i, \mu_i)} (1 - w_1)(1 - w_2)(1 - w_1 w_2)}{w_1^{3/2} w_2^{3/2}} \\ &\quad \times \frac{z^4(w_1 + w_2) - 2z^2(1 + w_1 w_2) + w_1 + w_2}{(1 - z^2)^2} + \mathcal{O}(q^{\frac{13}{12}}), \end{aligned} \quad (4.67)$$

where

$$f(\nu_i, \mu_i) = \frac{2}{3} \nu_1^3 + \frac{2}{3} \nu_2^3 + \mu_1^3 + \mu_2^3, \quad (4.68)$$

$$h(\nu_i, \mu_i) = -4\nu_1^2 - 4\nu_2^2 + 3\mu_1^2 + 3\mu_2^2. \quad (4.69)$$

One immediately notes that the leading term of \mathcal{F} is proportional to the square root of BAE. Thus,

$$\mathcal{F}|_{z=z_*} = \frac{\pm e^{\frac{2i}{\tau} f(\nu_i, \mu_i)}}{w_1^{7/12} w_2^{7/12}} + \mathcal{O}(q^{\frac{1}{2}}), \quad z_* \in S_{\pm} \setminus \{\pm 1\}. \quad (4.70)$$

Similarly, the leading term of \mathcal{H} is also simplified at each BAE solution as follows:

$$\begin{aligned} \mathcal{H}|_{z=z_*} &= q^{\frac{7}{12}} \times \frac{(3 \mp 1) e^{\frac{2i}{\tau} h(\nu_i, \mu_i)} (1 - w_1^2)(1 - w_2^2)(1 - w_1 w_2)}{w_1^{3/2} w_2^{3/2}} \\ &\quad + \mathcal{O}(q^{\frac{13}{12}}), \end{aligned} \quad (4.71)$$

where $z_* \in S_{\pm} \setminus \{\pm 1\}$ and we have used (4.66) for $z_* \in S_{\pm}$.

As a result, the generic leading term at $u_H/\tau = 1/2$ is given by

$$\begin{aligned} &\sum_{z_* \in S_{\pm} \setminus \{\pm 1\}} \mathcal{F}^p \mathcal{H}^{g-1} \Big|_{z=z_*} \\ &\sim q^{\frac{7}{12}(g-1)} \times [2^g + (-1)^p 4^g] e^{\frac{2i}{\tau} p f(\nu_i, \mu_i) + \frac{2i}{\tau} (g-1) h(\nu_i, \mu_i)} \\ &\quad \times \frac{[(1 - w_1^2)(1 - w_2^2)(1 - w_1 w_2)]^{g-1}}{w_1^{\frac{7}{12}p + \frac{3}{2}(g-1)} w_2^{\frac{7}{12}p + \frac{3}{2}(g-1)}}. \end{aligned} \quad (4.72)$$

Unless $g = 0$ and p is odd, this term does not vanish, and the leading q -exponent is given by

$$\mathcal{F}^p \mathcal{H}^{g-1} \Big|_{u_H/\tau=1/2} \sim q^{\frac{7}{12}(g-1)}. \quad (4.73)$$

On the other hand, if $g = 0$ and p is odd, this naive leading term cancels out. In such cases, we numerically find the true leading term, which turns out to be of order $q^{\frac{5}{12}}$.

At $u_H/\tau = 0$, on the other hand, the reduced BAE is given by

$$\frac{\prod_{i=1}^2 (z - y_i) \prod_{j=1}^2 (z^2 - w_j)^2}{\prod_{i=1}^2 (1 - zy_i) \prod_{j=1}^2 (1 - z^2 w_j)^2} = 1, \quad (4.74)$$

with the anomaly-free condition $y_1 y_2 w_1^4 w_2^4 = 1$. Since it is difficult to solve this equation analytically, instead, we tried numerical analysis for a given random phase values of y_i , w_j and found

$$\mathcal{F}^p \mathcal{H}^{g-1} \Big|_{u_H/\tau=0} \sim q^{-\frac{5}{12}(g-1)}, \quad (4.75)$$

which shows the exact agreement with the leading exponent at $u_H/\tau = 0$ predicted by (4.39). Thus, there is no cancellation of the leading terms at $u_H/\tau = 0$.

