

# Spectral Networks

Davide Gaiotto, Gregory W. Moore and Andrew Neitzke

**Abstract.** We introduce new geometric objects called *spectral networks*. Spectral networks are networks of trajectories on Riemann surfaces obeying certain local rules. Spectral networks arise naturally in four-dimensional  $\mathcal{N} = 2$  theories coupled to surface defects, particularly the theories of class  $S$ . In these theories, spectral networks provide a useful tool for the computation of BPS degeneracies; the network directly determines the degeneracies of solitons living on the surface defect, which in turn determines the degeneracies for particles living in the 4d bulk. Spectral networks also lead to a new map between flat  $\mathrm{GL}(K, \mathbb{C})$  connections on a two-dimensional surface  $C$  and flat abelian connections on an appropriate branched cover  $\Sigma$  of  $C$ . This construction produces natural coordinate systems on moduli spaces of flat  $\mathrm{GL}(K, \mathbb{C})$  connections on  $C$ , which we conjecture are cluster coordinate systems.

## 1. Introduction and Summary

In this paper, we study objects which we call *spectral networks*. A spectral network  $\mathcal{W}$  is made up of *walls* drawn on a punctured real surface  $C$ . Figure 1 gives a picture of one.

Each wall carries some extra discrete data: a pair  $ij$  and (roughly) an integer  $\mu$ . The labels  $i$  and  $j$  are drawn from the set of sheets of a finite branched covering  $\Sigma \rightarrow C$ . The walls and their discrete data obey certain local constraints; for example,

- each simple branch point of the covering  $\Sigma \rightarrow C$  gives birth to three walls;
- when an  $ij$  wall and a  $jk$  wall intersect, a new  $ik$  wall is born at their intersection point.

In Sects. 2–8 of this paper, we study some particular spectral networks which arise naturally in physics. In Sect. 9, we axiomatize the notion of spectral network. In Sect. 10, we describe a more mathematical application, to coordinate systems on moduli spaces of flat connections. Those sections are mostly self-contained.

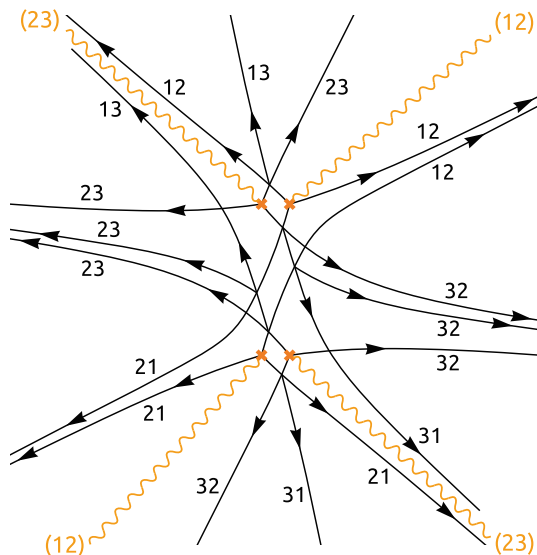


FIGURE 1. A spectral network, drawn on the stereographic projection of  $C = S^2$ , with a single puncture at infinity. All of the walls eventually asymptote to this puncture. The walls are labeled by pairs  $ij$ , where  $i, j$  are sheets of a threefold covering  $\Sigma \rightarrow C$ . The branch points of the covering are shown as *orange crosses*. We have trivialized the covering over the complement of some branch cuts, shown as *wavy orange lines* (color figure online)

Here is how spectral networks arise in physics. Fix an  $\mathcal{N} = 2$  supersymmetric theory  $T$  in  $d = 4$  and a point  $u$  of the Coulomb branch. Also fix a  $1/2$ -BPS surface defect  $\mathbb{S}_z$  which has a 1-complex-dimensional space  $C$  of UV parameters, and is massive in the IR, with finitely many vacua for any fixed  $z$ . These data determine a 1-parameter family of spectral networks  $\mathcal{W}_\vartheta$  drawn on  $C$ , labeled by phases  $\vartheta \in \mathbb{R}/2\pi\mathbb{Z}$ . Namely,  $\mathcal{W}_\vartheta$  is the locus of  $z \in C$  for which the surface defect  $\mathbb{S}_z$  carries a BPS soliton whose central charge  $Z$  is aligned with  $-e^{i\vartheta}$ . The discrete data  $\mu$  on the walls keep track of the degeneracies of these BPS solitons. The points of the finite cover  $\Sigma \rightarrow C$  over  $z \in C$  are the vacua of  $\mathbb{S}_z$ . The data  $ij$  on the walls keep track of which vacua of  $\mathbb{S}_z$  are being interpolated by the solitons.

In this paper, for concreteness, we focus on a particular (large) class of  $\mathcal{N} = 2$  theories, namely the theories of class  $S$  [1–3]. These are the theories  $S[\mathfrak{g}, C, D]$ , associated to a Lie algebra  $\mathfrak{g}$  (which we take to be  $\mathfrak{g} = A_{K-1}$ ), punctured Riemann surface  $C$ , and a collection  $D$  of defects placed at the punctures of  $C$ . They are obtained by a partially topologically twisted compactification of the six-dimensional  $(2, 0)$  theory  $S[\mathfrak{g}]$ . A point  $u$  of the Coulomb branch here means a tuple  $(\phi_2, \dots, \phi_K)$ , where each  $\phi_r$  is a meromorphic  $r$ -differential on  $C$ , with poles at the defects. In these theories, there is a canonical surface

defect  $\mathbb{S}_z$  whose parameter space is  $C$ . Moreover, in these theories, the covering  $\Sigma \rightarrow C$  can be identified with the Seiberg–Witten curve [4, 5] and is given concretely by (2) below.

The spectral networks  $\mathcal{W}_\vartheta$  arising from theories of class  $S$  can be described concretely. The walls of the network are solutions of differential equations, given in (3) below. The full network is built up by a continuous process: three walls are born from each branch point of the covering  $\Sigma$  and flow according to (3); whenever two walls intersect, they can give birth to an additional wall. This process is described in Sect. 5.4.

Here are some of the uses of the spectral networks  $\mathcal{W}_\vartheta$ .

*Framed BPS states.* By letting the parameter  $z$  of a surface defect vary along a path  $\wp$  in  $C$ , one can define a supersymmetric interface  $L_{\wp, \vartheta}$  [5]. Knowing  $\mathcal{W}_\vartheta$  and  $\mu$  allows one to determine the complete spectrum of BPS states of these interfaces, called “framed BPS states.” In fact, the framed BPS spectrum turns out to be *overdetermined*, so much so that one can use its consistency to compute  $\mu$ .

*Jumps of  $\mathcal{W}_\vartheta$  and the 4d BPS spectrum.* As the parameter  $\vartheta$  is varied,  $\mathcal{W}_\vartheta$  also varies. There are some critical phases  $\vartheta_c$  at which the topology of  $\mathcal{W}_\vartheta$  suddenly changes (in a precise sense explained in Sect. 6). These critical phases  $\vartheta_c$  are the phases of the central charges  $Z_\gamma$  of BPS states in the  $d = 4$  theory  $S[\mathfrak{g}, C, D]$ . Moreover, one can read off the degeneracies  $\Omega(\gamma)$  of these BPS states from the topology of  $\mathcal{W}_{\vartheta=\vartheta_c}$ .

We find in this way that BPS states in  $S[\mathfrak{g}, C, D]$  correspond to certain “finite webs” of strings on  $C$ , which appear inside the spectral network  $\mathcal{W}_\vartheta$  at the critical phase. Some pictures of finite webs appear in Fig. 2. The role of finite webs in the BPS spectrum was already expected [3, 6–8], but our analysis here gives a much more precise understanding of how to compute the corresponding BPS degeneracies  $\Omega(\gamma)$  than was previously available. It also provides a geometric argument (if not quite a proof) that the degeneracies so computed obey the wall-crossing formula of Kontsevich–Soibelman [9], as well as its extension to include the coupling between 2d and 4d BPS states, given in [5].

This story can be thought of as a broad generalization of parts of [3, 5], where we explained why the counts of finite webs obey the wall-crossing formulas in the case  $K = 2$ . In fact, our approach here leads to a simpler and more conceptual understanding even of that case.

The 4d BPS spectrum of  $\mathcal{N} = 2$  theories has been investigated by many authors recently; see in particular [10, 11]. One advantage of our approach here via spectral networks is that, given a point  $u$  of the Coulomb branch, one can determine any particular BPS degeneracy  $\Omega(\gamma)$  at  $u$  just by drawing the network  $\mathcal{W}_{\vartheta=\arg Z_\gamma}$  corresponding to  $u$ . The spectral network also gives more precise information—not only the 4d BPS degeneracies  $\Omega(\gamma)$  but also some enhancements  $\omega(\gamma, \cdot)$ , which keep track of the interaction between the 4d BPS state and the surface defects  $\mathbb{S}_z$ .

*Moduli of flat connections.* Suppose we compactify on  $S^1$  both the  $d = 4$  theory and the surface defects  $\mathbb{S}_z$ . The ground states of  $\mathbb{S}_z$  on  $S^1$  form a  $K$ -dimensional vector space  $E_z$ ; letting  $z$  vary we obtain a rank  $K$  vector bundle  $E$  over  $C$ . Now take a path  $\wp$  in  $C$  from  $z_1$  to  $z_2$ . The vacuum expectation value  $\langle L_{\wp, \vartheta} \rangle$  of the corresponding interface wrapped on  $S^1$  is an isomorphism from  $E_{z_1}$  to  $E_{z_2}$ . In other words, the bundle  $E$  is equipped with a natural connection [4], which moreover is actually flat.<sup>1</sup>

On the other hand, we could also take the perspective of the IR (abelian) theory. From that point of view, we would have surface defects labeled by points on  $\Sigma$  rather than on  $C$ , with one-dimensional spaces of vacua, thus forming a line bundle  $\mathcal{L}$  over  $\Sigma$ . Considering interfaces between these surface defects, we see that  $\mathcal{L}$  too is naturally equipped with a flat connection.

Therefore, to each vacuum of the compactification of the  $d = 4$  theory on  $S^1$ , we have assigned on the one hand a flat rank  $K$  connection over  $C$ , and on the other hand a flat rank 1 connection over the covering  $\Sigma$ . This induces a correspondence  $\Psi_{\mathcal{W}_\vartheta}$  between the two types of connection, which turns out to be (at least locally) an isomorphism of moduli spaces. Moreover, we can compute  $\Psi_{\mathcal{W}_\vartheta}$  concretely: the key is the framed BPS degeneracies mentioned above, which (roughly) gives the coefficients in the expansion of a UV interface  $L_{\wp, \vartheta}$  in terms of IR interfaces. Upon taking expectation values, this becomes an expansion of the vacuum expectation value  $\langle L_{\wp, \vartheta} \rangle$  in terms of “Darboux coordinates” as described, e.g. in Sect. 5.8 of [5]. In particular, the correspondence  $\Psi_{\mathcal{W}_\vartheta}$  is determined by the spectral network  $\mathcal{W}_\vartheta$ .

*Cluster coordinate systems.* Since the space of flat rank 1 connections over  $\Sigma$  is a product of copies of  $\mathbb{C}^\times$ , another way to describe the last item is to say that  $\Psi_{\mathcal{W}_\vartheta}$  is a local *coordinate system* on the moduli space  $\mathcal{M}$  of flat rank  $K$  connections over  $C$ . The construction of  $\Psi_{\mathcal{W}_\vartheta}$  from  $\mathcal{W}_\vartheta$  can be generalized: it uses only some general properties of  $\mathcal{W}_\vartheta$ , which we abstract into a definition of “spectral network”. Given any spectral network  $\mathcal{W}$  there is a corresponding  $\Psi_{\mathcal{W}}$ .

We conjecture that  $\Psi_{\mathcal{W}}$  is a particularly nice coordinate system: it is a *cluster coordinate system*, in the sense of [12]. Indeed, Fock and Goncharov proved [13] that the spaces  $\mathcal{M}$  we consider admit an atlas of cluster coordinate systems.<sup>2</sup> For  $K = 2$ , the cluster atlas consists precisely of the coordinate systems  $\Psi_{\mathcal{W}}$ . For  $K > 2$ , Fock and Goncharov did not give a completely explicit description of the atlas—rather they described some particular cluster coordinate systems. The most general cluster coordinate system would be obtained by beginning with one of these and performing some sequence of coordinate transformations known as “mutations”; each mutation generates a new coordinate system, and the cluster atlas is the set of all coordinate systems obtained

<sup>1</sup> We are oversimplifying slightly here: the precise story, laid out in the main text, involves a slight twisting of the notion of flat connection. This twisting does not modify the basic picture and can safely be ignored at first.

<sup>2</sup> This fact implies, e.g. that these spaces admit a natural quantization [12, 14].

in this way. We conjecture that our  $\Psi_{\mathcal{W}}$  are coordinate systems in the cluster atlas.<sup>3</sup>

How do the mutations arise in our story? Small deformations of the spectral network  $\mathcal{W}$  leave the coordinate system  $\Psi_{\mathcal{W}}$  invariant. However, there are some natural “degenerations” of spectral networks which interpolate between “inequivalent”  $\mathcal{W}$ . We have already seen these degenerations in this introduction: they occur in the families  $\mathcal{W}_{\vartheta}$  when the parameter  $\vartheta$  is adjusted to a critical phase  $\vartheta_c$ . When such a degeneration occurs,  $\Psi_{\mathcal{W}}$  jumps by an automorphism  $\mathcal{K}$  of the torus of flat rank 1 connections on  $\Sigma$ . This  $\mathcal{K}$  is simply determined by the degeneracies  $\Omega(\gamma)$  read off from the degenerate network (see (149) for its explicit form.)  $\mathcal{K}$  has a form very similar to that of a cluster mutation, and part of our conjecture is that for the simplest types of degenerations it is indeed a mutation.

Very recently Goncharov has also defined some new geometric objects which he calls *spectral webs* [15], and which correspond to coordinate systems in the cluster atlas. Despite the similar nomenclature, this definition is different from our definition of spectral networks. It should be interesting to compare the two constructions.

*WKB expansions and spectral networks as Stokes diagrams.* One of the key tools in [3] was the WKB approximation, applied to the  $\zeta \rightarrow 0$  behavior of families of flat  $SL(2, \mathbb{C})$  connections of the form

$$\nabla(\zeta) = R \frac{\varphi}{\zeta} + D + R\zeta \bar{\varphi}. \quad (1)$$

Such families are associated with solutions of Hitchin’s equations [16] and naturally arise in theories of class  $S$ .

Some aspects of the WKB approximation are difficult to generalize to rank  $K > 2$ , and for a long time, this proved a stumbling block to giving nontrivial illustrations of the general statements of [17, 5] in the higher rank case. As we briefly explain in Sect. 10.9, the spectral networks  $\mathcal{W}_{\vartheta}$  in theories of class  $S$  appear to be the key missing ingredient to solve this problem; they provide a way of constructing a basis of  $\nabla(\zeta)$ -flat sections in each connected component of the complement of  $\mathcal{W}_{\vartheta}$ , with “good WKB asymptotics” as  $\zeta \rightarrow 0$  in a half-plane  $\mathbb{H}_{\vartheta}$ . Thus, the walls of the spectral networks  $\mathcal{W}_{\vartheta}$  also have an interpretation as Stokes curves. Similar statements have appeared before in the mathematical literature, e.g. [18, 19] and especially [20] which contains some examples of spectral networks with  $K = 3$ .

## Open problems

Our work in this paper leaves many directions unexplored or incompletely explored. Here are a few:

1. We discuss the theories  $S[\mathfrak{g}, C, D]$  only for  $\mathfrak{g}$  of type  $A$ . There should be a closely parallel story for  $\mathfrak{g}$  of type  $D$  or  $E$ . Some parts of this story are easy

<sup>3</sup> Again, here we are glossing over some slight differences between our setup and that of Fock–Goncharov, and ignoring some extra discrete “flag data” attached to the connections.

to predict: the walls of the spectral networks will be labeled (locally) by roots of  $\mathfrak{g}$ , and spectral networks will correspond to coordinate systems on moduli spaces of  $\mathfrak{g}$ -connections over  $C$ . It would be desirable to work out this picture in detail.

2. Even in the case of type  $A$ , our story is not quite complete. We take the charges of BPS states to be valued in the lattice  $\Gamma = H_1(\Sigma; \mathbb{Z})$ , but according to the more precise picture in [3, 21], it is better to identify the charge lattice as a certain subquotient of  $H_1(\Sigma; \mathbb{Z})$ . This is related to the fact that our nonabelianization map  $\Psi_{\mathcal{W}_\vartheta}$  most directly produces connections with structure group  $\mathrm{GL}(K, \mathbb{C})$ , while the physically relevant objects should really be connections with structure group  $SL(K, \mathbb{C})$ , considered modulo some further discrete equivalences [21]. There are likely to be some topological subtleties here which deserve careful consideration.
3. Many of the results of [3] for  $K = 2$  are subsumed in the present paper and extended to  $K > 2$ . However, there is one important piece of [3] which we have not extended to  $K > 2$ . Namely, we gave a recipe for computing an object called the “spectrum generator,” which completely determines the whole 4d BPS spectrum of the theory at any point  $u$ . This recipe uses only the combinatorics of a single spectral network  $\mathcal{W}_\vartheta$ , for a fixed generic  $\vartheta$ . It would be desirable to extend it to the case  $K > 2$ , and also to types  $D$  and  $E$ . One could further ask for an extension of the spectrum generator to include the 2d spectrum.<sup>4</sup>
4. It would be desirable to generalize our discussion of BPS indices to include information about the spins of the BPS states. This means taking  $y \neq \pm 1$  in the various indices defined below. Some aspects of the “motivic 2d–4d wall-crossing formula” which governs these degeneracies were spelled out in [5], but important details remain to be filled in.
5. In Sect. 10.7, we paint a heuristic picture of the “space of all spectral networks,” which would be nice to spell out more precisely.
6. We conjecture in Sect. 10.8 that the coordinate systems  $\Psi_{\mathcal{W}}$  on the space  $\mathcal{M}$  of flat connections are cluster coordinate systems. It would be very interesting to prove this conjecture. It would be even more interesting if it turned out that the  $\Psi_{\mathcal{W}}$  actually exhaust the set of cluster coordinate systems. This might be of some internal use for the study of cluster varieties, e.g. for the positivity conjectures of [13].
7. In this paper, we always require that the curve  $C$  carries at least one defect  $D_n$ . This restriction (which we also imposed in [3]) leads to some simplifications. One reason for these simplifications is that the walls in  $\mathcal{W}_\vartheta$  are solutions of differential equations, and it seems that these equations generically imply that each wall is “attracted” asymptotically to some defect. In particular, any open path  $\wp$  on  $C$  meets  $\mathcal{W}_\vartheta$  only finitely many times.

This contrasts sharply with the case without defects: in that case, a typical open path on  $\wp$  may meet  $\mathcal{W}_\vartheta$  infinitely many times. We are

---

<sup>4</sup> In the case  $K = 2$  this has been worked out by Pietro Longhi.

hopeful that all of the constructions of this paper still make sense in that setting, but there will be issues of convergence to consider.

Fortunately, some things can be said about the asymptotic behavior of the walls even in the case without defects. For example, in the case  $K = 2$ , the walls in  $\mathcal{W}_\theta$  are horizontal trajectories of a holomorphic quadratic differential  $e^{2i\theta}\phi_2$ , and the leading asymptotic behavior of a generic such trajectory is classified by a virtual “limit cycle” in  $H_1(C, \mathbb{R})$ , with deviations governed by certain Lyapunov exponents; see [22] for a very useful review. One can hope that using these kinds of results (and their to-be-developed generalizations for  $K > 2$ ) will be possible to extend everything in this paper to the case without defects.

8. In this paper, we mainly work at a generic point of the Coulomb branch, where the gauge symmetry group in the IR is abelian. If the parameters at the defects are adjusted carefully (or if there are no defects at all), there may also be points where the unbroken gauge symmetry is nonabelian. The physics of this situation is much less explored than the fully abelian case. Nevertheless, much of what we have said should have an extension to this situation.

In particular, a spectral network  $\mathcal{W}$  associated to a  $K$ -fold covering  $\Sigma \rightarrow C$  should give not only a map  $\Psi_{\mathcal{W}}$  between moduli of flat  $\mathrm{GL}(1)$  connections on  $\Sigma$  and moduli of flat  $\mathrm{GL}(K)$  connections on  $C$ , but also more generally a map  $\Psi_{\mathcal{W}, N}$  between moduli of flat  $\mathrm{GL}(N)$  connections on  $\Sigma$  and moduli of flat  $\mathrm{GL}(NK)$  connections on  $C$ .<sup>5</sup> Our construction of  $\Psi_{\mathcal{W}}$  in Sect. 10 is set up in a way that should generalize directly to this setting. We understand that similar constructions will appear in upcoming work of Goncharov and Kontsevich.

9. In this paper, we encounter several tricky sign issues. One of these first pops up as an ambiguity in the notion of “fermion number,” which leads to an ambiguity in the sign of the 2d BPS degeneracies, and recurs many times thereafter. We have found a scheme for fixing this sign ambiguity, which we use systematically throughout the paper: very roughly speaking, it amounts to considering paths on  $C$  and  $\Sigma$  weighted by signs which keep track of the parity of the number of times the tangent direction to the path winds around the circle. Our slavish implementation of this scheme has various consequences, leading us, e.g., to consider *twisted* flat connections on  $C$  and  $\Sigma$  in Sect. 10 rather than ordinary flat connections. (Twisted connections on  $C$  also appeared in [13], and this was a useful clue which helped us to find our sign prescription.) While this scheme leads to a consistent picture both mathematically and physically, we cannot say that we have really understood from physical first principles *why* it works so well. It would be very good to have a better understanding of

---

<sup>5</sup> One way of thinking about this is that the  $K$ -fold covering  $\Sigma$  could arise as a non-reduced degeneration of an  $NK$ -fold covering, where the sheets coalesce in groups of  $N$ . Readers who prefer to think in terms of M-theory fivebranes might say that we consider the theory of  $NK$  fivebranes wrapped on  $C \subset T^*C$ , and then move to a point of the Coulomb branch represented by  $N$  fivebranes wrapped on  $\Sigma \subset T^*C$ .

how our sign rule arises from the physics of the six-dimensional theory  $S[\mathfrak{g}]$ .

10. In this paper, we consider spectral networks  $\mathcal{W}_\vartheta$  associated to a phase  $\vartheta$  and a point  $u$  of the Coulomb branch. It is natural to ask whether every spectral network  $\mathcal{W}$  (modulo the natural notion of equivalence described in Sect. 10.6) arises as  $\mathcal{W}_\vartheta$  for some  $(u, \vartheta)$ . If the answer is “no,” can we classify those spectral networks which do occur?
11. It is natural to ask whether the framed BPS degeneracies can be categorified. We have done some work along these lines with E. Witten, and we hope to return to it.
12. The networks  $\mathcal{W}_\vartheta$  we study on  $C$  are made up of “ $\mathcal{S}$ -walls,” the loci where framed 2d–4d BPS state degeneracies jump. They depend on a point  $u$  of the Coulomb branch  $\mathcal{B}$ , so really  $\mathcal{W}_\vartheta = \mathcal{W}_\vartheta(u)$  although we usually do not write this dependence explicitly. On the other hand, very similar networks  $\mathcal{D}_\vartheta$  appear directly on  $\mathcal{B}$ : they are made up of “ $\mathcal{K}$ -walls,” the loci where 4d framed BPS state degeneracies jump.  $\mathcal{D}_\vartheta$  can be studied by methods very similar to the methods we employ here for  $\mathcal{W}_\vartheta$ . In particular, this gives another scheme for computing the pure 4d BPS state degeneracies  $\Omega(\gamma)$ , parallel to what we do here for  $\mu(a)$ . (This has also been pointed out by Kontsevich–Soibelman.) Networks very similar to  $\mathcal{D}_\vartheta$  have appeared previously in the mirror symmetry literature, e.g. [23–27].

The family of networks  $\mathcal{W}_\vartheta(u)$  on  $C$  and the single network  $\mathcal{D}_\vartheta$  on  $\mathcal{B}$  can be unified into a network  $\mathcal{DW}_\vartheta$  on  $\mathcal{B} \times C$ . We believe that this is really the most natural perspective, although we do not adopt it explicitly in this paper.

13. We give a recipe for determining the BPS degeneracies  $\Omega(\gamma)$  from a degenerate spectral network, and work out several examples, but not many. It would be interesting to use spectral networks to study the spectrum of concrete theories, beyond the few examples we consider in Sect. 8. One obvious possibility would be to consider the standard  $SU(K)$  gauge theories coupled to fundamental hypermultiplets. Even the BPS spectrum of the pure  $SU(3)$  theory has been the source of some controversy; in this paper, we study it at strong coupling but do not analyze the more intricate weak-coupling spectrum. It would be interesting to compare results obtained from spectral networks with those in [28–31].
14. In the degenerate spectral networks, we do examine and we do not work out any example where the resulting  $|\Omega(\gamma)| > 2$ , which would suggest the possibility of higher spin states in the BPS spectrum. It would be interesting to find a degenerate spectral network which gives such higher spin states.
15. In this paper, we have made some significant progress in the determination of BPS spectra for a large class of  $\mathcal{N} = 2$  field theories. We hope that these techniques can shed light on some broader questions of general interest, such as to what extent the BPS spectrum uniquely determines a theory, and whether one can always engineer a theory to produce a desired (consistent) BPS spectrum.



Finally, we mention that in a companion paper [32], we apply the techniques of this paper to make a more direct connection to the work of Fock and Goncharov on higher Teichmüller theory [13]. We also find a formula for the spectrum generator for a special class of spectral networks, valid for all  $K \geq 2$ ; this gives an implicit determination of the BPS spectrum in the corresponding theories.

## 2. A Brief Review of Theories of Class S

In [1–3], a large and interesting class of  $\mathcal{N} = 2$ ,  $d = 4$  theories was studied. These theories, which we call “theories of class S,” can be constructed by compactification and twisting of  $\mathcal{N} = (2, 0)$ ,  $d = 6$  theories. Here, we quickly review the basic features of this construction.

Fix a compact Riemann surface  $C$ , with  $s$  punctures at points  $\mathfrak{s}_1, \dots, \mathfrak{s}_s$ . Also fix a Lie algebra  $\mathfrak{g}$  of  $ADE$  type. In this paper, we will focus on the case  $\mathfrak{g} = A_{K-1}$ . We consider the  $\mathcal{N} = (2, 0)$  theory  $S[\mathfrak{g}]$  compactified on  $C$ , and partially twisted as described in [3].  $S[\mathfrak{g}]$  admits half-BPS codimension-2 defects; we put one such defect  $D_n$  at each puncture  $\mathfrak{s}_n$ .

The main statements of this paper will be independent of the choice of which type of defect we put at each puncture, but it is useful to have an example in mind to fix ideas. There is a class of *regular* defects labeled by Young diagrams with  $K$  boxes [2, 3]. In particular, we can consider a *full* regular defect, corresponding to the Young diagram with a single row. Each full defect admits a natural mass deformation depending on  $K$  complex parameters  $m_n^{(i)}$ , with  $\sum_{i=1}^K m_n^{(i)} = 0$ , which we are free to fix arbitrarily.

In any case, whatever collection  $D$  of defects we choose, the compactification procedure yields an  $\mathcal{N} = 2$ ,  $d = 4$  theory  $S[\mathfrak{g}, C, D]$ . This theory is our main object of study.

The Coulomb branch of  $S[\mathfrak{g}, C, D]$  consists of tuples  $(\phi_2, \dots, \phi_K)$ , where  $\phi_r$  is an  $r$ -differential on  $C$  (i.e. a section of  $K_C^{\otimes r}$ ), which is holomorphic away from the punctures  $\mathfrak{s}_n$ , and has some prescribed singular behavior at the  $\mathfrak{s}_n$ . For example, if we choose  $D_n$  to be a full defect, then the prescription is that for each  $r$ ,  $\phi_r$  has a pole of order  $r$  at  $\mathfrak{s}_n$ , with residue determined by a combination of the parameters  $m_n^{(i)}$  [2, 3]. Having fixed  $(\phi_2, \dots, \phi_K)$ , the Seiberg–Witten curve is given by

$$\Sigma = \left\{ \lambda : \lambda^K + \sum_{r=2}^K \phi_r \lambda^{K-r} = 0 \right\} \subset T^*C. \quad (2)$$

$\Sigma$  is a  $K$ -fold branched cover of  $C$ . Since  $\Sigma$  sits inside  $T^*C$ , it carries a canonical 1-form, the restriction of the Liouville 1-form, which by slight abuse of notation we will also call  $\lambda$ .

In this paper, it will be crucial to introduce a half-BPS *surface defect* into  $S[\mathfrak{g}, C, D]$ . There is a canonical such defect  $\mathbb{S}_z$  [4, 5, 33], depending only on a point  $z \in C$  (and on a representation of  $\mathfrak{g}$ , but since we choose  $\mathfrak{g} = A_{K-1}$  we can just take the fundamental representation.) When  $z$  is generic, the defect  $\mathbb{S}_z$  has  $K$  distinct massive vacua, which corresponds to the  $K$  solutions of (2)

at  $z$ ; locally, we may denote these by  $z^{(1)}, \dots, z^{(K)}$ . From the point of view of the  $d = 6$  theory  $S[\mathfrak{g}]$ ,  $\mathbb{S}_z$  is obtained by inserting a surface defect at the point  $z$ .

### 3. Geometric Description of BPS States

#### 3.1. 4d BPS States

Now we recall the geometric description of BPS states in  $S[\mathfrak{g}, C, D]$ .

All such BPS states arise from BPS strings of the six-dimensional  $S[\mathfrak{g}]$ , which are extended along  $C$  and hence look like point particles in the remaining  $3 + 1$  dimensions. Roughly there are  $\binom{K}{2}$  distinct kinds of BPS string, but the topological twisting and compactification makes the story slightly trickier, as we now recall.

An oriented segment of string passing through a point  $z \in C$  is labeled by the choice of a pair of distinct sheets of the  $K$ -fold covering  $\Sigma \rightarrow C$ , i.e. two solutions  $\lambda^{(i)}, \lambda^{(j)} (i \neq j)$  of the degree- $K$  polynomial equation (2), in a neighborhood of  $z$ . To keep track of this discrete label, we use the term “ $ij$ -string” rather than just “string.” Reversing orientation exchanges  $ij$ -strings and  $ji$ -strings.

The  $\mathcal{N} = 2$  central charge  $Z$  of an  $ij$ -string is obtained by integrating the complex 1-form  $\frac{1}{\pi}(\lambda^{(i)} - \lambda^{(j)})$  along the string. The mass  $M$  of an  $ij$ -string is obtained by integrating the real density  $\frac{1}{\pi}|\lambda^{(i)} - \lambda^{(j)}|$ .

When is such a string BPS? Introduce a local coordinate on  $C$  by  $w^{(ij)} = \int \lambda^{(i)} - \lambda^{(j)}$ , and then define an  $ij$ -trajectory with phase  $\vartheta$  to be a straight line in the  $w^{(ij)}$ -coordinate, with inclination  $\vartheta$ , i.e. a line along which

$$\mathrm{Im}(e^{-i\vartheta} \dot{w}^{(ij)}) = 0. \quad (3)$$

An  $ij$ -trajectory is naturally oriented: the positive direction is the direction in which  $\mathrm{Re}(e^{-i\vartheta} w^{(ij)})$  increases. Reversing orientation of an  $ij$ -trajectory gives a  $ji$ -trajectory. An  $ij$ -string is BPS if and only if it is stretched along an  $ij$ -trajectory with some phase  $\vartheta$ ; we call such a string a “BPS string of phase  $\vartheta$ .” We see at once that for BPS strings  $|Z| = M$  as expected. The phase  $\vartheta$  determines which supercharges the BPS string preserves.

An  $ij$ -string can end in two ways. First, it can end on an  $(ij)$ -branch point, i.e. a point where  $\lambda^{(i)} - \lambda^{(j)} = 0$ .<sup>6</sup> Second, it can end on a junction where an  $ij$ -string,  $jk$ -string and  $ki$ -string meet (all oriented into the junction.) In the latter case, for the combined web of strings to be BPS, all three strings must be BPS strings with the *same* phase  $\vartheta$ .

BPS states in  $S[\mathfrak{g}, C, D]$  arise from webs of BPS strings, such that all strings in the web have finite total central charge. This condition means that the strings are either have closed loops or have both ends on branch points or junctions. We call these *finite webs*.

Some possible topologies for finite webs are shown in Fig. 2.

<sup>6</sup> We use the notation  $(ij)$  for the branch points but  $ij$  for the trajectories; the  $(ij)$  is meant to denote the *transposition* associated to the branch point, which exchanges sheet  $i$  and sheet  $j$  of the covering  $\Sigma \rightarrow C$ . In particular  $(ij) = (ji)$ .

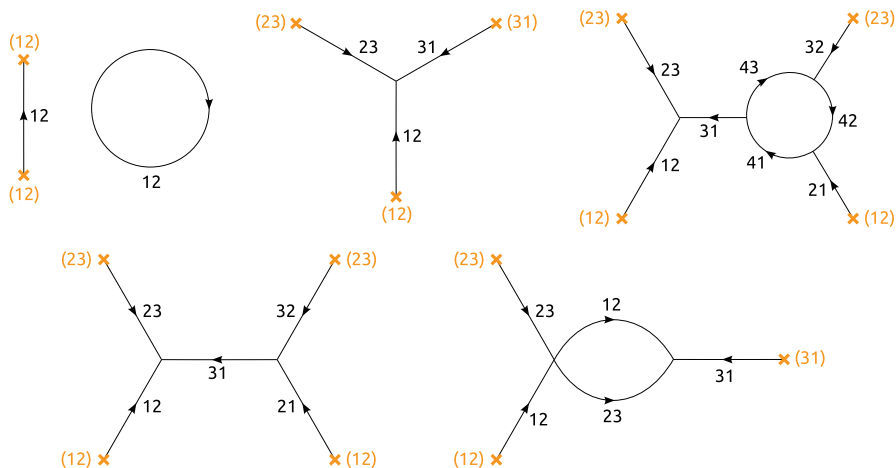


FIGURE 2. Some possible topologies for finite webs of BPS strings. An *orange cross* with label  $(ij)$  denotes an  $(ij)$ -branch point. Wherever a string with label  $ij$  appears, it could equally well have been represented by a string with label  $ji$  and the opposite orientation (color figure online)

The charges of the BPS states are determined by the topology of the webs, in the following way. An  $ij$ -string stretched along an oriented path  $p$  on  $C$  can be lifted in a canonical way to  $p_\Sigma$ , a union of oriented curves on  $\Sigma$ : namely,  $p_\Sigma$  is the union of the lift  $p^{(i)}$  of  $p$  to the  $i$ th sheet and the lift  $-p^{(j)}$  of  $-p$  ( $p$  with reversed orientation) to the  $j$ th sheet. Letting  $p$  run over the strings in a finite web  $N$ , the union of the  $p_\Sigma$  is a closed 1-cycle  $N_\Sigma$  on  $\Sigma$ .  $N_\Sigma$  has a homology class  $[N_\Sigma] \in \Gamma := H_1(\Sigma; \mathbb{Z})$ . This  $[N_\Sigma]$  is the charge of the BPS state.<sup>7</sup>

The central charge of a finite web  $N$  is the sum of the central charges of the strings in  $N$ ; this gives a simple result which depends only on  $\gamma = [N_\Sigma]$ ,

$$Z_\gamma = \frac{1}{\pi} \oint_\gamma \lambda. \quad (4)$$

To determine the BPS spectrum of the theory, in particular, the second helicity supertrace  $\Omega(\gamma)$ , one should in principle proceed by quantizing the zero modes of each finite web. This would give some definite formula for the contribution of each finite web to  $\Omega(\gamma)$ . In practice, such a quantization has not been completely carried out, and we will not do it here either.<sup>8</sup> Rather, we will

<sup>7</sup> The precise charge lattice of the theory  $S[\mathfrak{g}, C, D]$  is actually a subquotient of  $H_1(\Sigma; \mathbb{Z})$ , as explained in [3]. Nevertheless, in this paper, we consider  $\Gamma = H_1(\Sigma; \mathbb{Z})$  for simplicity.

<sup>8</sup> In [3] we used wall-crossing to determine  $\Omega(\gamma)$  in the case  $K = 2$ , reproducing earlier results of [6]. That case is particularly simple since there are only two possible topologies for finite webs: one can either have a single string connecting two branch points or a single closed loop. These topologies contribute  $\Omega(\gamma) = +1$  and  $\Omega(\gamma) = -2$ , respectively. We will reproduce this result yet again in Sect. 5.7 below.

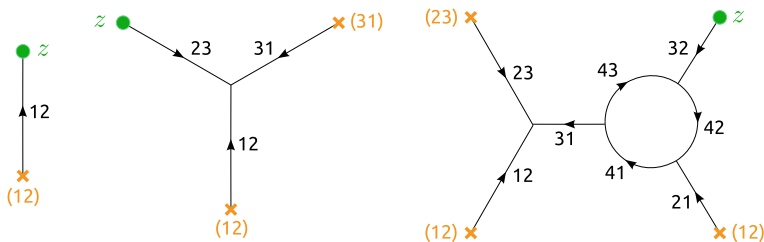


FIGURE 3. Some possible topologies for finite open webs of BPS strings, representing BPS solitons on the surface defect  $\mathbb{S}_z$ . Each finite open web includes one string that ends on the point  $z \in C$ . An *orange cross* with label  $(ij)$  denotes an  $(ij)$ -branch point (color figure online)

explain a more indirect route, which determines  $\Omega(\gamma)$  using the interactions between 2d and 4d BPS states.

One could also ask about extending  $\Omega(\gamma)$  to an object which keeps track of the spins of the BPS multiplets appearing. Such “refined BPS degeneracies” and “protected spin characters” have been considered in the  $\mathcal{N} = 2$  context e.g. in [9, 34–36, 21]. We will not consider that extension in this paper.

### 3.2. Solitons

In the presence of the canonical surface defect  $\mathbb{S}_z$ , there is a second kind of BPS state in the story: we can consider BPS particles which are bound to the defect and interpolate between distinct vacua. We call these particles *solitons*.

As described in [5], BPS solitons are also realized geometrically in terms of finite webs of strings. The main difference is that now we consider webs in which one of the strings ends on the point  $z$ , which we call *finite open webs*. See Fig. 3.

Let  $N$  denotes a finite open web. As for the pure 4d case, the charge of the corresponding BPS soliton is determined by the topology of the web as follows. Let  $N_\Sigma$  denotes the union of the lifts of all strings in  $N$  to  $\Sigma$ . Suppose the string of  $N$  ending on  $z$  is an  $ij$ -string oriented out of  $z$ . Let  $z^{(i)}$  and  $z^{(j)}$  be the preimages of  $z$  on the  $i$ th and  $j$ th sheets of  $\Sigma$ .  $N_\Sigma$  is a 1-chain with boundary:

$$\partial N_\Sigma = z^{(j)} - z^{(i)}. \quad (5)$$

We let  $\Gamma_{ij}(z, z)$  denotes the set of relative homology classes on  $\Sigma$  obeying (5), and

$$\Gamma(z, z) = \cup_{i,j} \Gamma_{ij}(z, z). \quad (6)$$

Thus, the charges  $[N_\Sigma]$  of BPS solitons on  $\mathbb{S}_z$  are elements of  $\Gamma(z, z)$ , and in fact of the smaller set  $\cup_{i,j, i \neq j} \Gamma_{ij}(z, z)$ .

The central charges of BPS solitons are given by a formula analogous to (4): for any  $\bar{a} \in \Gamma(z, z)$  we have<sup>9</sup>

$$Z_{\bar{a}} = \frac{1}{\pi} \int_{\bar{a}} \lambda. \quad (7)$$

There is a BPS index  $\mu(\bar{a}) \in \mathbb{Z}$  which counts 2d–4d BPS states of charge  $\bar{a}$ , defined as a trace over the Hilbert space of 1-particle BPS states [37, 38],

$$\mu(\bar{a}) = \text{Tr}_{\mathcal{H}_{\mathbb{S}_z, \bar{a}}^{1, \text{BPS}}} F e^{i\pi F}, \quad (8)$$

where  $F$  is a fermion number operator. In principle,  $\mu(\bar{a})$  could be determined by quantizing the zero modes of the BPS strings. In practice, as with  $\Omega(\gamma)$ , we will take a more indirect route in this paper.

