Seiberg–Witten: Monodromies, Confinement, and Elliptic Curves

Let me start by repeating some material from previous lectures about the $SU(2)$ gauge theory broken down to $U(1)$ by the VEV of a triplet Higgs. For generic values of the modulus $U = \mathrm{tr}(\Phi^2)$, all charged particles are massive, and the only massive particles are the $U(1)$ photon, the U modulus scalar, and their superpartners. The modulus-sependent abelian gauge coupling

$$\tau(U) \; = \; \frac{4\pi i}{e^2} \; + \; \frac{\Theta}{2\pi} \tag{1}$$

of the effective low-energy theory of the U and V superfields is more than just the Wilsonian coupling, it is the physical coupling of the $U(1)$ gauge fields at low momenta. It appears in the Hamiltonian for the $U(1)$ fields as

$$\mathcal{H} \; = \; \frac{|\vec{E}_{\mathrm{can}} + i\vec{B}_{\mathrm{can}}|^2}{2} \; = \; \frac{|\vec{\mathcal{C}} + \tau\vec{\mathcal{B}}|^2}{8\pi \,\mathrm{Im}\,\tau} \tag{2}$$

and to assure that this Hamiltonian density is positive, we absolutely must have

$$\mathrm{Im}\,\tau(U) \; > \; 0 \quad \forall U. \tag{3}$$

Fortunately, this positivity condition is preserved by all the S–dualities. Indeed, let

$$\tau_{\mathrm{dual}} \; = \; \frac{a\tau + b}{c\tau + d} \quad \text{for some} \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbf{Z}). \tag{4}$$

Then

$$\mathrm{Im}\,\tau_{\mathrm{dual}} \; = \; \frac{\mathrm{Im}\big((a\tau+b)(c\tau^*+d)\big)}{|c\tau+d|^2} \; = \; \frac{(ad-bc)\times \mathrm{Im}\,\tau}{|c\tau+d|^2} \; = \; \frac{\mathrm{Im}\,\tau}{|c\tau+d|^2}\,, \tag{5}$$

so it $\mathrm{Im}\,\tau > 0$ then $\mathrm{Im}\,\tau_{\mathrm{dual}} > 0$, and vice verse.

Second, suppose at some point U_0 of the moduli space, some generically massive particles become massless. We saw that is those particles have electric charges $\pm qe$, then $\tau(U)$ has a singularity, and in the vicinity of this singularity it behaves as

$$\tau(U) \;=\; \text{smooth}(U) \;-\; \frac{2iq^2}{2\pi}\log(U-U_0) \tag{6}$$

if the massless particles form chiral supermultiplets, but

$$\tau(U) \;=\; \text{smooth}(U) \;+\; \frac{4iq^2}{2\pi}\log(U-U_0) \tag{7}$$

is the supermultiplets are vector rather then chiral. But while the chiral case (6) is perfectly consistent with $\operatorname{Im}\tau > 0$, the vector case is not; instead, eq. (7) makes $\operatorname{Im}\tau$ negative in some small neighborhood of U_0, and that's not allowed. Consequently, all charged particles that becomes massless at special points of the moduli space must form chiral rather than vector multiplets.

The same rule applies to the magnetically charged particles that become massless at some point U_0. Indeed, using the electric-magnetic duality, we turn their magnetic charges $\mu = \pm p(4\pi/e)$ into electric charges $Q = \pm pe'$ while τ turns into $\tau_{\text{dual}} = -1/\tau$. The EM duality does not change the supermultiplet structures, hence in the vicinity of U_0 we have

$$\tau_{\text{dual}}(U) \;=\; \text{smooth}(U) \;-\; \frac{2ip^2}{2\pi}\log(U-U_0) \tag{8}$$

in the chiral case, and

$$\tau_{\text{dual}}(U) \;=\; \text{smooth}(U) \;+\; \frac{4ip^2}{2\pi}\log(U-U_0) \tag{9}$$

in the vector case. According to eq. (5), the requirement of $\operatorname{Im}\tau(U) > 0$ for all U is equivalent to $\operatorname{Im}\tau_{\text{dual}}(U) > 0$ for all U, and we know that the chiral formula (8) is consistent with this requirement but the vector formula (9) is not.