Combining these results at $u_H/\tau = 0$ and at $u_H/\tau = 1/2$, the leading term of the total partition function is given by

$$\Omega_{g,p} \sim \begin{cases} q^{-\frac{7}{12}}, & g = 0, \text{ p even,} \\ q^{-\frac{5}{12}(g-1)}, & \text{otherwise} \end{cases} \quad (4.76)$$

for $SU(2)$ theory with two fundamental chirals and two adjoint chirals.

Next, we move on to the physical gauge case. As advocated by Closset *et al.* [33], the integer quantization condition for R -charges can now be relaxed, and we can consider theories that flow to nontrivial superconformal points. The superconformal R -charge is then determined by the anomaly-free condition and the a maximization. For physical gauge, a canonical example with potential H -saddles is the ISS model, which is the $SU(2)$ model with a single isospin-3/2 matter. From the condition in Sec. III D, one can determine the H -saddles in the large radius limit as

$$\hat{u}_H/\tau = \frac{6}{35}, \quad \frac{3}{10}, \quad \frac{1}{2}, \quad \frac{7}{10}, \quad \frac{29}{35}. \quad (4.77)$$

Note that for physical gauge in the large radius limit, BAE does depend on the manifold because it contains $l_R = \frac{1-g}{p}$. For simplicity we stick to $l_R = 1$ cases.

As we mentioned, for $p = 1$, $g = 0$, i.e., the superconformal index, the partition function can be written as the unit circle contour integral, which predicts the Casimir energy

$$\begin{aligned} E_0 &= \sum_{\rho \in [3/2]} \frac{1}{2} \left(\frac{(\epsilon'_\rho)^3}{3} + (\epsilon'_\rho)^2 m'_\rho - \epsilon'_\rho m'_\rho - \frac{\epsilon'_\rho}{6} + \frac{m'_\rho}{6} \right) \Big|_{\epsilon'_\rho = \frac{3}{5}, m'_\rho = -1} \\ &+ \sum_{\alpha \in [1]} \frac{1}{2} \left(\frac{(\epsilon'_\alpha)^3}{3} + (\epsilon'_\alpha)^2 m'_\alpha - \epsilon'_\alpha m'_\alpha - \frac{\epsilon'_\alpha}{6} + \frac{m'_\alpha}{6} \right) \Big|_{\epsilon'_\alpha = 0, m'_\alpha = 1} \\ &= \frac{511}{1500}. \end{aligned} \quad (4.78)$$

On the other hand, for each H -saddle, the reduced BAE and the leading terms of $(\mathcal{F}^{\text{phys}})^p H^{g-1}$ are given by

$$\begin{aligned} -z^7 = 1 & \quad \text{at} \quad \hat{u}_H/\tau = \frac{6}{35}, \\ z^{-2} = 1 & \quad \text{at} \quad \hat{u}_H/\tau = \frac{3}{10}, \\ z^8 = 1 & \quad \text{at} \quad \hat{u}_H/\tau = \frac{1}{2}, \\ z^{-2} = 1 & \quad \text{at} \quad \hat{u}_H/\tau = \frac{7}{10}, \\ -z^7 = 1 & \quad \text{at} \quad \hat{u}_H/\tau = \frac{29}{35}, \end{aligned} \quad (4.79)$$

and

$$\begin{aligned} q^{-\frac{23}{10500}} \times \frac{1}{7z^2} & \quad \text{at} \quad \hat{u}_H/\tau = \frac{6}{35}, \\ q^{\frac{61}{1500}} \times \frac{z}{2} & \quad \text{at} \quad \hat{u}_H/\tau = \frac{3}{10}, \\ q^{\frac{211}{1500}} \times \frac{(1-z^2)^2}{8z^6} & \quad \text{at} \quad \hat{u}_H/\tau = \frac{1}{2}, \\ q^{\frac{61}{1500}} \times \frac{z}{2} & \quad \text{at} \quad \hat{u}_H/\tau = \frac{7}{10}, \\ q^{-\frac{23}{10500}} \times \left(-\frac{1}{7z^5} \right) & \quad \text{at} \quad \hat{u}_H/\tau = \frac{29}{35}, \end{aligned} \quad (4.80)$$