We hasten to warn the reader that there are two important subtleties in the definition (8). The first subtlety is addressed in Sect. 3.5: to define  $F$  properly, we will need to keep track of slightly more information about the 2d–4d BPS states, by extending the charge  $\bar{a}$  to a new “charge” denoted  $a$ . It will turn out that  $\mu$  really depends on  $a$ , not on  $\bar{a}$ . The second subtlety is that the space of 1-particle BPS states is only well defined when the parameters are not on a wall of marginal stability. At these walls,  $\mu(a)$  can jump. In particular, if  $\bar{a} \in \Gamma(z, z)$ , then generically, if we move  $z$  holding all other parameters fixed, and apply the natural parallel transport to the relative homology cycle  $\bar{a}$ ,  $\mu(a)$  remains constant. However, there are walls on  $C$  across which  $\mu(a)$  jumps, and in this case, we will need to define limits  $\mu^\pm(a)$  as  $z$  approaches the wall from either side. This will be important in Sect. 6.

### 3.3. Framed 2d–4d BPS States

Our approach to the BPS spectrum will involve an auxiliary device, the *framed* BPS states. These were introduced in the pure 4d context in [21], and in the 2d–4d context in [5]. Our interest in this paper is in the 2d–4d version.

Consider a pair of points  $z_1, z_2$  in  $C$ , and a path  $\wp$  in  $C$  from  $z_1$  to  $z_2$ . Also fix a parameter  $\vartheta \in \mathbb{R}$ . These data determine a pair of surface defects  $\mathbb{S}_{z_1}$  and  $\mathbb{S}_{z_2}$  in  $S[\mathfrak{g}, C, D]$  along with a supersymmetric interface  $L_{\wp, \vartheta}$  between the two surface defects. The interface  $L_{\wp, \vartheta}$  preserves 2 out of the 4 supercharges preserved by the surface defects; which 2 supercharges are preserved is determined by the parameter  $e^{i\vartheta}$ , as explained in [5].<sup>10</sup> It is generally believed that the defect  $L_{\wp, \vartheta}$  does not depend on the precise path  $\wp$  but only on its homotopy class. In the present paper, we will take this as an assumption, and will find a very consistent picture (although of course the quantities we study are somewhat protected by supersymmetry); see e.g. [39] for some related discussion. Indeed, we will find that the constraint of homotopy invariance is very strong.

<sup>9</sup> In [5] we denoted elements of  $\Gamma_{ij}(z, z)$  by  $\gamma_{ij}$ , but we are now deprecating that notation in favor of  $\bar{a}$ . See Appendix C for a summary of conventions.

<sup>10</sup> The parameter which was called  $\zeta$  in [5] is here given by  $\zeta = e^{i\vartheta}$ .



FIGURE 4. Two surface defects connected by an interface. (This picture lives in the three-dimensional space where the field theory  $S[\mathfrak{g}, C, D]$  is defined; we have factored out the time direction.)

Now we study the 4d theory  $S[\mathfrak{g}, C, D]$  with the defects  $\mathbb{S}_{z_1}$  and  $\mathbb{S}_{z_2}$  inserted on two half-lines  $\{x^1 = 0, x^2 = 0, x^3 > 0\}$  and  $\{x^1 = 0, x^2 = 0, x^3 < 0\}$  respectively, separated by the interface  $L_{\varphi, \vartheta}$  (see Fig. 4).

We then define (2d–4d) *framed BPS states* to be states of the 1-particle Hilbert space  $\mathcal{H}_{L_{\varphi, \vartheta}}^1$  of this combined system which are fully supersymmetric, i.e. preserve the 2 supercharges present in the system.

The 2d–4d framed BPS states should be thought of as different “vacuum states” of the interface  $L_{\varphi, \vartheta}$ . To make this statement sharp, however, we need to impose a constraint. We say  $\vartheta$  is *generic* if there is no charge  $\gamma \in \Gamma$  with  $e^{-i\vartheta} Z_\gamma \in \mathbb{R}_-$ . We say the pair  $(\varphi, \vartheta)$  is *generic* if  $\vartheta$  is generic and also there is no charge  $\bar{a} \in \Gamma(z_1, z_1)$  or  $\bar{a} \in \Gamma(z_2, z_2)$  with  $e^{-i\vartheta} Z_{\bar{a}} \in \mathbb{R}_-$ , where  $z_1, z_2$  are the initial and final points of  $\varphi$ . When the pair  $(\varphi, \vartheta)$  is generic, the 2d–4d framed BPS states are indeed localized near the interface, in the sense that they do not mix with the continua of unframed 2d–4d or 4d BPS states.

The classification of 2d–4d framed BPS states by charges is similar to that for the 2d–4d unframed BPS states in Sect. 3.2. Instead of paths on  $\Sigma$  between two lifts  $z^{(i)}, z^{(j)}$  of a single point  $z$ , now we consider paths on  $\Sigma$  from a lift  $z_1^{(i)}$  of  $z_1$  to a lift  $z_2^{(j)}$  of  $z_2$ . We let  $\Gamma_{ij}(z_1, z_2)$  denotes the set of relative homology classes represented by such paths, and  $\Gamma(z_1, z_2) = \cup_{i,j} \Gamma_{ij}(z_1, z_2)$ .<sup>11</sup>

There is a “framed BPS index”  $\overline{\Omega}(L_{\varphi, \vartheta}, \bar{a}; y)$  counting 2d–4d framed BPS states with charge  $\bar{a} \in \Gamma_{ij}(z_1, z_2)$ . It was defined in Sect. 4.4 of [5]:<sup>12</sup>

$$\overline{\Omega}(L_{\varphi, \vartheta}, \bar{a}; y) = \text{Tr}_{\mathcal{H}_{L_{\varphi, \vartheta}}^{1, \text{BPS}}, \bar{a}} e^{i\pi F} (-y)^{\mathcal{J}}. \quad (9)$$

In this paper, we will concentrate on the indices at  $y = 1$ , which will be denoted by  $\overline{\Omega}'(L_{\varphi, \vartheta}, \bar{a})$ . We leave the generalization of our story to arbitrary  $y$  as an important open problem.

<sup>11</sup> In [5], we denoted elements of  $\Gamma(z_1, z_2)$  by  $\gamma_{ij'}$ , but we are now deprecating that notation in favor of  $\bar{a}$ . See Appendix C for a summary of conventions.

<sup>12</sup> As in [5],  $F$  denotes a generator of  $u(1)_V$  of the 2d (2, 2) supersymmetry algebra preserved by  $L_{\varphi, \vartheta}$ , and  $\mathcal{J} = 2J_{12} + 2I_{12}$  is a linear combination of rotation and  $su(2)_R$  generators of the 4d  $\mathcal{N} = 2$  supersymmetry. There is a subtlety here which we address in Sect. 3.5: to define  $F$  properly we will need to keep track of slightly more information about the 2d–4d framed BPS states, by extending the charge  $\bar{a}$  to a new “charge” denoted  $a$ . It will turn out that  $\overline{\Omega}$  really depends on  $a$ , not  $\bar{a}$ .

Framed 2d–4d BPS states admit a geometric description somewhat similar to those given above for unframed 2d–4d BPS states. We will not describe it explicitly here.<sup>13</sup>

### 3.4. Enhanced Degeneracies

In the presence of a surface defect  $\mathbb{S}_z$ , there is an important enhancement to the 4d BPS degeneracies  $\Omega(\gamma)$  [5]: they are replaced by numbers  $\omega(\gamma, \bar{a}) \in \mathbb{Z}$  for any  $\gamma \in \Gamma$  and  $\bar{a} \in \Gamma(z_1, z_2)$ .  $\omega$  is “linear” in its second argument, i.e. it obeys

$$\omega(\gamma, \bar{a} + \bar{b}) = \omega(\gamma, \bar{a}) + \omega(\gamma, \bar{b}) \quad (10)$$

when  $\bar{a} + \bar{b}$  is defined (that is, when the end of  $\bar{a}$  coincides with the start of  $\bar{b}$ ). Moreover,  $\omega$  obeys

$$\omega(\gamma, \bar{a} + \gamma') - \omega(\gamma, \bar{a}) = \Omega(\gamma) \langle \gamma, \gamma' \rangle \quad (11)$$

for  $\gamma' \in \Gamma$ . In particular, this equation is sufficient to determine  $\Omega$  if we know  $\omega$ . (More precisely, it determines  $\Omega(\gamma)$  for all  $\gamma$  not in the kernel of  $\langle \cdot, \cdot \rangle$ , i.e. all  $\gamma$  which are not pure flavor charges.)

The  $\omega(\gamma, \bar{a})$  are a bit subtle to interpret directly in terms of traces over Hilbert spaces. In [5], the most general interpretation we found was in terms of a Hilbert space of “halo states” (analogues of the ones studied by Frederik Denef<sup>14</sup>), which induces 2d–4d wallcrossing. The individual states in this Hilbert space may be interpreted either as 4d particles carrying charge  $\gamma$  or as 2d particles living on the surface defect and carrying the same charge. This fact will become relevant in one of our concrete examples, in Sect. 8.3.

### 3.5. A Problem of Signs

We must now confront a pesky but important detail.

There is an ambiguity in (9): the generators  $F$  and  $\mathcal{J}$  are well defined only up to  $c$ -number shifts. As a result, the index  $\overline{\Omega}'(L_{\varnothing, \vartheta}, \bar{a})$  suffers from some potential ambiguity (which even depends on  $\varnothing$ ). We can partially fix this ambiguity by requiring that  $\overline{\Omega}'(L_{\varnothing, \vartheta}, \bar{a})$  is real, but this still leaves the possibility of an integer shift of  $F$  or  $\mathcal{J}$ . Such a shift would reverse the sign of  $\overline{\Omega}'(L_{\varnothing, \vartheta}, \bar{a})$ . Therefore, a priori, we would expect that we need some additional data in order to fix  $\overline{\Omega}'(L_{\varnothing, \vartheta}, \bar{a})$  uniquely. A similar ambiguity afflicts the 2d–4d BPS degeneracies  $\mu(\bar{a})$ : the definition (8) depends on the choice of generator  $F$ , and changing this choice can change  $\mu(\bar{a})$  by a sign.

There is no difficulty in fixing these ambiguities locally in any particular corner of parameter space. In this paper, though, we will mainly be concerned with phenomena which occur when parameters are varied, sometimes over long distances in parameter space. It is, therefore, desirable to have a global way of fixing these sign ambiguities, which is consistent with all of the physical constraints, and ideally one that does not depend on any arbitrary choices.

<sup>13</sup> In the case  $K = 2$  (i.e.  $\mathfrak{g} = A_1$ ) we did explain the relevant objects in [21], where we called them “millipedes.”

<sup>14</sup> See [21, 40–42] for a description of the halo approach to wall-crossing.

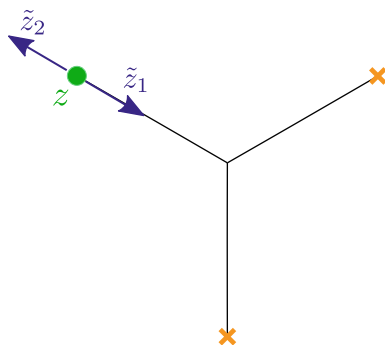


FIGURE 5. Two tangent directions at  $z$  determined by a finite web  $N$ :  $\tilde{z}_1$  points “into” the finite web while  $\tilde{z}_2$  points “away”

We have found such a rule, which we now describe. It would be desirable to give a first-principles *derivation* of this rule from the physics of the  $(2, 0)$  theory  $S[\mathfrak{g}]$  in six dimensions.

It will be useful to keep track of a bit more information about the finite webs  $N$  representing 2d–4d BPS states. For any real surface  $S$ , let  $\tilde{S}$  denote the circle bundle of tangent directions to  $S$ . For  $z \in S$  we will let  $\tilde{z}$  denote a lift of  $z$  to  $\tilde{S}$ , that is, a choice of tangent direction at  $z$ . Any smooth path  $\wp$  on  $S$  carries a natural tangent direction field and hence has a canonical lift to a path  $\tilde{\wp}$  on  $\tilde{S}$ ; we will use this lift often. There is a distinguished class  $H \in H_1(\tilde{\Sigma}; \mathbb{Z})$ , represented by a path which winds once around a fiber of  $\tilde{\Sigma}$  (the choice of fiber does not matter).

By smoothing out the junctions slightly, we can deform  $N_\Sigma$  into a finite union of *smooth* paths on  $\Sigma$ , which thus has a canonical lift to a path  $N_{\tilde{\Sigma}}$  on  $\tilde{\Sigma}$ . Let  $\tilde{z}_1$  and  $\tilde{z}_2$  denote the tangent directions to  $C$  at  $z$  shown in Fig. 5. Let  $\tilde{z}_1^{(i)}$  be the lift of  $\tilde{z}_1$  to the  $i$ th sheet of  $\tilde{\Sigma}$ , and similarly define  $\tilde{z}_2^{(j)}$ . The path  $N_{\tilde{\Sigma}}$  then has

$$\partial N_{\tilde{\Sigma}} = \tilde{z}_2^{(j)} - \tilde{z}_1^{(i)}. \quad (12)$$

Let  $\tilde{\Gamma}(\tilde{z}_1, \tilde{z}_2)$  denote the set of relative homology classes on  $\tilde{\Sigma}$  obeying (12), *modulo* shifts by the class  $2H$ . There is an exact sequence

$$0 \rightarrow \mathbb{Z}/2\mathbb{Z} \rightarrow \tilde{\Gamma}(\tilde{z}_1, \tilde{z}_2) \rightarrow \Gamma(z_1, z_2) \rightarrow 0. \quad (13)$$

The relative homology class  $[N_{\tilde{\Sigma}}] \in \tilde{\Gamma}(\tilde{z}_1, \tilde{z}_2)$  thus keeps track of one extra  $\mathbb{Z}/2\mathbb{Z}$  worth of information beyond that in  $[N_\Sigma] \in \Gamma(z, z)$ . There can be different open finite BPS webs  $N$  carrying the same charge  $[N_\Sigma]$  but with different lifts  $[N_{\tilde{\Sigma}}]$ ; see Fig. 6 for an example.

Now we can explain our proposal for how the sign of  $\mu$  behaves. Given a charge  $\bar{a} \in \Gamma(z, z)$ ,  $\mu(\bar{a})$  is not well defined (although it is well defined up to sign). In order to make it well defined, we propose that we must choose a class  $a \in \tilde{\Gamma}(\tilde{z}_1, \tilde{z}_2)$  which projects to  $\bar{a}$  in (13). Having done so, there should be a way of fixing the ambiguity of  $F$  to obtain a well-defined BPS degeneracy,



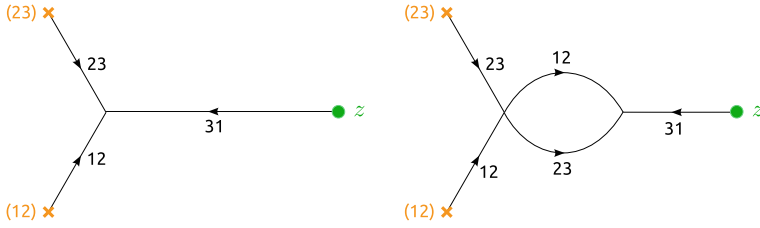


FIGURE 6. Two open finite BPS webs  $N$  on the same patch of  $C$ , drawn displaced from one another for clarity. These two webs carry the same charge  $[N_\Sigma]$  but have different lifts  $[N_\Sigma]$ . Informally, one can say that the lifts of these two webs “differ by one unit of winding”

which we call  $\mu(a)$ . Two choices  $a, a'$  differ by a winding number  $w(a, a')$ ; the corresponding  $\mu$  should obey

$$\mu(a)/\mu(a') = (-1)^{w(a, a')}. \quad (14)$$

(Of course, knowing  $|\mu(a)|$  for all  $a$  and knowing (14) is still not enough by itself to determine  $\mu$ . Later in this paper we will fix one more convention, in (23), after which we will be able to calculate  $\mu(a)$  for all  $a$ .)

A similar discussion applies to  $\overline{\Omega}'(L_{\varphi, \vartheta}, \bar{a})$ , but with the extra complication that now we have to discuss both the dependence on the path  $\varphi$  and on the charge  $\bar{a}$ . First suppose we *fix* the path  $\varphi$  from  $z_1$  to  $z_2$ , with initial tangent vector  $\tilde{z}_1$  and final tangent vector  $\tilde{z}_2$ . Then, we propose that the situation is strictly parallel to our discussion of  $\mu$  above: the degeneracies  $\overline{\Omega}'(L_{\varphi, \vartheta}, \bar{a})$  are not well defined (although they are well defined up to sign), and what is really defined is an integer  $\overline{\Omega}'(L_{\varphi, \vartheta}, a)$ , where  $a \in \tilde{\Gamma}(\tilde{z}_1, \tilde{z}_2)$  is a lift of  $\bar{a} \in \Gamma(z_1, z_2)$ . The dependence on the choice of lift is given by

$$\overline{\Omega}'(L_{\varphi, \vartheta}, a)/\overline{\Omega}'(L_{\varphi, \vartheta}, a') = (-1)^{w(a, a')}. \quad (15)$$

Now let us also consider the dependence on  $\varphi$ . Recall from Sect. 3.3 that the line defect  $L_{\varphi, \vartheta}$  depends only on the homotopy class of the path  $\varphi$ . Nevertheless, the prescription for fixing the sign of  $\overline{\Omega}'(L_{\varphi, \vartheta}, a)$  can depend on more information than just the homotopy class. We propose that the sign actually depends on the homotopy class of the lift of  $\varphi$  from  $C$  to the bundle of tangent directions  $\tilde{C}$ . For two paths  $\varphi, \varphi'$  whose lifts to  $\tilde{C}$  have the same initial and final endpoints, there is a mod-2 winding number  $w(\varphi, \varphi')$  (defined similarly to the winding number we considered above for paths on  $\Sigma$ ). We propose that for such paths we have

$$\overline{\Omega}'(L_{\varphi, \vartheta}, a)/\overline{\Omega}'(L_{\varphi', \vartheta}, a) = (-1)^{w(\varphi, \varphi')}. \quad (16)$$

We summarize this proposal by saying that the framed 2d–4d BPS degeneracies of the interface  $L_{\varphi, \vartheta}$  are not quite homotopy invariants of  $\varphi$ , but rather are “twisted homotopy invariants” of  $\varphi$ .

## 4. Basics of Framed 2d–4d Indices

In this section, we study the basic properties of the indices  $\overline{\Omega}'(L_{\wp,\vartheta}, a)$  counting framed 2d–4d BPS states. We will organize these indices into a natural generating function  $F(\wp, \vartheta)$  which is the main player in this paper.

Throughout this section we assume  $(\wp, \vartheta)$  is generic, in the sense explained in Sect. 3.3.

### 4.1. Generating Functions of Framed 2d–4d Indices

The fundamental properties of  $\overline{\Omega}'(L_{\wp,\vartheta}, a)$  are most elegantly expressed in terms of a formal generating function  $F(\wp, \vartheta)$ .

The idea is to introduce a formal variable  $X_{\bar{a}}$  for each charge  $\bar{a} \in \Gamma(z_1, z_2)$ , and then define  $F(\wp, \vartheta) = \sum_{\bar{a} \in \Gamma(z_1, z_2)} \overline{\Omega}'(L_{\wp,\vartheta}, \bar{a}) X_{\bar{a}}$ . As we have noted in Sect. 3.5, though, we need to take some care here to deal with sign ambiguities. Therefore, more precisely, we introduce formal variables  $X_a$  for each  $a \in \tilde{\Gamma}(\tilde{z}_1, \tilde{z}_2)$ , subject to the relation that if  $a$  and  $a'$  project to the same class  $\bar{a} \in \Gamma(z_1, z_2)$  then we have

$$X_a/X_{a'} = (-1)^{w(a,a')}. \quad (17)$$

We then choose one representative  $a \in \tilde{\Gamma}(\tilde{z}_1, \tilde{z}_2)$  for each  $\bar{a} \in \Gamma(z_1, z_2)$ , and define

$$F(\wp, \vartheta) := \sum_{\bar{a} \in \Gamma(z_1, z_2)} \overline{\Omega}'(L_{\wp,\vartheta}, a) X_a. \quad (18)$$

$F(\wp, \vartheta)$  is independent of our choices of representatives  $a$ , thanks to (15), (17).

### 4.2. Formal Products and Composition

Suppose  $a$  and  $b$  are the relative homology classes of two open paths on  $\tilde{\Sigma}$ . If the end of  $a$  coincides with the start of  $b$ , we let  $a + b$  denote the relative homology class of the concatenation of  $a$  and  $b$ . Then we introduce a product law on our formal variables:

$$X_a X_b = \begin{cases} X_{a+b} & \text{if the end of } a \text{ is the start of } b, \\ 0 & \text{otherwise.} \end{cases} \quad (19)$$

We can now state one of the key properties of the generating functions  $F(\wp, \vartheta)$ : if  $\wp$  and  $\wp'$  are paths on  $C$  which can be concatenated to make a smooth path  $\wp\wp'$  (with the end of  $\wp$  attached to the start of  $\wp'$ ), and if both  $(\wp, \vartheta)$  and  $(\wp', \vartheta)$  are generic, then

$$F(\wp, \vartheta) F(\wp', \vartheta) = F(\wp\wp', \vartheta). \quad (20)$$

Here is the physical reason for (20). We consider three surface defects  $\mathbb{S}_{z_1}, \mathbb{S}_{z_2}, \mathbb{S}_{z_3}$ , connected by two interfaces  $L_{\wp,\vartheta}$  and  $L_{\wp',\vartheta}$ , as shown in Fig. 7. The Hilbert space  $\mathcal{H}$  of framed BPS states in this situation should be independent of the separation between the interfaces. For large separations, since the framed BPS states are localized near the interfaces,  $\mathcal{H}$  is a tensor product between a space of framed BPS states for the interface  $\wp$  and one for the interface  $\wp'$ . On the other hand, by considering the limit of zero separation, we see



FIGURE 7. Three surface defects connected by two interfaces. (This picture lives in the three-dimensional space where the field theory  $S[\mathfrak{g}, C, D]$  is defined; we have factored out the time direction)

that  $\mathcal{H}$  is the space of framed BPS states for the interface  $\wp\wp'$ . Equating these two descriptions of  $\mathcal{H}$  gives (20).

(One could wonder whether there should be a  $\pm$  sign on the right side of (20); this amounts to asking whether our rules for fixing the signs of  $F(\wp, \vartheta)$ ,  $F(\wp', \vartheta)$  and  $F(\wp\wp', \vartheta)$  are compatible with one another. Fortunately, it will follow from our explicit rules below that there is no sign needed.)

### 4.3. The Spectral Network $\mathcal{W}_\vartheta$

Our next aim is to explain how  $F(\wp, \vartheta)$  can actually be computed. As it turns out, the answer depends crucially on how  $\wp$  meets a certain codimension-1 locus  $\mathcal{W}_\vartheta \subset C$ , which we call a *spectral network*.

We say that a point  $z \in C$  *supports* those charges  $\bar{a} \in \Gamma(z, z)$  for which  $Z_{\bar{a}}(z)/e^{i\vartheta} \in \mathbb{R}_-$ . Define  $\mathcal{W}_\vartheta$  to be the set of  $z \in C$  such that  $z$  supports some  $\bar{a} \in \Gamma(z, z)$  with  $\mu(a) \neq 0$  (for either lift  $a$  of  $\bar{a}$ ).

$\mathcal{W}_\vartheta$  is a codimension-1 network on  $C$ , the union of segments which we call  $\mathcal{S}$ -walls. The  $\mathcal{S}$ -walls can end at branch points or at special points which we call *joints*, and can also asymptote to the punctures of  $C$ . We will assume that along any  $\mathcal{S}$ -wall, a generic point supports exactly one charge  $\bar{a}$ .

There are two ways in which this genericity could be violated. One is for  $\vartheta$  to be non-generic: in that case, some  $\mathcal{S}$ -walls will support both a charge  $\bar{a} \in \Gamma_{ij}(z, z)$  and a charge  $\bar{b} \in \Gamma_{ji}(z, z)$ . This phenomenon is crucial for our story and will be analyzed in Sect. 6. For now, however, we are assuming  $\vartheta$  generic, so we explicitly exclude this possibility. The second way in which genericity could be violated is less common: it might happen accidentally that e.g.  $\lambda_i - \lambda_j = \lambda_k - \lambda_l$  on a whole patch of  $C$ . In this case a single wall could support both a charge  $\bar{a} \in \Gamma_{ij}(z, z)$  and  $\bar{b} \in \Gamma_{kl}(z, z)$ . If  $i, j, k, l$  are all distinct, all our discussion in this section has a straightforward extension to that case; if they are not all distinct the situation is more subtle. In any case, from now on we assume that a generic point along an  $\mathcal{S}$ -wall supports exactly one charge  $\bar{a}$ .

In this case, all points of any single  $\mathcal{S}$ -wall support “the same” charge  $\bar{a}$ , in the sense that if  $z, z'$  are generic points on a common wall, the natural parallel transport along the wall takes the charge  $\bar{a} \in \Gamma(z, z)$  supported at  $z$  into the charge  $\bar{a}' \in \Gamma(z', z')$  supported at  $z'$ . If we choose a charge  $a \in \tilde{\Gamma}(\tilde{z}, \tilde{z})$  lifting  $\bar{a}$  as discussed in Sect. 3.5, then the parallel transport also takes  $a$  to an  $a' \in \tilde{\Gamma}(\tilde{z}', \tilde{z}')$  lifting  $\bar{a}'$ .

Now recall the soliton degeneracies  $\mu(a) \in \mathbb{Z}$ , which are defined for any  $a \in \tilde{\Gamma}(\tilde{z}, \tilde{z})$ , so long as  $z \in C$  does not lie on a wall of marginal stability.<sup>15</sup>  $\mu(a)$  does not depend on the parameter  $\vartheta$ , and its definition does not involve the spectral network  $\mathcal{W}_\vartheta$ . In our computation of  $F(\varphi, \vartheta)$  below, though, we will find that the  $\mu(a)$  which are really important are the ones where  $\bar{a} \in \Gamma(z, z)$  is supported at  $z$ —or said otherwise, we will be mainly interested in evaluating  $\mu(a)$  *along the  $\mathcal{S}$ -walls supporting  $\bar{a}$* . We will also find below that  $\mu(a)$  is *constant* along each  $\mathcal{S}$ -wall supporting  $\bar{a}$ ; with this in mind we immediately simplify our notation by letting  $\mu(a, p)$  denote the constant value of  $\mu(a)$  along a wall  $p$  supporting  $\bar{a}$ .

#### 4.4. Computing $F(\varphi, \vartheta)$

Now we can describe how  $F(\varphi, \vartheta)$  is computed. The recipe we will summarize here follows from the 2d–4d wall-crossing formula of [5].

The simplest situation occurs when  $\varphi$  does not cross  $\mathcal{W}_\vartheta$  anywhere. Define

$$\mathcal{D}(\varphi) = \sum_{i=1}^K X_{\varphi^{(i)}}, \quad (21)$$

where the  $\varphi^{(i)}$  are the canonical lifts of  $\varphi$  to the  $K$  sheets of  $\tilde{\Sigma}$ . Then if  $\varphi \cap \mathcal{W}_\vartheta = \emptyset$  we have simply

$$F(\varphi, \vartheta) = \mathcal{D}(\varphi). \quad (22)$$

In other words, interfaces corresponding to short enough paths  $\varphi$  on  $C$  just support  $K$  framed BPS states, one for each vacuum of the surface operator.

The more interesting question is how to compute  $F(\varphi)$  if  $\varphi$  does cross  $\mathcal{W}_\vartheta$ . Because of the composition property (20), it is enough to answer this question in the case where  $\varphi$  crosses  $\mathcal{W}_\vartheta$  exactly once.

So suppose  $\varphi$  crosses  $\mathcal{W}_\vartheta$  at a point  $z$  supporting a charge  $\bar{a} \in \Gamma_{ij}(z, z)$ . In this case,  $\varphi$  is divided into two subpaths  $\varphi_+$  and  $\varphi_-$ , which we deform slightly to obtain  $\varphi'_+$  and  $\varphi'_-$  as shown in Fig. 8.

Then, we have<sup>16</sup>

$$F(\varphi, \vartheta) = \mathcal{D}(\varphi) + \mathcal{D}(\varphi'_+)(\mu(a)X_a)\mathcal{D}(\varphi'_-). \quad (23)$$

The second term is the interesting one: it says that when  $\varphi$  crosses  $\mathcal{W}_\vartheta$ ,  $F(\varphi, \vartheta)$  includes paths which are segments of lifts of  $\varphi$  combined with “detours” along the lifts of BPS solitons. (The point of our deformation from  $\varphi_\pm$  to  $\varphi'_\pm$  here was to arrange that the second term makes sense, i.e. that  $\mathcal{D}(\varphi'_+)$  and  $\mathcal{D}(\varphi'_-)$  can be concatenated with the soliton charge  $a$ .)

For later convenience, we introduce a second notation for (23). Let  $\tilde{z}$  denote the tangent vector to  $\varphi$  at  $z$ . Let  $t_1$  be the shortest arc running from  $\tilde{z}$

<sup>15</sup> To reduce potential confusion we emphasize that the walls of marginal stability are *not* the same thing as the  $\mathcal{S}$ -walls.

<sup>16</sup> The factor  $\mu(a)X_a$  appearing in (23) is independent of the choice of lift  $a$  of  $\bar{a}$ , because of (14), (17).

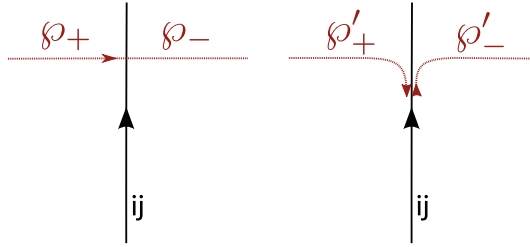


FIGURE 8. A path  $\wp$  crossing an  $\mathcal{S}$ -wall is locally divided into two pieces  $\wp_{\pm}$ . We deform  $\wp_{\pm}$  slightly to paths  $\wp'_{\pm}$ , so that the final tangent vector of  $\wp'_{+}$  and the initial tangent vector of  $\wp'_{-}$  point along the  $\mathcal{S}$ -wall, in opposite directions. The paths  $\wp'_{\pm}$  are shown slightly displaced from the wall for clarity

to the initial point of  $a$ , in the fiber of  $\tilde{\Sigma}$  over  $z$ . Similarly, let  $t_2$  be the shortest arc running from the final point of  $a$  to  $\tilde{z}$ , in the fiber of  $\tilde{\Sigma}$  over  $z$ ; and define

$$a_{\tilde{z}} = t_1 + a + t_2. \quad (24)$$

Note that  $t_1, t_2$  are well-defined because the intersection of  $\wp$  and  $\mathcal{W}_{\vartheta}$  is assumed to be transverse. With this definition (23) can be rewritten as

$$F(\wp, \vartheta) = \mathcal{D}(\wp_{+})(1 + \mu(a)X_{a_{\tilde{z}}})\mathcal{D}(\wp_{-}). \quad (25)$$

The formula (25) can be interpreted as a kind of wall-crossing formula for the framed 2d–4d BPS spectrum: it implies that when an endpoint of the path  $\wp$  is moved across an  $\mathcal{S}$ -wall,  $F(\wp, \vartheta)$  jumps by multiplication with a factor  $(1 + \mu(a)X_{a_{\tilde{z}}})$ . This is the way that this formula appeared in [5].

## 5. The Spectral Network $\mathcal{W}_{\vartheta}$ at Fixed $\vartheta$

In Sect. 4, we have given a recipe for computing the generating functions  $F(\wp, \vartheta)$ . The recipe depends on the data of the spectral network  $\mathcal{W}_{\vartheta}$  and the framed 2d–4d BPS degeneracies  $\mu(\cdot, p)$  along each wall  $p$  of  $\mathcal{W}_{\vartheta}$ . To make this recipe explicit, then, we need to be able to determine  $\mathcal{W}_{\vartheta}$  and  $\mu$ . In this section we explain how this can be done.

Throughout this section, we continue to assume that  $\vartheta$  is generic, in the sense explained in Sect. 3.3.

### 5.1. Walls as Trajectories

Suppose the  $\mathcal{S}$ -wall  $p$  supports the charge  $\bar{a} \in \Gamma_{ij}$  (by which we mean more precisely that every point  $z \in p$  supports a charge  $\bar{a}(z) \in \Gamma_{ij}(z, z)$ ). Then,  $e^{-i\vartheta}Z_{\bar{a}}$  is real everywhere along  $p$ . In particular, since  $dZ_{\bar{a}} = \frac{1}{\pi}(\lambda^{(j)} - \lambda^{(i)})$ , this means that  $p$  is an  $ij$ -trajectory with phase  $\vartheta$ , in the sense of (3).

Therefore, the walls in  $\mathcal{W}_{\vartheta}$  obey differential equations, and in fact exactly the *same* equations which are obeyed by the BPS strings which make up both the 2d and the 4d BPS states. This is not a coincidence, as we will see below.

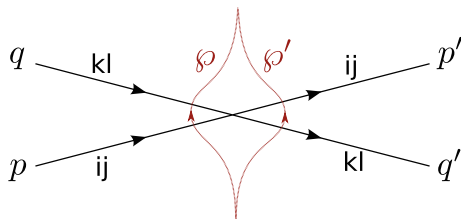


FIGURE 9. The local picture around a joint where exactly two  $\mathcal{S}$ -walls meet. We also show two paths  $\wp, \wp'$  on  $C$  which are related to one another by regular homotopy across the joint

By virtue of being an  $ij$ -trajectory,  $p$  is naturally oriented. As we move along  $p$  in the positive direction, the BPS mass  $|Z_{\bar{a}}| = -e^{-i\vartheta} Z_{\bar{a}}$  increases.

## 5.2. Joints and Wall-Crossing for $\mu$

Define a *joint* to be a point  $z \in C$  where at least two  $\mathcal{S}$ -walls intersect. A joint thus supports at least two charges, say  $\bar{a} \in \Gamma_{ij}(z, z)$  and  $\bar{b} \in \Gamma_{kl}(z, z)$ .

The simplest situation arises if the joint  $z$  supports *only* these two charges, so exactly two  $\mathcal{S}$ -walls meet there. This can only occur if  $i \neq l$  and  $j \neq k$ . In that case, the local picture around  $z$  is as shown in Fig. 9.

Now let us consider what happens to the soliton degeneracies  $\mu$  as we move across a joint. As it turns out, we can answer this question completely by considering the defect operators attached to the paths  $\wp, \wp'$  on  $C$  shown in Fig. 9. The requirement of homotopy invariance says that

$$F(\wp, \vartheta) = F(\wp', \vartheta). \quad (26)$$

To understand what this really means, let us evaluate both sides using the rules of Sect. 4.4, and then deform all the resulting paths on  $\tilde{\Sigma}$  so that they run directly into the joint  $z$ , with a common tangent vector  $\tilde{z}$  at  $z$ . Then we obtain

$$\begin{aligned} &\mathcal{D}(\wp_+)(1 + \mu(a, p)X_{a_{\tilde{z}}})(1 + \mu(b, q)X_{b_{\tilde{z}}})\mathcal{D}(\wp_-) \\ &= \mathcal{D}(\wp_+)(1 + \mu(b, q')X_{b_{\tilde{z}}})(1 + \mu(a, p')X_{a_{\tilde{z}}})\mathcal{D}(\wp_-). \end{aligned} \quad (27)$$

The paths  $a_{\tilde{z}}$  and  $b_{\tilde{z}}$  are not composable in either direction (since  $a_{\tilde{z}}$  runs from  $\tilde{z}^{(i)}$  to  $\tilde{z}^{(j)}$  while  $b_{\tilde{z}}$  runs from  $\tilde{z}^{(k)}$  to  $\tilde{z}^{(l)}$ ), so we have  $X_{a_{\tilde{z}}}X_{b_{\tilde{z}}} = X_{b_{\tilde{z}}}X_{a_{\tilde{z}}} = 0$ . Considering the terms linear in  $X_{a_{\tilde{z}}}$  and  $X_{b_{\tilde{z}}}$  in (27) then gives

$$\mu(a, p) = \mu(a, p'), \quad (28)$$

$$\mu(b, q) = \mu(b, q'). \quad (29)$$

In short, when an  $\mathcal{S}$ -wall of type  $ij$  and an  $\mathcal{S}$ -wall of type  $kl$  meet transversely, they cross without any changes in  $\mu$ .

A more interesting situation arises when an  $\mathcal{S}$ -wall of type  $ij$  and an  $\mathcal{S}$ -wall of type  $jk$  meet transversely. Suppose the two walls support charges  $\bar{a} \in \Gamma_{ij}(z, z)$  and  $\bar{b} \in \Gamma_{jk}(z, z)$ . (Note that in this case we must have  $i \neq k$ ; otherwise the two walls would obey the *same* differential equation, and it would

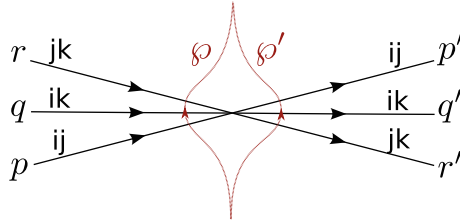


FIGURE 10. The local picture around a generic collision between three  $\mathcal{S}$ -walls

be impossible for them to intersect transversely.) In this case, the joint  $z$  supports both  $\bar{a}$  and  $\bar{b}$ , which are *composable* since  $z^{(j)}$  is both the end of  $\bar{a}$  and the start of  $\bar{b}$ . Hence the joint also supports a third charge  $\bar{c} = \bar{a} + \bar{b} \in \Gamma_{ik}(z, z)$ , and so there could be a third  $\mathcal{S}$ -wall meeting  $z$ . When  $\mu$  is generic enough, the picture is as shown in Fig. 10.<sup>17</sup>

The condition

$$F(\wp, \vartheta) = F(\wp', \vartheta) \quad (30)$$

becomes

$$\begin{aligned} & \mathcal{D}(\wp_+)(1 + \mu(a, p)X_{a_{\bar{z}}})(1 + \mu(c, q)X_{c_{\bar{z}}})(1 + \mu(b, r)X_{b_{\bar{z}}})\mathcal{D}(\wp_-) \\ &= \mathcal{D}(\wp_+)(1 + \mu(b, r')X_{b_{\bar{z}}})(1 + \mu(c, q')X_{c_{\bar{z}}})(1 + \mu(a, p')X_{a_{\bar{z}}})\mathcal{D}(\wp_-). \end{aligned} \quad (31)$$

Using  $X_{a_{\bar{z}}}X_{b_{\bar{z}}} = (-1)^{w(a_{\bar{z}}+b_{\bar{z}}, c_{\bar{z}})}X_{c_{\bar{z}}}$  and  $X_{b_{\bar{z}}}X_{a_{\bar{z}}} = 0$ , (31) implies

$$\mu(a, p') = \mu(a, p), \quad (32)$$

$$\mu(b, r') = \mu(b, r), \quad (33)$$

$$\mu(c, q') = \mu(c, q) + (-1)^{w(a_{\bar{z}}+b_{\bar{z}}, c_{\bar{z}})}\mu(a, p)\mu(b, r). \quad (34)$$

The result (34) says that the number of solitons carrying charge  $\bar{c}$  changes as we move the modulus  $z$  of the surface defect  $\mathbb{S}_z$  across the joint. This reflects the phenomenon of decay/formation of bound states between solitons of charges  $\bar{a}$  and  $\bar{b}$ . Indeed (34) is the same wall-crossing formula which was discovered in [38] in the context of pure 2d theories, and was reinterpreted in the 2d–4d context in [5]. Our derivation of it here, using consistency of the framed 2d–4d BPS spectrum, is essentially the same as the one given in [5].

There are a few important special cases which deserve separate discussion. One arises when  $\mu(c, q) = 0$ , so that we actually have only two  $\mathcal{S}$ -walls intersecting, of types  $ij$  and  $jk$ . In this case, the situation is as shown in Fig. 11: the wall  $q'$  is born from the joint, and (34) reduces to  $\mu(c, q') = (-1)^{w(a_{\bar{z}}+b_{\bar{z}}, c_{\bar{z}})}\mu(a, p)\mu(b, r)$ .