with $z = e^{2\pi i \sigma}$. Note that $z = \pm 1$ at $\hat{u}_H/\tau = 1/2$ are Weyl invariant and again excluded from the solution set. One can see that those leading contributions all vanish for $p = 1$ as expected. We also confirmed numerically that the subleading terms with q -exponents less than $\frac{511}{1500}$ are all canceled out such that the true leading term of the total partition function is of order $q^{\frac{511}{1500}}$. Note that the value happens to coincide with the would-be exponent at $\hat{u}_H/\tau = 0$, even though the tower sits at the $\hat{u}_H/\tau = 1/2$ saddle. This is one example of cancellations for SCI's which was advertised previously.

On the other hand, such cancellations in favor of the would-be exponent at $u_H = 0$ does not necessarily happen for general values of p . For $p = -2$, $g = 3$, as the second example, the leading $(\mathcal{F}^{\text{phys}})^p H^{g-1}$ is given by

$$\begin{aligned}
 q^{\frac{23}{5250}} \times 49z^4 & \text{ at } \hat{u}_H/\tau = \frac{6}{35}, \\
 q^{-\frac{61}{750}} \times \frac{4}{z^2} & \text{ at } \hat{u}_H/\tau = \frac{3}{10}, \\
 q^{-\frac{211}{750}} \times \frac{64z^{12}}{(1-z^2)^4} & \text{ at } \hat{u}_H/\tau = \frac{1}{2}, \\
 q^{-\frac{61}{750}} \times \frac{4}{z^2} & \text{ at } \hat{u}_H/\tau = \frac{7}{10}, \\
 q^{\frac{23}{5250}} \times 49z^{10} & \text{ at } \hat{u}_H/\tau = \frac{29}{35}. \tag{4.81}
 \end{aligned}$$

The locations of H -saddles are the same as those of $p = 1$, $g = 0$ because the two geometries share the same $l_R = 1$. Substituting the BAE solutions at each H -saddle, the leading terms are evaluated as

$$\begin{aligned}
 q^{\frac{23}{5250}} \times 0 & \text{ at } \hat{u}_H/\tau = \frac{6}{35}, \\
 q^{-\frac{61}{750}} \times 8 & \text{ at } \hat{u}_H/\tau = \frac{3}{10}, \\
 q^{-\frac{211}{750}} \times 72 & \text{ at } \hat{u}_H/\tau = \frac{1}{2}, \\
 q^{-\frac{61}{750}} \times 8 & \text{ at } \hat{u}_H/\tau = \frac{7}{10}, \\
 q^{\frac{23}{5250}} \times 0 & \text{ at } \hat{u}_H/\tau = \frac{29}{35}. \tag{4.82}
 \end{aligned}$$

Thus, the leading q -exponent $-\frac{211}{750}$ persists in this case, meaning that the nontrivial $\hat{u}_H/\tau = 1/2$ saddle is dominant and the leading exponent there suffers no cancellations, in contrast to the $p = 1$ case.

5. Anomaly or not

Several observations relating these asymptotic coefficients to the axial and to the conformal anomalies were made recently [18,20–22,48]. One well-known example is a relation between the Cardy exponent and the sum of $U(1)_R$ charges of fermions, on par with (4.40), which, for superconformal cases, translates to “ $a - c$ ” where a and c are the usual conformal anomaly coefficients.

One main consequence of our investigation is that such a connection cannot be trusted in general. Whenever the matter content involves gauge representation beyond the simplest ones, H -saddles will tend to appear at $u_H \neq 0$, some of which could dominate the naive one at $u_H = 0$ easily. The canonical example of SQCD escapes this, since the chiral multiplets are all in the fundamental representation. It probably explains why this rather generic phenomenon has so far failed to be noticed. For SCI’s, such a deviation from $\frac{(1-g)}{12} \cdot \text{tr}_f R$ has been observed first by Ardehali [16] and subsequently by Di Pietro and Honda [17] for a handful of examples, but, as we saw, this deviation is more of a rule than an exception.