The reverse situation is also allowed: if (34) gives  $\mu(c, q') = 0$ , then the wall  $q$  “dies” at the joint, as shown in Fig. 12.

<sup>17</sup> One might think that there is a second, inequivalent possibility: one could have exchanged the  $ij$  and  $jk$  labels in Fig. 10. This gives a new picture, which is not related to Fig. 10 by a rotation, but is related to Fig. 10 by an orientation-reversing map. Fortunately, our rules for computing  $F(\wp, \vartheta)$  do not use the orientation of  $C$ . So our analysis is fully general.

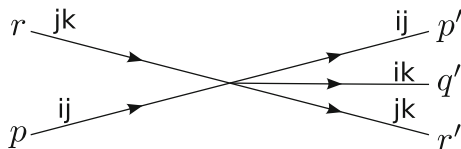


FIGURE 11. A collision between  $\mathcal{S}$ -walls, at which a new  $\mathcal{S}$ -wall is born

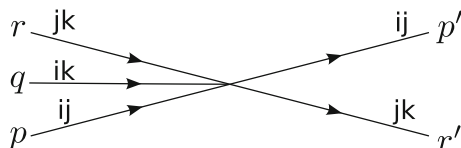


FIGURE 12. A collision between  $\mathcal{S}$ -walls, at which an  $\mathcal{S}$ -wall dies

So far we have described some local properties of  $\mathcal{W}_\vartheta$  and  $\mu$ . In the next few sections, we describe a recipe for explicitly *constructing* them.

### 5.3. The Mass Filtration; $\mathcal{W}_\vartheta[\Lambda]$ and $\mu[\Lambda]$ for Small $\Lambda$

We begin with the observation that  $\mathcal{W}_\vartheta$  carries a useful filtration. Namely, for any  $\Lambda > 0$  (with dimensions of mass), we can define a new object  $\mathcal{W}_\vartheta[\Lambda]$  by truncating all the  $\mathcal{S}$ -walls: for a wall  $w$  supporting a charge  $\bar{a}$ ,  $\mathcal{W}_\vartheta[\Lambda]$  includes only the portion of  $w$  with  $|Z_{\bar{a}}| < \Lambda$ . Therefore,

$$\mathcal{W}_\vartheta[\Lambda] \subset \mathcal{W}_\vartheta[\Lambda'] \quad \text{for } \Lambda \leq \Lambda' \quad (35)$$

and

$$\mathcal{W}_\vartheta = \lim_{\Lambda \rightarrow \infty} \mathcal{W}_\vartheta[\Lambda]. \quad (36)$$

We also define a truncated version  $\mu[\Lambda]$  of the soliton degeneracies  $\mu$ :  $\mu[\Lambda](a)$  is defined only for charges  $a$  with  $|Z_{\bar{a}}| < \Lambda$ , and for such charges it agrees with  $\mu(a)$ .

For small enough  $\Lambda$ , we can describe  $\mathcal{W}_\vartheta[\Lambda]$  and  $\mu[\Lambda]$  simply and explicitly. The reason is that charges  $\bar{a} \in \Gamma(z, z)$  with  $|Z_{\bar{a}}| < \Lambda$  are represented by very short paths on  $\Sigma$  between distinct lifts  $z^{(i)}$  and  $z^{(j)}$  of  $z$ . The only way to get such a short path is for  $z$  to be close to an  $(ij)$  branch point  $\mathfrak{b}$ : then if we take  $\bar{a} \in \Gamma(z, z)$  to be a path running from  $z^{(i)}$  to the ramification point over  $\mathfrak{b}$  and returning to  $z^{(j)}$ , we indeed have  $|Z_{\bar{a}}| \rightarrow 0$  as  $z \rightarrow \mathfrak{b}$ . We could also have considered the charge  $-\bar{a} \in \Gamma(z, z)$ , corresponding to a path running in the opposite direction. These two charges are exchanged by the monodromy when  $z$  goes once around  $\mathfrak{b}$ . They are the only charges which become massless at  $\mathfrak{b}$ .

Therefore, to determine  $\mathcal{W}_\vartheta[\Lambda]$  for  $\Lambda$  small, we just have to describe the  $\mathcal{S}$ -walls supporting these light charges. Letting  $z$  be a local coordinate with  $z = 0$  at  $\mathfrak{b}$ , we have  $Z_{\bar{a}} \sim z^{3/2}$  and  $Z_{-\bar{a}} \sim -z^{3/2}$  (the square-root branch cut



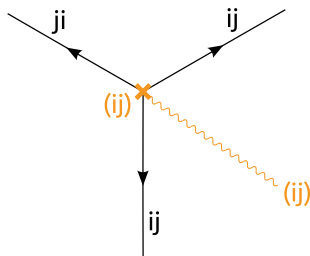


FIGURE 13. The  $3\mathcal{S}$ -walls emerging from a branch point. Because the two relevant sheets of the covering  $\Sigma \rightarrow C$  are exchanged by monodromy around the branch point, we cannot assign labels to the trajectories globally; instead we have chosen a branch cut and a labeling of the sheets  $i$  and  $j$  on the complement of the cut. We also label the cut with the transposition  $(ij)$  which relates the sheets on the two sides of the cut

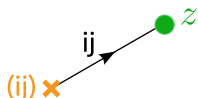


FIGURE 14. A finite open web  $N(z)$  consisting of a single short  $ij$ -trajectory, with one end on the  $(ij)$ -branch point  $\mathfrak{b}$  and one end on  $z$ . This finite open web represents a light BPS soliton on the surface defect  $\mathbb{S}_z$

arises from the monodromy  $\bar{a} \rightarrow -\bar{a}$  noted above.) It follows that there are  $3\mathcal{S}$ -walls emerging from the branch point, as shown in Fig. 13.

If  $z$  lies on one of these three  $\mathcal{S}$ -walls near  $\mathfrak{b}$ , then there is a light 2d–4d BPS soliton on the surface defect  $\mathbb{S}_z$ , with central charge in  $\mathbb{R}_+ e^{i\vartheta}$ .<sup>18</sup> This BPS soliton is represented by a short finite open web  $N(z)$ , consisting of a single BPS string connecting  $z$  to  $\mathfrak{b}$ , as illustrated in Fig. 14. In accordance with the general rules of Sect. 3.2, the charge  $\bar{a}$  of this BPS state is  $\bar{a} = [N(z)_\Sigma]$ , where  $N(z)_\Sigma$  is the lift of  $N(z)$  to  $\Sigma$ , i.e. a short path on  $\Sigma$  running from  $z^{(i)}$  to the ramification point over  $\mathfrak{b}$  and then back to  $z^{(j)}$ .

As we have mentioned in Sect. 3.5, however, to fix the sign of the BPS degeneracy for this soliton, we need to choose a lift from the charge  $\bar{a}$  to a class  $a$ . Suppose we make the most obvious choice, namely the one determined by the web  $N(z)$ :  $a = [N(z)_\Sigma]$ . Let  $p$  denote the  $\mathcal{S}$ -wall on which  $z$  sits. Then finally  $\mu(a, p)$  is a well-defined integer, and it is meaningful to ask what it

<sup>18</sup> In a sense it would be better to refer to these light states simply as “2d” solitons rather than “2d–4d,” since their existence does not depend much on the coupling to the 4d theory. Indeed, the existence of these light states can be deduced from a universal computation involving the simplest nontrivial Landau–Ginzburg model. See Section 8.1 of [5] for further discussion.

is. We have a single isolated 2d–4d BPS web here, so it should contribute just a single state; the only question is whether we will have  $\mu(a, p) = +1$  or  $\mu(a, p) = -1$ . We claim that the correct answer is

$$\mu(a, p) = +1. \quad (37)$$

Indeed, this is forced on us by the requirement of homotopy invariance, which we discuss in Sect. 5.6 below.

#### 5.4. $\mathcal{W}_\vartheta[\Lambda]$ and $\mu[\Lambda]$ for General $\Lambda$

We have determined  $\mathcal{W}_\vartheta[\Lambda]$  and  $\mu[\Lambda]$  for small  $\Lambda$ . Now we can ask how  $\mathcal{W}_\vartheta[\Lambda]$  and  $\mu[\Lambda]$  evolve as  $\Lambda$  increases. Using the properties of  $\mathcal{W}_\vartheta$  and  $\mu$  we have already determined, it is straightforward to deduce the answer.

The three  $\mathcal{S}$ -walls emerging from each branch point  $\mathfrak{b}$  flow according to the differential equation (3), for a distance determined by the cutoff  $\Lambda$ . When we increase  $\Lambda$  enough, it might happen that an  $\mathcal{S}$ -wall of type  $ij$  intersects another  $\mathcal{S}$ -wall of type  $jk$ . At this point a new  $\mathcal{S}$ -wall of type  $ik$  is born, as shown in Fig. 11. The soliton degeneracy  $\mu$  on this new wall is determined by the soliton degeneracies on its parents, as in (34).<sup>19</sup> This new “secondary”  $\mathcal{S}$ -wall, in turn, also evolves according to (3). As  $\Lambda$  increases the secondary walls can intersect with other  $\mathcal{S}$ -walls and give birth to yet more progeny, and so on. In Fig. 15, we give an illustration of the growth of  $\mathcal{W}_\vartheta[\Lambda]$  with  $\Lambda$  in one particular theory  $S[\mathfrak{g}, C, D]$ .

We may also have intersections where three  $\mathcal{S}$ -walls meet at a point, as in Fig. 10. (The reader might feel that this phenomenon should not occur generically; indeed, three arbitrary trajectories of types  $ij$ ,  $jk$ ,  $ik$  would be unlikely to intersect at a single point, but the  $\mathcal{S}$ -walls are not arbitrary trajectories. We include an example in Fig. 16.)

There is one more complicated phenomenon which one can imagine: what would happen if an  $\mathcal{S}$ -wall of type  $ij$  ran directly into an  $(ij)$  branch point? In this case, it would not be immediately clear how to continue the network  $\mathcal{W}_\vartheta[\Lambda]$  and the BPS degeneracies  $\mu[\Lambda]$ . This puzzling-looking situation cannot occur when  $\vartheta$  is generic. It does occur for non-generic  $\vartheta$ , and this fact plays a crucial role in the considerations of Sect. 6 below.

We have now given a recipe for constructing  $\mathcal{W}_\vartheta[\Lambda]$  and  $\mu[\Lambda]$  for any  $\Lambda > 0$  and generic  $\vartheta$ .  $\mathcal{W}_\vartheta[\Lambda]$  consists of a finite number of  $\mathcal{S}$ -walls for any  $\Lambda$ . Taking the limit  $\Lambda \rightarrow \infty$ , we obtain the full  $\mathcal{W}_\vartheta$  and  $\mu$ . So we have now managed to determine *all* of the soliton degeneracies of the theory (including their tricky signs) using only the constraint of homotopy invariance!

More precisely, so far we have fixed some  $\vartheta$  and determined all  $\mu(a, p)$ —i.e. we determined  $\mu(a)$  when the parameter  $z$  of the surface defect lies on an  $\mathcal{S}$ -wall supporting the charge  $a$ . This is not what one would usually mean by “determining the soliton degeneracies”: what one would usually mean is that we fix some  $z$  once and for all, and then compute all  $\mu(a)$  for  $a \in \tilde{\Gamma}(\tilde{z}, \tilde{z})$ .

<sup>19</sup> Note that if two  $\mathcal{S}$ -walls carrying charges  $\bar{a} \in \Gamma_{ij}$ ,  $\bar{b} \in \Gamma_{jk}$  meet at a joint  $z \in C$ , the joint is visible in  $\mathcal{W}_\vartheta[\Lambda]$  beginning at  $\Lambda = \text{Max}(|Z_{\bar{a}}(z)|, |Z_{\bar{b}}(z)|)$ , but the  $\mathcal{S}$ -wall born from the joint does not appear until  $\Lambda = |Z_{\bar{a}}(z)| + |Z_{\bar{b}}(z)|$ .

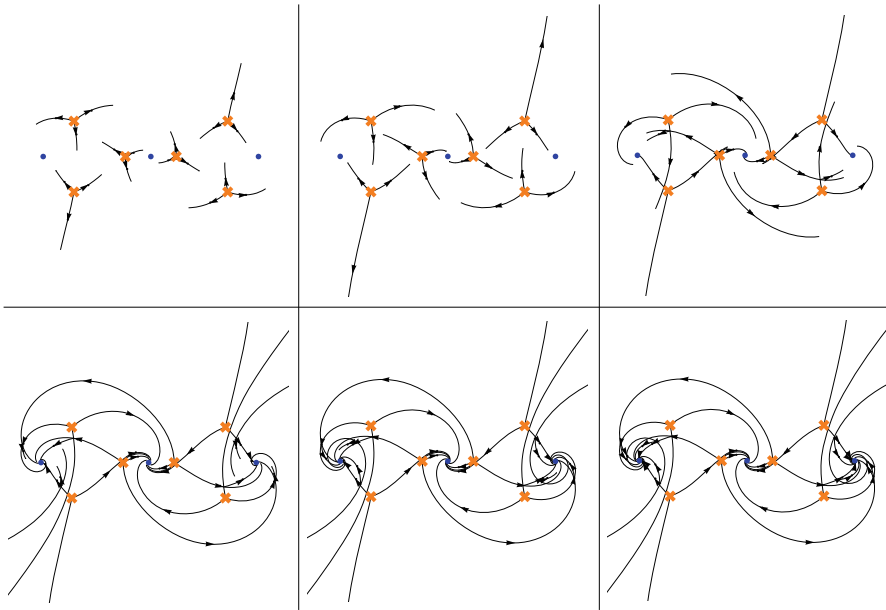


FIGURE 15. Sample  $\mathcal{W}_\vartheta[\Lambda]$  in the theory  $S[A_2, \mathbb{CP}^1, D]$  where the  $D_n$  are 3 full defects (blue dots at  $z = -1, 0, 1$ ). We choose  $\vartheta = \pi/4$ ,  $\phi_2 = \frac{2}{(z-1)^2} + \frac{2}{(z+1)^2} - \frac{4}{z^2}$ ,  $\phi_3 = \frac{1-3(z-1)+6(z-1)^2}{(z-1)^3} + \frac{1+3(z+1)+6(z+1)^2}{(z+1)^3} - \frac{2+12z^2}{z^3}$ . The plots shown are at  $\Lambda = 1, 2, 5, 12, 20, \infty$ . At  $\Lambda = 0$  three  $\mathcal{S}$ -walls are born at each of the 6 branch points (*orange crosses*); at larger values of  $\Lambda$ , additional  $\mathcal{S}$ -walls are born at intersections. As  $\Lambda \rightarrow \infty$  all  $\mathcal{S}$ -walls asymptotically approach the defects. Some  $\mathcal{S}$ -walls are cut off by the boundary of the plot; the visible range is a square with side length 2.6. See [43] for an animated version of this figure (color figure online)

The point is that, for any  $a \in \tilde{\Gamma}(\tilde{z}, \tilde{z})$ , there is some  $\vartheta$  for which  $z$  *does* lie on an  $\mathcal{S}$ -wall supporting  $a$ : namely,  $\vartheta = \arg -Z_{\tilde{a}}$ . So for each  $a$  we can draw the corresponding network  $\mathcal{W}_{\vartheta=\arg -Z_{\tilde{a}}}$  and use it to compute  $\mu(a)$ .

### 5.5. $\mathcal{W}_\vartheta$ Near Full Defects

On a close look at Fig. 15, one notices that as  $\Lambda \rightarrow \infty$ , all of the  $\mathcal{S}$ -walls in  $\mathcal{W}[\Lambda]$  asymptotically approach the full defects  $D_n$ .

To understand this concretely, let us consider the general behavior of  $ij$ -trajectories around a full defect. In a local coordinate  $z$  where the defect is at  $z = 0$ , we have

$$\lambda^{(i)} = m^{(i)} \frac{dz}{z} + \dots \quad (38)$$

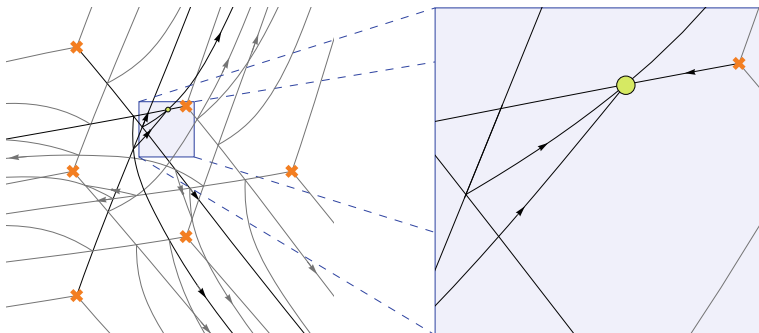


FIGURE 16. A portion of a sample  $\mathcal{W}_\vartheta$ , exhibiting an intersection between three  $\mathcal{S}$ -walls, marked by a *yellow dot*. The region in the *blue box* is blown up at right. The  $\mathcal{S}$ -walls which appear somewhere in the genealogy of the three intersecting  $\mathcal{S}$ -walls are shown in *black*; other  $\mathcal{S}$ -walls are shown in *gray*. There are six branch points (*orange crosses*). This example arises in the theory  $S[A_3, C = \mathbb{CP}^1, D]$  with  $\phi_2 = -10z \, dz^2$ ,  $\phi_3 = 4dz^3$ ,  $\phi_4 = 9z^2 \, dz^4$ . We took  $\vartheta = 1.841$ , but we emphasize that this triple intersection persists for nearby  $\vartheta$  as well (color figure online)

where  $\dots$  denotes regular terms.  $ij$ -trajectories near  $z = 0$  thus behave asymptotically like

$$z(t) = z_0 \exp(\xi^{(ij)} t), \quad (39)$$

where we defined

$$\xi^{(ij)} = \frac{e^{i\vartheta}}{m^{(i)} - m^{(j)}}. \quad (40)$$

As  $t \rightarrow \infty$ , we have  $z(t) \rightarrow 0$  if and only if  $\operatorname{Re} \xi^{(ij)} < 0$ . This suggests a natural ordering on the sheets in a neighborhood of  $z = 0$ : we say that  $i < j$  if  $\operatorname{Re} e^{-i\vartheta} m^{(i)} < \operatorname{Re} e^{-i\vartheta} m^{(j)}$ .  $ij$ -trajectories near  $z = 0$  asymptote to  $z = 0$  if and only if  $i < j$  in this ordering.

In the coordinate  $z$ , these infalling trajectories asymptotically approach logarithmic spirals. Passing to the covering coordinate  $w = \log z$ , the  $ij$ -trajectories are straight lines,

$$w(t) = w_0 + \xi^{(ij)} t. \quad (41)$$

Now suppose we have an infalling asymptotic  $\mathcal{S}$ -wall of type  $ij$  and another of type  $jk$  (so  $i < j < k$ .) Assuming  $\arg \xi^{(ij)} \neq \arg \xi^{(jk)}$ , these two walls intersect at infinitely many points as they spiral into  $z = 0$ . Each intersection gives birth to a new  $\mathcal{S}$ -wall of type  $ik$ , which also spirals into  $z = 0$ . This new  $\mathcal{S}$ -wall can in turn intersect other inspiraling  $\mathcal{S}$ -walls and give birth to yet more progeny. The strict ordering of the sheets ensures that the progeny cannot commit incest with one another. Nevertheless, for  $K > 2$  there are an infinite number of joints accumulating at the full defect. See Fig. 17 for an illustration.

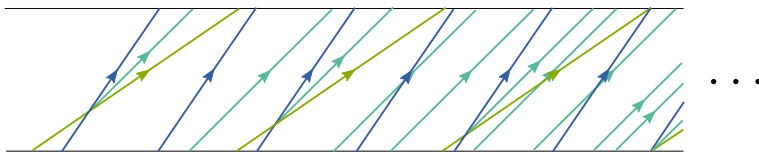


FIGURE 17.  $\mathcal{S}$ -walls falling into a full defect, shown in the covering coordinate  $w = \log z$ . There is a single  $ij$  wall and a single  $jk$  wall. We have  $i < j < k$  in the ordering described in the text. Each intersection between these two walls generates a new  $ik$  wall. There are infinitely many such intersections, accumulating at the defect, which generate infinitely many  $ik$  walls

Similar (but more involved) remarks should apply for more general types of defect.

Based on computer experimentation, we expect that the behavior observed in Fig. 15 is indeed generic: whenever there is a defect with sufficiently generic mass parameters, and  $\vartheta$  is generic, all  $\mathcal{S}$ -walls in  $\mathcal{W}_\vartheta$  should asymptotically approach punctures. In the case  $K = 2$ , this follows directly from known mathematical results on the trajectories of quadratic differentials [44] (see [3] for an account of this.) For  $K > 2$ , the analogous foundational results are not yet available as far as we know. It would be very desirable to work this out.

### 5.6. Homotopy Invariance

Now let us consider a key consistency check of our story so far. We have claimed on general physical grounds that  $F(\wp, \vartheta)$  should be a homotopy invariant of  $\wp$ , or more precisely a twisted homotopy invariant as described in Sect. 3.5. On the other hand, we have also given a recipe which completely determines  $F(\wp, \vartheta)$ . So we can ask whether this recipe indeed obeys the necessary twisted homotopy invariance.

To check this twisted homotopy invariance, it is enough to check the invariance under a few elementary moves, which we now consider in turn. This is the first place where the tricky minus signs mentioned in Sect. 3.5 play a decisive role.

First, consider a pair of paths  $\wp$  and  $\wp'$ , neither of which meets any  $\mathcal{S}$ -walls, and which are not only homotopic but related by a *regular* homotopy, i.e. a homotopy through immersions. In this case, the constraint of twisted homotopy invariance requires that

$$F(\wp, \vartheta) = F(\wp', \vartheta). \quad (42)$$

But since neither path meets any  $\mathcal{S}$ -walls, this reduces to

$$\mathcal{D}(\wp) = \mathcal{D}(\wp'), \quad (43)$$

which is indeed true: the regular homotopy of  $\wp$  to  $\wp'$  lifts to a regular homotopy of each  $\wp^{(i)}$  to  $\wp'^{(i)}$ .

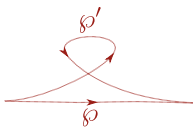


FIGURE 18. A pair of paths that are homotopic but differ by a unit of winding

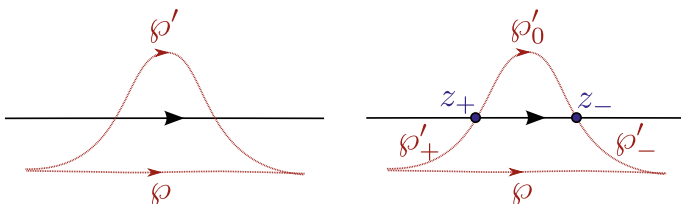


FIGURE 19. Two open paths  $\varphi, \varphi'$  which differ by a homotopy across an  $\mathcal{S}$ -wall

Next, consider the pair of paths in Fig. 18. In this case the constraint of twisted homotopy invariance requires that

$$F(\varphi, \vartheta) = -F(\varphi', \vartheta). \quad (44)$$

But neither of these paths meets any  $\mathcal{S}$ -walls, so this reduces to

$$\mathcal{D}(\varphi) = -\mathcal{D}(\varphi'), \quad (45)$$

which indeed follows by lifting the homotopy of  $\varphi$  to  $\varphi'$  to homotopies of  $\varphi^{(i)}$  to  $\varphi'^{(i)}$  and using (17). We stress that this is a non-regular homotopy, which accounts for the sign change.

Next, consider two paths  $\varphi, \varphi'$  which differ by homotopy across an  $\mathcal{S}$ -wall as shown in Fig. 19. The constraint of twisted homotopy invariance says

$$F(\varphi, \vartheta) = F(\varphi', \vartheta). \quad (46)$$

Let  $\bar{a}$  be the charge supported on the  $\mathcal{S}$ -wall in the figure. The two intersections  $z_{\pm}$  between  $\varphi'$  and the wall divide  $\varphi'$  into  $\varphi'_+ \varphi'_0 \varphi'_-$ ; let  $\tilde{z}_{\pm}$  be the tangent vectors to  $\varphi'$  at the intersection points. Then evaluating both sides directly, letting  $a$  be the charge supported along the wall, and using the fact that  $\mu(a)$  is constant along the wall, (46) becomes

$$\begin{aligned} \mathcal{D}(\varphi) &= \mathcal{D}(\varphi'_+)(1 + \mu(a)X_{a_{\tilde{z}_+}})\mathcal{D}(\varphi'_0)(1 + \mu(a)X_{a_{\tilde{z}_-}})\mathcal{D}(\varphi'_-) \\ &= \mathcal{D}(\varphi') + \mu(a) \cdot \mathcal{D}(\varphi'_+) \left( X_{a_{\tilde{z}_+}} \mathcal{D}(\varphi'_0) + \mathcal{D}(\varphi'_0) X_{a_{\tilde{z}_-}} \right) \mathcal{D}(\varphi'_-) \\ &\quad + \mu(a)^2 \mathcal{D}(\varphi'_+) X_{a_{\tilde{z}_+}} \mathcal{D}(\varphi'_0) X_{a_{\tilde{z}_-}} \mathcal{D}(\varphi'_-). \end{aligned} \quad (47)$$

The two soliton paths  $a_{\tilde{z}_+}$  and  $a_{\tilde{z}_-}$  are both of type  $ij$ , so  $X_{a_{\tilde{z}_+}} \mathcal{D}(\varphi'_0) X_{a_{\tilde{z}_-}} = 0$ . Using the definition (24), one can check that the two terms in the parentheses correspond to paths which differ by one unit of winding; according to the rule

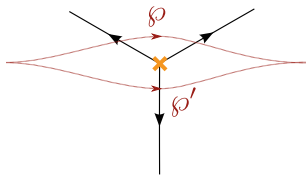


FIGURE 20. Two paths which differ by homotopy across a branch point

(17), these two terms thus differ by a minus sign, and so cancel one another. Thus (47) reduces to

$$\mathcal{D}(\varphi) = \mathcal{D}(\varphi') \quad (48)$$

which is indeed true since the obvious homotopy of  $\varphi$  into  $\varphi'$  also gives homotopies between each  $\varphi^{(i)}$  and  $\varphi'^{(i)}$ . Note that this would not have worked out without using the fact that  $\mu(a)$  is constant along the wall: indeed this shows that homotopy invariance requires  $\mu(a)$  to be constant along walls, thus making good on a promise we made in Sect. 4.3.

Next, consider two paths  $\varphi, \varphi'$  which differ by homotopy across a joint, as in Fig. 9 or Fig. 10. We have already verified in Sect. 5.2 that in this case  $F(\varphi, \vartheta) = F(\varphi', \vartheta)$ : indeed we used this constraint as part of our system for determining  $\mu$ .

Finally we reach the most interesting case. Consider two paths  $\varphi, \varphi'$  which differ by homotopy across a branch point, as shown in Fig. 20. The constraint of twisted homotopy invariance says

$$F(\varphi, \vartheta) = F(\varphi', \vartheta). \quad (49)$$

Evaluating both sides of (49), we obtain the explicit sums of paths shown in Fig. 21. Naively we see  $K + 3$  terms in  $F(\varphi, \vartheta)$  and  $K + 1$  in  $F(\varphi', \vartheta)$ . The last two terms in  $F(\varphi, \vartheta)$ , running from sheet  $j$  to sheet  $j$ , are of the form  $X_a + X_{a'}$ , where  $a, a'$  both project to the same charge  $\bar{a}$ , but differ by one unit of winding. These two terms thus cancel one another thanks to our sign rule (17). The other  $K + 1$  terms precisely match the  $K + 1$  terms in  $F(\varphi', \vartheta)$ .

So the constraint (49) is indeed satisfied.

As an aside, note that this last constraint would *not* have worked out if there were no  $\mathcal{S}$ -walls emerging from the branch point: in other words, homotopy invariance of the framed 2d–4d BPS spectrum really *requires* that spectrum to undergo wall-crossing. This is closely analogous to the considerations of monodromy invariance that led Seiberg and Witten to discover wall-crossing in the pure 4d BPS spectrum [45]. The constraint of monodromy invariance also would not have worked out if we had taken  $\mu = -1$  instead of  $\mu = +1$  in (37).

### 5.7. $K = 2$ Theories and Ideal Triangulations

In the special case  $K = 2$ ,  $\mathcal{W}_\vartheta$  is enormously simplified: transverse intersections of  $\mathcal{S}$ -walls as discussed in Sect. 5.2 require at least three distinct sheets

FIGURE 21. The explicit sums of paths which occur in (49). In the equations shown in the figure, we represent the formal variable  $X_a$  by an explicit picture of a smooth path on  $\Sigma$ , whose canonical lift to  $\tilde{\Sigma}$  represents the relative homology class  $a$ . The labels next to path segments show which sheet of the covering  $\Sigma \rightarrow C$  the segments are on. We have labeled the sheets in the same way as we did in Fig. 13

$i, j, k$ , and there is no room for this if  $K = 2$ . Therefore, in this case, there are no joints where new  $\mathcal{S}$ -walls could be born, and so  $\mathcal{W}_\vartheta$  consists simply of three  $\mathcal{S}$ -walls emerging from each branch point. Assuming that the parameters  $m_a$  at the punctures  $D_n$  are generic enough, each of these  $\mathcal{S}$ -walls asymptotes to one of the punctures  $D_n$ .

When  $K = 2$ , then, the dual to the network  $\mathcal{W}_\vartheta$  is an ideal triangulation  $T(\mathcal{W}_\vartheta)$  of  $C$  (determined up to isotopy) (see Fig. 22). Combinatorially speaking, the ideal triangulation  $T(\mathcal{W}_\vartheta)$  and the spectral network  $\mathcal{W}_\vartheta$  contain the same information.

It has been natural to wonder what is the appropriate higher- $K$  generalization of the notion of ideal triangulation. In Sect. 9, we propose a general definition of “spectral network” which we believe is the right answer to this question. The  $\mathcal{W}_\vartheta$  which we have been discussing so far are examples of spectral networks.

## 6. Varying $\vartheta$

In Sect. 5, we held  $\vartheta$  fixed and studied the generating functions  $F(\wp, \vartheta)$ . In this section, we consider what happens when  $\vartheta$  is allowed to vary. In so doing, we will uncover new phenomena, associated with the 4d BPS states in the theory  $S[\mathfrak{g}, C, D]$ .

$F(\wp, \vartheta)$  is piecewise constant as a function of  $\vartheta \in \mathbb{R}$ : if the pair  $(\wp, \vartheta)$  is generic and  $\delta\vartheta$  small enough,  $F(\wp, \vartheta) = F(\wp, \vartheta + \delta\vartheta)$ . However,  $F(\wp, \vartheta)$  jumps at some special values  $\vartheta = \vartheta_c$ . These special values arise for two distinct reasons, to be described in the next two sections.



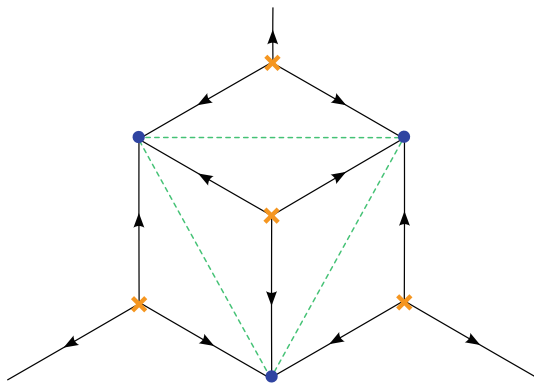


FIGURE 22. A portion of a spectral network  $\mathcal{W}_\vartheta$  with  $K = 2$ , and one face in the corresponding ideal triangulation  $T(\mathcal{W}_\vartheta)$  up to isotopy (green dashed lines). Each wall of  $\mathcal{W}_\vartheta$  begins on a branch point (orange cross) and asymptotes to a defect (blue dot); these defects are also the vertices of  $T(\mathcal{W}_\vartheta)$ . Each branch point is contained in a unique face of  $T(\mathcal{W}_\vartheta)$  (color figure online)

At the special values, we define for convenience<sup>20</sup>

$$F(\wp, \vartheta_c^\pm) = \lim_{\vartheta \rightarrow \vartheta_c^\pm} F(\wp, \vartheta) \quad (50)$$

(think of  $\vartheta_c^\pm$  as representing  $\vartheta_c \pm \varepsilon$  for  $\varepsilon \rightarrow 0^+$ .) What we will explain in the rest of this section is how to determine the relation between  $F(\wp, \vartheta_c^+)$  and  $F(\wp, \vartheta_c^-)$ , and how to extract from this relation the 4d BPS degeneracies.

### 6.1. Endpoints Crossing $\mathcal{S}$ -Walls

First, there are “simple” jumps which occur when an  $\mathcal{S}$ -wall moves across one of the endpoints of  $\wp$ . Namely,  $F(\wp, \vartheta)$  jumps whenever there is some  $\bar{a} \in \Gamma(z_1, z_1)$  or  $\bar{a} \in \Gamma(z_2, z_2)$  with  $e^{-i\vartheta_c} Z_{\bar{a}} \in \mathbb{R}_-$  and  $\mu(\bar{a}) \neq 0$ . If  $\bar{a} \in \Gamma(z_1, z_1)$  and  $\wp$  intersects the  $\mathcal{S}$ -wall for  $\vartheta_c + \varepsilon$  then the jump is of the form

$$F(\wp, \vartheta_c^+) = (1 + \mu(a)X_{a_{\bar{z}_1}})F(\wp, \vartheta_c^-). \quad (51)$$

Similarly, if  $\bar{a} \in \Gamma(z_2, z_2)$  and  $\wp$  intersects the  $\mathcal{S}$ -wall for  $\vartheta_c + \varepsilon$  then the jump is of the form

$$F(\wp, \vartheta_c^+) = F(\wp, \vartheta_c^-)(1 + \mu(a)X_{a_{\bar{z}_2}}). \quad (52)$$

The jumps (51) and (52) are easily deduced from our rule (25) for computing  $F(\wp, \vartheta)$  in Sect. 4.4.

<sup>20</sup> The meaning of the limit (50) is easy to understand if there are no other special values  $\vartheta'_c$  in some neighborhood of  $\vartheta_c$ . However, it does sometimes happen that the special values accumulate. In that case, we have to explain what kind of limit we mean in (50). One possibility would be to fix some  $R > 0$  and work with a “ $Z$ -adic” norm where  $\|\sum c_a X_a\| = \sum |c_a| e^{-R|Z_a|}$ . We expect that in this norm the limit (50) indeed exists. At any rate, in what follows we work formally, assuming that the limits make sense.

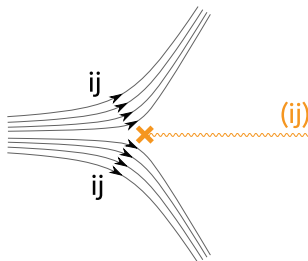


FIGURE 23. The behavior of  $ij$ -trajectories approaching a branch point of type  $(ij)$

These equations can be compared to (2.27) of [5]. The signs appear different because in [5] there is a cocycle  $\sigma$  which was undetermined in that section.

## 6.2. $\mathcal{K}$ -walls and The Degenerate Spectral Networks $\mathcal{W}_{\vartheta_c}$

There is also a second more interesting kind of jump which can occur at a critical phase  $\vartheta = \vartheta_c$ . In contrast to the previous case, these jumps do not arise because of a change in the interaction between  $\mathcal{W}_{\vartheta}$  and  $\wp$ . Rather, they arise because of a topology change in  $\mathcal{W}_{\vartheta}$  itself.

The claim that a topology change in  $\mathcal{W}_{\vartheta}$  could occur as we continuously vary  $\vartheta$  might at first seem strange. After all, we have given a recipe for  $\mathcal{W}_{\vartheta}$  in Sect. 5,  $\vartheta$  enters this recipe only through the differential equation (3), and this equation evidently depends continuously on  $\vartheta$ . The point is that the *solutions* of (3) can nevertheless exhibit discontinuous behavior, because of the phenomenon of bifurcation near the branch points:  $ij$ - or  $ji$ -trajectories coming close to an  $(ij)$  branch point can veer off in one of two directions, as indicated in Fig. 23. This figure can be considered to depict either a foliation of  $C$  by  $ij$ -trajectories at fixed  $\vartheta$  or the evolution of an  $\mathcal{S}$ -wall in  $\mathcal{W}_{\vartheta}$  as  $\vartheta$  varies. From the latter viewpoint, it follows that the critical phases  $\vartheta_c$  are those at which some  $\mathcal{S}$ -wall of type  $ij$  or  $ji$  runs directly into a branch point of type  $(ij)$ . We call the  $\vartheta_c$  where this occurs  $\mathcal{K}$ -walls. A simple example is shown in Fig. 24.

To describe the relation between  $F(\wp, \vartheta_c^+)$  and  $F(\wp, \vartheta_c^-)$ , it is convenient to work directly with the limiting spectral network  $\mathcal{W}_{\vartheta_c}$ . This limiting network has some features not seen for the generic  $\mathcal{W}_{\vartheta}$ . In particular, several (possibly infinitely many)  $\mathcal{S}$ -walls, supporting different charges, might coalesce into a single segment as  $\vartheta \rightarrow \vartheta_c$ ; hence a wall  $p$  of  $\mathcal{W}_{\vartheta_c}$  might support several distinct charges. Moreover, the  $\mathcal{S}$ -walls coalescing onto  $p$  might not all be oriented in the same direction; in this case, we call  $p$  a *two-way street*.<sup>21</sup> Again see Fig. 24 for an example: the saddle connection in the middle of the figure

<sup>21</sup> We now have two metaphors for the segments in a spectral network. They are *walls* because they are the loci where framed 2d–4d BPS degeneracies jump. On the other hand we will find that the metaphor of *streets* is also very useful when tracking the solitons. Rather than insisting on one term, we will use them as synonyms. We feel no inclination to go to wall-street.

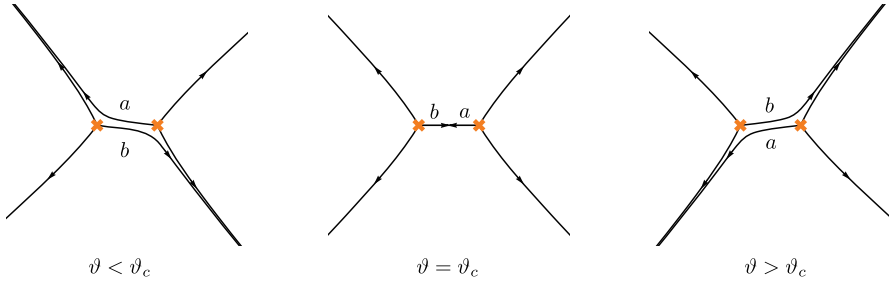


FIGURE 24. The simplest way in which  $\mathcal{W}_\vartheta$  can jump at a  $\mathcal{K}$ -wall. As  $\vartheta \rightarrow \vartheta_c$ , two  $\mathcal{S}$ -walls merge into a single two-way street running between two branch points. The two-way street supports two distinct charges  $\bar{a}, \bar{b}$

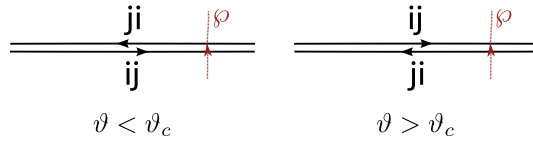


FIGURE 25. A path  $\wp$  crossing a pair of nearly coincident  $\mathcal{S}$ -walls, which coalesce into a two-way street at  $\vartheta = \vartheta_c$

is a two-way street. Indeed,  $\mathcal{W}_{\vartheta_c}$  always contains at least one two-way street, since we get a two-way street whenever an  $\mathcal{S}$ -wall runs into a branch point.

Now let us consider a path  $\wp$  which crosses a two-way street  $p$ . A representative picture of the situation when  $\vartheta$  is near  $\vartheta_c$  is indicated in Fig. 25.<sup>22</sup> We define BPS soliton degeneracies  $\mu^\pm(a, p)$  along the walls of  $\mathcal{W}_{\vartheta_c}$  by taking limits of  $\mu(a, p)$  as  $\vartheta \rightarrow \vartheta_c^\pm$ . Applying the rules of Sect. 4.4 and taking the limit  $\vartheta \rightarrow \vartheta_c^\pm$ , we obtain formulas for  $F(\wp, \vartheta_c^\pm)$ :

$$F(\wp, \vartheta_c^-) = \mathcal{D}(\wp_+) \left( 1 + \sum_{\bar{b} \in \Gamma_{ij}^s} \mu^-(b) X_{b_{\bar{z}}} \right) \left( 1 + \sum_{\bar{a} \in \Gamma_{ji}^s} \mu^-(a) X_{a_{\bar{z}}} \right) \mathcal{D}(\wp_-), \quad (53)$$

$$F(\wp, \vartheta_c^+) = \mathcal{D}(\wp_+) \left( 1 + \sum_{\bar{a} \in \Gamma_{ji}^s} \mu^+(a) X_{a_{\bar{z}}} \right) \left( 1 + \sum_{\bar{b} \in \Gamma_{ij}^s} \mu^+(b) X_{b_{\bar{z}}} \right) \mathcal{D}(\wp_-). \quad (54)$$

Here  $\Gamma_{ij}^s \subset \Gamma_{ij}(z, z)$  denotes the set of charges supported along the two-way street.

Note that the order in these products really matters, since  $X_{a_{\bar{z}}}$  and  $X_{b_{\bar{z}}}$  do not commute. A convenient way to think about this is to regard the limiting network  $\mathcal{W}_{\vartheta_c}$  as equipped with a bit of extra structure: each two-way street is resolved into two infinitesimally separated and oppositely oriented “lanes”.

<sup>22</sup> In general the picture might be a bit more complicated than Fig. 25, because there might be several walls of type  $ij$  and/or of type  $ji$  which all coalesce at  $\vartheta = \vartheta_c$ . An example appears in Fig. 27 below.

For  $\vartheta_c^-$  the division of lanes is according to the American rule (drive on the right), while for  $\vartheta_c^+$  it is according to the British rule (drive on the left). We can determine  $F(\wp, \vartheta_c^\pm)$  completely from  $\mathcal{W}_{\vartheta_c}$  and  $\mu^\pm$ , using either (53) or (54) as appropriate.

### 6.3. The Jump of $F(\wp, \vartheta)$ at a $\mathcal{K}$ -wall

The jump of  $F(\wp, \vartheta)$  at  $\vartheta_c$  is given by a certain universal substitution  $\mathcal{K}$  acting on the formal variables  $X_a$ :

$$F(\wp, \vartheta_c^+) = \mathcal{K} \left( F(\wp, \vartheta_c^-) \right). \quad (55)$$

$\mathcal{K}$  is determined by the degenerate spectral network  $\mathcal{W}_{\vartheta_c}$ . In this section, we describe what  $\mathcal{K}$  is, deferring the proof of (55) to Sect. 6.6.

We are going to combine the soliton degeneracies on any two-way street  $p$  into a new generating function  $Q(p)$ . Unlike the generating functions  $F(\wp, \vartheta)$  we have considered before,  $Q(p)$  is written in terms of formal variables  $X_{\tilde{\gamma}}$  with  $\tilde{\gamma} \in \tilde{\Gamma}$ , where  $\tilde{\Gamma} = H_1(\tilde{\Sigma}; \mathbb{Z})$  modulo shifts by  $2H$  (cf. the definition of  $\tilde{\Gamma}(\tilde{z}_1, \tilde{z}_2)$  in Sect. 3.5). So these formal variables are associated to closed paths, rather than the open ones we have encountered up to now. We extend our multiplication rules,

$$X_{\tilde{\gamma}} X_a = X_{\tilde{\gamma}+a}, \quad X_{\tilde{\gamma}} X_{\tilde{\gamma}'} = X_{\tilde{\gamma}+\tilde{\gamma}'} \quad (56)$$

(where the  $+$  denotes the obvious action of  $\tilde{\Gamma}$  on  $\tilde{\Gamma}(\tilde{z}_1, \tilde{z}_2)$  or on  $\tilde{\Gamma}$ ). We also extend our sign rules by imposing

$$X_H = -1, \quad (57)$$

so the two  $X_{\tilde{\gamma}}$  corresponding to different lifts of a single  $\gamma \in \Gamma$  differ only by a sign.

One can build closed paths from open ones: given  $a \in \tilde{\Gamma}_{ii}(\tilde{z}, \tilde{z})$ , there is a corresponding  $\text{cl}(a) \in \tilde{\Gamma}$  which is obtained just by forgetting the basepoint  $\tilde{z}^{(i)}$ . Using this, we define

$$Q(p) = 1 + \sum_{\bar{a} \in \Gamma_{ij}^s, \bar{b} \in \Gamma_{ji}^s} \mu^-(a, p) \mu^-(b, p) X_{\text{cl}(a+b)}. \quad (58)$$

(We will see later that we would have gotten the same  $Q(p)$  if we had used  $\mu^+$  instead of  $\mu^-$  on the right.)  $Q(p)$  is a power series in the variables  $X_{\tilde{\gamma}}$ , of a constrained sort: every  $\gamma$  that occurs is a sum  $\bar{a} + \bar{b}$  of charges supported at  $z$ , and hence has  $e^{-i\vartheta_c} Z_\gamma \in \mathbb{R}_-$ .

Now we make a new genericity assumption. Let  $\Gamma_c \subset \Gamma$  be the set of all  $\gamma \in \Gamma$  with  $e^{-i\vartheta_c} Z_\gamma \in \mathbb{R}_-$ ; we assume that  $\Gamma_c$  is generated by a single element  $\gamma_0$ . (This condition holds automatically if our chosen  $u = (\phi_2, \dots, \phi_K)$  is not on a wall of marginal stability in the Coulomb branch.)  $Q(p)$  is then a power series in a single variable: choosing a lift  $\tilde{\gamma}_0$ , each  $X_{\text{cl}(a+b)} = \pm X_{\tilde{\gamma}_0}^n$  for some  $n > 0$ .

Our next aim is to extract the BPS degeneracies by rewriting this power series as a product, (60) below.

For each  $\gamma \in \Gamma$  we can define a preferred lift  $\tilde{\gamma}$  by the following rule:<sup>23</sup> represent  $\gamma$  as a union of smooth closed curves  $c_m$  on  $\Sigma$ ; then  $\tilde{\gamma}$  is the sum of the canonical lifts of the  $c_m$  to  $\tilde{\Sigma}$ , shifted by  $\left(\sum_{m \leq n} \delta_{mn} + \#(c_m \cap c_n)\right) H$  (of course since we work modulo  $2H$ , all that matters here is whether this sum is odd or even.) One can check directly that  $\tilde{\gamma}$  so defined is independent of the choice of how we represent  $\gamma$  as a union of  $c_m$  (this requirement is what forced us to add the tricky-looking shift.) Moreover, with this definition one has

$$X_{\tilde{\gamma}+\tilde{\gamma}'} = (-1)^{\langle \gamma, \gamma' \rangle} X_{\widetilde{\gamma+\gamma'}}. \quad (59)$$

We are ready to factorize  $Q(p)$ : there exist exponents  $\alpha_\gamma(p) \in \mathbb{Z}$  such that

$$Q(p) = \prod_{\gamma \in \Gamma_c} (1 - X_{\tilde{\gamma}})^{\alpha_\gamma(p)}. \quad (60)$$

(Indeed, the equations determining the  $\alpha_\gamma(p)$  from  $Q(p)$  are upper-triangular and hence can be solved.) Then, for any  $\gamma \in \Gamma_c$ , define a 1-chain  $L(\gamma)$  on  $\Sigma$  by

$$L(\gamma) = \sum_p \alpha_\gamma(p) p_\Sigma, \quad (61)$$

where  $p$  runs over the walls in  $\mathcal{W}_{\partial_c}$ , and  $p_\Sigma$  is the oriented 1-chain obtained by lifting  $p$  as in Sect. 3.1.

It is a crucial fact (proven in Sect. 6.6 below) that  $L(\gamma)$  so defined is actually a 1-cycle. This allows us to define<sup>24</sup> an intersection number  $\langle \bar{c}, L(\gamma) \rangle$  for any  $\bar{c} \in \Gamma(z_1, z_2)$ . Moreover,  $e^{-i\vartheta_c} \int_{p_\Sigma} \lambda \in \mathbb{R}_-$ , so  $e^{-i\vartheta_c} Z_{[L(\gamma)]} \in \mathbb{R}$ ; under our genericity assumption this implies that the homology class  $[L(\gamma)]$  is a multiple of  $\gamma$  (though not necessarily a positive multiple). This fact will be useful below.

Finally we can define our universal substitution  $\mathcal{K}$ :

$$\mathcal{K}(X_a) = \prod_{\gamma \in \Gamma_c} (1 - X_{\tilde{\gamma}})^{\langle \bar{a}, L(\gamma) \rangle} X_a. \quad (62)$$

#### 6.4. 4d BPS Degeneracies

According to the analysis of [21, 5], the jump  $\mathcal{K}$  of the framed 2d–4d degeneracies captures the degeneracies of 4d BPS particles.<sup>25</sup> Since we have now given a formula for  $\mathcal{K}$ , it follows that we can use it to determine the 4d BPS degeneracies. Comparing our formula (62) to those of [5, 21], we find that the

<sup>23</sup> In the first preprint version of this paper we proposed a different lifting rule based on the principle that the product (60) should be finite. More recently, in studying more complicated examples, it has turned out that sometimes this finiteness is violated [46]. For completeness, we have back-ported the corrected rule from [46] to here.

<sup>24</sup> Different representatives of the class  $\bar{c}$  differ by addition of 1-boundaries, and the 1-cycle  $L(\gamma)$  has zero intersection with any 1-boundary; this would not have worked if  $L(\gamma)$  were merely a 1-chain.

<sup>25</sup> This is possible because the jump occurs when framed 2d–4d BPS bound states form/decay by binding/releasing 4d BPS particles.

enhanced BPS degeneracies reviewed in Sect. 3.4 are given here by<sup>26</sup>

$$\omega(\gamma, \bar{a}) = \langle L(\gamma), \bar{a} \rangle. \quad (63)$$

This result contains much more information than the ordinary 4d BPS degeneracies: it knows not only the total number of 4d BPS states but also some local information about where they sit on  $C$ , as measured by their interaction with surface defects. Nevertheless, it is also interesting to see how we can recover simpler invariants: using (63) and (11), we find that the 4d BPS degeneracy is

$$\Omega(\gamma) = [L(\gamma)]/\gamma. \quad (64)$$

In examples below, we will illustrate how this formula determines  $\Omega(\gamma)$  in various concrete situations.

### 6.5. Wall-Crossing Formula

In Sect. 2.3 of [5], we wrote a *2d–4d wall-crossing formula* which should be obeyed by the degeneracies  $\mu$  and  $\omega$  in any coupled 2d–4d system. We have given an explicit description of  $\mu$  and  $\omega$  in the theories we are now considering. Therefore, it is natural to ask whether one can show directly that the  $\mu$  and  $\omega$  defined here indeed obey this formula. In a sense, our definition of  $\omega$  above was engineered so that this will be true. We will not be able to give a complete proof here, but let us at least explain the idea.

The basic strategy of proof was already explained in [5], as follows. Let  $u = (\phi_2, \dots, \phi_K)$  denote a point of the Coulomb branch. Let  $\wp$  denote a path in  $C$  which both begins and ends at some point  $z$ . As we vary the basepoint  $z$ , we can likewise deform the path  $\wp$  (in a unique way up to homotopy), giving a family of paths  $\wp(z)$ . Now let  $z$ ,  $\vartheta$  and  $u$  vary along small contractible loops  $z(t)$ ,  $\vartheta(t)$ ,  $u(t)$  ( $0 \leq t \leq 1$ ) in parameter space, with  $z(0) = z$ , and consider the corresponding generating functions  $F(\wp(z(t)), \vartheta(t), u(t))$ ,<sup>27</sup> which we denote simply as  $F(\wp, t)$ .

Since  $t$  is a closed contractible loop we have

$$F(\wp, t = 0) = F(\wp, t = 1). \quad (65)$$

On the other hand, as  $t$  varies,  $F(\wp, t)$  jumps at various critical  $t_c$ . These jumps occur either when  $z(t)$  crosses an  $\mathcal{S}$ -wall in  $\mathcal{W}_{\vartheta(t)}$  or when  $\vartheta(t)$  crosses a  $\mathcal{K}$ -wall.

<sup>26</sup> More precisely, (62) should be compared with (2.30) of [5], except for the detail that in this paper we are working with framed protected spin characters at  $y = 1$  rather than  $y = -1$ . The needed modification for  $y = 1$  has not quite appeared anywhere before, although in (3.26) of [21] we did give the jump of the framed degeneracies (without surface defects) at  $y = 1$ . We also noted there (in Sect. 6.4) that under the assumption that there are no “exotic BPS states,” the jump becomes somewhat simpler (because then all of the factors  $(-1)^m$  appearing in (3.26) are the same). Strictly speaking then, (62) should be compared with the most obvious combination of (2.30) of [5] and (3.26) of [21], taking into account this simplification. A final detail: in (2.30) of [5] there appeared an undetermined  $\pm 1$ -valued cocycle  $\sigma$ ; in (62) this sign is encoded in the choice of lift  $\tilde{\gamma}$  of  $\gamma$ .

<sup>27</sup> Up until now we have usually represented these generating functions as  $F(\wp, \vartheta)$ , holding  $u$  fixed and implicit; but to recover the usual statement of the wall-crossing formula we have to let  $u$  vary.

At each critical  $t_c$ , the jump of  $F(\wp, t)$  is given by an explicit transformation  $J(t_c)$  of the formal variables  $X_a$ . Writing the total jump as  $\mathbf{J} = \prod_{t_c} J(t_c)$  (with the product taken in order of increasing  $t_c$ ), (65) thus says

$$\mathbf{J}[F(\wp, t = 0)] = F(\wp, t = 0). \quad (66)$$

Now, the statement of the 2d–4d wall-crossing formula is that  $\mathbf{J}$  is the identity, or more concretely,

$$\mathbf{J}[X_a] = X_a \quad \text{for every } a \in \Gamma(z, z). \quad (67)$$

The most direct way to obtain (67) from (66) would be to show that any  $X_a$  for  $a \in \Gamma(z, z)$  can be obtained as a linear combination of the  $F(\wp, t = 0)$  for various paths  $\wp$  from  $z$  to  $z$ . We have not proven this, and indeed (as one sees by considering simple examples) it cannot literally be true except under some restrictions on the type of surface defects  $D_n$  we allow; in general, we expect to have to extend the set of allowed  $\wp$  to include some paths which run into the surface defects, as we did in [21].

Even after extending the set of allowed  $\wp$  appropriately, it does not appear to be straightforward to show that any  $X_a$  can be obtained as a linear combination of the  $F(\wp, t = 0)$ . We believe this is an interesting and important question (related to conjectures of Fock–Goncharov on the relation between universally positive Laurent polynomials and tropical points; we discussed the  $K = 2$  case of this connection in [21].) However, we can also propose an alternative “poor man’s” approach to proving (67). The idea is that even if we cannot show that we can express  $X_a$  literally as a function of the  $F(\wp, t = 0)$ , we can at least do so up to at most a finite ambiguity. More precisely, we claim that the group  $G$  of automorphisms obeying the equation (66) is *finite*. This point will be explained in Sect. 10.5 below. Assuming it for now, we conclude in particular that  $\mathbf{J}$  is of finite order. But the automorphisms  $J(t_c)$  are all “upper triangular”, in the sense that each  $J(t_c)$  is of the form  $1 + X$  where all of the  $X$  belong to a common pronilpotent group. It follows that  $\mathbf{J}$  is also of this form; but then  $\mathbf{J}$  cannot have finite order without being the identity.

## 6.6. Proof of the $\mathcal{K}$ -wall Formula

Here, we provide the proofs omitted in Sect. 6.3.

First let us explain why  $L(\gamma)$  as defined in (61) is indeed a 1-cycle, i.e. has no boundary. For any two-way street  $p$ , the boundary of  $p_\Sigma$  lies over the boundary of  $p$ . More precisely, a boundary point of  $p$  may be either a branch point or a joint;  $p_\Sigma$  has no boundary over a branch point, but does have a boundary over a joint. So we need to check that  $L(\gamma)$  has no boundary over a joint. To establish this we consider Fig. 26.

Let  $z$  denote the joint in the figure. The coefficient of  $z^{(i)}$  in  $\partial L(\gamma)$  is  $\alpha_\gamma(p_1) + \alpha_\gamma(p_2) - \alpha_\gamma(p_3) - \alpha_\gamma(p_4)$ , and we would like to show that this vanishes. Consider the open paths  $\wp$  and  $\wp'$ . The constraint of twisted homotopy invariance says that  $F(\wp, \vartheta_c^-) = F(\wp', \vartheta_c^-)$ . In particular, we can look at the pieces on both sides which involve paths which both begin and end on sheet  $i$ :

$$F(\wp, \vartheta_c^-)_{ii} = F(\wp', \vartheta_c^-)_{ii}. \quad (68)$$

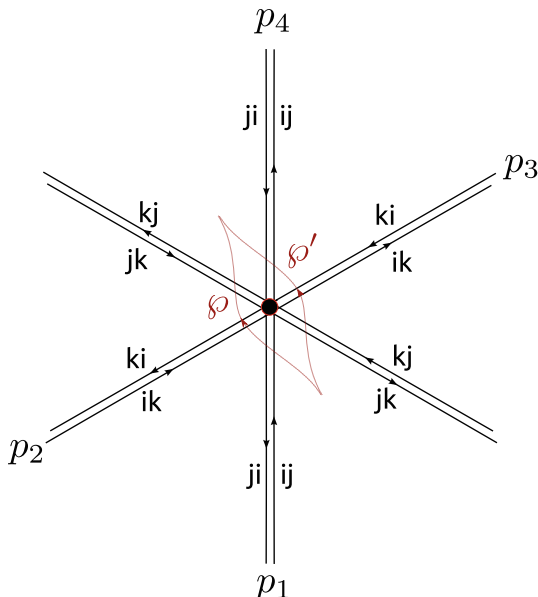


FIGURE 26. A joint where two-way streets meet. Each two-way street is shown slightly “resolved”, as explained at the end of Sect. 6.2. The resolution here is the “American resolution,” as appropriate since we are considering the limit  $\vartheta \rightarrow \vartheta_c^-$

Directly computing the two sides of (68) we obtain

$$Q(p_1)Q(p_2) = Q(p_3)Q(p_4). \quad (69)$$

This amounts to

$$\alpha_\gamma(p_1) + \alpha_\gamma(p_2) = \alpha_\gamma(p_3) + \alpha_\gamma(p_4), \quad (70)$$

which is what we wanted to show.

Now let us explain why the jump formula (55) is true. We will show that the following hold both for  $\mathcal{F}(\varphi) = F(\varphi, \vartheta_c^+)$  and  $\mathcal{F}(\varphi) = \mathcal{K}(F(\varphi, \vartheta_c^-))$ :

**P1.**  $\mathcal{F}(\varphi)$  is a twisted homotopy invariant of  $\varphi$  (in the sense explained in Sect. 3.5).

**P2.** If  $\varphi$  does not meet  $\mathcal{W}_{\vartheta_c}$ , then

$$\mathcal{F}(\varphi) = \mathcal{D}(\varphi). \quad (71)$$

**P3.** If  $\varphi$  and  $\varphi'$  have endpoints off  $\mathcal{W}_{\vartheta_c}$ ,  $\mathcal{F}$  obeys the composition law

$$\mathcal{F}(\varphi)\mathcal{F}(\varphi') = \mathcal{F}(\varphi\varphi'). \quad (72)$$



- P4.** If  $\wp$  crosses  $\mathcal{W}_{\vartheta_c}$  exactly once at a point  $z$  on a one-way street  $p$ , then  $\mathcal{F}(\wp)$  is of the form

$$\mathcal{F}(\wp) = \mathcal{D}(\wp_+) \left( 1 + \sum_{\bar{a} \in \Gamma_{ij}^s} \mu(a) X_{a\bar{z}} \right) \mathcal{D}(\wp_-), \quad (73)$$

for some  $\mu(a) \in \mathbb{Z}$ .

- P5.** If  $\wp$  crosses  $\mathcal{W}_{\vartheta_c}$  exactly once at a point  $z$  on a two-way street  $p$ , and the intersection between  $\wp$  and  $p$  is positive (with respect to the orientation of  $p$  as a  $ji$  trajectory and the underlying orientation of  $C$  as a complex curve), then  $\mathcal{F}(\wp)$  is of the form

$$\mathcal{F}(\wp) = \mathcal{D}(\wp_+) \left( 1 + \sum_{\bar{a} \in \Gamma_{ji}^s} \mu(a) X_{a\bar{z}} \right) \left( 1 + \sum_{\bar{b} \in \Gamma_{ij}^s} \mu(b) X_{b\bar{z}} \right) \mathcal{D}(\wp_-) \quad (74)$$

for some  $\mu(a) \in \mathbb{Z}$  and  $\mu(b) \in \mathbb{Z}$ .

For  $\mathcal{F}(\wp) = F(\wp, \vartheta_c^+)$ , these properties follow directly from what we have already said; the most nontrivial one is P5, which is (54). To prove them for  $\mathcal{F}(\wp) = \mathcal{K}(F(\wp, \vartheta_c^-))$  is slightly harder. P1–3 are true of  $F(\wp, \vartheta_c^-)$  and clearly preserved by  $\mathcal{K}$ . P4 is also true of  $F(\wp, \vartheta_c^-)$ , but we have to show it is preserved by  $\mathcal{K}$ : this follows simply from the fact that  $\mathcal{K}$  just multiplies each term by a function of the  $X_{\tilde{\gamma}}$ , which indeed preserves the form (73). P5 is the only really nontrivial one. We begin with the formula (53) for  $F(\wp, \vartheta_c^-)$ , and need to show that the action of  $\mathcal{K}$  transforms it into the form (74). Expanding out (53) we find various classes of terms:

$$F(\wp, \vartheta_c^-)_{ii} = \wp_+^{(i)} \left( 1 + \sum_{\bar{a}, \bar{b}} \mu^-(a) \mu^-(b) X_b X_a \right) \wp_-^{(i)} = Q(p) \wp_-^{(i)}, \quad (75)$$

$$F(\wp, \vartheta_c^-)_{ij} = \wp_+^{(i)} \left( \sum_{\bar{b}} \mu^-(b) X_b \right) \wp_-^{(j)}, \quad (76)$$

$$F(\wp, \vartheta_c^-)_{ji} = \wp_+^{(j)} \left( \sum_{\bar{a}} \mu^-(a) X_a \right) \wp_-^{(i)}, \quad (77)$$

$$F(\wp, \vartheta_c^-)_{jj} = \wp_+^{(j)}, \quad (78)$$

$$F(\wp, \vartheta_c^-)_{kk} = \wp_-^{(k)} \quad \text{for } k \notin \{i, j\}. \quad (79)$$

Within each class, the terms in the sum over  $\bar{a}, \bar{b}$  differ only by multiplication by factors  $X_{\tilde{\gamma}}$ . Since  $L(\gamma)$  is a multiple of  $\gamma$ , it follows that  $\langle \gamma, L(\gamma) \rangle = 0$ . Hence the action of  $\mathcal{K}$  on each class of terms is independent of the particular term. We consider them in turn. For the  $ii$  terms,  $\mathcal{K}$  acts by multiplication by  $\prod_{\gamma \in \Gamma_c} (1 + X_{\tilde{\gamma}})^{\langle \wp^{(i)}, L(\gamma) \rangle}$ , but using (61) we see that  $\langle \wp^{(i)}, L(\gamma) \rangle = -\alpha_{\gamma}(p)$  (since  $p$  is the only edge of  $\mathcal{W}$  crossed by  $\wp$ ) and hence this reduces to multiplication by  $Q(p)^{-1}$ . On the  $jj$  terms,  $\mathcal{K}$  similarly acts by multiplication by

$Q(p)$ . On the  $ij$  terms, the action of  $\mathcal{K}$  is more complicated; it is multiplication by a new function  $H = \prod_{\gamma \in \Gamma_c} (1 + X_{\tilde{\gamma}})^{\langle \wp_+^{(i)} + \bar{b} + \wp_-^{(j)}, L(\gamma) \rangle}$ . But we have  $\langle \wp_+^{(i)} + \bar{b} + \wp_-^{(j)}, L(\gamma) \rangle = -\langle \wp_+^{(j)} + \bar{a} + \wp_-^{(i)}, L(\gamma) \rangle$  as one readily sees using the facts that  $\text{cl}(\bar{a} + \bar{b})$  is proportional to  $L(\gamma)$  and  $\langle \wp^{(i)}, L(\gamma) \rangle = -\langle \wp^{(j)}, L(\gamma) \rangle$ . It follows that the action of  $\mathcal{K}$  on the  $ji$  terms is multiplication by  $H^{-1}$ . These facts together are sufficient to imply the desired P5.<sup>28</sup>

Finally we must explain why these properties actually *determine*  $\mathcal{F}(\wp)$ . As a warm-up, suppose that there are only one-way streets. In that case we have shown in Sects. 4 and 5 above that the properties P1–P4 are sufficient to determine all of the  $\mu(a)$ , and hence also to fix  $\mathcal{F}(\wp)$ . The same kind of argument can be made in the presence of two-way streets; the only new complication is that the local structure around joints and branch points is more complicated. Nevertheless, a direct computation using only the fact that  $\mathcal{F}(\wp)$  obey P1–P5 determines the  $\mu(a)$  appearing there for all charges  $a$  oriented out of a joint, in terms of the  $\mu(a)$  for charges  $a$  oriented into the same joint. (The resulting explicit formulae are recorded in Appendix A, e.g. (163).) Inductively we can thus determine the 2d–4d degeneracies  $\mu(a)$  everywhere on  $\mathcal{W}_{\vartheta_c}$ , just as we did in the (generic) case with only one-way streets. Having done so the  $\mathcal{F}(\wp)$  are also determined by P1–P5. This completes the proof.

## 7. Examples of $\mathcal{K}$ -walls

In this section, we illustrate the general discussion of Sect. 6 with some examples.

### 7.1. Saddle Connections

The simplest possibility has already appeared in Fig. 24 above. It involves two  $\mathcal{S}$ -walls, colliding along a two-way street  $p$  running between two branch points. We call such a two-way street a “saddle connection,” following the standard terminology for trajectories of quadratic differentials (which is literally what we are considering here in the case  $K = 2$ ).

The two  $\mathcal{S}$ -walls support charges  $\bar{a}$  and  $\bar{b}$ , with natural lifts  $a, b$ , and we have

$$\mu^-(a) = 1, \quad \mu^-(b) = 1, \quad (80)$$

with all other  $\mu^-(\cdot)$  vanishing. In this case, (58) reads

$$Q(p) = 1 + X_{\text{cl}(a+b)}. \quad (81)$$

If we let  $\gamma = \text{cl}(\bar{a} + \bar{b})$ , the preferred lift is  $\tilde{\gamma} = \text{cl}(a + b) + H$ ; we thus have

$$Q(p) = 1 - X_{\tilde{\gamma}}. \quad (82)$$

<sup>28</sup> Incidentally, they also imply that  $\mathcal{F}(\wp)_{jj} = Q(p)$ ; this and the fact  $\mathcal{F}(\wp) = F(\wp, \vartheta_c^+)$  (which we are in the process of showing) together prove our remark under (58).

This can indeed be decomposed according to (60), taking  $\alpha_\gamma(p) = 1$ , and  $\alpha_{n\gamma}(p) = 0$  for  $n \neq 1$ . We thus obtain the simple result

$$L(\gamma) = p_\Sigma \quad (83)$$

and  $L(n\gamma) = 0$  for  $n \neq 1$ . As we have seen above, this result completely determines the wall-crossing at this  $\mathcal{K}$ -wall, through the formula (62).

In particular, we have  $[L(\gamma)] = \gamma$ , so the 4d BPS degeneracy is

$$\Omega(\gamma) = 1. \quad (84)$$

We have thus found that a saddle connection represents a BPS hypermultiplet. This recovers a result of [6, 3] for the case  $K = 2$ , and extends it to arbitrary  $K$ .

## 7.2. Closed Loops

A more interesting possibility is shown in Fig. 27. This case involves an  $\mathcal{S}$ -wall which winds around a cylinder in  $C$  many times, more and more tightly as  $\vartheta \rightarrow \vartheta_c$ . In this case, two two-way streets appear simultaneously at  $\vartheta = \vartheta_c$ ; let  $p$  denote one of them.  $p$  begins and ends at the *same* branch point, and the soliton degeneracies are a bit more interesting than the previous case: there are infinitely many nonvanishing  $\mu^-(\cdot)$ , coming from the infinite set of windings of the  $\mathcal{S}$ -wall around the cylinder, which have all coalesced onto  $p$ . Explicitly

$$\mu^-(a) = 1, \quad \mu^-(b + n(a + b)) = 1 \text{ for all } n \geq 0, \quad (85)$$

with all other  $\mu^-(\cdot)$  vanishing. Hence by (58),

$$Q(p) = 1 + X_{\text{cl}(a+b)} + X_{\text{cl}(2(a+b))} + \cdots \quad (86)$$

$$= (1 - X_{\text{cl}(a+b)})^{-1}. \quad (87)$$

Define  $\gamma = \text{cl}(\bar{a} + \bar{b})$ . The preferred lift of  $\gamma$  is  $\tilde{\gamma} = \text{cl}(a + b)$ , so we have  $Q(p) = (1 - X_{\tilde{\gamma}})^{-1}$ , which means  $\alpha_\gamma(p) = -1$ . Now, to compute  $L(\gamma)$  we must sum the contributions from both two-way streets—call them  $p^1$  and  $p^2$ —so we obtain

$$L(\gamma) = -p_\Sigma^1 - p_\Sigma^2. \quad (88)$$

Both  $p_\Sigma^1$  and  $p_\Sigma^2$  are in the homology class  $\gamma$ , so  $[L(\gamma)] = -2\gamma$ , or

$$\Omega(\gamma) = -2. \quad (89)$$

Thus we have recovered the result of [6, 3]: this pair of closed trajectories represents a BPS vectormultiplet.

In [3], we derived the result  $\Omega(\gamma) = -2$  in this situation (in the special case  $K = 2$ ), by a rather delicate analysis of the jumps of the vacuum expectation values of line defects associated to *closed* paths. This analysis gave the right answer but depended on the assumption that at  $\vartheta = \vartheta_c$  the third trajectory emerging from the branch points in Fig. 27 ends on a puncture. As Ivan Smith pointed out to us, this assumption can be violated for higher genus  $C$ : the cylinder of closed trajectories could be dividing  $C$  into two pieces, one of which contains no puncture. Our present analysis using line defects associated to *open* paths is much simpler and is free of such extra assumptions.

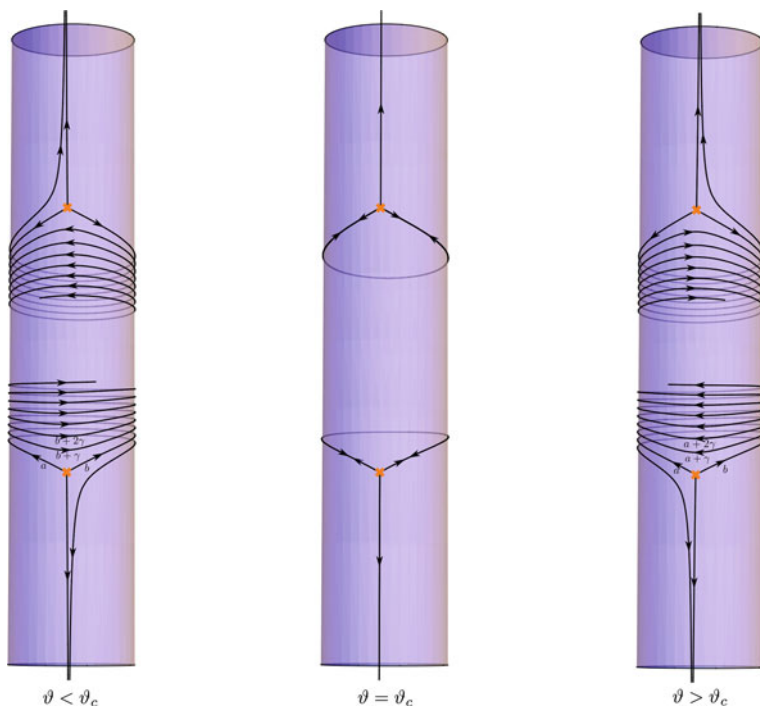


FIGURE 27. A more complicated way in which  $\mathcal{S}$ -walls can collide head-on. At  $\vartheta < \vartheta_c$ , the  $\mathcal{S}$ -wall supporting charge  $a$  winds once clockwise around the cylinder and then disappears toward the bottom, while the  $\mathcal{S}$ -wall supporting charge  $b$  winds many times counterclockwise around the cylinder and eventually disappears toward the top. In the figure, for clarity we truncate the latter  $\mathcal{S}$ -wall after it has wound around a few times. At the critical  $\vartheta = \vartheta_c$ , all of the windings of the  $\mathcal{S}$ -wall supporting charge  $b$  coalesce onto a single closed trajectory. At  $\vartheta > \vartheta_c$  these two  $\mathcal{S}$ -walls exchange roles

Incidentally, from the point of view of the theory  $S[\mathfrak{g}, C, D]$ , these two closed trajectories would not seem to be the whole story. Indeed, these two  $ij$ -trajectories are actually the two ends of a one-parameter *family* of closed  $ij$ -trajectories, which sweep out a cylinder on  $C$ . Any member of this family is a BPS string. So physically speaking, we should regard this  $\Omega(\gamma) = -2$  as giving the contribution from this whole one-parameter family, not just from the two ends of the family. From our current point of view, though the ends play a privileged role, in that they are  $\mathcal{S}$ -walls, while the trajectories in the interior of the cylinder are not.

Finally, although we have hidden it up to this point, the behavior of the networks  $\mathcal{W}_\vartheta$  as  $\vartheta \rightarrow \vartheta_c$  is actually rather complicated. In Fig. 27, at  $\vartheta < \vartheta_c$ , we see two  $\mathcal{S}$ -walls which wind many times around the cylinder. There we

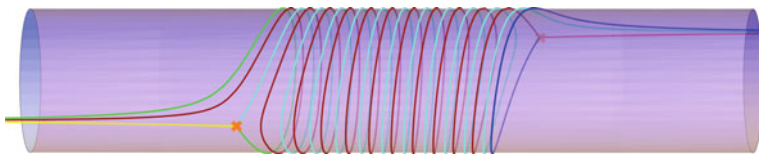


FIGURE 28. Another picture of the  $\mathcal{S}$ -walls on the cylinder for  $\vartheta$  near the critical phase  $\vartheta_c$ . Here we do *not* truncate the walls, and we distinguish different walls by giving them different colors

truncated them at some finite distance in order not to make the figure too confusing. In Fig. 28, we show the full  $\mathcal{S}$ -walls at some particular  $\vartheta < \vartheta_c$ . Note that the two winding  $\mathcal{S}$ -walls thread past one another many times before they escape the cylinder. As we vary  $\vartheta$  slightly, the angle at which they are threaded changes, and there are infinitely many critical phases  $\vartheta_n$  at which they actually collide head-to-head. The critical phases  $\vartheta_n$  accumulate at the value  $\vartheta_c$ . At each of these critical phases we have a saddle connection and the network  $\mathcal{W}_\vartheta$  jumps, just as we discussed in Sect. 7.1.

We analyzed this limiting process rather closely in [3]; but we emphasize that in our current analysis, if all we want to know is the BPS degeneracy  $\Omega(\gamma)$  which appears exactly at the critical phase  $\vartheta_c$ , it is *not* necessary to study this infinite sequence of jumps.

### 7.3. Three-String Webs

When we go beyond  $K = 2$ , we encounter many new and varied phenomena.

The simplest new possibility is shown in Fig. 29. Near the critical phase  $\vartheta = \vartheta_c$ , we see *three* pairs of  $\mathcal{S}$ -walls nearly colliding head-on. Each pair consists of one “primary”  $\mathcal{S}$ -wall, born at a branch point, and one “secondary”  $\mathcal{S}$ -wall, born at the intersection between two primary walls. At the critical phase  $\vartheta = \vartheta_c$  each such pair of walls collides along a two-way street  $p$ . Each of the three two-way streets  $p$  supports a different pair of two 2d–4d charges, with all three pairs summing to  $\gamma = \text{cl}(\bar{a} + \bar{b} + \bar{c})$ .

We would like to compute the 1-cycle  $L(\gamma)$ . Let us focus attention on the two-way street  $p$  supporting the 2d–4d charges  $\bar{a}$  and  $\bar{b} + \bar{c}$ . At  $\vartheta < \vartheta_c$ , the  $\mathcal{S}$ -wall supporting charge  $\bar{a}$  is emerging directly from a branch point, and thus has a natural lift  $a$  with  $\mu(a) = 1$ . The  $\mathcal{S}$ -wall supporting  $\bar{b} + \bar{c}$ , on the other hand, is emerging from a joint where walls supporting  $\bar{b}$  and  $\bar{c}$  intersect. These two walls in turn emerge from branch points, so at the joint  $\mu^-(b) = 1$ ,  $\mu^-(c) = 1$ . It follows from (34) that the emerging wall has  $\mu^-(b + c) = 1$ . Since these are the only two  $\mathcal{S}$ -walls which collide along  $p$ , all other  $\mu^-(\cdot)$  vanish along  $p$ . Plugging into (58) we obtain

$$Q(p) = 1 + X_{\text{cl}(a+b+c)} \quad (90)$$

and hence, letting  $\gamma = \text{cl}(\bar{a} + \bar{b} + \bar{c})$  with preferred lift  $\tilde{\gamma} = \text{cl}(a + b + c) + H$ , for suitable  $a, b, c$ , we find  $\alpha_\gamma(p) = 1$ . The same result holds for the other two two-way streets  $p^n$ , so  $L(p)$  is the sum:

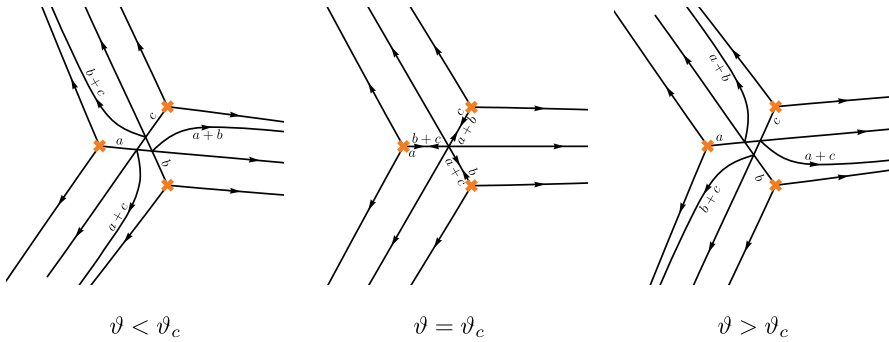


FIGURE 29. A new way in which  $\mathcal{S}$ -walls can collide, which appears only in theories with  $K > 2$ . The picture shows a patch of  $C$  containing three branch points.  $\mathcal{W}_\vartheta$  involves both “primary”  $\mathcal{S}$ -walls born from the branch points (supporting charges  $a, b, c$ ) and “secondary”  $\mathcal{S}$ -walls born from intersections between primary  $\mathcal{S}$ -walls (supporting charges  $a + b, b + c, a + c$ .) At the critical phase  $\vartheta = \vartheta_c$ , primary and secondary walls collide head-on, along a locus which forms a three-string finite web

$$L(\gamma) = p_\Sigma^1 + p_\Sigma^2 + p_\Sigma^3. \quad (91)$$

In other words, letting  $N$  denote the finite web made up of the three two-way streets  $p^n$ , we have

$$L(\gamma) = N_\Sigma. \quad (92)$$

Moreover,  $[N_\Sigma] = \gamma$ , so the 4d BPS degeneracy here is

$$\Omega(\gamma) = 1. \quad (93)$$

Therefore, we have found that the finite web  $N$ , made up of three strings which meet at a junction, corresponds to a BPS hypermultiplet.

#### 7.4. The Setting Sun

Next we briefly consider a more complicated example, in which several overlapping finite webs appear simultaneously. Unlike the previous examples, we just draw the degenerate network of interest, and not the nondegenerate ones at nearby phases (which would be terribly cluttered in this example). See Fig. 30.