We also saw that something similar happens with the Casimir limit as well. We again find that the notion of H -saddles would be valid even in the large radius limit provided that there are two circles in the spacetime, at least for computation of the partition functions. This will generally complicate the asymptotics of typical partition functions, just as in the Cardy limit. For this Casimir energy side, however, the connection to the global anomaly [20,22] is a little more robust than the Cardy side, although somewhat dependent on the background geometry; SCI’s, in particular, turned out to enjoy a rather special structure such that this naive Casimir energy, apparently from $u_H = 0$, stands uncorrected even though nontrivial H -saddles exist, and, more surprisingly, even when the naive $u_H = 0$ saddle is absent, due to magical cancellations between BAE vacua or even between H -saddles. For general partition functions, for example with $p > 1$, such cancellations are more scarce.

Much of this section explored such diverse forms of the Casimir energies and the Cardy exponents, and gave precise methods for isolating these, albeit with no obvious universal formula.

V. SUMMARY

We have introduced the notion of the holonomy saddle, or H -saddle, and explored how the phenomenon manifests in $d = 4$ $\mathcal{N} = 1$ massive gauge theories.

Certain discrete values of the gauge holonomy are found to support $d = 3$ $\mathcal{N} = 2$ supersymmetric gauge theories. When the space is taken to be noncompact, the existence of multiple H -saddles means that a theory G compactified on a small circle admits multiple superselection sectors at discrete holonomy values u_H ’s, where supersymmetric vacua are clustered which are in turn attributable to an effective 3d theory H . Such an H -theory tends to have generally a smaller light field content than the naive dimensional reduction, due to the symmetry breaking by the Wilson line, although one also typically finds the naive saddle at $u_H = 0$ as well. This observation dovetails nicely against some of the existing studies of 4d-to-3d and 3d-to-2d reduction of dualities [13,14], where one finds that a single dual pair typically generates multiple dual pairs in the lower dimensions.

Its manifestation in the compact spacetime equipped with a circle, on the other hand, implies precise relations between the twisted partition function of the G theory and those of the subsequent H theories, to which we have devoted the bulk of the computations. As such, the Witten index of the G theory would be generally a sum of Witten indices of the H theories, which explains, in part, how the number of the supersymmetric vacua differs between two theories in the adjacent dimensions even with the same supermultiplet content. This also offers a definite method for reconstructing one from the others.

We also investigated the consequences for supersymmetry-preserving torus-fibered compact spacetimes by observing how the twisted partition functions behave in

a small radius limit of one of the two circle fibers. The 4d twisted partition function reduces to a sum of 3d partition functions in those limits, modulo exponential prefactors which are interpreted either as the Cardy or as the Casimir behavior, depending on which direction is taken to be the Euclidean time. The results on such exponents are generally different from the existing claims, as the latter tends to focus, effectively, on the naive saddle at $u_H = 0$.

In the current examples of partition functions and theories, which admit the BAE description, H -saddles are located by asking which subset of chiral matter fields become light at which discrete values of the holonomy. FI constants and Chern-Simons levels, generated by KK modes, can further complicate the pattern, which we also delineated in much detail. The characterization of H -saddles should be a bit more general, however: an H -saddle would appear in the holonomy space wherever the dimensionally reduced theory admits supersymmetric vacua, normalizable or non-normalizable [2]. This general criterion for H -saddles should be valid for any supersymmetric gauge theories, as long as the gauge holonomy is not exactly flat at the quantum level.

What we have not explored here is how this phenomenon relates to and interacts with the matter of disconnected holonomy sectors, well known in the context of the 4d Witten index computations of pure Yang-Mills theories. Since our H -saddles would occur already for the holonomy on S^1 and since the corresponding discrete choices u_H arise from the dynamics rather than from the topology, it is clear that the topological consideration must be separately considered as well for a more general gauge group \mathcal{G} . An immediate question is how the discussion here should be generalized when \mathcal{G} is not simply connected [40] or when the so-called “triple” is relevant [36–39]. We suspect we will encounter more issues related to such holonomy saddles and holonomy islands in the near future.

ACKNOWLEDGMENTS

We would like to thank Cyril Closset, Richard Eager, Heeyon Kim, Nati Seiberg, and Edward Witten for useful conversations. The research of S. L. is supported in part by the National Research Foundation of Korea (NRF) Grant No. NRF-2017R1C1B1011440.