Label the three two-way streets in this figure as  $p^1, p^2, p^3$  from top to bottom. Using the rules of Appendix A, we can directly compute the soliton degeneracies  $\mu(\cdot, \cdot)$  on all the streets in the figure. The nonzero degeneracies on  $p^1, p^2, p^3$  come out to

$$\mu(b, p^1) = 1, \quad \mu(c + a + n(b + c + a), p^1) = 1 \quad \text{for } n \geq 0, \quad (94)$$

$$\mu(c, p^2) = 1, \quad \mu(a + b, p^2) = 1, \quad (95)$$

$$\mu(a + n(b + c + a), p^3) = 1 \quad \text{for } n \geq 0, \quad \mu(b + c, p^3) = 1. \quad (96)$$

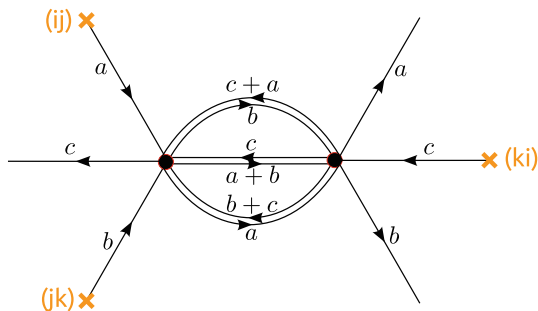


FIGURE 30. The “setting sun”: a picture which could occur as part of a degenerate spectral network at some  $\vartheta = \vartheta_c$ . We show the two-way streets slightly resolved, with the “American resolution”, which would occur in the limit  $\vartheta \rightarrow \vartheta_c^-$ . Each street supports the charges indicated as well as their shifts by positive multiples of  $\gamma = \text{cl}(\bar{a} + \bar{b} + \bar{c})$

(All other streets turn out to be one-way, as anticipated in Fig. 30, and so do not contribute to  $L$  below.) Define  $\gamma = \text{cl}(\bar{a} + \bar{b} + \bar{c})$ , and let  $\tilde{\gamma}$  be its preferred lift. Plugging into (58) and keeping careful track of windings, we find

$$Q(p^1) = Q(p^3) = (1 - X_{\tilde{\gamma}})^{-1}, \quad (97)$$

$$Q(p^2) = 1 - X_{\tilde{\gamma}}. \quad (98)$$

This gives

$$L(\gamma) = -p_{\Sigma}^1 - p_{\Sigma}^3 + p_{\Sigma}^2. \quad (99)$$

This class is homologically trivial,  $[L(\gamma)] = 0$ . (It could hardly be otherwise, since the projection of  $L(\gamma)$  to  $C$  lies in a contractible region containing no branch points, which means  $L(\gamma)$  itself lies in the disjoint union of three contractible open sets on  $\Sigma$ .) Therefore, in particular, this degenerate network does not contribute to the 4d BPS spectrum:

$$\Omega(\gamma) = 0. \quad (100)$$

Nevertheless the enhanced degeneracies  $\omega(\gamma, \cdot)$  are certainly *not* zero in this example; they are given as usual by (63).

How should we understand the result (100)? First note that there is a 1-parameter family of finite BPS webs here, parameterized by an interval, as indicated in Fig. 31. The two ends of the family are built from the  $\mathcal{S}$ -walls appearing in the degenerate network  $\mathcal{W}_{\vartheta_c}$ , while the other finite webs involve a “bubble” made out of BPS strings which are not  $\mathcal{S}$ -walls. This is the same phenomenon we had in Sect. 7.2 above, where we considered a cylinder swept out by BPS strings; the boundaries of the cylinder were  $\mathcal{S}$ -walls, but the generic BPS strings inside the cylinder were not.

In the present case, one end of the family (at left) looks like the “three-string web” we encountered in Sect. 7.3 above. If this three-string web occurred

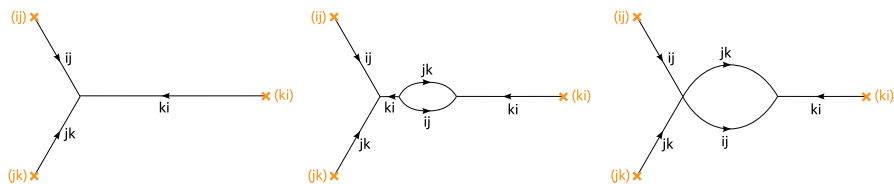


FIGURE 31. Three finite webs, belonging to a 1-parameter family of finite webs parameterized by an interval: as we move along the interval, the “bubble” in the middle of the web expands from zero size (*left*) to a finite maximum size (*right*)

in isolation it would give rise to a BPS hypermultiplet, with  $\Omega(\gamma) = +1$ . However, when it sits in this 1-parameter family its contribution cannot be evaluated in isolation: rather we must quantize the whole family at once. What we have seen is that the contribution to  $\Omega(\gamma)$  from this family vanishes.

## 8. Some BPS Spectra

In this section, we finally show how the spectral networks  $\mathcal{W}_\vartheta$  can be used to determine the 4d BPS spectrum in some simple examples of theories  $S[\mathfrak{g} = A_{K-1}, C, D]$ .

A first comment is that all of the examples described in Sect. 9 of [3] are also examples of the structures considered here: more precisely they are examples of the case  $K = 2$ , where (as explained in Sect. 5.7) studying the spectral networks  $\mathcal{W}_\vartheta$  is equivalent to studying some special ideal triangulations of  $C$ . We therefore regard those examples as incorporated here by reference, and move on to the really new phenomena.

### 8.1. The Pentagon Theory Revisited

The first new example, we consider is obtained by taking  $K = 3$  and  $C = \mathbb{CP}^1$ , with a single defect at  $z = \infty$ , imposing the boundary conditions that  $\phi_2$  has a pole of order 4 and  $\phi_3$  one of order 8. These conditions imply that after rescaling and shifting the coordinate  $z$ , one can put  $\phi_2$  and  $\phi_3$  in the form

$$\phi_2 = 3\Lambda^2 dz^2, \quad \phi_3 = (z^2 + u) dz^3. \quad (101)$$

Here,  $\Lambda$  is a parameter and  $u$  parameterizes the 1-dimensional Coulomb branch  $\mathcal{B}$ . For any particular  $u$ , we have a corresponding 3-fold cover of  $C$  (Seiberg–Witten curve) given by (2),

$$\Sigma = \{\lambda^3 + (3\Lambda^2 dz^2)\lambda + (z^2 + u) dz^3 = 0\} \subset T^*C, \quad (102)$$

or if we write more concretely  $\lambda = x dz$ ,

$$\Sigma = \{x^3 + 3\Lambda^2 x + z^2 + u = 0\}. \quad (103)$$

We can now study the BPS spectrum, by scanning through the spectral networks  $\mathcal{W}_\vartheta$  as  $\vartheta$  varies between 0 and  $\pi$  and looking for critical phases where the topology of  $\mathcal{W}_\vartheta$  jumps.



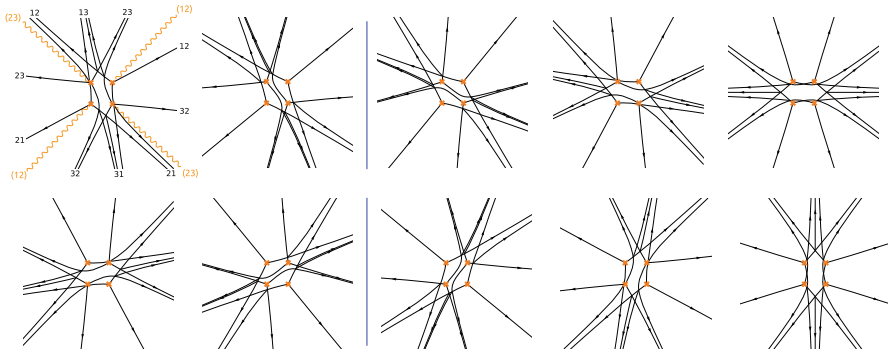


FIGURE 32. The network  $\mathcal{W}_\vartheta$  when  $u = 0$ , at the phases  $\vartheta = \pi n/10$ , with  $1 \leq n \leq 10$ ; the first row begins with  $n = 1$  and the second with  $n = 6$ . We indicate the labelings of the  $\mathcal{S}$ -walls in the  $n = 1$  figure: all walls in each indicated group carry the same label (e.g. there are three walls carrying the label 13 exiting the figure to the north.) The two critical phases at which a  $\mathcal{K}$ -wall occurs are indicated by *blue lines*; one is between  $n = 2$  and  $n = 3$ , the other between  $n = 7$  and  $n = 8$ . See [43] for an animated version of this figure (color figure online)

For example, suppose  $u = 0$ . In this case,  $\mathcal{W}_\vartheta$  jumps 2 times as we vary the phase, as we show in Fig. 32. Both of these jumps are of the type we discussed in Sect. 7.1: at the critical phase there is a saddle connection, i.e. an  $ij$ -trajectory running between two  $(ij)$  branch points. We depict these two saddle connections in Fig. 33. Following the recipe of Sect. 6.3, we see that they correspond to closed loops  $L_1, L_2$  on  $\Sigma$  (just obtained by lifting the saddle connections to  $\Sigma$ ) with corresponding charges  $\gamma_1 = [L_1]$ ,  $\gamma_2 = [L_2]$ , and the 4d BPS degeneracies are

$$\Omega(\gamma_1) = 1, \quad \Omega(\gamma_2) = 1. \quad (104)$$

(We also have  $\Omega(-\gamma_1) = \Omega(-\gamma_2) = 1$ ; these other two BPS multiplets would be encountered in varying  $\vartheta$  between  $\pi$  and  $2\pi$ . The network  $\mathcal{W}_{\vartheta+\pi}$  is obtained from  $\mathcal{W}_\vartheta$  just by transposing the labels  $ij$  on all walls, so we do not need to draw new figures for this range of phases.) The intersection pairing between these charges is  $\langle \gamma_2, \gamma_1 \rangle = 1$ ; this reflects the fact that the two saddle connections cross at a single point, at which they have a single sheet in common (sheet 2 in the notation of Fig. 33.)

On the other hand, suppose we take  $u/\Lambda = 5$ . In this case, the picture looks somewhat different:  $\mathcal{W}_\vartheta$  jumps times times as we vary  $\vartheta$  from 0 to  $\pi$ , as shown in Fig. 34. These three jumps correspond to three finite webs as indicated in Fig. 35.

Again following Sect. 6.3, we find that all three of these lift to closed loops on  $\Sigma$ , now with corresponding charges  $\gamma_2, \gamma_1 + \gamma_2$  and  $\gamma_1$  in order, and we have

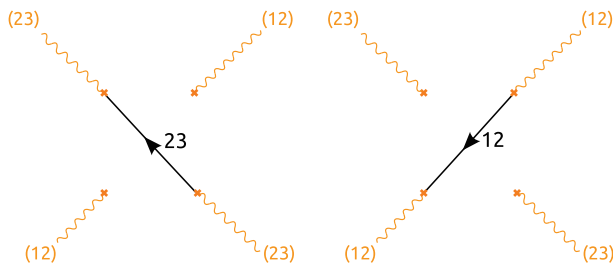


FIGURE 33. The two saddle connections which appear at the critical phases, when  $u = 0$ . They could equally well have been represented by segments with the opposite orientation and the labels  $ij$  transposed

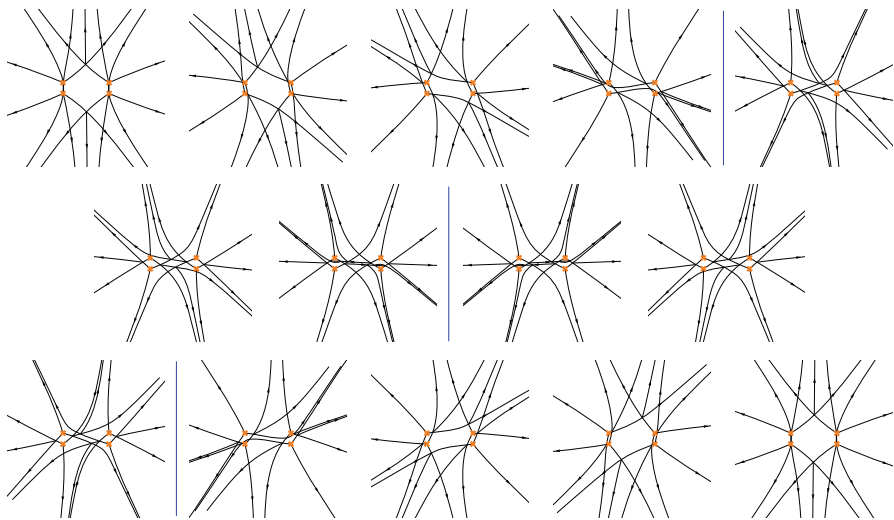


FIGURE 34. The network  $\mathcal{W}_\vartheta$  when  $u = 4$  and  $\Lambda = 1$ , at the phases  $\vartheta = \pi n/200$ , with (reading from left to right and top to bottom)  $n = 1, 23, 45, 68, 82, 90, 97, 103, 110, 118, 132, 155, 177, 199$ . The three critical phases at which a  $\mathcal{K}$ -wall occurs are indicated by *blue lines*. See [43] for an animated version of this figure

$$\Omega(\gamma_1) = 1, \quad \Omega(\gamma_2) = 1, \quad \Omega(\gamma_1 + \gamma_2) = 1 \tag{105}$$

(along with  $\Omega(-\gamma_1) = \Omega(-\gamma_2) = \Omega(-\gamma_1 - \gamma_2) = 1$ , as before).

So the 4d BPS spectrum changes as we vary  $u$ . Of course, this is not unexpected: it is the wall-crossing phenomenon, and occurs exactly as predicted by the wall-crossing formula [9, 3, 17, 40]. At large  $|u|$  the two BPS multiplets of charges  $\gamma_1$  and  $\gamma_2$  form a bound multiplet of charge  $\gamma_1 + \gamma_2$ .

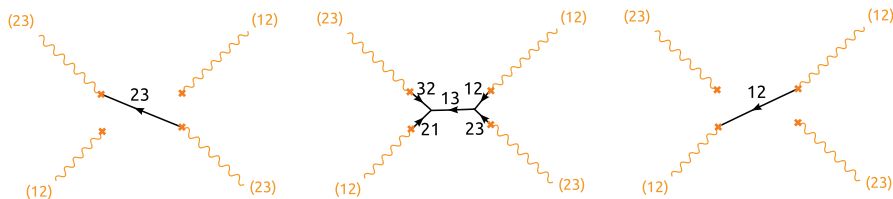


FIGURE 35. The two saddle connections and one more complicated finite web which appear at the critical phases, when  $u = 4$  and  $\Lambda = 1$ . Each segment could equally well be replaced by one with the opposite orientation and the labels  $ij$  transposed

The BPS spectrum here (two hypermultiplets with symplectic product 1 at small  $|u|$ , three hypermultiplets at large  $|u|$ ) might look familiar to the reader: it is just the same structure one meets in the first nontrivial Argyres–Douglas theory. Wall-crossing in Argyres–Douglas theories was first studied in [47]; we also studied this particular theory in Section 9.4.4 of [3]. There we took  $K = 2$  and  $\phi_2 = z^3 - 3\Lambda^2 z + u$ , thus obtaining the Seiberg–Witten curve

$$\Sigma = \{\lambda^2 + (z^3 - 3\Lambda^2 z + u)dz^2 = 0\}, \quad (106)$$

or writing  $\lambda = x dz$ ,

$$\Sigma = \{x^2 + z^3 - 3\Lambda^2 z + u = 0\}. \quad (107)$$

Now comes the point: the change of variables  $x \rightarrow -x$ ,  $z \rightarrow x$  transforms this into (103)! This change of variables does not quite preserve  $\lambda$ , but it takes  $\lambda \rightarrow \lambda - d(xz)$ , so as far as the periods of  $\lambda$  over closed cycles are concerned, these two Seiberg–Witten curves are fully equivalent. This constitutes strong evidence that the two 4d theories  $S[A_1, C, D]$  and  $S[A_2, C, D']$  which we have considered are actually the same. These two descriptions of the theory however privilege different classes of surface defect, one with  $K = 2$  vacua and one with  $K = 3$ . Moreover they lead to rather different-looking representations of the BPS states: in the  $K = 2$  picture, all three BPS hypermultiplets at large  $u$  are represented by saddle connections, quite unlike the situation depicted in Fig. 35. We regard the fact that the BPS degeneracies nevertheless agree as a useful consistency check of our story.

This “duality” between two different descriptions of the same theory is an example of a more general phenomenon. There is a class of theories discussed in [48], labeled by pairs  $(G, G')$  of Dynkin diagrams, and the theory with diagram  $(G, G')$  is the same as the one with diagram  $(G', G)$ . The example we considered here is the case of  $(A_1, A_2)$ .

## 8.2. The Pure $SU(3)$ Theory at Strong Coupling

For our next example, we take again  $K = 3$  and  $C = \mathbb{CP}^1$ , but this time with defects both at  $z = 0$  and  $z = \infty$ , with each defect imposing the boundary conditions that  $\phi_2$  has a pole of order at most 2 and  $\phi_3$  one of order 4. These

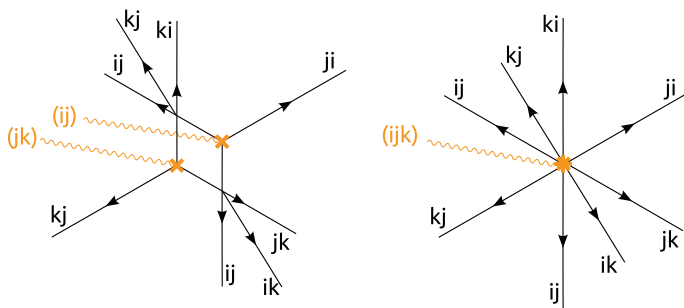


FIGURE 36. *Left* the spectral network  $\mathcal{W}_\vartheta$  in a neighborhood of a pair of nearby branch points of types  $(ij)$  and  $(jk)$ ; *right* the limit where the branch points coalesce to form a branch point of type  $(ijk)$ . This situation occurs in the pure  $SU(3)$  theory as  $u_2 \rightarrow 0$

conditions imply that after rescaling the coordinate  $z$ , one can put  $\phi_2$  and  $\phi_3$  in the form

$$\phi_2 = -3u_2 \left( \frac{dz}{z} \right)^2, \quad \phi_3 = \left( \frac{\Lambda}{z} + u_3 + \Lambda z \right) \left( \frac{dz}{z} \right)^3. \quad (108)$$

This corresponds to the pure  $SU(3)$  theory [3]. Here  $\Lambda$  is a parameter (the dynamical scale) and  $(u_3, u_2)$  parameterize the Coulomb branch. In this paper we will not attempt a complete study of the BPS spectrum in this theory: we just describe what happens at the locus where  $u_3 = 0$  and  $|u_2| \ll |\Lambda|^{\frac{2}{3}}$ . This locus is in the “strongly coupled” region of the theory.

We could proceed immediately to the pictures of  $\mathcal{W}_\vartheta$  as  $\vartheta$  varies, but to calibrate our expectations, it is useful to make some preliminary exploration. At a generic point  $(u_3, u_2)$ , the Seiberg–Witten curve  $\Sigma$  is a threefold covering of  $C$  with 4 simple branch points. If  $u_2 = 0$  then the 4 branch points coalesce in pairs. Each pair consists of an  $(ij)$  and a  $(jk)$  branch point, which at  $u_2 = 0$  coalesce to a single branch point, with cyclic monodromy  $(ijk)$ . There is nothing singular about this situation from the point of view of the IR four-dimensional physics (in contrast with the case where two  $(ij)$  branch points coalesce, in which case the mass of a BPS hypermultiplet goes to zero.) For small  $u_2$  and generic  $\vartheta$ , we can work out (and verify by computer calculation) what the spectral network  $\mathcal{W}_\vartheta$  around such a pair looks like: see Fig. 36. Taking the limit  $u_2 \rightarrow 0$  we find that the branch points with monodromy  $(ijk)$  emit 8 walls. As  $\vartheta$  is increased continuously through an angle  $\pi/3$ , these 8 walls rotate by one unit counterclockwise.

Now fixing  $u_3 = 0$  and  $u_2$  small (or even  $u_2 = 0$ ) we can draw  $\mathcal{W}_\vartheta$  as  $\vartheta$  varies from 0 to  $\pi$ . In fact, the picture has a further approximate symmetry under a shift  $\vartheta \mapsto \vartheta + \frac{\pi}{3}$  (which becomes exact at  $u_2 = 0$ ), so we only need to look at the variation over some range of length  $\frac{\pi}{3}$ . We show this variation in Fig. 37.

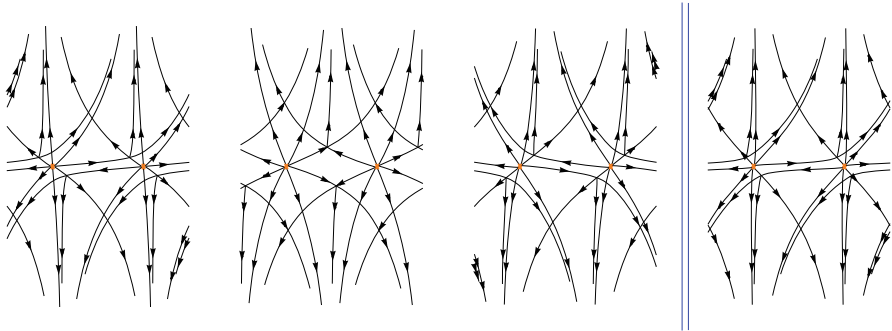


FIGURE 37. The evolution of  $\mathcal{W}_\vartheta$  from  $\vartheta = \frac{\pi}{30}$  to  $\vartheta = \frac{11\pi}{30}$ , at  $u_3 = 0$  and very small  $u_2$ . Each figure is drawn on a flattened-out cylinder, so the left and right sides should be identified. Note that the last figure looks essentially identical to the first. There are two critical phases  $\vartheta_c$  in this range, very close together, marked by the two blue lines. See [43] for an animated version of this figure

We find two critical phases  $\vartheta_c$  very close together. At each critical phase a BPS hypermultiplet appears, represented by a two-way street connecting the two pairs of branch points. (One of these is evident in the figure; the other is harder to spot since it exits the right side and re-appears on the left.) As  $\vartheta$  varies from 0 to  $\pi$  this picture is repeated twice more, giving 2 more BPS states each time. Thus altogether we find that the strong-coupling spectrum of the  $SU(3)$  theory consists of 6 distinct BPS hypermultiplets (or 12 if we include the antiparticles). This agrees with the recent result of [10, 11] where the same spectrum is obtained using quiver representations.

### 8.3. The Theory of 9 Free Hypermultiplets

Next let us consider  $K = 3$  and  $C = \mathbb{CP}^1$ , with three defects, two “full” (at  $z = \pm 1$ ) and one “simple” (at  $z = 0$ ). This means that at  $z = \pm 1$  we impose the condition that  $\phi_2$  has a pole of order 2 and  $\phi_3$  one of order 3, while at  $z = 0$  we require that the discriminant  $\Delta = 27\phi_3^2 - 4\phi_2^3$  has a pole of order 4. Concretely this means we take

$$\phi_2 = \frac{-3m^2 + az + bz^2}{z^2(z+1)^2(z-1)^2}dz^2, \quad \phi_3 = \frac{2m^3 - amz - cz^2 + dz^3}{z^3(z+1)^3(z-1)^3}dz^3. \quad (109)$$

Here  $m, a, b, c, d$  are complex parameters, related to the flavor masses which will appear below.

According to [2], the corresponding theory  $S[A_2, C, D]$  is a theory of 9 free hypermultiplets, transforming in the  $(\mathbf{3}, \mathbf{3}, +1)$  of an  $SU(3) \times SU(3) \times U(1)$  flavor symmetry. The mass parameter for the  $U(1)$  flavor symmetry is  $m$ , while those for the two  $SU(3)$  are the residues  $r_1^{(i)}, r_{-1}^{(i)}$  of  $\lambda^{(i)}$  at 1,  $-1$  respectively (these are some functions of  $m, a, b, c, d$ ). As a test of this statement (and of our whole picture) one can choose some arbitrary values for  $a, b, c, d, m$  and study

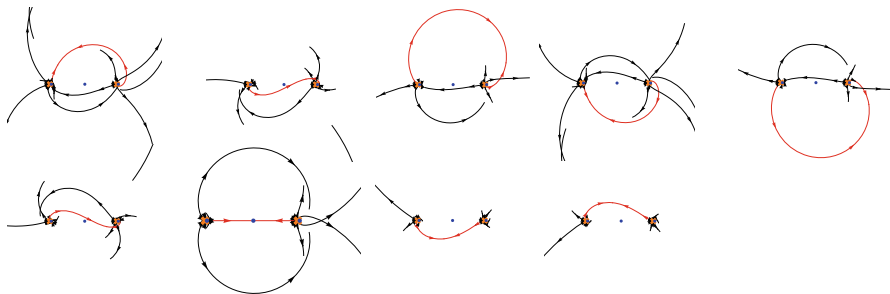


FIGURE 38. The cutoff spectral networks  $\mathcal{W}_\vartheta[\Lambda]$  where  $\Lambda e^{i\vartheta}$  runs over the 9 values  $\frac{1}{2}Z_{\gamma(i,j)}$ , and we have fixed parameters such that  $|m| \gg |r_{\pm 1}^{(i)}|$ . In each network, we see a pair of  $\mathcal{S}$ -walls meeting head-to-head, thus combining into a single path, highlighted in red; this corresponds to a BPS hypermultiplet with central charge  $Z_{\gamma(i,j)}$

the BPS spectrum. We should expect to find 9 BPS multiplets, corresponding to the quanta of the elementary hypermultiplet fields, carrying charges  $\gamma(i,j)$  for  $1 \leq i, j \leq 3$ . The expected central charges of these BPS multiplets are determined by the flavor mass parameters:

$$Z_{\gamma(i,j)} = 2i(-m + r_1^{(i)} + r_{-1}^{(j)}). \quad (110)$$

We found it simplest to study the spectrum in the regime where  $|m| \gg |r_{\pm 1}^{(i)}|$ . In this regime the 4 branch points coalesce into two pairs, each pair sitting very close to one of the two full punctures. For several chosen values of parameters in this regime, we indeed found 9 BPS multiplets with exactly the predicted central charges. See Fig. 38 for an example.

In addition, we found 6 extra BPS multiplets, 3 associated to each of the 2 full punctures, with central charges

$$Z_{\gamma_{\pm 1}(i,j)} = 2i(r_{\pm 1}^{(i)} - r_{\pm 1}^{(j)}) \quad (i > j). \quad (111)$$

(Taking  $i < j$  in this formula gives the central charges for the corresponding antiparticles.) In the regime we are considering, these multiplets are represented by small loops around the punctures. They might at first seem unexpected, but they have a natural explanation: they do not represent 4d particles at all but rather 2d particles living on the surface defect  $\mathbb{S}_z$ . Indeed, as we emphasized in [5], the quantity  $\omega(\gamma, a)$  in general must be interpreted as a sum of contributions from 4d particles carrying charge  $\gamma$  and 2d particles carrying the same charge. The phenomenon that a closed loop around a puncture represents a 2d particle carrying flavor charge also arose there, in the context of the  $\mathbb{CP}^1$  sigma model.

## 9. General Spectral Networks and Path Lifting

As we have remarked in Sect. 5.7, in the theories  $S[A_1, C, D]$ , each spectral network  $\mathcal{W}_\vartheta$  corresponds naturally to an ideal triangulation of  $C$ . Ideal triangulations are rather flexible objects and one might study them without regard for whether they arise from any  $\mathcal{W}_\vartheta$ . In an analogous way, we now generalize the  $\mathcal{W}_\vartheta$  we have worked with thus far to some purely topological objects  $\mathcal{W}$ , which we will refer to as *spectral networks*.

A spectral network  $\mathcal{W}$  is associated to a branched cover  $\Sigma \rightarrow C$ . In contrast to the case of  $\mathcal{W} = \mathcal{W}_\vartheta$  constructed in Sect. 5, now we do not require that  $\Sigma \subset T^*C$ . In particular, we do not have the canonical 1-form  $\lambda$  on  $\Sigma$  anymore. In consequence, the  $\mathcal{S}$ -walls making up a spectral network are not solutions to any differential equation; rather, locally they are arbitrary paths. We thus gain some flexibility in the  $\mathcal{S}$ -walls, but at the same time we lose some topological data which were previously induced by the 1-form  $\lambda$  (Seiberg–Witten differential). We will build substitutes for those data explicitly into our definition of “spectral network.”

### 9.1. General Spectral Networks

Let  $C$  be an oriented real surface with a (perhaps empty) boundary. Each connected component of the boundary is a copy of  $S^1$ ; on each such component fix a (nonempty) set of marked points. Also fix a (perhaps empty) set of marked points in the interior of  $C$ . We refer to the interior marked points as *punctures* and we refer to all the marked points as *singular points* (although they are not singularities of  $C$ ) and denote them as  $\mathfrak{s}_n$ . We require that there is at least one singular point.

Let  $\Sigma \rightarrow C$  be a  $K$ -fold branched covering which is unramified over the boundaries and the singular points. Let  $C'$  be  $C$  minus the branch points. For simplicity we assume the branch points are all simple, so the monodromy around each branch point just exchanges two sheets of  $\Sigma$ . (This condition can likely be relaxed at the price of a more cumbersome definition.)

A *spectral network subordinate to the covering  $\Sigma$*  is a collection

$$\mathcal{W} = (o(\mathfrak{s}_n), \{z_\mu\}, \{p_c\}) \quad (112)$$

where the symbols refer to the following data.

- D1.** For each singular point  $\mathfrak{s}_n$ ,  $o(\mathfrak{s}_n)$  is a *partially ordered subset* of the set of sheets of  $\Sigma$  over a neighborhood of  $\mathfrak{s}_n$ .  $o(\mathfrak{s}_n)$  must contain at least two elements, and if  $\mathfrak{s}_n$  is a puncture,  $o(\mathfrak{s}_n)$  must contain *all* of the sheets over a neighborhood of  $\mathfrak{s}_n$ .
- D2.**  $\{z_\mu\}$  is a locally finite collection of points on  $C'$ , called *joints*.
- D3.**  $\{p_c\}$  is a finite or countable collection of closed segments (i.e. images of *embeddings* of  $[0, 1]$  into  $C$ ), called *walls* or *streets* (depending which metaphor is more useful in a given context, cf. footnote 21). For each orientation  $o$  of the street  $p_c$ ,  $p_c$  is labeled with an ordered pair of distinct sheets of the covering  $\Sigma \rightarrow C$  over  $p_c$ . Reversing the orientation reverses this ordered pair of sheets. So  $p_c$  comes with two labels which we could write as  $(o, ij)$  and  $(-o, ji)$ .

The data must satisfy the following conditions.

- C1.** The segments  $p_c$  cannot cross one another (but they are allowed to have common tangents). Each  $p_c$  must begin on a branch point or a joint, and must end on a joint or a singular point. Any compact subset of  $C'$  intersects only finitely many segments.
- C2.** Around each branch point  $\mathfrak{b}$  there is a neighborhood where  $\mathcal{W}$  looks like Fig. 13. That is, each branch point of type  $(ij)$  is an endpoint of three streets which carry labels  $(o, ij)$  or  $(o, ji)$ , and the streets encountered consecutively traveling around a loop around  $\mathfrak{b}$  have oppositely ordered sheets.
- C3.** Around each joint  $z_\mu$  there is a neighborhood where  $\mathcal{W}$  looks like Fig. 9, Fig. 10, or Fig. 11.
- C4.** If a segment with label  $ij$  ends at a singular point  $\mathfrak{s}_n$ , then  $i$  and  $j$  lie in the ordered subset  $o(\mathfrak{s}_n)$ , and with respect to the ordering of  $o(\mathfrak{s}_n)$  we have  $i < j$ .

Our definition of “spectral network” is somewhat provisional. With an eye toward the future let us mention one natural generalization: we could have relaxed the requirement that there is at least one singular point  $\mathfrak{s}_n$ , and allowed the segments to be infinite in one direction. This generalization would be needed if we want our definition to encompass the networks  $\mathcal{W}_\vartheta$  in theories  $S[\mathfrak{g}, C]$  where the set of defects is actually *empty*. The resulting spectral networks would be expected to look much more complicated—e.g. the streets may well be dense on  $C$ . Nevertheless, as we described in the introduction, we think that it should be possible to extend everything described in this paper to this case.

## 9.2. Canonical Examples: the $\mathcal{W}_\vartheta$

The  $\mathcal{W}_\vartheta$  discussed in previous sections, which arose naturally from the physics of theories of class  $S$ , are essentially examples of spectral networks.

The data D2 and D3 above appeared in Sect. 5. The datum D1 is determined by the behavior of the  $\lambda^{(i)}$  near  $\mathfrak{s}_n$ . We will not describe it explicitly here, except in the basic case where  $\mathfrak{s}_n$  is a full defect: in that case  $o(\mathfrak{s}_n)$  consists of all the sheets, ordered by  $\text{Re}(e^{-i\vartheta} m^{(i)})$ . As we saw in Sect. 5.5, the  $\mathcal{W}_\vartheta$  indeed obey our condition C4 with this ordering.

We say  $\mathcal{W}_\vartheta$  are “essentially” spectral networks because of two technical points:

1. In the networks  $\mathcal{W}_\vartheta$  as we defined them, it is possible for a wall to “die” at a joint, as indicated in Fig. 12. In contrast, our present definition of spectral network we do not allow this: the walls always continue through joints.

This difference arises because in  $\mathcal{W}_\vartheta$  we included a wall  $p$  only if it supports a charge  $a$  with  $\mu(a, p) \neq 0$ . We could have dropped the requirement  $\mu(a, p) \neq 0$ , thus including some additional “invisible” walls. This would not have changed anything in previous sections, except to make the notation a bit more cumbersome.





FIGURE 39. The American (*left*) and British (*right*) resolutions of a street

2.  $\mathcal{W}_\partial$  can fail to be a spectral network because it has a wall which is not of type  $ij$  or  $ji$  but accidentally runs into a branch point of type  $(ij)$ . It would be possible to extend the definition of spectral network to include this situation, but for simplicity we have avoided it.

### 9.3. Soliton Content

To a spectral network, we can associate some additional data having to do with *solitons*, which we now define.

A *soliton*  $s(z)$ , where  $z \in p_c$ , is an immersion of  $[0, 1]$  into  $\Sigma$ , which begins and ends on preimages of  $z$ , and such that its projection to  $C$  lies in the spectral network  $\mathcal{W}$  (or more precisely in a very small neighborhood of  $\mathcal{W}$ ; this correction is necessary because at joints we smooth out the sharp corners.) If  $p_c$  carries the label  $(o, ij)$ , then we say  $s(z)$  is *compatible with*  $o$  if it begins on  $z^{(i)}$  and ends on  $z^{(j)}$ , and the projection of  $s(z)$  to  $C$  begins with orientation  $-o$  and ends with orientation  $o$ . The soliton content is, for each street  $p_c$  and each point  $z \in p_c$ , a pair of sets of solitons  $\mathcal{S}_c^o(z)$ ,  $\mathcal{S}_c^{-o}(z)$ , such that the solitons in  $\mathcal{S}_c^o(z)$  are compatible with  $o$  and those in  $\mathcal{S}_c^{-o}(z)$  are compatible with  $-o$ .

The soliton sets must satisfy some rules, which we will refer to as *solitonic traffic rules*. Actually the rules come in two variants: either *American* or *British*. In order to state the rules we introduce a resolution of the streets of the spectral network, regarding each street as resolved into two oriented “lanes” infinitesimally displaced from one another, either in the “American” or the “British” fashion. See Fig. 39.

Note that this definition uses the orientation of  $C$ . The American and British traffic rules are thus related to one another by a reflection in the plane. In what follows we show the British rules only; to get the American rules simply requires a little reflection.

- ST1.** As  $z$  moves continuously along a street  $p_c$ , the soliton sets  $\mathcal{S}_c^{\pm o}(z)$  evolve continuously, by the natural parallel transport; in other words, the soliton sets do not “jump.” With this in mind, we abuse notation by writing the soliton sets simply as  $\mathcal{S}_c^{\pm o}$ , suppressing the trivial  $z$  dependence.
- ST2.** Let  $\mathfrak{b}$  be a branch point. The network  $\mathcal{W}$  then looks like Fig. 40 in the vicinity of  $\mathfrak{b}$ . There are three streets  $p_c$  emerging from  $\mathfrak{b}$ . We denote the set of orientations of each street by  $\{\text{in}, \text{out}\}$ . For each  $c$ , define the *light soliton*  $s_c$  as follows: if  $p_c$  carries the label  $(\text{out}, ij)$ , then  $s_c(z)$  begins on  $z^{(i)}$ , travels along  $p_c^{(i)}$  back to the ramification point on  $\Sigma$  covering  $\mathfrak{b}$ , and returns along  $p_c^{(j)}$  to  $z^{(j)}$ . Then, the soliton sets are related by

$$\mathcal{S}_1^{\text{out}} = \mathcal{S}_3^{\text{in}} \cup \{s_1\}, \quad \mathcal{S}_2^{\text{out}} = \mathcal{S}_1^{\text{in}} \cup \{s_2\}, \quad \mathcal{S}_3^{\text{out}} = \mathcal{S}_2^{\text{in}} \cup \{s_3\}. \quad (113)$$

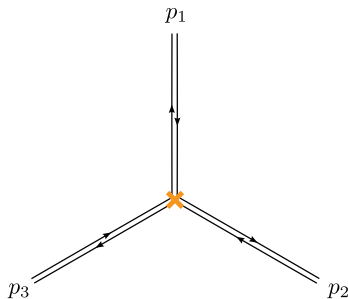


FIGURE 40. A spectral network in the vicinity of a branch point

(In this equation, the solitons are understood to be evolved continuously from one street to the other. Since there is a branch point at  $\mathfrak{b}$ , this continuous evolution depends on which path we follow to go from one street to the next; we follow the *short* path, i.e. we go around an arc of length  $2\pi/3$ , not  $4\pi/3$ .)

**ST3.** In a sufficiently small neighborhood of a joint of the “four-way junction” type shown in Fig. 9 (where  $ij$  and  $kl$  walls meet), the soliton sets vary continuously, i.e. the outgoing soliton sets are equal to the incoming ones. In a sufficiently small neighborhood of a joint of the “six-way junction” type (where  $ij$ ,  $jk$  and  $ki$  walls meet), the outgoing soliton sets are determined by the incoming ones according to the rules shown in Figs. 41, 42. (A motivation for this peculiar-looking rule is explained in Appendix A.)

Importantly, for any spectral network  $\mathcal{W}$ , the solitonic traffic rules (including the choice of American or British) uniquely determine the soliton content. To prove this claim, one can consider a discrete version of the mass filtration of the spectral networks  $\mathcal{W}_\vartheta$ : just define the *length* of a soliton to be the number of walls in its projection to  $C$ , and construct the soliton content by induction on length using the solitonic traffic rules.

It is convenient to distinguish two different possibilities for the soliton content. Consider the two soliton sets  $\mathcal{S}_c^o$ ,  $\mathcal{S}_c^{-o}$  on a street  $p_c$ . If both are nonempty, we refer to  $p_c$  as a *two-way street*. If exactly one is empty,  $p_c$  is a *one-way street*, and in that case we write  $\mathcal{S}_c$  for the single nonempty soliton set. (It never happens that both are empty.) If there are no two-way streets we say that  $\mathcal{W}$  is a *nondegenerate* spectral network; otherwise we say it is *degenerate*.

If  $\mathcal{W}$  is nondegenerate, the solitonic traffic rules simplify considerably, to the following. Consider the joint of Fig. 10. We continue the  $ij$  and  $jk$  soliton sets continuously through the joint. For the  $ik$  wall, after passing through the joint we add to the set of  $ik$  solitons all new solitons of the following description: the new soliton begins at  $z^{(i)}$ , projects on  $C$  to the path going back to the joint, then follows one of the  $ij$  solitons from sheet  $i$  to sheet  $j$  above the joint, then follows one of the  $jk$  solitons from sheet  $j$  to sheet  $k$  above the joint,

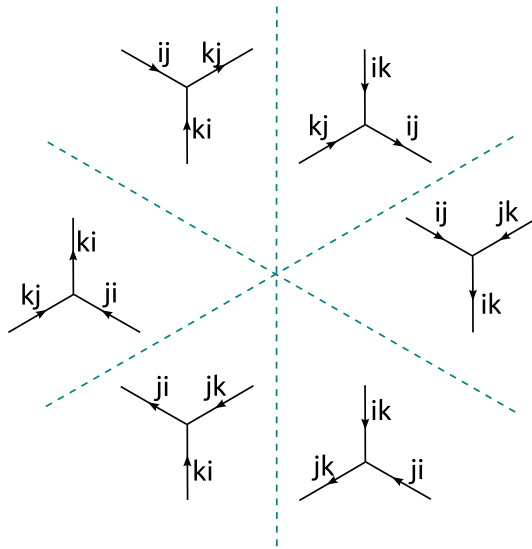


FIGURE 41. The rule for constructing outgoing solitons from a six-way junction, with the British resolution. We show six allowed types of junction, one in each “Weyl chamber” around the joint. The most general outgoing soliton is constructed by a diagrammatic prescription as follows. We fix any set of incoming solitons; for each one we draw an incoming line on the diagram, coming in along the appropriate direction, and on the left side of the median. We then combine these incoming solitons with junctions of the allowed types, to make a single outgoing line. Every diagram so obtained determines an outgoing soliton in a natural way, by concatenation of the incoming solitons

then returns on sheet  $k$  on the trajectory projecting to the  $ik$  wall ending at  $z^{(k)}$ . In particular, these simplified rules do not depend on whether we started out with American or British traffic rules (as we should expect since we are considering the case where there are no two-way streets.)

So for nondegenerate spectral networks, the soliton content is unique, while for degenerate ones there are two possible soliton contents, one following American traffic rules and one British.

In previous sections, we considered integers  $\mu(p, a)$  attached to the spectral networks  $\mathcal{W}_\vartheta$ . These integers are determined by the soliton content of  $\mathcal{W}_\vartheta$ —indeed they are simply counts of the solitons with appropriate signs:  $\mu(p_c, a)X_a = \sum_{\nu \in \mathcal{S}_c} X_{[\tilde{s}_\nu]}$ . So the soliton content is a slight extension of the  $\mu(p, a)$  to keep track of the actual solitons, not only their number. This extension will actually not be used for anything in this paper (all of our constructions really depend only on  $\mu$ ) but we believe it may be useful in the future.

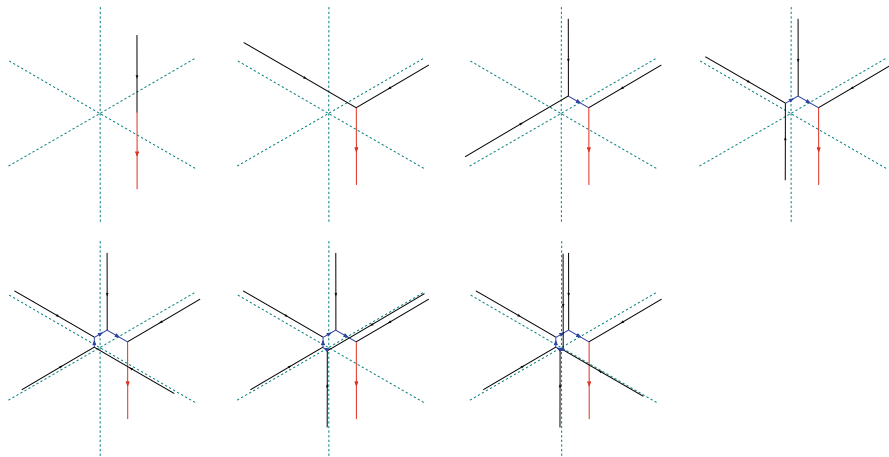


FIGURE 42. Some examples of diagrams representing outgoing solitons. Incoming solitons are represented by *black lines*, the outgoing path by a *red line*. (The first example is exceptional in that the incoming line is the same as the outgoing one.) (color figure online)

#### 9.4. Path Lifting

In Sects. 4–5, we studied at great length the generating functions  $F(\wp, \vartheta)$  of framed 2d–4d BPS degeneracies. Let us set aside the physical meaning of these functions for a moment and just think of them as some interesting mathematical objects. We found in Sect. 4.4 that  $F(\wp, \vartheta)$  could be completely constructed from the datum of the spectral network  $\mathcal{W}_\vartheta$  together with its soliton content. One of the main motivations of our definition of spectral network is that, given *any* spectral network, we can make a very similar construction. In this section, we describe that construction. For notational convenience, we consider only the case of a nondegenerate spectral network, but what we write has an obvious extension to the degenerate case.

We will be rather brief since everything is parallel to what we did in Sect. 4.4. However, we slightly modify Sect. 4.4, in two respects. First, with an eye toward future applications, instead of *homology* classes of open paths on  $\tilde{\Sigma}$ , we will keep track of *homotopy* classes. This turns out to be no more difficult. We will generally denote the homotopy objects with a bold letter to distinguish them from their homology cousins. Second, instead of considering smooth paths  $\wp$  in  $C$  and using their canonical lift to  $\tilde{C}$ , it will be convenient to work with arbitrary paths  $\tilde{\wp}$  on  $\tilde{C}$  from the beginning. This will involve lifting many objects from  $C$  to  $\tilde{C}$  or from  $\Sigma$  to  $\tilde{\Sigma}$ ; we always use a tilde to denote the lifted objects.

For any homotopy class  $\mathbf{a}$  of open paths on  $\tilde{\Sigma}$ , we introduce a corresponding formal variable  $\mathbf{X}_{\mathbf{a}}$ , subject to the relation that if  $\mathbf{a}$  and  $\mathbf{a}'$  project to the same homotopy class on  $\Sigma$  then we have

$$\mathbf{X}_{\mathbf{a}}/\mathbf{X}_{\mathbf{a}'} = (-1)^{w(\mathbf{a}, \mathbf{a}')}. \quad (114)$$

Now, given a spectral network  $\mathcal{W}$ , let  $\tilde{\mathcal{W}}$  be the preimage of  $\mathcal{W}$  on  $\tilde{C}$ . Given a path  $\tilde{\varphi}$  on  $\tilde{C}$  whose endpoints are not on  $\tilde{\mathcal{W}}$ , we are going to define a formal sum of these variables, with integer coefficients:

$$\mathbf{F}(\tilde{\varphi}, \mathcal{W}) = \sum_{\mathbf{a}} \overline{\Omega}'(\tilde{\varphi}, \mathcal{W}, \mathbf{a}) \mathbf{X}_{\mathbf{a}}. \quad (115)$$

(In Sect. 4.1, we met a similar expansion for which the coefficients  $\overline{\Omega}'$  were interpreted as framed 2d–4d BPS degeneracies. In the more general setting of this section, we do not know what the physical interpretation of  $\overline{\Omega}'$  should be.)

The assignment  $\mathbf{F}(\cdot, \mathcal{W})$  will obey two important properties:

- For two concatenatable paths  $\tilde{\varphi}, \tilde{\varphi}'$  on  $\tilde{C}$ ,

$$\mathbf{F}(\tilde{\varphi}\tilde{\varphi}', \mathcal{W}) = \mathbf{F}(\tilde{\varphi}, \mathcal{W})\mathbf{F}(\tilde{\varphi}', \mathcal{W}). \quad (116)$$

- If  $\tilde{\varphi}$  and  $\tilde{\varphi}'$  are two paths on  $\tilde{C}$  which project to the same homotopy class on  $C$ ,

$$\mathbf{F}(\tilde{\varphi}, \mathcal{W}) = (-1)^{w(\tilde{\varphi}, \tilde{\varphi}')} \mathbf{F}(\tilde{\varphi}', \mathcal{W}). \quad (117)$$

(In particular, if  $\tilde{\varphi}$  and  $\tilde{\varphi}'$  are homotopic, then  $\mathbf{F}(\tilde{\varphi}, \mathcal{W}) = \mathbf{F}(\tilde{\varphi}', \mathcal{W})$ .)

Given any path  $\tilde{\varphi}$  on  $\tilde{C}'$  we first define

$$\mathbf{D}(\tilde{\varphi}) = \sum_{i=1}^K \mathbf{X}_{\tilde{\varphi}^{(i)}}. \quad (118)$$

where  $\tilde{\varphi}^{(i)}$  is the open path given by lifting the initial point of  $\tilde{\varphi}$  to the  $i$ th sheet and then using the canonical connection on  $\tilde{\Sigma} \rightarrow \tilde{C}$  to lift the path. For any  $\tilde{\varphi}$  which does not cross  $\tilde{\mathcal{W}}$  we have simply

$$\mathbf{F}(\tilde{\varphi}, \tilde{\mathcal{W}}) = \mathbf{D}(\tilde{\varphi}). \quad (119)$$

For any  $\tilde{\varphi}$  which crosses  $\tilde{\mathcal{W}}$  exactly once at a wall  $p_c$ , we have

$$\mathbf{F}(\tilde{\varphi}, \tilde{\mathcal{W}}) = \mathbf{D}(\tilde{\varphi}_+) \left( 1 + \sum_{\nu \in \mathcal{S}_c} \mathbf{X}_{\mathbf{a}(\nu, \tilde{\varphi})} \right) \mathbf{D}(\tilde{\varphi}_-), \quad (120)$$

where  $\mathbf{a}(\nu, \tilde{\varphi})$  is a particular lift of the soliton  $s_\nu$  to  $\tilde{\Sigma}$ , described in the next paragraph.

Let  $\varphi$  be the projection of  $\tilde{\varphi}$  to  $C$ . Let  $z$  be the point of intersection between  $\varphi$  and  $\mathcal{W}$ . Let  $\tilde{z}$  be the point of  $\tilde{\varphi}$  lying over  $z$ , dividing  $\tilde{\varphi}$  into  $\tilde{\varphi}_+ \tilde{\varphi}_-$ . Let  $o_+$  be a tangent vector at  $z$  oriented along  $\mathcal{W}$ , and  $o_-$  a tangent vector at  $z$  oriented oppositely to  $\mathcal{W}$ . Let  $A_+$  be some path from  $\tilde{z}$  to  $o_+$  in the circle fiber  $\tilde{C}_z$ , and  $A_-$  a path from  $o_-$  to  $\tilde{z}$  in  $\tilde{C}_z$ . The concatenation  $A = A_- + A_+$  is thus a path from  $o_-$  to its antipode  $o_+$  in  $\tilde{C}_z$ ; we fix our choices of  $A_\pm$  so that  $A$  is homotopic to a simple arc which covers half of  $\tilde{C}_z$ , and this arc includes the tangent vector  $\varphi'(z)$ . Each soliton  $s_\nu(z)$  has a canonical lift to a

smooth path  $\tilde{s}_\nu(z)$  on  $\tilde{\Sigma}$  from  $o_+^{(i)}$  to  $o_-^{(j)}$ . The path  $\mathbf{a}(\nu, \tilde{\varphi})$  on  $\tilde{\Sigma}$  from  $\tilde{z}^{(i)}$  to  $\tilde{z}^{(j)}$  is defined by

$$\mathbf{a}(\nu, \tilde{\varphi}) = A_+^{(i)} + \tilde{s}_\nu(z) + A_-^{(j)}. \quad (121)$$

Finally, to define  $\mathbf{F}(\tilde{\varphi}, \mathcal{W})$  for an arbitrary path  $\tilde{\varphi}$  on  $\tilde{C}$  whose endpoints are not on  $\tilde{\mathcal{W}}$ , we perturb  $\tilde{\varphi}$  slightly so that its intersections with  $\tilde{\mathcal{W}}$  are all transverse, break it into pieces which meet  $\tilde{\mathcal{W}}$  at most once, and use (119), (120) and the composition property (116).

Our construction makes the composition property (116) manifest. The same arguments given in Sect. 5.6 apply to our current construction and show the twisted homotopy invariance (117).

The path-lifting rule  $\mathbf{F}$  will be put to some mathematical use in Sect. 10 below. As mentioned in Sect. 1, we are hopeful that it will also have some interesting physical applications, as well as connections to upcoming work of Goncharov and Kontsevich.

## 10. Coordinates for Moduli of Flat Connections

In this last section, we discuss a mathematical application of spectral networks. Given a spectral network  $\mathcal{W}$  subordinate to a  $K$ -fold covering  $\Sigma \rightarrow C$ , we will construct a map

$$\Psi_{\mathcal{W}} : \mathcal{M}(\Sigma, \mathrm{GL}(1); \mathfrak{m}) \rightarrow \mathcal{M}_F(C, \mathrm{GL}(K); \mathfrak{m}) \quad (122)$$

where  $\mathcal{M}(\Sigma, \mathrm{GL}(1); \mathfrak{m})$  is a moduli space of twisted flat  $\mathrm{GL}(1)$ -connections  $\nabla^{\mathrm{ab}}$  on  $\Sigma$  and  $\mathcal{M}_F(C, \mathrm{GL}(K); \mathfrak{m})$  is a moduli space of twisted flat  $\mathrm{GL}(K)$ -connections  $\nabla$  on  $C$ , decorated by some “flag data” as we explain below. This map is a local symplectomorphism, and conjecturally 1–1 onto its image.

Roughly speaking,  $\nabla = \Psi_{\mathcal{W}}(\nabla^{\mathrm{ab}})$  is obtained by a two-step process. We first push forward the  $\mathrm{GL}(1)$  connection  $\nabla^{\mathrm{ab}}$  from  $\Sigma$  to  $C$ . This gives a flat  $\mathrm{GL}(K)$  connection on the complement of the branch locus in  $C$ , which is everywhere diagonal. This connection however cannot be extended over the branch points, because it has monodromy around them. To deal with this problem, we cut  $C$  into pieces along the network  $\mathcal{W}$ , and then reglue the connection with a nontrivial (and non-diagonal) transition function, controlled by the soliton content of  $\mathcal{W}$ . This process eliminates the monodromy around branch points, while not introducing unwanted monodromy anywhere else. We call this construction “nonabelianization.”

The space  $\mathcal{M}(\Sigma, \mathrm{GL}(1); \mathfrak{m})$  is a torsor for  $(\mathbb{C}^\times)^{2g_\Sigma}$ , where  $\bar{\Sigma}$  is the closure of  $\Sigma$ , so another way to read  $\Psi_{\mathcal{W}}$  is as a *local coordinate system* on  $\mathcal{M}_F(C, \mathrm{GL}(K); \mathfrak{m})$ . These coordinate systems are closely related to the “cluster coordinates” introduced by Fock and Goncharov [13]. In the case  $K = 2$ , it is straightforward to see that our coordinates in fact coincide with the cluster coordinates. In the case  $K > 2$ , a fully explicit description of the most general cluster coordinate system is not available in the literature; we conjecture that the  $\Psi_{\mathcal{W}}$  actually give the most general cluster coordinates, thus filling this gap.

One point in favor of this conjecture is that when  $\mathcal{W}$  is of a special form,  $\Psi_{\mathcal{W}}$  yields a coordinate system which was explicitly described by Fock and Goncharov; this will be described in [32].

Although we are emphasizing its mathematical content in this section, this construction also has a natural physical meaning, which appears most clearly when the theory is compactified from four to three dimensions on  $S^1$ . Indeed, the expansion (127) of the parallel transport of  $\nabla$  as a linear combination of parallel transports  $\mathbf{X}_a(\nabla^{\text{ab}})$  is precisely the “Darboux expansion” of the vacuum expectation value of the supersymmetric interface associated with  $\tilde{\phi}$ . This point of view was described at length in [5].

### 10.1. Nonabelianization

We now explain how, given a nondegenerate spectral network  $\mathcal{W}$  subordinate to a covering  $\Sigma \rightarrow C$ , and given also a twisted flat rank 1 connection over  $\Sigma$ , we construct a corresponding twisted flat rank  $K$  connection over  $C$ .

By a *twisted flat rank  $K$  connection* over a real surface  $S$  we mean a pair of a complex rank  $K$  vector bundle on  $\tilde{S}$  and a flat connection therein, such that the holonomy around each fiber of  $\tilde{S}$  is  $-1$ . This is a very small modification of the usual notion of a rank  $K$  connection over  $S$ . In particular, given a spin structure on  $S$ , there is a canonical isomorphism between the moduli space of twisted flat connections and that of ordinary connections.<sup>29</sup> Nevertheless, this slight twisting is important in the construction we are about to describe.

Now suppose given a rank 1 twisted flat connection  $(\mathcal{L}, \nabla^{\text{ab}})$  over  $\Sigma$ . Recall that  $C'$  denotes  $C$  minus the branch points of the covering  $\Sigma \rightarrow C$ . Let  $\pi : \tilde{\Sigma} \rightarrow \tilde{C}'$  be the projection. Then define the complex rank  $K$  vector bundle

$$E = \pi_*(\mathcal{L}). \quad (123)$$

The fiber of  $E$  at any  $\tilde{z} \in \tilde{C}'$  is simply

$$E_{\tilde{z}} = \bigoplus_i \mathcal{L}_{\tilde{z}(i)}. \quad (124)$$

Given the flat rank 1 connection  $\nabla^{\text{ab}}$  on  $\mathcal{L}$ , there is a canonical “push-forward” flat connection  $\pi_*(\nabla^{\text{ab}})$  on  $E$ . But the reader should beware that  $\pi_*(\nabla^{\text{ab}})$  is *not* the connection we are trying to construct. Indeed, while  $\pi_*(\nabla^{\text{ab}})$  is flat on  $\tilde{C}'$ , it has nontrivial holonomy around any small loop  $\ell_{\mathfrak{b}}$  in  $\tilde{C}'$  linking the fiber over a branch point  $\mathfrak{b}$ ; this monodromy is induced by the permutation of sheets attached to  $\mathfrak{b}$ . It follows that  $\pi_*(\nabla^{\text{ab}})$  cannot be extended to a flat connection over  $\tilde{C}$ , since on  $\tilde{C}$  the loop  $\ell_{\mathfrak{b}}$  would be contractible. Our construction will modify  $(E, \pi_*(\nabla^{\text{ab}}))$  in a way which eliminates this holonomy.

The key is the path-lifting rule  $\mathbf{F}$  defined in Sect. 9.4. Let  $C'_{\mathcal{W}} := C' \setminus \mathcal{W}$ . Given any path  $\tilde{\phi}$  from  $\tilde{z}_1 \in \tilde{C}'_{\mathcal{W}}$  to  $\tilde{z}_2 \in \tilde{C}'_{\mathcal{W}}$  we have built a formal sum of

<sup>29</sup> To obtain this isomorphism we use the fact that a spin structure is the same thing as a fiberwise double covering of  $\tilde{S}$ ; then given a twisted flat connection, pulling back to this double cover gives a new flat connection with holonomy  $+1$  around the fibers, which then descends to a flat connection over  $S$ .

paths on  $\tilde{\Sigma}$ , given by (115):

$$\mathbf{F}(\tilde{\varphi}, \mathcal{W}) = \sum_{\mathbf{a}} \overline{\Omega}'(\tilde{\varphi}, \mathcal{W}, \mathbf{a}) \mathbf{X}_{\mathbf{a}}. \quad (125)$$

Each  $\mathbf{a}$  in the sum (125) is a homotopy class of paths on  $\tilde{\Sigma}$  from  $\tilde{z}_1^{(i)}$  to  $\tilde{z}_2^{(j)}$  for some  $i, j$ . Let

$$\mathbf{X}_{\mathbf{a}}(\nabla^{\text{ab}}) \in \text{Hom}(\mathcal{L}_{\tilde{z}_1^{(i)}}, \mathcal{L}_{\tilde{z}_2^{(j)}}) \quad (126)$$

denote the parallel transport of  $\nabla^{\text{ab}}$  along any path in the class  $\mathbf{a}$ . The fact that  $\nabla^{\text{ab}}$  is a flat connection implies that  $\mathbf{X}_{\mathbf{a}}(\nabla^{\text{ab}})$  depends only on the class  $\mathbf{a}$  and not on the particular path, and the fact that  $\nabla^{\text{ab}}$  has holonomy  $-1$  around the fibers of  $\tilde{\Sigma}$  implies that  $\mathbf{X}_{\mathbf{a}}(\nabla^{\text{ab}})$  obeys the relation (114). Hence the map  $\mathbf{X}_{\mathbf{a}} \rightarrow \mathbf{X}_{\mathbf{a}}(\nabla^{\text{ab}})$  is well defined.

We now define a new operator  $\mathbf{F}(\tilde{\varphi}, \mathcal{W}, \nabla^{\text{ab}})$  just by replacing  $\mathbf{X}_{\mathbf{a}} \rightarrow \mathbf{X}_{\mathbf{a}}(\nabla^{\text{ab}})$  in (125):

$$\mathbf{F}(\tilde{\varphi}, \mathcal{W}, \nabla^{\text{ab}}) := \sum_{\mathbf{a}} \overline{\Omega}'(\tilde{\varphi}, \mathcal{W}, \mathbf{a}) \mathbf{X}_{\mathbf{a}}(\nabla^{\text{ab}}) \in \text{Hom}(E_{\tilde{z}_1}, E_{\tilde{z}_2}). \quad (127)$$

$\mathbf{F}(\tilde{\varphi}, \mathcal{W}, \nabla^{\text{ab}})$  has precisely the formal properties one would expect if it were the parallel transport operator in some flat vector bundle: it is homotopy invariant and behaves properly under composition. The key to our construction is that indeed  $\mathbf{F}(\tilde{\varphi}, \mathcal{W}, \nabla^{\text{ab}})$  is the parallel transport operator from  $\tilde{z}_1$  to  $\tilde{z}_2$  for a twisted rank  $K$  flat connection  $(\hat{E}, \nabla)$  over  $C$ .  $(\hat{E}, \nabla)$  will be constructed in an essentially tautological way to make this sentence true.

The most direct way of building  $(\hat{E}, \nabla)$  is to construct its sheaf  $\mathbb{F}$  of flat sections, as follows. For any open set  $\mathcal{U} \subset \tilde{C}$ ,  $\mathbb{F}(\mathcal{U})$  consists of all sections  $\psi$  of  $E$  over  $\mathcal{U} \cap \tilde{C}'_{\mathcal{W}}$  such that for any path  $\tilde{\varphi} \subset \mathcal{U}$  with endpoints  $\tilde{z}_1, \tilde{z}_2 \in \tilde{C}'_{\mathcal{W}}$ , we have

$$\psi(\tilde{z}_1) \mathbf{F}(\tilde{\varphi}, \mathcal{W}, \nabla^{\text{ab}}) = \psi(\tilde{z}_2). \quad (128)$$

(Note that the parallel transport acts from the right, in accordance with Appendix C.) Any  $\psi \in \mathbb{F}(\mathcal{U})$  is obviously determined by its value at a single  $\tilde{z}_0 \in \mathcal{U} \cap \tilde{C}'_{\mathcal{W}}$ . Conversely, if  $\mathcal{U}$  is contractible, then any chosen  $\psi(\tilde{z}_0) \in E_{\tilde{z}_0}$  can be extended to  $\psi \in \mathbb{F}(\mathcal{U})$  (to prove this one uses the invariance of  $\mathbf{F}(\tilde{\varphi}, \mathcal{W}, \nabla^{\text{ab}})$  under homotopies of  $\tilde{\varphi}$  and the composition law (116).) Therefore,  $\mathbb{F}$  is a locally constant sheaf of  $K$ -dimensional vector spaces. Giving such a sheaf is equivalent to giving a rank  $K$  vector bundle with connection; this is our desired  $(\hat{E}, \nabla)$ .

It will be useful in what follows to have a more concrete description of  $(\hat{E}, \nabla)$ . The stalk of  $\mathbb{F}$  at any  $\tilde{z} \in \tilde{C}'_{\mathcal{W}}$  is just  $E_{\tilde{z}}$ , so on  $\tilde{C}'_{\mathcal{W}}$  we have a canonical isomorphism  $\hat{E} \simeq E$ . Moreover, for paths  $\tilde{\varphi} \subset \tilde{C}'_{\mathcal{W}}$  we have  $\mathbf{F}(\tilde{\varphi}, \mathcal{W}) = \mathbf{D}(\tilde{\varphi})$ , from which it follows that  $\mathbf{F}(\tilde{\varphi}, \mathcal{W}, \nabla^{\text{ab}})$  is just the parallel transport of  $\pi_*(\nabla^{\text{ab}})$ . So on  $\tilde{C}'_{\mathcal{W}}$  we have simply  $(\hat{E}, \nabla) \simeq (E, \pi_* \nabla^{\text{ab}})$ . Now,  $\tilde{C}'_{\mathcal{W}}$  is divided into various connected components  $\mathcal{U}_{\alpha}$ , separated from one another by the lifts of walls of  $\mathcal{W}$ , which are topologically cylinders in  $\tilde{C}$ . We want to know how



to patch together the bundle  $\hat{E}$  across these cylinders. Given a wall  $p_c$  whose lift  $\tilde{p}_c$  separates components  $\mathcal{U}_\alpha$  and  $\mathcal{U}_\beta$ , we define a section of  $\text{End}(E)|_{\tilde{p}_c}$  by

$$\mathcal{T}_c^{\alpha,\beta}(\tilde{z}) = 1 + \sum_{\nu \in \mathcal{S}_c(z)} \mathbf{X}_{\mathbf{a}(\nu, \tilde{\varphi})}(\nabla^{\text{ab}}) \quad (129)$$

where  $\tilde{\varphi}$  denotes a short path which crosses from  $\tilde{z}_1 \in \mathcal{U}_\alpha$  to  $\tilde{z}_2 \in \mathcal{U}_\beta$ , and we take the limit as  $\tilde{z}_1$  and  $\tilde{z}_2$  approach the common point  $\tilde{z}$  of  $\tilde{p}_c$ . Recall here that  $\mathcal{S}_c(z)$  is the soliton set of  $\mathcal{W}$  at  $z$  ( $\mathcal{W}$  is nondegenerate, so only one of the soliton sets is nonempty.) The bundle  $\hat{E}$  can be realized by extending  $E$  to the closure of each component and then gluing along the boundary cylinders:

$$\hat{E} = \left( \bigsqcup_{\alpha} E|_{\bar{\mathcal{U}}_\alpha} \right) / \sim \quad (130)$$

where the equivalence relation  $\sim$  identifies  $(\psi, \tilde{z}) \in E|_{\bar{\mathcal{U}}_\alpha}$  with  $(\mathcal{T}_c^{\alpha,\beta}\psi, \tilde{z}) \in E|_{\bar{\mathcal{U}}_\beta}$  whenever  $\tilde{z}$  lies on the cylinder  $\tilde{p}_c$ .

## 10.2. Singularities and Flags

The twisted flat connection  $(\hat{E}, \nabla)$  produced by our nonabelianization construction carries a bit of extra structure which will be important in what follows. In this section, we briefly describe it.

Let us consider  $(\hat{E}, \nabla)$  in a neighborhood  $\mathcal{U} \subset \tilde{C}$  of one of the singular points  $\mathfrak{s}_n$ . On  $\mathcal{U}$  we can trivialize the covering  $\tilde{\Sigma} \rightarrow \tilde{C}$  and hence there is a decomposition into line bundles (cf. (124)),

$$E = \bigoplus_i \mathcal{L}^{(i)}, \quad (131)$$

preserved by the connection  $\pi_* \nabla^{\text{ab}}$ . We do not generally have such a decomposition for  $\hat{E}$ , because the gluing transformations  $\mathcal{T}_c^{\alpha,\beta}$  mix the different  $\mathcal{L}^{(i)}$ . Still, the  $\mathcal{T}_c^{\alpha,\beta}$  do preserve some structure, as follows. Recall that the walls  $p_c$  of  $\mathcal{W}$  which end at  $\mathfrak{s}_n$  carry labels  $ij$  which obey a constraint: both  $i$  and  $j$  must belong to  $o(\mathfrak{s}_n)$  and we must have  $i < j$ . Using (129), it then follows that all the  $\mathcal{T}_c^{\alpha,\beta}$  are “upper triangular”: for  $i \notin o(\mathfrak{s}_n)$  they preserve  $\mathcal{L}^{(i)}$ , and for  $i \in o(\mathfrak{s}_n)$  they preserve the subspace

$$F^{(i)} = \bigoplus_{j \geq i} \mathcal{L}^{(j)}. \quad (132)$$

Therefore, we find that  $\hat{E}$  decomposes over  $\mathcal{U}$  as

$$\hat{E} = \left( \bigoplus_{i \notin o(\mathfrak{s}_n)} \mathcal{L}^{(i)} \right) \oplus \hat{E}_{\mathfrak{s}_n} \quad (133)$$

where each  $\mathcal{L}^{(i)}$  is  $\nabla$ -invariant, and  $\hat{E}_{\mathfrak{s}_n}$  has rank  $|o(\mathfrak{s}_n)|$  and carries various  $\nabla$ -invariant subbundles  $F^{(i)}$ . The ranks of these subbundles and of their intersections are determined by the structure of the partial ordering in  $o(\mathfrak{s}_n)$ .

In the simplest case,  $o(\mathfrak{s}_n)$  is a totally ordered set containing all  $K$  sheets. In that case we may as well identify the labels with integers  $1 \leq i \leq K$ , and

write the ordering as  $1 < 2 < \dots < K$ . Then the extra structure just discussed is simply a  $\nabla$ -invariant filtration of  $\hat{E}$ ,

$$0 = F^{(K)} \subset F^{(K-1)} \subset F^{(K-2)} \subset \dots \subset F^{(0)} = \hat{E}, \quad (134)$$

where  $F^{(i)}$  has dimension  $K-i$ . If the monodromy of  $\nabla$  around  $\mathfrak{s}_n$  is diagonalizable with all eigenvalues distinct, then there exist  $K!$  such invariant filtrations, and our construction picks one out of those  $K!$ .

What we have found is that our construction naturally produces not just  $(\hat{E}, \nabla)$ , but a bit of extra structure around each  $\mathfrak{s}_n$ , namely the decomposition (133) and the subbundles  $F^{(i)}$ . We refer to this extra structure as *flag data*.

### 10.3. The Nonabelianization Map

Given a nondegenerate spectral network  $\mathcal{W}$  subordinate to a covering  $\Sigma \rightarrow C$ , we have defined a nonabelianization operation which takes a twisted flat rank 1 connection over  $\Sigma$  to a twisted flat rank  $K$  connection over  $C$  with flag data. In this section, we discuss the corresponding map  $\Psi_{\mathcal{W}}$  between *moduli spaces* of twisted flat connections. For simplicity, we take the case where  $C$  only has punctures, not boundaries, and restrict attention to  $\mathcal{W}$  for which each  $o(\mathfrak{s}_n)$  is a total ordering of the set of all sheets.<sup>30</sup>

Let us first describe the relevant moduli space of rank 1 connections  $\nabla^{\text{ab}}$ . The covering surface  $\Sigma$  has punctures  $\mathfrak{s}_n^{(i)}$ , preimages of the  $\mathfrak{s}_n$  on  $C$ . Fix a complex line bundle  $\mathcal{L}$  over  $\tilde{\Sigma}$ ; we will consider flat connections on  $\mathcal{L}$  with fixed holonomy around the punctures. Namely, fix parameters  $\mathfrak{m}_n^{(i)} \in \mathbb{R}/\mathbb{Z}$ , with  $\mathfrak{m}_n^{(i)} \neq \mathfrak{m}_n^{(j)}$  for  $i \neq j$ . Let  $\ell_{n^{(i)}}$  be a small counterclockwise loop around  $\mathfrak{s}_n^{(i)}$ , and  $\tilde{\ell}_{n^{(i)}}$  its canonical lift to  $\tilde{\Sigma}$ . We require that the holonomy of  $\nabla^{\text{ab}}$  around  $\tilde{\ell}_{n^{(i)}}$  is  $\exp(2\pi i \mathfrak{m}_n^{(i)})$ . As usual, we consider connections  $\nabla^{\text{ab}}$  only up to gauge equivalence; our gauge group is the group of smooth maps  $\tilde{\Sigma} \rightarrow \mathbb{C}^\times$ , where  $\tilde{\Sigma}$  is the closure of  $\Sigma$ . Let  $\mathcal{M}(\Sigma, \text{GL}(1); \mathfrak{m})$  be the resulting moduli space of twisted rank 1 flat connections.  $\mathcal{M}(\Sigma, \text{GL}(1); \mathfrak{m})$  is a torsor for  $H^1(\tilde{\Sigma}, \mathbb{C}^\times) \simeq (\mathbb{C}^\times)^{2g}$ , where  $g$  is the genus of  $\tilde{\Sigma}$ .

Next, what can we say about the rank  $K$  flat connections  $\nabla$  we obtain by nonabelianization? Fix a small clockwise loop  $\ell_n$  around  $\mathfrak{s}_n$ , and let  $\tilde{\ell}_n$  be its canonical lift to  $\tilde{C}$ . The holonomy of  $\nabla$  around  $\ell_n$  is the product of several factors: the holonomy of  $\nabla^{\text{ab}}$  along the pieces of  $\ell_n$  running between walls of  $\mathcal{W}$ , and the transformations  $\mathcal{T}_p$  attached to the walls. With respect to the decomposition (131), the former are diagonal matrices whose product is  $\text{diag}\{\exp[2\pi i \mathfrak{m}_n^{(i)}]\}$ , while the latter are unipotent upper triangular. Their product is thus an upper-triangular matrix, with diagonal elements  $\exp[2\pi i \mathfrak{m}_n^{(i)}]$ . Since we have assumed these values distinct, it follows that the monodromy of  $\nabla$  around  $\tilde{\ell}_n$  is semisimple, with eigenvalues  $\exp[2\pi i \mathfrak{m}_n^{(i)}]$ . So, let

<sup>30</sup> Our discussion can be generalized to include marked points on boundaries; in that case the relevant moduli spaces are spaces of connections with *irregular* singularities, on the surface obtained by shrinking each boundary component of  $C$  to a point. In the special case  $K = 2$  this was discussed in [3]. Examples with irregular singularities and  $K > 2$  are discussed in [32].

$\mathcal{M}(C, \mathrm{GL}(K); \mathfrak{m})$  denote the moduli space of twisted rank  $K$  flat connections on  $C$  (with fixed topology of the principal bundle), such that the monodromy around  $\tilde{\ell}_n$  is conjugate to

$$\mathrm{diag}\{\exp[2\pi i \mathfrak{m}_n^{(1)}], \dots, \exp[2\pi i \mathfrak{m}_n^{(K)}]\}. \quad (135)$$

As we have noted, nonabelianization produces not only a twisted flat connection over  $C$ , but a twisted flat connection over  $C$  equipped with flag data. Thus we also define an extended moduli space  $\mathcal{M}_F(C, \mathrm{GL}(K); \mathfrak{m})$ , consisting of elements of  $\mathcal{M}(C, \mathrm{GL}(K); \mathfrak{m})$  with flag data.  $\mathcal{M}_F(C, \mathrm{GL}(K); \mathfrak{m})$  is a finite cover of  $\mathcal{M}(C, \mathrm{GL}(K); \mathfrak{m})$ , as explained at the end of Sect. 10.2. Nonabelianization gives a map

$$\Psi_{\mathcal{W}} : \mathcal{M}(\Sigma, \mathrm{GL}(1); \mathfrak{m}) \rightarrow \mathcal{M}_F(C, \mathrm{GL}(K); \mathfrak{m}). \quad (136)$$

One simple property of  $\Psi_{\mathcal{W}}$  is worth noting at once. Given a spectral network  $\mathcal{W}$ , let  $\mathcal{W}^*$  be obtained by reversing the labels on all walls and the orderings at all punctures. Then we have

$$\Psi_{\mathcal{W}^*}(\mathcal{L}, \nabla^{\mathrm{ab}}) = \Psi_{\mathcal{W}}(\mathcal{L}^*, (\nabla^{\mathrm{ab}})^*)^*. \quad (137)$$

This fact may be of some practical use: the reason is that  $\mathcal{W}_{\vartheta}^* = \mathcal{W}_{\vartheta+\pi}$ , and determining the relation between  $\Psi_{\mathcal{W}_{\vartheta}}$  and  $\Psi_{\mathcal{W}_{\vartheta+\pi}}$  is equivalent to determining the “spectrum generator” which captures the full BPS spectrum.

For the rest of this section, we will study  $\Psi_{\mathcal{W}}$  as defined above. However, we must note that in the physical applications the actual moduli space  $\mathcal{M}$  of the physical theory  $S[\mathfrak{g}, C, D]$  compactified on  $S^1$  is not quite  $\mathcal{M}(C, \mathrm{GL}(K); \mathfrak{m})$ ; it is more closely related to the space of  $SL(K)$  connections (but even this is a slight oversimplification, see [21]). We expect that there is a variant of our nonabelianization map for which the codomain is precisely  $\mathcal{M}$ . The domain of this variant map should not be quite  $\mathcal{M}(\Sigma, \mathrm{GL}(1); \mathfrak{m})$ ; rather it should be a space of twisted rank 1 connections on  $\Sigma$  subject to some additional constraints and/or carrying some additional structure. This is related to the fact that the IR charge lattice of  $S[\mathfrak{g}, C, D]$  is not  $H_1(\Sigma; \mathbb{Z})$ , but rather an appropriate subquotient of it [3]. (See [49, 50] for a careful discussion of subtle issues of this sort that arise in abelianization of Higgs bundles for a general group  $G$ ; we expect that, as far as these topological issues are concerned, the story for moduli of twisted flat connections will be similar.) We leave the proper extension of our construction to account for these variations as an open problem. It is related to open problem 2 of Sect. 1.

#### 10.4. Holomorphic Symplectic Structures

In this section, we explain one of the important properties of  $\Psi_{\mathcal{W}}$ : both its domain and codomain are holomorphic symplectic, and  $\Psi_{\mathcal{W}}$  is a holomorphic symplectic map,

$$\Psi_{\mathcal{W}}^*(\varpi_C) = \varpi_{\Sigma}. \quad (138)$$

We begin with  $\mathcal{M}(\Sigma, \mathrm{GL}(1); \mathfrak{m})$ . The easiest way to understand its holomorphic symplectic form is to recall that on the space of *untwisted* connections

there is a standard such form [51, 52]. This form, formally speaking, is obtained by symplectic quotient from

$$\varpi_{\Sigma} = \int_{\Sigma} \delta\alpha \wedge \delta\alpha \quad (139)$$

where  $\alpha$  denotes the abelian connection 1-form.<sup>31</sup> Using a choice of spin structure on  $\Sigma$  to relate twisted and untwisted connections we can transfer this holomorphic symplectic structure to  $\mathcal{M}(\Sigma, \mathrm{GL}(1); \mathfrak{m})$ ; by abuse of notation, we denote that structure also as  $\varpi_{\Sigma}$ .

Similarly,  $\mathcal{M}_F(C, \mathrm{GL}(K); \mathfrak{m})$  has symplectic form obtained by symplectic quotient from

$$\varpi_C = \int_C \mathrm{Tr} \delta\mathcal{A} \wedge \delta\mathcal{A} \quad (140)$$

with  $\mathcal{A}$  the nonabelian connection 1-form. Again we use a spin structure on  $C$  to pass between twisted and untwisted connections.

To see why (138) is true, first consider variations  $\delta\alpha$  which have support away from the spectral network  $\mathcal{W}$ . In this case the corresponding variation of  $\mathcal{A}$  is simply  $\delta\mathcal{A} = \pi_*\delta\alpha$ . It follows easily that for such variations we have  $\varpi_{\Sigma}(\delta\alpha, \delta\alpha') = \varpi_C(\delta\mathcal{A}, \delta\mathcal{A}')$ . If  $\delta\alpha$  has support intersecting  $\mathcal{W}$  then the situation is slightly more involved: there is a distribution-valued contribution to  $\delta\mathcal{A}$  supported on  $\mathcal{W}$ . By a gauge transformation  $\delta\alpha \rightarrow \delta\alpha + d\chi$  we may assume however that  $\delta\alpha$  vanishes in neighborhoods of joints. Therefore, it is enough to consider what happens in a patch intersecting only a single wall  $p_c$  of  $\mathcal{W}$ . Taking local coordinates  $(x, y)$  on  $C$  where  $p_c$  is the locus  $y = 0$ , one finds the form

$$\delta\mathcal{A} = \pi_*\delta\alpha + s(x)\delta(y)dy, \quad (141)$$

where in the decomposition (131)  $s(x)$  is off-diagonal and  $\pi_*\delta\alpha$  diagonal. One thus has  $\mathrm{Tr}(s(x)dy \wedge \pi_*\delta\alpha') = 0$ , so the last term in (141) does not contribute to  $\varpi$ ; we get once again  $\varpi_{\Sigma}(\delta\alpha, \delta\alpha') = \varpi_C(\delta\mathcal{A}, \delta\mathcal{A}')$ .