-
- [1] K. G. Wilson, Confinement of quarks, *Phys. Rev. D* **10**, 2445 (1974).
 - [2] C. Hwang and P. Yi, Twisted partition functions and H -saddles, *J. High Energy Phys.* **06** (2017) 045.
 - [3] V. G. Kac and A. V. Smilga, Normalized vacuum states in $N = 4$ supersymmetric Yang-Mills quantum mechanics with any gauge group, *Nucl. Phys.* **B571**, 515 (2000).
 - [4] M. Staudacher, Bulk Witten indices and the number of normalizable ground states in supersymmetric quantum mechanics of orthogonal, symplectic and exceptional groups, *Phys. Lett. B* **488**, 194 (2000).
 - [5] V. Pestun, $N = 4$ SYM matrix integrals for almost all simple gauge groups (except E(7) and E(8)), *J. High Energy Phys.* **09** (2002) 012.
 - [6] P. Yi, Witten index and threshold bound states of D-branes, *Nucl. Phys.* **B505**, 307 (1997).
 - [7] S. Sethi and M. Stern, D-brane bound states redux, *Commun. Math. Phys.* **194**, 675 (1998).
 - [8] M. B. Green and M. Gutperle, D particle bound states and the D instanton measure, *J. High Energy Phys.* **01** (1998) 005.
 - [9] G. W. Moore, N. Nekrasov, and S. Shatashvili, D particle bound states and generalized instantons, *Commun. Math. Phys.* **209**, 77 (2000).
 - [10] S. J. Lee and P. Yi, Witten index for noncompact dynamics, *J. High Energy Phys.* **06** (2016) 089.
 - [11] S. J. Lee and P. Yi, D-particles on orientifolds and rational invariants, *J. High Energy Phys.* **07** (2017) 046.
 - [12] E. Witten, Dynamical breaking of supersymmetry, *Nucl. Phys.* **B188**, 513 (1981).
 - [13] O. Aharony, S. S. Razamat, N. Seiberg, and B. Willett, 3d dualities from 4d dualities, *J. High Energy Phys.* **07** (2013) 149.
 - [14] O. Aharony, S. S. Razamat, and B. Willett, From 3d duality to 2d duality, *J. High Energy Phys.* **11** (2017) 090.
 - [15] M. Aganagic, K. Hori, A. Karch, and D. Tong, Mirror symmetry in $(2 + 1)$ -dimensions and $(1 + 1)$ -dimensions, *J. High Energy Phys.* **07** (2001) 022.
 - [16] A. A. Ardehali, High-temperature asymptotics of supersymmetric partition functions, *J. High Energy Phys.* **07** (2016) 025.
 - [17] L. Di Pietro and M. Honda, Cardy Formula for 4d SUSY theories and localization, *J. High Energy Phys.* **04** (2017) 055.
 - [18] L. Di Pietro and Z. Komargodski, Cardy formulae for SUSY theories in $d = 4$ and $d = 6$, *J. High Energy Phys.* **12** (2014) 031.
 - [19] C. Closset, H. Kim, and B. Willett, $\mathcal{N} = 1$ supersymmetric indices and the four-dimensional A-model, *J. High Energy Phys.* **08** (2017) 090.
 - [20] N. Bobev, M. Bullimore, and H. C. Kim, Supersymmetric Casimir energy and the anomaly polynomial, *J. High Energy Phys.* **09** (2015) 142.
 - [21] D. Martelli and J. Sparks, The character of the supersymmetric Casimir energy, *J. High Energy Phys.* **08** (2016) 117.
 - [22] B. Assel, D. Cassani, L. Di Pietro, Z. Komargodski, J., and D. Martelli, The Casimir energy in curved space and its supersymmetric counterpart, *J. High Energy Phys.* **07** (2015) 043.