## 10.5. Dimension Counts

Now let us compare the dimensions of the moduli spaces related by  $\Psi_{\mathcal{W}}$ .

First, we consider  $\mathcal{M}(C, \mathrm{GL}(K); \mathfrak{m})$ . This space can be represented as the space of  $\mathrm{GL}(K, \mathbb{C})$  matrices  $A_1, \dots, A_{g_C}, B_1, \dots, B_{g_C}, C_1, \dots, C_s$ , subject to some relations: first, the eigenvalues of the  $C_n$  are fixed; second, we impose

$$\prod_i [A_i, B_i] \prod_n C_n = 1; \quad (142)$$

third, we divide out by overall conjugation. The condition on the eigenvalues of the  $C_n$  eliminates  $Ks$  degrees of freedom. The equation (142) imposes only  $K^2 - 1$  independent conditions: if  $\prod_n \det C_n \neq 1$  there are no solutions, but if

<sup>31</sup> Here and below, we use the freedom to fix a gauge so that we consider only variations  $\delta\alpha$  which vanish near the punctures. This makes the integral (139) convergent, despite the fact that the connections we consider are singular.

$\prod_n \det C_n = 1$ , while there are solutions, the determinant equation is redundant. Similarly, the center of the gauge group acts trivially, so the quotient by the gauge group reduces the dimension by  $K^2 - 1$ . Altogether we get

$$\dim_{\mathbb{C}} \mathcal{M}(C, \mathrm{GL}(K); \mathfrak{m}) = (2g_C + s)K^2 - Ks - 2(K^2 - 1). \quad (143)$$

On the other hand, using the Riemann-Hurwitz relation

$$\dim_{\mathbb{C}} \mathcal{M}(\Sigma, \mathrm{GL}(1); \mathfrak{m}) = 2g_{\Sigma} = K(2g_C - 2) + B + 2, \quad (144)$$

where  $B$  is the branching number (which is simply the number of branch points, since we assume all branch points are simple.) It follows that

$$\begin{aligned} \dim_{\mathbb{C}} \mathcal{M}(C, \mathrm{GL}(K); \mathfrak{m}) - \dim_{\mathbb{C}} \mathcal{M}(\Sigma, \mathrm{GL}(1); \mathfrak{m}) \\ = (K^2 - K)(2g_C + s - 2) - B. \end{aligned} \quad (145)$$

Since  $\Psi_{\mathcal{W}}$  is symplectic it is in particular locally injective, so the quantity in (145) must be nonnegative. We thus get a topological restriction on the branching number of a branched cover  $\Sigma \rightarrow C$  of degree  $K$  admitting a spectral network:

$$B \leq (K^2 - K)(2g_C + s - 2). \quad (146)$$

If this bound is saturated,  $\Psi_{\mathcal{W}}$  is a local symplectomorphism. Note that this does happen for the spectral curves (2), to which the spectral networks  $\mathcal{W}_{\vartheta}$  are subordinate.

Equation (146) is a purely topological statement. It says that there cannot be too many branch points in a spectral network. It would be interesting to give a more direct topological proof of this assertion.

In the case, where (146) is saturated and  $\Psi_{\mathcal{W}}$  is a covering map, we have constructed not only a local isomorphism between  $\mathcal{M}(C)$  and  $\mathcal{M}(\Sigma)$ , but also a local identification between  $(\pi_* \mathcal{L}, \pi_* \nabla^{\mathrm{ab}})$  and  $(\hat{E}, \nabla)$ . We can use this to tie up a loose end from Sect. 6.5. Letting  $(\mathcal{L}, \nabla)$  vary we get a *universal bundle*  $\pi_*(\mathcal{L})$  over  $\mathcal{M}(\Sigma) \times C$ . Restricting this to some point  $\tilde{z} \in \tilde{C}$  gives a line bundle over  $\mathcal{M}(\Sigma)$ , which we call  $\pi_*(\mathcal{L})_{\tilde{z}}$ . Similarly we have a universal bundle  $\hat{E}_{\tilde{z}}$  over  $C$ .<sup>32</sup> Now let  $A(\Sigma)$  be the algebra of global rational sections of  $\mathrm{End}(\pi_*(\mathcal{L})_{\tilde{z}})$  (i.e. the algebra of sections over the generic point), and similarly  $A(C)$  the algebra of global rational sections of  $\mathrm{End}(\hat{E}_{\tilde{z}})$ .  $A(\Sigma)$  is a central simple algebra of degree  $K$  over  $\mathcal{R}(\mathcal{M}(\Sigma))$ , the field of rational functions on  $\mathcal{M}(\Sigma)$ , and similarly  $A(C)$  is a central simple algebra of degree  $K$  over  $\mathcal{R}(\mathcal{M}(C))$ . Our nonabelianization construction gives an embedding of  $A(C)$  in  $A(\Sigma)$  and a compatible embedding of  $\mathcal{R}(\mathcal{M}(C))$  in  $\mathcal{R}(\mathcal{M}(\Sigma))$ . Now, we have just shown that this embedding realizes  $\mathcal{M}(\Sigma)$  as a *cover* of  $\mathcal{M}(C)$ . It follows that  $\mathcal{R}(\mathcal{M}(\Sigma))$  is a finite extension of  $\mathcal{R}(\mathcal{M}(C))$  (since it is an extension and the two have the same dimension, hence the same transcendence degree over  $\mathbb{C}$ , and are both finitely generated). Thus, the group  $\mathrm{Aut}(\mathcal{R}(\mathcal{M}(\Sigma)) : \mathcal{R}(\mathcal{M}(C)))$  is finite. But now it follows by dimension counting that  $\mathrm{Aut}(A(\Sigma) : A(C))$  is also finite.

<sup>32</sup> Actually this universal bundle does not quite exist owing to subtleties involving the center of  $\mathrm{GL}(K)$ ; but its bundle of endomorphisms does exist, which is all we will use below.

### 10.6. Equivalence of Spectral Networks

Given a nondegenerate spectral network  $\mathcal{W}$  we have constructed a corresponding nonabelianization map  $\Psi_{\mathcal{W}}$ . It is natural to ask how this map depends on  $\mathcal{W}$ .

We can get some intuition by remembering previous results about the case  $K = 2$ . In that case, as already mentioned in Sect. 5.7,  $\mathcal{W}$  determines an ideal triangulation  $T(\mathcal{W})$  of  $C$  up to isotopy.  $T(\mathcal{W})$  does not change if we vary  $\mathcal{W}$  by an isotopy (while holding  $\Sigma$  fixed). Moreover, the map  $\Psi_{\mathcal{W}}$  in this case only depends on  $T(\mathcal{W})$ . So we find that  $\Psi_{\mathcal{W}}$  is an isotopy invariant of  $\mathcal{W}$ .

We now extend this discussion to general  $K$ . We begin by defining a notion of equivalence between nondegenerate spectral networks. Our experience from the case  $K = 2$  suggests that equivalence should at least include isotopy. In fact, for  $K > 2$  it will be natural to allow even some moves which are *not* isotopies. Given two nondegenerate spectral networks  $\mathcal{W}$  and  $\mathcal{W}'$ , an *equivalence* between  $\mathcal{W}$  and  $\mathcal{W}'$  is a one-parameter family of path-lifting rules  $\mathbf{F}(\cdot, t)$  to coverings  $\Sigma(t)$ , such that:

**E1.**  $\mathbf{F}(\cdot, 0) = \mathbf{F}(\cdot, \mathcal{W})$  and  $\mathbf{F}(\cdot, 1) = \mathbf{F}(\cdot, \mathcal{W}')$ .

**E2.** For any point  $\tilde{z}$  on  $\tilde{C}$  and pair  $0 \leq t_1 \leq t_2 \leq 1$ , there exists a formal sum of paths  $R(\tilde{z}, t_1, t_2)$ , such that for any path  $\wp$  from  $\tilde{z}_1$  to  $\tilde{z}_2$  we have

$$\mathbf{F}(\wp, t_2) = R(\tilde{z}_1, t_1, t_2) \mathbf{F}(\wp, t_1) R(\tilde{z}_2, t_1, t_2)^{-1}. \quad (147)$$

The point of this definition is that if  $\mathcal{W}$  and  $\mathcal{W}'$  are connected by an equivalence then we have

$$\Psi_{\mathcal{W}} = \Psi_{\mathcal{W}'}. \quad (148)$$

(The meaning of the equal signs in (147), (148) has to be clarified in case the coverings  $\Sigma(t)$  vary with  $t$ : in this case we identify the different coverings using the “Gauss-Manin” parallel transport induced by the family  $\{\Sigma(t)\}$ .)

One natural way of getting an equivalence is to consider a 1-parameter family of nondegenerate spectral networks  $\mathcal{W}(t)$  and take  $\mathbf{F}(\cdot, t) = \mathbf{F}(\cdot, \mathcal{W}(t))$ . Then  $R(\tilde{z}, t_1, t_2)$  is a product, with one factor of the form  $(1 + \sum_{\nu \in \mathcal{S}_c(t)} \mathbf{X}_{\mathbf{a}(\nu)})$  for each  $t$  with  $t_1 \leq t \leq t_2$  such that  $z$  lies on a  $\mathcal{S}$ -wall  $p_c(t)$  in  $\mathcal{W}(t)$ .

Here are some examples of equivalences:

- M1.** An *isotopy*. This is a one-parameter family of spectral networks  $\mathcal{W}(t)$  subordinate to coverings  $\Sigma(t)$ , such that the  $\Sigma(t)$  vary continuously in the obvious sense (in particular the branch points move continuously), each wall  $p_c$  varies by isotopy, and the joints  $z_\mu$  move continuously. The combinatorics of  $\mathcal{W}(t)$  do not change during an isotopy.
- M2.** A *crossing at a defect*. This is a family  $\mathcal{W}(t)$  which implements a move as illustrated in Fig. 43. Here the combinatorics of the network do change at the critical value  $t = t_c$ . Nevertheless the family gives an equivalence.
- M3.** A *bubble*. This is a family  $\mathcal{W}(t)$  which implements a move as illustrated in Fig. 44.
- M4.** A *traversal of a branch point*. This is a family  $\mathcal{W}(t)$  which implements a move as illustrated in Fig. 45. The member  $\mathcal{W}(t_c)$  strictly speaking is not quite a spectral network according to our definition, because of the extra

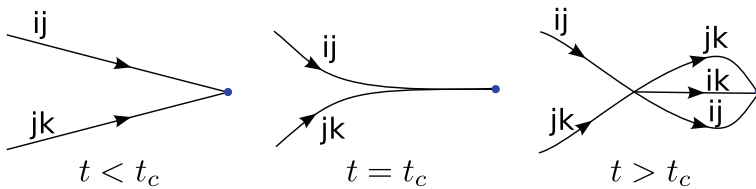


FIGURE 43. Crossing at a defect: an equivalence which relates two non-isotopic spectral networks

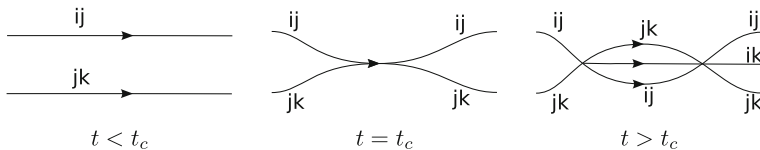


FIGURE 44. *Bubble*: an equivalence which relates two non-isotopic spectral networks

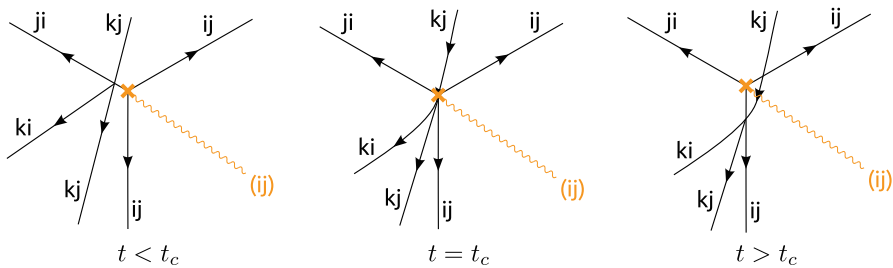


FIGURE 45. Branch point traversal: an equivalence which relates two non-isotopic spectral networks

$kj$  lines passing through the branch point; nevertheless one can define  $\mathbf{F}(\cdot, t_c)$  by a straightforward extension of our rules, and after so doing, E1, E2 above are obeyed.

All of these equivalences really occur in practice, e.g. in families  $\mathcal{W}_\vartheta$  as  $\vartheta$  varies. We emphasize that these equivalences are *not* the jumps of  $\mathcal{W}_\vartheta$  which we studied in Sect. 6.2: those jumps generally connect spectral networks which are *inequivalent* in our sense.

### 10.7. A Picture of the Set of All Spectral Networks

Fix a curve  $C$  and singularities  $\mathfrak{s}_n$  and orderings  $o(\mathfrak{s}_n)$ . Exploration of some examples leads to a rough conjecture about the set  $X = X(C, \mathfrak{s}_n, o(\mathfrak{s}_n))$  of all spectral networks with these data fixed.

There should be a natural topology on  $X$ , such that  $X$  is decomposed into connected cells, separated by codimension-1 loci where the spectral networks become degenerate. (Ideally the cells should even be contractible, but this might require extending our definition of spectral network to allow more

non-generic phenomena which occur in codimension greater than 1, e.g. allowing several joints to coalesce.) Moving around in a single cell corresponds to varying  $\mathcal{W}$  by equivalences in the sense of Sect. 10.6. Crossing one of the boundaries between cells corresponds to one of the more interesting “ $\mathcal{K}$ -wall” jumps that occur when the network degenerates, which we discussed in Sect. 6.

We leave it as an interesting open problem to formulate this picture on a more rigorous and precise basis.

### 10.8. Coordinate Systems and a Cluster Conjecture

Let us consider a spectral network  $\mathcal{W}$  for which (146) is saturated. In this case, as we have noted,  $\Psi_{\mathcal{W}}$  is at least locally one-to-one. Thus the holonomies  $\tilde{\mathcal{Y}}_{\tilde{\gamma}}$  of  $\nabla^{\text{ab}} = \Psi_{\mathcal{W}}^{-1}(\nabla)$ , where  $\tilde{\gamma}$  runs over a basis of  $H_1(\tilde{\Sigma}; \mathbb{Z})$ , provide *local coordinates* on  $\mathcal{M}_F(C, \text{GL}(K); \mathfrak{m})$ .

It is natural to ask whether  $\Psi_{\mathcal{W}}$  might be even globally one-to-one, so that it would give global coordinates on its whole image  $\mathcal{U}_{\mathcal{W}}$ . We conjecture that this is indeed the case.<sup>33</sup> At least in the case  $K = 2$ , this conjecture is true: in this case the coordinate system in question is essentially the Fock-Goncharov coordinate system [13, 3]. Some further special cases in which the conjecture is true will be treated in [32].

As we noted above, if  $\mathcal{W}$  and  $\mathcal{W}'$  are equivalent nondegenerate spectral networks then  $\Psi_{\mathcal{W}}$  and  $\Psi_{\mathcal{W}'}$  are related by parallel transport, in the sense explained under (147). Thus, we can roughly say that  $\Psi_{\mathcal{W}}$  is constant as  $\mathcal{W}$  varies within one of the cells of  $X$ . If  $\mathcal{W}$  and  $\mathcal{W}'$  are in neighboring cells, then the story is more interesting: taking a path  $\mathcal{W}(t)$  from  $\mathcal{W}$  to  $\mathcal{W}'$ , there is a critical  $t = t_c$  where  $\mathcal{W}(t)$  becomes a degenerate spectral network. At this moment  $\Psi_{\mathcal{W}(t)}$  jumps discontinuously. This jump should be determined by the arguments of Sect. 6.3. Indeed, by the same formulas we used there (and assuming the same conjectures we assumed there), the degenerate network  $\mathcal{W}(t_c)$  determines integers  $\Omega(\gamma)$  and distinguished lifts  $\tilde{\gamma}$ , for  $\gamma$  lying in a semilattice  $\Gamma_c \subset H_1(\Sigma(t_c); \mathbb{Z})$  with a single generator. Using these data we define a (birational) automorphism  $\mathcal{K}$  of  $\mathcal{M}(\Sigma(t_c), \text{GL}(1); \mathfrak{m})$ , by its action on the holonomies  $\tilde{\mathcal{Y}}_{\tilde{\gamma}}$  of  $\nabla^{\text{ab}}$  around 1-cycles  $\tilde{\gamma}$ :

$$\mathcal{K}(\tilde{\mathcal{Y}}_{\tilde{\gamma}}) = \tilde{\mathcal{Y}}_{\tilde{\gamma}} \prod_{\gamma' \in \Gamma_c} (1 + \tilde{\mathcal{Y}}_{\tilde{\gamma}'}^{\langle \gamma, \gamma' \rangle \Omega(\gamma')}). \quad (149)$$

When  $t$  crosses  $t_c$ , the map  $\Psi_{\mathcal{W}(t)}$  jumps by composition with  $\mathcal{K}$ , so

$$\Psi_{\mathcal{W}}(\nabla^{\text{ab}}) = \Psi_{\mathcal{W}'}(\mathcal{K}^{\pm 1}(\nabla^{\text{ab}})) \quad (150)$$

(where the sign  $\pm 1$  is determined by which direction we cross the cell boundary, and we must bear in mind the comment under (148).)

<sup>33</sup> This is the moment where it is important that we use the moduli space  $\mathcal{M}_F(C, \text{GL}(K); \mathfrak{m})$  with flag data rather than just  $\mathcal{M}(C, \text{GL}(K); \mathfrak{m})$ . If we used  $\mathcal{M}(C, \text{GL}(K); \mathfrak{m})$ , then  $\Psi_{\mathcal{W}}$  would be only finite-to-one.



Finally, we conjecture that the coordinate systems we have defined are actually *cluster* coordinates.<sup>34</sup> The spaces  $\mathcal{M}(C)$  which we are considering are indeed cluster varieties, as shown by Fock–Goncharov in [13]. However, relatively few of the cluster coordinate systems on  $\mathcal{M}(C)$  have been described explicitly. We believe that for each  $\mathcal{W}$ , the coordinate system induced by  $\Psi_{\mathcal{W}}$  should be an element of the cluster atlas. So in particular, we conjecture that there is an algorithm which determines from a  $\mathcal{W}$  a finite set of charges  $\gamma_{\mathcal{W}}^n \in \Gamma$ . The  $\gamma_{\mathcal{W}}^n$  should be the elements of a cluster seed, with exchange matrix given by the intersection pairings:  $\epsilon^{nm} = \langle \gamma^n, \gamma^m \rangle$ . Informally speaking, the  $\gamma_{\mathcal{W}}^n$  should be the charges of finite webs which can appear inside of  $\mathcal{W}$  when  $\mathcal{W}$  degenerates; said otherwise, the  $\gamma_{\mathcal{W}}^n$  label faces of the boundary of the cell in  $X$  containing  $\mathcal{W}$ . Finally, the  $\mathcal{K}$ -transformation (149), in the simplest case where there is only a single  $\Omega(\gamma) = 1$  contributing, should be identified with the action of a mutation on cluster variables.

This conjecture is true when  $K = 2$ , and in [32], we will give some additional evidence for it in some special cases with  $K > 2$ . In general, though, we do not know how to prove it, or even how to read off the  $\gamma_{\mathcal{W}}^n$  from  $\mathcal{W}$ .

It is natural to wonder further whether spectral networks might give *all* of the coordinate systems in the cluster atlas on  $\mathcal{M}$ .

### 10.9. WKB Asymptotics

So far in this section, we have been considering the map  $\Psi_{\mathcal{W}}$  associated to an arbitrary spectral network  $\mathcal{W}$ . These maps have especially interesting asymptotic properties if we choose  $\mathcal{W}$  to be the particular network  $\mathcal{W}_{\mathfrak{g}}$  which we studied in the rest of the paper. In this section, we briefly explain this point. (We will gloss over the role of the twisting in this section, and use spin structures to pass freely between twisted and untwisted connections.)

Equip  $C$  with a complex structure. Suppose given a  $\mathrm{GL}(K)$  Higgs bundle over  $C$ , i.e. a triple  $(V, \bar{\partial}, \varphi)$  where  $V$  is a complex vector bundle of rank  $K$ ,  $\bar{\partial}$  a holomorphic structure on  $V$ , and  $\varphi$  a meromorphic  $\mathrm{End}(V)$ -valued 1-form, with poles at the punctures on  $C$ . Let  $\Sigma$  be the spectral cover determined by  $\varphi$ ,

$$\Sigma = \{\det(\varphi - \lambda) = 0\} \subset T^*C. \quad (151)$$

Now assume that  $\Sigma$  is smooth (this is the case for generic  $\varphi$ .) Then we have a corresponding spectral line bundle over  $\Sigma$ ,

$$\mathcal{L} = \ker(\varphi - \lambda), \quad (152)$$

and the Higgs bundle  $(V, \bar{\partial}, \varphi)$  is the pushforward of  $(\mathcal{L}, \lambda)$ .<sup>35</sup>

<sup>34</sup> It was noted in [9, 21] that the  $\mathcal{K}$ -transformations which appeared there could be viewed as cluster transformations in an appropriate sense; see also [10, 38, 48, 53, 54] for further discussion of the relation of cluster varieties to four-dimensional  $\mathcal{N} = 2$  theory.

<sup>35</sup> The passage from  $(V, \bar{\partial}, \varphi)$  to  $\mathcal{L}$  is often called abelianization and has been exploited heavily in the study of “nonabelian theta functions”; see e.g. [55] where it was introduced, and [50] for a very precise description of the abelianization map in the more general setting of an arbitrary Lie group  $G$ . Abelianization for Higgs bundles is simpler than for flat connections in one important respect: for Higgsbundles the monodromy around branch points causes no

Now, given a Higgs bundle there is a corresponding solution of Hitchin's equations [16, 56–59]. In [5] we proposed a new method for constructing that solution. The key ingredient is a set of integral equations written in Sect. 5.6 of [5]. The equations are determined by the data of:

- the Higgs bundle  $(V, \bar{\partial}, \varphi)$ ,
- the BPS degeneracies  $\mu$  and  $\omega$ , which (as we have explained in Sects. 4, 5, 6) are computed from the spectral networks  $\mathcal{W}_\vartheta$  determined by  $\Sigma$ ,
- a real parameter  $R > 0$ .

These equations are expected to have a unique solution for large enough  $R$  (constructed e.g. by iteration). The solution produces a family of flat connections  $\nabla(\zeta)$  in  $V$ , together with a family of flat connections  $\nabla^{\text{ab}}(\zeta)$  in  $\mathcal{L}$ , and a family of isomorphisms

$$g(\zeta) : \Psi_{\mathcal{W}_{\vartheta=\arg \zeta}}(\mathcal{L}, \nabla^{\text{ab}}(\zeta)) \rightarrow (V, \nabla(\zeta)). \quad (153)$$

The integral equations give some control over the analytic structure and  $\zeta \rightarrow 0, \infty$  asymptotic behavior of  $g(\zeta)$  and  $\nabla^{\text{ab}}(\zeta)$ . From this we deduce that the family  $\nabla(\zeta)$  is of the form

$$\nabla(\zeta) = R\zeta^{-1}\varphi + D + R\zeta\bar{\varphi}, \quad (154)$$

where  $D$  is a connection in  $V$ , unitary with respect to some Hermitian metric  $h$  in  $V$  (“harmonic metric”), and  $\bar{\varphi}$  is the adjoint of  $\varphi$  in the metric  $h$ . The fact that  $\nabla(\zeta)$  is flat for all  $\zeta \in \mathbb{C}^\times$  is equivalent to the statement that  $(V, \varphi, D, \bar{\varphi})$  constitute a solution of *Hitchin's equations* on  $C$ . (The natural conjecture is that it is the unique such solution corresponding to the given Higgs bundle  $(V, \bar{\partial}, \varphi)$ ). So the integral equations are a machine for producing solutions to Hitchin's equations.

In addition, the integral equations give some more interesting asymptotic information. This information is most naturally formulated in terms of the objects

$$\mathcal{Y}_a(\zeta) = X_a(\nabla^{\text{ab}}(\zeta)) \in \text{Hom}(V_{\tilde{z}_1}, V_{\tilde{z}_2}). \quad (155)$$

If we think of  $\mathcal{Y}_a(\zeta)$  as a function of  $\tilde{z}_2$  then it is flat with respect to the connection  $\nabla(\zeta)$ , except when  $z_2$  meets the network  $\mathcal{W}_{\vartheta=\arg \zeta}$ : when  $z_2$  crosses the  $\mathcal{S}$ -walls,  $\mathcal{Y}_a(\zeta)$  jumps discontinuously. Analogous remarks hold for the dependence of  $\mathcal{Y}_a$  on  $\tilde{z}_1$ . From the integral equations we learn that the  $\zeta \rightarrow 0$  asymptotics of  $\mathcal{Y}_a(\zeta)$  are given by

$$\mathcal{Y}_a(\zeta) \sim \exp \left[ \pi \frac{R}{\zeta} Z_{\bar{a}} \right]. \quad (156)$$

Indeed, we can say something a bit stronger: suppose that we define a new section  $\mathcal{Y}_a^\vartheta(\zeta)$  by *analytic continuation* of  $\mathcal{Y}_a(\zeta)$  from the locus  $\arg \zeta = \vartheta$ . In this case, we have the asymptotics

---

Footnote 35 continued

problem and no spectral network is needed. There is thus only *one* abelianization map for Higgs bundles, in contrast to the various  $\Psi_{\mathcal{W}}$  we found in this paper for flat connections.

$$\mathcal{Y}_a^\vartheta(\zeta) \sim \exp \left[ \pi \frac{R}{\zeta} Z_a \right], \quad (157)$$

so long as  $\zeta \rightarrow 0$  while remaining in the half-plane  $\mathbb{H}_\vartheta$  centered on the ray  $\arg \zeta = \vartheta$ . This is rather sharp asymptotic information about the flat sections of the family of flat connections  $\nabla(\zeta)$ .

Our way of describing this asymptotic information is perhaps a bit unfamiliar, so let us relate it to something better known: the WKB analysis of 1-parameter families of differential equations. The flatness equation  $\nabla(\zeta)s = 0$  takes the general form

$$\left[ \zeta \frac{d}{dz} + (R\varphi + \zeta(\cdots)) \right] s = 0. \quad (158)$$

The WKB analysis of this equation involves studying formal solutions of the form

$$s_{\text{fo}}^{(i)}(\zeta) = \exp \left[ \frac{1}{\zeta} \sum_{n=0}^{\infty} S_n^{(i)} \zeta^n \right] \sum_{n=0}^{\infty} T_n^{(i)} \zeta^n \quad (159)$$

where each  $S_n^{(i)}$  is a function and  $T_n^{(i)}$  is a section of  $V$ . One important difficulty in the WKB method is that these solutions really are only formal: the series (159) typically has zero radius of convergence. So one should interpret (159) as an asymptotic series. It is then natural to ask, are there actual  $\nabla(\zeta)$ -flat sections  $s^{(i)}(\zeta)$  which have these asymptotics?

Our discussion in this section provides the answer: such sections do indeed exist, in the following sense. Fix a phase  $\vartheta$ , a basepoint  $\tilde{z}_1 \in \tilde{C}'_{\mathcal{W}_\vartheta}$ , and basis vectors  $e^{(i)} \in \mathcal{L}_{\tilde{z}_1}^{(i)}$ . Then letting  $\tilde{\varphi}$  be a path in  $\tilde{C}'_{\mathcal{W}_\vartheta}$  from  $\tilde{z}_1$  to  $\tilde{z}$ , we can define

$$s^{(i)}(\zeta) = e^{(i)} \mathcal{Y}_{\tilde{\varphi}^{(i)}}^\vartheta(\zeta). \quad (160)$$

$s^{(i)}(\zeta)$  is a solution of  $\nabla(\zeta)s^{(i)}(\zeta) = 0$ , and (157) implies that  $s^{(i)}$  indeed has an asymptotic expansion of the form (159), as  $\zeta \rightarrow 0$  in the half-plane  $\mathbb{H}_\vartheta$ . So, in each connected component of  $\tilde{C}'_{\mathcal{W}_\vartheta}$ , we have a basis of  $\nabla(\zeta)$ -flat sections which have the asymptotics predicted by WKB, as  $\zeta \rightarrow 0$  in  $\mathbb{H}_\vartheta$ . As we move from one component of  $\tilde{C}'_{\mathcal{W}_\vartheta}$  to another, these bases change. Thus, the walls of  $\mathcal{W}_\vartheta$  have an interpretation as Stokes lines.

Finally, we should say that some of the structures which appeared in this paper have appeared before in the WKB literature (although we arrived at them independently.) In the case  $K = 2$ , our constructions here and in [3] seem to be closely related to a line of development pursued by Voros and others (e.g. [60, 61, 19, 18]). For  $K > 2$ , the phenomenon that new Stokes lines can be born at intersections between old ones was apparently first noticed in [62], and has been followed up in a few works since then; we note in particular the reference [20], which contains examples of Stokes diagrams in the case  $K = 3$  which look identical to our  $\mathcal{W}_\vartheta$ . What we have here called *joints* are there called *virtual turning points*. There are also results of Simpson on asymptotics

of monodromy which seem likely to be related to ours, e.g. [63, 64]. We have not understood the precise relation.

## Acknowledgements

We would like to thank P. Boalch, T. Bridgeland, A. Goncharov, N. Hitchin, S. Keel, I. Smith and E. Witten for discussions. GM and AN thank the KITP for hospitality, where some of this work was done. GM would also like to thank the Institute for Advanced Study and the Ambrose Monell Foundation for partial support during part of this research. This research was supported in part by DARPA under grant HR0011-09-1-0015 and by the NSF under grant PHY05-51164. The work of DG is supported in part by NSF grant PHY-0969448 and in part by the Roger Dashen membership in the Institute for Advanced Study. The work of GM is supported by the DOE under grant DE-FG02-96ER40959. The work of AN is supported by the NSF under grant DMS-1006046.

## Appendix A. Joint Rules for Two-Way Streets

In this appendix, we describe the rules governing the behavior of the solitons near a rather general kind of joint, where we have solitons oriented into the joint on up to six distinct trajectories. This kind of joint does not occur in a nondegenerate spectral network, but it can occur for a degenerate one. In the language of Sect. 5, this means that this kind of joint does not occur in  $\mathcal{W}_\vartheta$  at generic values of  $\vartheta$ , but it can occur for non-generic  $\vartheta$ . For generic values of  $\vartheta$ , the simpler rules of Sect. 5 suffice.

### A.1. Rules for Soliton degeneracies

We consider six two-way streets  $p_n$  entering a joint as in Fig. 46. It is convenient to summarize the soliton spectrum by defining generating functions for the soliton degeneracies on these six streets:

$$\nu_n = \sum_{\bar{a}} \mu(a, p_n) X_{\bar{a}}, \quad \tau_n = \sum_{\bar{b}} \mu(b, p_{n+3}) X_{\bar{b}}, \quad (161)$$

where in  $\nu_n$  the sum runs over all ingoing charges  $\bar{a}$  supported on  $p_n$  (i.e. charges for which  $|Z_{\bar{a}}|$  increases as we go into the joint), in  $\tau_n$  the sum runs over all outgoing charges  $\bar{b}$  supported on  $p_{n+3}$ , and  $n$  is always taken mod 6. Applying the constraint of homotopy invariance to artfully chosen paths one shows

$$\begin{aligned} \tau_1 &= \nu_1 + \nu_6 \tau_2, & \tau_2 &= \nu_2 + \tau_3 \nu_1, & \tau_3 &= \nu_3 + \nu_2 \tau_4, \\ \tau_4 &= \nu_4 + \tau_5 \nu_3, & \tau_5 &= \nu_5 + \nu_4 \tau_6, & \tau_6 &= \nu_6 + \tau_1 \nu_5. \end{aligned} \quad (162)$$

(These equations have a symmetry under a cyclic shift of the index by 1 combined with reversing order of all products, corresponding to the symmetry of Fig. 46 under a rotation by  $\pi/3$  combined with reversing the order of the labels

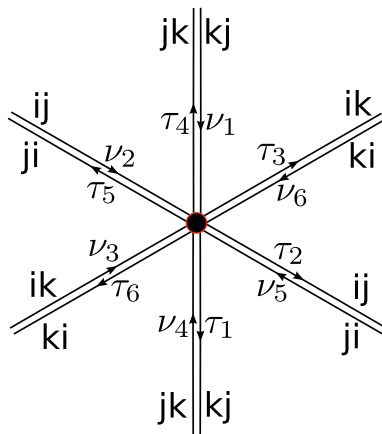


FIGURE 46. The British resolution of a joint with two-way streets. Incoming soliton degeneracies are described by generating functions  $\nu_n$ ,  $n = 1, \dots, 6$ . Outgoing soliton degeneracies are described by generating functions  $\tau_n$ ,  $n = 1, \dots, 6$

on all lines.) Using (162) repeatedly we obtain the outgoing degeneracy  $\tau_1$  in terms of the incoming degeneracies  $\nu_i$ ,

$$\begin{aligned} \tau_1 &= \nu_1 + \nu_6\nu_2 + \nu_6\nu_3\nu_1 + \nu_6\nu_2\nu_4\nu_1 + \nu_6\nu_2\nu_5\nu_3\nu_1 + \nu_6\nu_2\nu_4\nu_6\nu_3\nu_1 \\ &\quad + \nu_6\nu_2\nu_4\nu_1\nu_5\nu_3\nu_1 + \nu_6\nu_2\nu_4\nu_6\nu_2\nu_5\nu_3\nu_1 + \nu_6\nu_2\nu_4\nu_6\nu_3\nu_1\nu_5\nu_3\nu_1 + \dots \\ &= \frac{\nu_1 + \nu_6\nu_2 + \nu_6\nu_3\nu_1 + \nu_6\nu_2\nu_4\nu_1 + \nu_6\nu_2\nu_5\nu_3\nu_1 + \nu_6\nu_2\nu_4\nu_6\nu_3\nu_1}{1 - \overrightarrow{\nu_6\nu_2\nu_4} \otimes \overleftarrow{\nu_5\nu_3\nu_1}}. \end{aligned} \quad (163)$$

The arrows indicate that the denominator is to be expanded as a geometric series and then ordered so that factors  $(\nu_6\nu_2\nu_4)$  are to be multiplied successively on the left and the factors  $(\nu_5\nu_3\nu_1)$  are to be multiplied successively on the right.

Note that if we put  $\nu_3 = \nu_4 = \nu_5 = 0$ , so that we have only three incoming streets, then we find that  $\tau_3 = \tau_4 = \tau_5 = 0$ , and

$$\tau_2 = \nu_2, \quad \tau_1 = \nu_1 + \nu_6\nu_2, \quad \tau_6 = \nu_6. \quad (164)$$

This reproduces the rules for one-way streets which we obtained in Sect. 5.2, as expected.

For the record, the analog of (162) with the American resolution is

$$\begin{aligned} \tau_1 &= \nu_1 + \tau_6\nu_2, & \tau_2 &= \nu_2 + \nu_3\tau_1, & \tau_3 &= \nu_3 + \tau_2\nu_4, \\ \tau_4 &= \nu_4 + \nu_5\tau_3, & \tau_5 &= \nu_5 + \tau_4\nu_6, & \tau_6 &= \nu_6 + \nu_1\tau_5, \end{aligned} \quad (165)$$

which leads to

$$\begin{aligned} \tau_1 &= \frac{1}{1 - \overrightarrow{\nu_1\nu_5\nu_3} \otimes \overleftarrow{\nu_4\nu_6\nu_2}} (\nu_1 + \nu_6\nu_2 + \nu_1\nu_5\nu_2 \\ &\quad + \nu_1\nu_4\nu_6\nu_2 + \nu_1\nu_5\nu_3\nu_6\nu_2 + \nu_1\nu_5\nu_2\nu_4\nu_6\nu_2). \end{aligned} \quad (166)$$

For a branch point with two-way streets, we also have new rules. As explained in Sect. 5.3 each of the three streets  $p_i$  emerging from the branch point carries a charge  $\bar{a}_i$  with a natural lift  $a_i$ , and  $\mu(a_i, p_i) = 1$ . What changes in the two-way case is that there may also be other charges supported on these streets. Nevertheless, the outgoing degeneracies  $\tau_n$  are fully determined in terms of the incoming ones  $\nu_n$ : the general statement is

$$\tau_n = X_{a_n} + \nu_n. \quad (167)$$

(To make sense of this equation, we have to say something about how we continue from one street to another, since  $\tau_n$  and  $\nu_n$  are defined on different streets. Unlike the case of the joint above, we cannot simply continue from both streets to the branch point, because there is monodromy there; we have to specify which way we go around the branch point. We follow the short path around, traversing an angle  $2\pi/3$  rather than  $4\pi/3$ .) If we take all  $\nu_n = 0$  then we simply recover

$$\tau_n = X_{a_n} \quad (168)$$

which we found in Sect. 5.3.

## A.2. Joint Rules for Soliton Sets

In the last subsection, we discussed the rules for the soliton degeneracies  $\mu$  at a joint. These rules can be determined by the constraint of homotopy invariance. However, the strange-looking series (163) seems to be crying out for some more geometric interpretation. Here we provide one. As a bonus, this interpretation also motivates a natural set of rules for soliton sets (not only soliton degeneracies); these rules were included in Sect. 9.3 as part of our definition of the soliton content of a general spectral network.

We consider a joint of the type shown in Fig. 46 which arises in some spectral network  $\mathcal{W}_{\vartheta_c}$ . In this picture, by assumption, each of the three  $\mathcal{S}$ -walls supports charges oriented in *both* directions, e.g. the vertically oriented wall supports charges  $\bar{a} \in \Gamma_{jk}$  and  $\bar{a}' \in \Gamma_{kj}$ . Concatenating these two gives a closed cycle,  $\text{cl}(\bar{a} + \bar{a}') \in \Gamma$ , with  $e^{-i\vartheta_c} Z_{\text{cl}(\bar{a} + \bar{a}')} \in \mathbb{R}_-$ . Now, if the situation is generic enough, the semilattice of  $\gamma \in \Gamma$  with the property  $e^{-i\vartheta_c} Z_\gamma \in \mathbb{R}_-$  is generated by a single element; let us assume we are in that situation, and let  $\gamma$  denote the generator. Then without loss of generality the charges supported on the vertical wall can be parameterized as  $\{\bar{a}, \bar{a} + \gamma, \bar{a} + 2\gamma, \dots\} \cup \{\bar{a}', \bar{a}' + \gamma, \bar{a}' + 2\gamma, \dots\}$ , where  $\text{cl}(\bar{a} + \bar{a}') = \gamma$ . Similarly for the other two walls, we replace  $\bar{a}$  with  $\bar{b} \in \Gamma_{ji}$  and  $\bar{c} \in \Gamma_{ik}$ , and  $\bar{a}'$  with  $\bar{b}' \in \Gamma_{ij}$  and  $\bar{c}' \in \Gamma_{ki}$ . Both  $\text{cl}(\bar{a} + \bar{b} + \bar{c})$  and  $\text{cl}(\bar{a}' + \bar{b}' + \bar{c}')$  are positive multiples of  $\gamma$ , from which it follows that one of them is  $\gamma$  and the other is  $2\gamma$ . Again without loss of generality, let us assume  $\text{cl}(\bar{a} + \bar{b} + \bar{c}) = \gamma$  (Fig. 47).

Now, suppose we perturb  $\vartheta$  away from  $\vartheta_c$  slightly, in the positive direction. After the perturbation, these charges are no longer all supported on three walls: rather, each of the three breaks into an infinite set of walls, each supporting a single charge. We thus obtain a further “resolution” of the spectral network in a small neighborhood of the joint, pictured in Fig. 48.

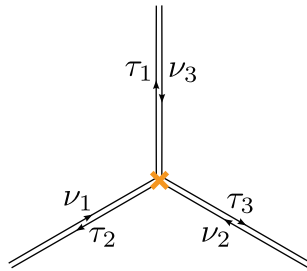


FIGURE 47. Three two-way streets around a branch point

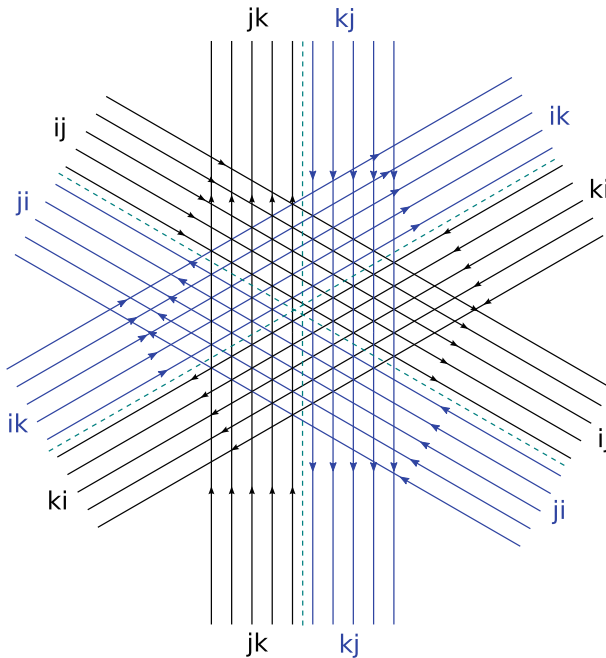


FIGURE 48. The further resolution of the joint of Fig. 46, replacing each two-way street by infinitely many parallel one-way ones. (For clarity we only indicate five one-way streets in each direction.) Note that there is a vertical “highway divider” separating the  $kj$  and  $jk$  walls, and similarly for the other two pairs. Accordingly, the plane is separated into 6 regions, resembling the Weyl chambers of  $su(3)$

Following the soliton lines in this picture, we arrive at a natural interpretation of the solitonic traffic rule illustrated in Figures 41, 42. Indeed, the “highway divides” pictured in Fig. 48 partition a neighborhood of the joint into six chambers, and in each chamber there is a unique way in which two solitons can merge to produce a third.

Now we are ready to explain the series (163). The first few terms correspond to the pictures in Fig. 42. For example, the first term  $\nu_1$  corresponds to a single soliton which just travels through the vicinity of the joint unmolested, while the second term  $\nu_2\nu_6$  corresponds to a pair of incoming solitons which merge to produce an outgoing soliton. To continue the series we build up longer and longer soliton trajectories by concatenating successive mergings. At each step this amounts to replacing a factor  $\nu_n$  by a pair of factors  $\nu_{n+1}\nu_{n-1}$  or  $\nu_{n-1}\nu_{n+1}$ , alternating between these possibilities. Each factor successively replaced is (again alternately) one of the two factors inserted at the previous step. Thus, we begin with  $\nu_1$  and then replace  $\nu_1 \rightarrow \nu_6\nu_2$ ; then in the product  $\nu_6\nu_2$  we replace  $\nu_2 \rightarrow \nu_3\nu_1$ ; then in the product  $\nu_3\nu_1$  we replace  $\nu_3 \rightarrow \nu_2\nu_4$ ; then in the product  $\nu_2\nu_4$  we replace  $\nu_4 \rightarrow \nu_5\nu_3$ ; and so on. We recognize the seventh term as  $(\nu_6\nu_2\nu_4)\nu_1(\nu_5\nu_3\nu_1)$ , and then the process repeats itself beginning with the factor  $\nu_1$  in the middle.

## Appendix B. A Categorical Approach

In this appendix, we introduce some categorical constructions which allow us to summarize some of the main results of the paper in a language which, we hope, some of our mathematically inclined readers will find congenial.

We begin in Sects. B.1 to B.3 with a number of definitions. Then in Sect. B.4, we state a theorem regarding the “formal parallel transport,” or “path-lifting,” and its homotopy properties. Finally, in Sect. B.5, we indicate how one can work with paths on  $C$  and  $\Sigma$ , rather than on  $\tilde{C}$  and  $\tilde{\Sigma}$ , at the price of introducing some tricky signs in the multiplication laws.

### B.1. The Ring of a Category

To any category  $\mathcal{C}$  we may associate a ring  $R(\mathcal{C})$ . As an abelian group,  $R(\mathcal{C})$  is the free group on the space of morphisms:

$$R(\mathcal{C}) = \bigoplus_{f \in \text{Mor}(\mathcal{C})} \mathbb{Z} \cdot \ell_f. \quad (169)$$

The ring structure in  $R(\mathcal{C})$  is defined by

$$\ell_{f_1} \cdot \ell_{f_2} := \begin{cases} 0 & \text{if } f_1 \text{ and } f_2 \text{ are not composable,} \\ \ell_{f_1 f_2} & \text{if } f_1 \text{ and } f_2 \text{ are composable.} \end{cases} \quad (170)$$

Any functor  $F : \mathcal{C} \rightarrow \mathcal{D}$  induces a canonical ring homomorphism  $F : R(\mathcal{C}) \rightarrow R(\mathcal{D})$ .

More generally, we could define a twisted version of  $R(\mathcal{C})$  by writing

$$\ell_{f_1} \cdot \ell_{f_2} = \begin{cases} 0 & \text{if } f_1 \text{ and } f_2 \text{ are not composable,} \\ b(f_1, f_2) \ell_{f_1 f_2} & \text{if } f_1 \text{ and } f_2 \text{ are composable,} \end{cases} \quad (171)$$

where  $b(f_1, f_2)$  is  $\mathbb{Z}_2$ -valued. Associativity requires that  $b$  is a cocycle. A change of basis changes it by a coboundary. We denote the cohomology class of  $b$  by  $\sigma$  and the corresponding twisted ring (up to isomorphism) by  $R(\mathcal{C}, \sigma)$ .



## B.2. Categories of Paths

Let  $X$  be any topological space. By a *path in  $X$*  we mean a continuous map  $\varphi : [T_1, T_2] \rightarrow X$  for some interval  $[T_1, T_2] \subset \mathbb{R}$ . We define the *path groupoid*  $\mathcal{P}(X)$  as follows: the objects of  $\mathcal{P}(X)$  are points of  $X$ , and the morphism space  $\mathcal{P}(X)(x_1, x_2)$  is the set of all paths from  $x_1$  to  $x_2$ , with the obvious composition.

By taking a quotient on the morphism spaces in  $\mathcal{P}(X)$  we can reduce to the *fundamental groupoid*  $\pi_{\leq 1}(X)$ . This is the groupoid whose objects are points of  $X$  and morphism spaces  $\pi_{\leq 1}(X)(x_1, x_2)$  are homotopy classes of paths from  $x_1$  to  $x_2$ .

In Sects. B.4 and B.5, we will also make use of a further “quotient” of the fundamental groupoid, which we call the *first homology groupoid* and denote  $H_{\leq 1}(X)$ . It is a groupoid whose objects are again points of  $X$ . The morphism space between two points  $x_1, x_2$  in  $X$ , denoted  $H_{\leq 1}(x_1, x_2)$ , is defined as follows. First let  $C_1(x_1, x_2)$  be the set of all 1-chains  $\mathbf{c}$  in  $X$  with  $\partial \mathbf{c} = x_2 - x_1$ . Note that if  $x_1 = x_2$  these are simply 1-cycles. In general  $C_1(x_1, x_2)$  is a torsor for the group of 1-cycles. We identify

$$H_{\leq 1}(x_1, x_2) := C_1(x_1, x_2) / \sim \quad (172)$$

where  $\mathbf{c}_1 \sim \mathbf{c}_2$  if  $\mathbf{c}_1 - \mathbf{c}_2$  is a 1-boundary. The morphism space  $H_{\leq 1}(X)(x_1, x_2)$  is an affine subspace of the relative homology  $H_1(X, \{x_1, x_2\}; \mathbb{Z})$ , and is a torsor for  $H_1(X; \mathbb{Z})$ . Composition of morphisms is induced by addition of chains. The automorphism group of any object is canonically  $H_1(X; \mathbb{Z})$ . Note that there are natural functors  $\mathcal{P}(X) \rightarrow \pi_{\leq 1}(X) \rightarrow H_{\leq 1}(X)$ .

## B.3. The Winding Ideal

Let  $\pi : \tilde{X} \rightarrow X$  be a circle bundle. There is an exact sequence

$$\pi_1(S^1) \xrightarrow{\iota} \pi_1(\tilde{X}) \xrightarrow{\pi_*} \pi_1(X) \rightarrow \pi_0(S^1). \quad (173)$$

Choosing the standard generator  $L$  of  $\pi_1(S^1)$ , let  $W$  be the group generated by  $\iota(L)$ . Then (173) induces an exact sequence

$$1 \rightarrow W \rightarrow \pi_1(\tilde{X}) \xrightarrow{\pi_*} \pi_1(X) \rightarrow 1. \quad (174)$$

So any class  $[\varphi] \in \pi_1(\tilde{X})$  with  $\pi_*([\varphi]) = 1$  has a natural  $W$ -valued “winding”  $w([\varphi])$ .

In the case where  $X = S$  is a surface and  $\tilde{X} = \tilde{S}$  is its bundle of tangent directions,  $W$  is either  $\mathbb{Z}$  (if  $X$  is punctured) or  $\mathbb{Z}/\chi(X)\mathbb{Z}$  (if  $X$  is unpunctured). In either case, there is a unique nontrivial map  $W \rightarrow \mathbb{Z}/2\mathbb{Z}$ . Applying this map to  $w([\varphi])$  we obtain a  $\mathbb{Z}_2$ -valued “winding” which by abuse of notation we also call  $w([\varphi])$ . Given a pair of classes  $[\varphi_1], [\varphi_2] \in \pi_{\leq 1}(\tilde{X})(x_1, x_2)$  with  $\pi_*([\varphi_1 \varphi_2^{-1}]) = 1$ , we define the “relative winding”  $w([\varphi_1], [\varphi_2]) = w([\varphi_1 \varphi_2^{-1}])$ . Now let the *winding ideal*  $I \subset R(\pi_{\leq 1}(\tilde{X}))$  be the ideal generated by  $[\varphi_1] - (-1)^{w([\varphi_1, \varphi_2])} [\varphi_2]$  for all pairs  $[\varphi_1], [\varphi_2] \in \pi_{\leq 1}(\tilde{X})(x_1, x_2)$  such that  $\pi_*([\varphi_1 \circ \varphi_2^{-1}]) = 1$ . This construction also descends to homology, giving an ideal  $I \subset R(H_{\leq 1}(\tilde{X}))$ .

### B.4. Formal Parallel Transport For the Path Groupoid

Now let  $\Sigma \rightarrow C$  be a  $K$ -fold branched cover of an oriented surface with at least one puncture, and let  $\mathcal{W}$  be a spectral network subordinate to that cover, as defined in Sect. 9.1. The construction of Sect. 9.4 defines a notion of *formal parallel transport*, or *path lifting*, along paths in  $C''$ , where  $C''$  is  $C$  with all joints of  $\mathcal{W}$  and branch points of  $\Sigma \rightarrow C$  removed. The formal parallel transport is a ring homomorphism

$$\mathbf{F}(\cdot, \mathcal{W}) : R(\mathcal{P}(\tilde{C}'')_{\text{off}}) \rightarrow R(\mathcal{P}(\tilde{\Sigma})). \quad (175)$$

Here  $\tilde{C}''$  and  $\tilde{\Sigma}$  denote the circle bundles of tangent directions as usual, and  $\mathcal{P}(\tilde{C}'')_{\text{off}}$  is the subcategory whose objects do not lie on  $\tilde{\mathcal{W}}$ .

The formal parallel transport is closely related to flat connections, as we saw in Sect. 10.1. The hallmark of a flat connection is that its parallel transport depends only on the *homotopy class* of the path. In order to see the homotopy invariance we must project  $R(\mathcal{P}(\tilde{\Sigma})) \rightarrow R(\pi_{\leq 1}(\tilde{\Sigma})) \rightarrow R(\pi_{\leq 1}(\tilde{\Sigma}))/I$ : after doing so, and only after doing so,  $\mathbf{F}(\varphi, \mathcal{W})$  depends only on the homotopy class of  $\varphi$  in  $\tilde{C}''$ . Moreover, a check of several special cases (equivalent to the computations of Sect. 5.6) shows that in fact  $\mathbf{F}(\varphi, \mathcal{W})$  only depends on the homotopy class in  $\tilde{C}$ . In establishing this, it is crucial that we divide the codomain by the winding ideal  $I$ , which enables various homotopically distinct paths in  $\mathbf{F}(\varphi, \mathcal{W})$  to cancel one another. Hence one can summarize the result of Sect. 5.6 as

**Theorem:**  $\mathbf{F}(\cdot, \mathcal{W})$  descends to a homomorphism

$$R(\pi_{\leq 1}(\tilde{C})_{\text{off}}) \rightarrow R(\pi_{\leq 1}(\tilde{\Sigma}))/I. \quad (176)$$

Moreover, the homomorphism (176) actually factors through a homomorphism

$$R(\pi_{\leq 1}(\tilde{C})_{\text{off}})/I \rightarrow R(\pi_{\leq 1}(\tilde{\Sigma}))/I. \quad (177)$$

Passing from homotopy to homology in the codomain we obtain a map:

$$F(\cdot, \mathcal{W}) : R(\pi_{\leq 1}(\tilde{C})_{\text{off}})/I \rightarrow R(H_{\leq 1}(\tilde{\Sigma}))/I. \quad (178)$$

Expanding  $F([\varphi], \mathcal{W})$  on a  $\mathbb{Z}$ -basis  $X_a$  for  $R(H_{\leq 1}(\tilde{\Sigma}))/I$ , where  $a$  runs over  $\tilde{\Gamma}(\tilde{z}_1, \tilde{z}_2)$ , the coefficients are the degeneracies  $\underline{\overline{Q}}'(L_\varphi, a)$ .

### B.5. Cocycles

We would now like to relate the formalism of this paper to the formalism using the  $X_{\gamma_{ij}'}$  used in our previous paper [5]. The essential problem here is one of signs. In this section, we explain how one can construct a twisted ring, which allows us to work directly with paths on  $C$  and  $\Sigma$ , at the cost of introducing some subtle cocycles in the multiplication rules. The precise statement is equation (186) below. What we actually prove is the closely related statement (184), but to pass to (186), we rely on a conjecture stated below (185).

In order to define the framed 2d–4d degeneracies in terms of paths on  $C$  rather than on  $\tilde{C}$ , we will need a way of lifting from  $C$  to  $\tilde{C}$ . To do this, we introduce  $\mathcal{I}(C)$ , the *immersion category* of  $C$ . The objects are points on  $C$  together

with tangent directions, i.e. points of  $\tilde{C}$ , and the morphisms are immersions of intervals  $[T_1, T_2]$  into  $C$ . Note that there is no identity morphism so this “category” is a category without identity morphisms, i.e., a *semicategory*. We can view it as a subsemicategory of  $\mathcal{P}(\tilde{C})$ . If we mod out by regular homotopy (i.e. homotopy through immersions), we obtain a semicategory  $\pi_{\leq 1}^{\text{Reg}}(C)$ . We also have an obvious functor  $\pi_{\leq 1}^{\text{Reg}}(C) \rightarrow \mathcal{P}(\tilde{C})$ , hence a ring homomorphism

$$\text{Lift} : R(\pi_{\leq 1}^{\text{Reg}}(C)) \rightarrow R(\pi_{\leq 1}(\tilde{C})). \quad (179)$$

Composing with the projection  $R(\pi_{\leq 1}(\tilde{C})) \rightarrow R(\pi_{\leq 1}(\tilde{C}))/I$  we get

$$R(\pi_{\leq 1}^{\text{Reg}}(C)) \rightarrow R(\pi_{\leq 1}(\tilde{C}))/I. \quad (180)$$

Recall that the objects of  $\pi_{\leq 1}^{\text{Reg}}(C)$  are points of  $\tilde{C}$ . Thus, a point  $z \in C$  with tangent vector corresponds to a point  $\tilde{z} \in \tilde{C}$ . A little closed loop beginning at  $\tilde{z}$  and wrapping around the fiber of  $\tilde{C}$  over  $z$  maps to  $-1_{\tilde{z}}$ .

The statement we are aiming for, (186) below, concerns homotopy classes of paths, rather than regular homotopy classes. Therefore, the first step is to find a homomorphism  $R(\pi_{\leq 1}(C)) \rightarrow R(\pi_{\leq 1}^{\text{Reg}}(C))$ . We could attempt to find such a homomorphism by choosing, for each homotopy class, an immersion in the same homotopy class. In general this will *not* produce anything like a homomorphism into  $R(\pi_{\leq 1}^{\text{Reg}}(C))$ . First of all, composable paths in  $\pi_{\leq 1}(C)$  will, in general, not map to composable immersions in  $\pi_{\leq 1}^{\text{Reg}}(C)$ . Thus, the first thing we should do is choose a nowhere-zero vector field on  $C$  and, for each homotopy class in  $\pi_{\leq 1}(C)$ , choose an immersion whose initial and final tangent vectors match the tangent vector at a point. However, even once we have done this, if we lift two composable homotopy classes  $[\tilde{\varphi}_1]$  and  $[\tilde{\varphi}_2]$  to regular homotopy classes of immersions  $[\iota_1]$  and  $[\iota_2]$  there is no guarantee that the lift  $[\iota_{12}]$  of  $[\tilde{\varphi}_1 \circ \tilde{\varphi}_2]$  is the same as  $[\iota_1 \circ \iota_2]$ . That is,

$$[\iota_{12}] \neq [\iota_1 \circ \iota_2] := [\iota_1] \cdot [\iota_2] \quad (181)$$

so we still do not get a homomorphism. However, if we compose our non-canonical non-homomorphism with the map to the quotient (180) then we *will* get a homomorphism from a twisted ring:

$$R(\pi_{\leq 1}(C), \sigma) \rightarrow \text{Lift} \left( R(\pi_{\leq 1}^{\text{Reg}}(C)) \right) / I \quad (182)$$

The cocycle in question is

$$\sigma([\tilde{\varphi}_1], [\tilde{\varphi}_2]) = (-1)^{w[\iota_1 \circ \iota_2 \circ \iota_{12}^{-1, a}]} \quad (183)$$

where  $\iota_{12}^{-1, a}$  is the *anti-lift* of the inverse immersion  $\iota_{12}^{-1}$ . (Given an immersion, we can lift it in the usual way, but then we can compose with the antipodal map in the fiber. Call that the *anti-lift*. Given an immersion  $\tilde{\varphi}$  in  $C$  with canonical lift  $\varphi$  in  $\tilde{C}$ ,  $\varphi^{-1}$  is the antilift of  $\tilde{\varphi}^{-1}$ .) It would be nice to have a more conceptual and invariant description of this cocycle.

Now let us reconsider the framed 2d–4d BPS degeneracies. We can compose the map (182) with the formal monodromy and finally project to the homology groupoid to get a homomorphism of the form

$$\bar{F} : R(\pi_{\leq 1}(C), \sigma) \rightarrow R(H_{\leq 1}(\tilde{\Sigma}))/I \quad (184)$$

The whole discussion above for the curve  $C$  can be repeated for  $\Sigma$  to construct a map

$$R(H_{\leq 1}(\Sigma), \sigma^{\text{ir}}) \rightarrow R(H_{\leq 1}(\tilde{\Sigma}))/I \quad (185)$$

where the cocycle  $\sigma^{\text{ir}}$  is analogous to that appearing in (184). It is natural to conjecture that (184) in fact factors through a homomorphism

$$\bar{F} : R(\pi_{\leq 1}(C), \sigma^{\text{uv}}) \rightarrow R(H_{\leq 1}(\Sigma), \sigma^{\text{ir}}) \quad (186)$$

where  $\sigma^{\text{uv}}$  is the cocycle (183) appearing in (184). This would be a gauge-invariant formulation of the 2d–4d degeneracies directly involving paths on  $C$  and  $\Sigma$ .

## Appendix C. Convention Convention

We summarize here some conventions used in this and our preceding papers.

1. Homology classes on the IR (Seiberg–Witten) curve  $\Sigma$  are denoted in this paper by  $\bar{a} \in \Gamma_{ij'}(z, z')$ . In our previous paper [5], they were denoted by  $\gamma_{ij'}$ . They are represented by paths oriented from  $z^{(i)}$  to  $z'^{(j')}$ . Thus,  $\partial\gamma_{ij'} = z'^{(j')} - z^{(i)}$ . The sum  $\gamma_{ij'} + \gamma_{j'k''} \in \Gamma_{ik''}$  is thus oriented from  $z^{(i)}$  to  $z''^{(k'')}$ . Note that the  $+$  is not commutative: the sum in the opposite order  $\gamma_{j'k''} + \gamma_{ij'}$  in general does not make sense.
2. Oriented open paths on the UV curve  $C$  are generally denoted by  $\wp$ , and often the initial point is  $z_1$  and the final point is  $z_2$ . If  $\wp'$  is another such path, with initial and final points  $z'_1$  and  $z'_2$ , respectively, and  $z_2 = z'_1$  and moreover the canonical lifts of  $\wp$  and  $\wp'$  to  $\tilde{C}$  can be concatenated, then we denote by  $\wp\wp'$  the composed path, which goes from  $z_1$  to  $z'_2$ . Note the “later” path is written to the right.
3.  $Z_\gamma = \frac{1}{\pi} \oint_\gamma \lambda$  is the central charge of a 4d state,  $Z_{\bar{a}} = \frac{1}{\pi} \int_{\bar{a}} \lambda$  is the central charge of a 2d soliton for  $\bar{a} \in \Gamma_{ij}(z, z)$ ,  $i \neq j$ , and the same expression is also the central charge of a framed 2d–4d state with  $\bar{a} \in \Gamma(z_1, z_2)$ . If  $\bar{a} \in \Gamma_{ij}(z_1, z_2)$  we can regard  $Z_{\bar{a}}$  as a function on  $\mathcal{U}_i \times \mathcal{U}_j$ , where  $\mathcal{U}_i$  is a sufficiently small open region around  $z_1^{(i)}$  and  $\mathcal{U}_j$  is a sufficiently small open region around  $z_2^{(j)}$ . In this case  $dZ_{\bar{a}} = \frac{1}{\pi}(\lambda^{(j)} - \lambda^{(i)})$ .
4. For an ordered pair  $(i, j)$  of sheets, BPS walls of type  $ij$  are defined by  $e^{-i\vartheta} Z_{\gamma_{ij}} < 0$  with  $\mu(\gamma_{ij}) \neq 0$ . Therefore, they are oriented on  $C$  so that  $e^{-i\vartheta} \langle \lambda^{(i)} - \lambda^{(j)}, \partial_t \rangle > 0$ , where  $\partial_t$  is a tangent vector to the wall in the direction of the orientation. Note that this implies that the integral  $e^{-i\vartheta} \int_{\gamma_{ij}}^{z(t)} \lambda^{(i)} - \lambda^{(j)} = -e^{-i\vartheta} \int_{\gamma_{ij}} \lambda$  increases as  $z(t)$  moves along the direction of the orientation of the wall.
5. The mass of a soliton with charge  $\bar{a} \in \Gamma_{ij}(z, z)$  is  $M(z) = -\frac{e^{-i\vartheta}}{\pi} \int_{\bar{a}} \lambda$ .
6. Lifting to the circle bundle of directions in the tangent bundle is denoted by a tilde. Thus  $\tilde{\Sigma}$  is the circle bundle of tangent directions over  $\Sigma$ , and so forth.

7.  $F(\wp, \vartheta)$ ,  $\mathbf{F}(\wp, \mathcal{W})$  and  $\mathbf{F}(\tilde{\wp}, \mathcal{W}, \nabla^{\text{ab}})$  all compose in the same way we compose paths in item 2 above:  $F(\wp, \cdot)F(\wp', \cdot) = F(\wp\wp', \cdot)$ .
8. The parallel transport  $\mathbf{F}(\tilde{\wp}, \mathcal{W}, \nabla^{\text{ab}}) \in \text{Hom}(E_{\tilde{z}_1}, E_{\tilde{z}_2})$  if  $\wp$  goes from  $z_1$  to  $z_2$ . In [5]  $\hat{E}$  was denoted by  $V$ ,  $\mathcal{L}_i$  was denoted by  $V_i$ , and  $F(\wp, \mathcal{W}_\vartheta, \nabla^{\text{ab}})$  was denoted  $\langle L_\wp \rangle$  and expanded in  $\mathcal{Y}_{\gamma_{ij'}} \in \text{Hom}(\mathcal{L}_i, \mathcal{L}_{j'})$ . (The conventions of Sect. 5 of [5] are opposite to those of Sect. 7–9 of [5]; we use the conventions of Sect. 7–9.)
9. We compose linear transformations on the *right* so that the composition of linear operators  $T_1 \in \text{Hom}(V_1, V_2)$  and  $T_2 \in \text{Hom}(V_2, V_3)$  is written  $T_1 T_2 \in \text{Hom}(V_1, V_3)$ . The parallel transport of a section  $s(z)$  from  $z_1$  to  $z_2$  by the connection  $\nabla = \Psi_{\mathcal{W}}(\nabla^{\text{ab}})$  is thus given by  $s(z_2) = s(z_1)\mathbf{F}(\tilde{\wp}, \mathcal{W}, \nabla^{\text{ab}})$ . This differs from the more usual convention and in particular differs from [5]. In a local trivialization we can write this as

$$s(z_2) = s(z_1) \text{Pexp} \left( - \int_{z_1}^{z_2} \mathcal{A} \right) \quad (187)$$

where  $\mathcal{A}$  is the  $\mathfrak{gl}(K, \mathbb{C})$ -valued 1-form representing  $\nabla$ .

10. In [5], equations (5.31) and (5.32), we introduced flat sections  $\mathcal{Y}_{\gamma_{ij'}}(z, z')$  of the vector bundle  $\hat{E}^* \otimes \hat{E}$  on  $C \times C$  obtained by projection and parallel transport. For a homology class  $a \in \Gamma(\tilde{z}, \tilde{z}')$  connecting  $\tilde{z}^{(i)}$  and  $\tilde{z}'^{(j')}$  we define  $\mathcal{Y}_a$  to be parallel transport with respect to  $\nabla^{\text{ab}}$  along  $a$  from  $\tilde{z}^{(i)}$  to  $\tilde{z}'^{(j')}$ . The two notions coincide for  $z, z'$  away from  $\mathcal{S}$ -walls if we make use of the isomorphism (131). From either point of view we have, in local coordinates

$$\begin{aligned} \partial_z \mathcal{Y}_{\gamma_{ij'}} - \mathcal{A}_z \mathcal{Y}_{\gamma_{ij'}} &= 0, \\ \partial_{z'} \mathcal{Y}_{\gamma_{ij'}} + \mathcal{Y}_{\gamma_{ij'}} \mathcal{A}_{z'} &= 0. \end{aligned} \quad (188)$$

11. The map between solutions  $(\varphi, A)$  of Hitchin equations and flat connections  $\mathcal{A}$  (used in [17, 3, 21, 5] although not in this paper) is  $\mathcal{A} = \frac{R}{\zeta} \varphi + A + R\zeta \bar{\varphi}$ .
12. The WKB asymptotics of flat sections are formally given by

$$\mathcal{Y}_{\gamma_{ij}} \sim \exp \frac{\pi R}{\zeta} Z_{\gamma_{ij}} + \cdots \quad (189)$$

This implies that if  $\zeta$  lies in the half-plane  $\mathbb{H}_\vartheta$ , then  $\mathcal{Y}_{\gamma_{ij}}(z)$  is exponentially *small* as  $\zeta \rightarrow 0$  when  $z$  lies on a wall of type  $ij$  supporting the charge  $\gamma_{ij}$ .

## References

- [1] Witten, E.: Solutions of four-dimensional field theories via M-theory. Nucl. Phys. **B500**, 3–42 (1997). <http://www.arXiv.org/abs/hep-th/9703166>
- [2] Gaiotto, D.:  $\mathcal{N} = 2$  dualities. <http://www.arXiv.org/abs/0904.2715>

- [3] Gaiotto, D., Moore, G.W., Neitzke, A.: Wall-crossing, Hitchin Systems, and the WKB Approximation. <http://www.arXiv.org/abs/0907.3987>
- [4] Gaiotto, D.: Surface Operators in  $\mathcal{N} = 2$  4d Gauge Theories. <http://www.arXiv.org/abs/0911.1316>
- [5] Gaiotto, D., Moore, G.W., Neitzke, A.: Wall-Crossing in Coupled 2d–4d Systems. <http://www.arXiv.org/abs/1103.2598v1>
- [6] Klemm, A., Lerche, W., Mayr, P., Vafa, C., Warner, N.P.: Self-dual strings and  $\mathcal{N} = 2$  supersymmetric field theory. Nucl. Phys. **B477**, 746–766 (1996). <http://www.arXiv.org/abs/hep-th/9604034>
- [7] Brandhuber, A., Stieberger, S.: Self-dual strings and stability of BPS states in  $\mathcal{N} = 2SU(2)$  gauge theories. Nucl. Phys. **B488**, 199–222 (1997). <http://www.arXiv.org/abs/hep-th/9610053>
- [8] Aharony, O., Hanany, A., Kol, B.: Webs of  $(p, q)$  5-branes, five dimensional field theories and grid diagrams. JHEP **01**, 002 (1998). <http://www.arXiv.org/abs/hep-th/9710116>
- [9] Kontsevich, M., Soibelman, Y.: Stability structures, motivic Donaldson–Thomas invariants and cluster transformations. <http://www.arXiv.org/abs/0811.2435>
- [10] Alim, M., Cecotti, S., Cordova, C., Espahbodi, S., Rastogi, A., Vafa, C.: BPS Quivers and Spectra of Complete  $\mathcal{N} = 2$  Quantum Field Theories. <http://www.arXiv.org/abs/1109.4941>
- [11] Alim, M., Cecotti, S., Cordova, C., Espahbodi, S., Rastogi, A., Vafa, C.:  $\mathcal{N} = 2$  Quantum Field Theories and Their BPS Quivers. <http://www.arXiv.org/abs/1112.3984>
- [12] Fock, V.V., Goncharov, A.B.: Cluster ensembles, quantization and the dilogarithm. Ann. Sci. Éc. Norm. Supér. (4) **42**(6), 865–930 (2009). <http://www.arXiv.org/abs/math/0311245>
- [13] Fock, V., Goncharov, A.: Moduli spaces of local systems and higher Teichmüller theory. Publ. Math. Inst. Hautes Études Sci. **103**, 1–211 (2006). <http://www.arXiv.org/abs/math/0311149>
- [14] Fock, V.V., Goncharov, A.B.: The quantum dilogarithm and representations of quantum cluster varieties. <http://www.arXiv.org/abs/math/0702397>
- [15] Goncharov, A.: To appear
- [16] Hitchin, N.J.: The self-duality equations on a Riemann surface. Proc. Lond. Math. Soc. (3) **55**(1), 59–126 (1987)
- [17] Gaiotto, D., Moore, G.W., Neitzke, A.: Four-dimensional wall-crossing via three-dimensional field theory. Commun. Math. Phys. **299**, 163–224 (2010). <http://www.arXiv.org/abs/0807.4723>
- [18] Delabaere, É., Dillinger, H., Pham, F.: Résurgence de Voros et périodes des courbes hyperelliptiques. Ann. Inst. Fourier (Grenoble) **43**(1), 163–199 (1993)
- [19] Voros, A.: The return of the quartic oscillator: the complex WKB method. Ann. Inst. H. Poincaré Sect. A (N.S.) **39**(3), 211–338 (1983)
- [20] Aoki, T., Kawai, T., Sasaki, S., Shudo, A., Takei, Y.: Virtual turning points and bifurcation of Stokes curves for higher order ordinary differential equations. J. Phys. A **38**(15), 3317–3336 (2005). <http://www.arXiv.org/abs/math-ph/0409005>

- [21] Gaiotto, D., Moore, G.W., Neitzke, A.: Framed BPS States. <http://www.arXiv.org/abs/1006.0146v1>
- [22] Zorich, A.: Flat surfaces. <http://www.arXiv.org/abs/math/0609392>
- [23] Kontsevich, M., Soibelman, Y.: Affine structures and non-Archimedean analytic spaces. In: The unity of mathematics. Progr. Math., vol. 244, pp. 321–385. Birkhäuser, Boston (2006). <http://www.arXiv.org/abs/math.ag/0406564>
- [24] Auroux, D.: Mirror symmetry and  $T$ -duality in the complement of an anticanonical divisor. J. Gökova Geom. Topol. GGT **1**, 51–91 (2007)
- [25] Gross, M., Siebert, B.: From real affine geometry to complex geometry. <http://www.arXiv.org/abs/math/0703822>
- [26] Gross, M., Pandharipande, R., Siebert, B.: The tropical vertex. <http://www.arXiv.org/abs/0902.0779>
- [27] Gross, M., Hacking, P., Keel, S.: Mirror symmetry for log Calabi–Yau surfaces I. <http://www.arXiv.org/abs/1106.4977>
- [28] Fraser, C., Hollowood, T.J.: On the weak coupling spectrum of  $N = 2$  supersymmetric  $SU(n)$  gauge theory. Nucl. Phys. **B490**, 217–238 (1997). <http://www.arXiv.org/abs/hep-th/9610142>
- [29] Taylor, B.J.: On the strong coupling spectrum of pure  $SU(3)$  Seiberg–Witten theory. JHEP **0108**, 031 (2001). <http://www.arXiv.org/abs/hep-th/0107016>
- [30] Taylor, B.J.: On the moduli space of  $SU(3)$  Seiberg–Witten theory with matter. JHEP **0212** 040 (2002). <http://www.arXiv.org/abs/hep-th/0211086>
- [31] Chen, H.-Y., Dorey, N., Petunin, K.: Moduli space and wall-crossing formulae in higher-rank gauge theories. JHEP **1111**, 020 (2011). <http://www.arXiv.org/abs/1105.4584>
- [32] Gaiotto, D., Moore, G.W., Neitzke, A.: Spectral networks and snakes (2013, to appear)
- [33] Alday, L.F., Gaiotto, D., Gukov, S., Tachikawa, Y., Verlinde, H.: Loop and surface operators in  $N = 2$  gauge theory and Liouville modular geometry. JHEP **1001** (2010). <http://www.arXiv.org/abs/0909.0945>
- [34] Dimofte, T., Gukov, S., Soibelman, Y.: Quantum Wall Crossing in  $N = 2$  Gauge Theories. <http://www.arXiv.org/abs/0912.1346>
- [35] Cecotti, S., Vafa, C.: BPS Wall Crossing and Topological Strings. <http://www.arXiv.org/abs/0910.2615>
- [36] Cecotti, S., Vafa, C.: 2d Wall-Crossing, R-twisting, and a Supersymmetric Index. <http://www.arXiv.org/abs/1002.3638>
- [37] Cecotti, S., Fendley, P., Intriligator, K.A., Vafa, C.: A new supersymmetric index. Nucl. Phys. **B386**, 405–452 (1992). <http://www.arXiv.org/abs/hep-th/9204102>
- [38] Cecotti, S., Vafa, C.: On classification of  $\mathcal{N} = 2$  supersymmetric theories. Commun. Math. Phys. **158**, 569–644 (1993). <http://www.arXiv.org/abs/hep-th/9211097>
- [39] Gaiotto, D., Witten, E.: Knot invariants from four-dimensional gauge theory. <http://www.arXiv.org/abs/1106.4789>
- [40] Denef, F., Moore, G.W.: Split states, entropy enigmas, holes and halos. JHEP **1111** (2011), <http://www.arXiv.org/abs/hep-th/0702146>

- [41] Andriyash, E., Denef, F., Jafferis, D.L., Moore, G.W.: Bound state transformation walls. <http://www.arXiv.org/abs/1008.3555>
- [42] Andriyash, E., Denef, F., Jafferis, D.L., Moore, G.W.: Wall-crossing from supersymmetric galaxies. JHEP **1201** (2012). <http://www.arXiv.org/abs/1008.0030>
- [43] Gaiotto, D., Moore, G.W., Neitzke, A.: Spectral network movies. <http://www.ma.utexas.edu/users/neitzke/spectral-network-movies/>
- [44] Strebel, K.: Quadratic differentials. *Ergebnisse der Mathematik und ihrer Grenzgebiete (3)*, vol. 5 [Results in Mathematics and Related Areas(3)]. Springer, Berlin (1984)
- [45] Seiberg, N., Witten, E.: Electric-magnetic duality, monopole condensation, and confinement in  $\mathcal{N} = 2$  supersymmetric Yang–Mills theory. Nucl. Phys. **B426**, 19–52 (1994). <http://www.arXiv.org/abs/hep-th/9407087>
- [46] Galakhov, D., Longhi, P., Mainiero, T., Moore, G. W., Neitzke, A.: To appear (2013)
- [47] Shapere, A.D., Vafa, C. BPS structure of Argyres–Douglas superconformal theories. <http://www.arXiv.org/abs/hep-th/9910182>
- [48] Cecotti, S., Neitzke, A., Vafa, C.: R-Twisting and 4d/2d Correspondences. <http://www.arXiv.org/abs/1006.3435>
- [49] Donagi, R.: Spectral covers. In: Current topics in complex algebraic geometry (Berkeley, CA, 1992/93). Math. Sci. Res. Inst. Publ., vol. 28, pp. 65–86. Cambridge Univ. Press, Cambridge (1995)
- [50] Donagi, R., Gaiotto, D.: The gerbe of Higgs bundles. <http://www.arXiv.org/abs/math/0005132>
- [51] Atiyah, M.F., Bott, R.: The Yang–Mills equations over Riemann surfaces. Philos. Trans. R. Soc. Lond. Ser. A **308**(1505), 523–615 (1983)
- [52] Boalch, P.P.: Symplectic Geometry and Isomonodromic Deformations. PhD thesis, Wadham College, Oxford (1999)
- [53] Alexandrov, S., Persson, D., Pioline, B.: Wall-crossing, Rogers dilogarithm, and the QK/HK correspondence. JHEP **1112** (2011). <http://www.arXiv.org/abs/1110.0466>
- [54] Cecotti, S., Del Zotto, M.: On Arnold’s 14 ‘exceptional’  $\mathcal{N} = 2$  superconformal gauge theories. JHEP **1110**, 099 (2011). <http://www.arXiv.org/abs/1107.5747>
- [55] Atiyah, M.F.: The geometry and physics of knots. *Lezioni Lincee*. [Lincei Lectures]. Cambridge University Press, Cambridge (1990)
- [56] Corlette, K.: Flat  $G$ -bundles with canonical metrics. J. Differ. Geom. **28**(3), 361–382 (1988)
- [57] Donaldson, S.K.: Twisted harmonic maps and the self-duality equations. Proc. Lond. Math. Soc. (3) **55**(1), 127–131 (1987)
- [58] Simpson, C.: Harmonic bundles on noncompact curves. J. Am. Math. Soc. **3**, 713–770 (1990)
- [59] Biquard, O., Boalch, P.: Wild nonabelian Hodge theory on curves. <http://www.arXiv.org/abs/math/0111098>
- [60] Bender, C.M., Wu, T.T.: Anharmonic oscillator. Phys. Rev. **184**, 1231–1260 (1969)
- [61] Zinn-Justin, J.: Instantons in quantum mechanics: numerical evidence for a conjecture. J. Math. Phys. **25**, 549 (1984)



- [62] Berk, H.L., Nevins, W.M., Roberts, K.V.: New Stokes' line in WKB theory. J. Math. Phys. **23**(6), 988–1002 (1982)
- [63] Simpson, C.: Asymptotics for general connections at infinity. <http://www.arXiv.org/abs/math/0311531>
- [64] Simpson, C.: Asymptotic behavior of monodromy. In: Lecture Notes in Mathematics, vol. 1502. Springer, Berlin (1991)

Davide Gaiotto  
School of Natural Sciences  
Institute for Advanced Study  
Princeton, NJ 08540, USA  
e-mail: [dgaiotto@ias.edu](mailto:dgaiotto@ias.edu)

Gregory W. Moore  
NHETC and Department of Physics and Astronomy  
Rutgers University  
Piscataway, NJ 08855–0849, USA  
e-mail: [gmoore@physics.rutgers.edu](mailto:gmoore@physics.rutgers.edu)

Andrew Neitzke  
Department of Mathematics  
University of Texas at Austin  
Austin, TX 78712, USA  
e-mail: [neitzke@math.utexas.edu](mailto:neitzke@math.utexas.edu)

Communicated by Marcos Marino.

Received: October 9, 2012.

Accepted: December 7, 2012.