- [23] C. Romelsberger, Counting chiral primaries in $N = 1$, $d = 4$ superconformal field theories, *Nucl. Phys.* **B747**, 329 (2006).
- [24] J. Kinney, J. M. Maldacena, S. Minwalla, and S. Raju, An index for 4 dimensional super conformal theories, *Commun. Math. Phys.* **275**, 209 (2007).
- [25] F. Benini, R. Eager, K. Hori, and Y. Tachikawa, Elliptic genera of two-dimensional $N = 2$ gauge theories with rank-one gauge groups, *Lett. Math. Phys.* **104**, 465 (2014).
- [26] F. Benini, R. Eager, K. Hori, and Y. Tachikawa, Elliptic genera of 2d $\mathcal{N} = 2$ gauge theories, *Commun. Math. Phys.* **333**, 1241 (2015).
- [27] K. Hori, H. Kim, and P. Yi, Witten Index and Wall Crossing, *J. High Energy Phys.* **01** (2015) 124.
- [28] C. Hwang, J. Kim, S. Kim, and J. Park, General instanton counting and 5d SCFT, *J. High Energy Phys.* **07** (2015) 063; Addendum, **04** (2016) 094.
- [29] K. Intriligator and N. Seiberg, Aspects of 3d $N = 2$ Chern-Simons-matter theories, *J. High Energy Phys.* **07** (2013) 079.
- [30] F. Benini and A. Zaffaroni, A topologically twisted index for three-dimensional supersymmetric theories, *J. High Energy Phys.* **07** (2015) 127.
- [31] F. Benini and A. Zaffaroni, Supersymmetric partition functions on Riemann surfaces, *Proc. Symp. Pure Math.* **96**, 13 (2017).
- [32] C. Closset and H. Kim, Comments on twisted indices in 3d supersymmetric gauge theories, *J. High Energy Phys.* **08** (2016) 059.
- [33] C. Closset, H. Kim, and B. Willett, Supersymmetric partition functions and the three-dimensional A-twist, *J. High Energy Phys.* **03** (2017) 074.
- [34] J. L. Cardy, Operator content of two-dimensional conformally invariant theories, *Nucl. Phys.* **B270**, 186 (1986).
- [35] N. A. Nekrasov and S. L. Shatashvili, Supersymmetric vacua and Bethe ansatz, *Nucl. Phys. B, Proc. Suppl.* **192–193**, 91 (2009).
- [36] E. Witten, Toroidal compactification without vector structure, *J. High Energy Phys.* **02** (1998) 006.
- [37] A. Keurentjes, Nontrivial flat connections on the 3 torus I: $G(2)$ and the orthogonal groups, *J. High Energy Phys.* **05** (1999) 001.
- [38] A. Keurentjes, Nontrivial flat connections on the three torus. 2. The exceptional groups F_4 and E_6 , E_7 , E_8 , *J. High Energy Phys.* **05** (1999) 014.
- [39] V. G. Kac and A. V. Smilga, Vacuum structure in supersymmetric Yang-Mills theories with any gauge group, in *The Many Faces of the Superworld*, edited by M. A. Shifman (World Scientific, Singapore, 2000), pp. 185–234, DOI: 10.1142/9789812793850_0014.
- [40] E. Witten, Supersymmetric index in four-dimensional gauge theories, *Adv. Theor. Math. Phys.* **5**, 841 (2001).
- [41] K. Hori and C. Vafa, Mirror symmetry, [arXiv:hep-th/0002222](https://arxiv.org/abs/hep-th/0002222).
- [42] K. Hori and D. Tong, Aspects of Non-Abelian gauge dynamics in two-dimensional $N = (2, 2)$ theories, *J. High Energy Phys.* **05** (2007) 079.
- [43] E. Witten, Topological sigma models, *Commun. Math. Phys.* **118**, 411 (1988).
- [44] L. Alvarez-Gaume, S. Della Pietra, and G. W. Moore, Anomalies and odd dimensions, *Ann. Phys. (N.Y.)* **163**, 288 (1985).
- [45] E. Witten, Fermion path integrals and topological phases, *Rev. Mod. Phys.* **88**, 035001 (2016).
- [46] E. Witten, Supersymmetric index of three-dimensional gauge theory, in *The Many Faces of the Superworld*, edited by M. A. Shifman (World Scientific, Singapore, 2000), pp. 156–184, DOI: 10.1142/9789812793850_0013.
- [47] K. A. Intriligator, N. Seiberg, and S. H. Shenker, Proposal for a simple model of dynamical SUSY breaking, *Phys. Lett. B* **342**, 152 (1995).
- [48] S. M. Hosseini, A. Nedelin, and A. Zaffaroni, The Cardy limit of the topologically twisted index and black strings in AdS_5 , *J. High Energy Phys.* **04** (2017) 014.