Finally, suppose some dyons with (magnetic,electric) charges $\pm(p,q)$ become massless at $U = U_0$. For any combination of charges (p,q) there is some S–duality that would make them purely electric,

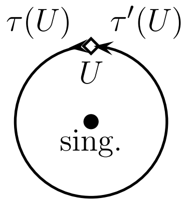
$$\exists M \in SL(2,\mathbf{Z}) \text{ such that } M \begin{pmatrix} q \\ -p \end{pmatrix} = \begin{pmatrix} r \\ 0 \end{pmatrix}, \qquad r = \gcd(p,q). \tag{10}$$

After this duality, we have equations similar to (8) and (9) for the dual coupling $\tau_{\text{dual}}(U)$, the only difference from the purely magnetic case being the relation between the τ_{dual} and τ. However, for any S–duality, the $\tau_{\text{dual}}(U)$ must have positive imaginary part for all U, and that's consistent with the chiral multiplets becoming massless but not the vector multiplets.

The bottom line is, *the particles becoming massless at special points of the moduli space can have electric charges, or magnetic charges, or both, but they should form chiral rather then vector multiplets of the $\mathcal{N}=1$ SUSY.* Similarly, for the $\mathcal{N}=2$ SUSY, the particles becoming massless at a special point should form hyper-multiplets rather then vector multiplets. But for the $\mathcal{N}=4$ SUSY the vector multiplets are OK because their masses do not affect the gauge coupling τ.

Monodromies

The holomorphic coupling τ of the un-broken $U(1)$ gauge symmetry is a multi-valued function of the modulus U. As we move in the moduli space along a closed path enclosing some singularity, the $\tau'(U)$ at the end of the path may be different from the $\tau(U)$ we started from at the same point U:



$$\begin{aligned} \tau'(U) &\neq \tau(U), \\ \text{but } \tau'(U) &\equiv \tau(U) \text{ modulo an S-duality.} \end{aligned} \tag{11}$$

Instead, $\tau'(U)$ should equivalent to $\tau(U)$ modulo an S-duality,

$$\tau'(U) = \frac{a\tau(U)+b}{c\tau(U)+d} \qquad \text{for some } SL(2,\mathbf{Z}) \text{ matrix } M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}. \tag{12}$$

This S-duality transform — which acts not only on the τ but also on all the electric and magnetic charges of the $U(1)$ theory — is called the *monodromy* of the path in question.

Note: the monodromy is a topological, or rather a homotopical property of a path — all paths circling the same singularities the same numbers of times (and in the same order) will have the same monodromy.

Electric example: Suppose at some point U_0 in the moduli space, a pair of electrically charged ($Q = \pm q \times e$) chiral superfields $E^{\pm}$ becomes massless. In the vicinity of U_0 we have

$$\tau(U) \;=\; \mathrm{smooth}(U) \;-\; \frac{2iq^2}{2\pi}\,\log(U - U_0), \tag{6}$$

and when we circle counterclockwise around the singularity U_0, the logarithm $\log(U - U_0)$ changes by $2\pi i$, hence

$$\tau(U) \;\mapsto\; \tau(U) \;+\; 2q^2. \tag{13}$$

This monodromy is described by the $SL(2\mathbf{Z})$ matrix

$$M_{\mathrm{el}}(q) \;=\; \begin{pmatrix} 1 & 2q^2 \\ 0 & 1 \end{pmatrix}, \tag{14}$$

which acts not only on $\tau(U)$ but also on (electric,magnetic) charges (n, m) of all the particles, massless or massive. Specifically, the purely electric charges $(n, 0)$ remain unchanged, but the magnetic monopoles and dyons with $m \neq 0$ change their electric charges n by $2q^2 \times m$.

Magnetic example: Now suppose that a pair of magnetic monopoles with charges $\mu = \pm p(4\pi/e)$ becomes massless for $U = U_0$. In this case we have eq. (8) for the dual coupling $\tau_{\mathrm{dual}}(U)$, which yields monodromy

$$\tau_{\mathrm{dual}}(U) \;\mapsto\; \tau_{\mathrm{dual}}(U) \;+\; 2p^2. \tag{15}$$

Translating this monodromy in terms of $\tau = -1/\tau_{\mathrm{dual}}$, we have

$$\frac{-1}{\tau(U)} \;\mapsto\; \frac{-1}{\tau(U)} \;+\; 2p^2 \quad\Longrightarrow\quad \tau(U) \;\mapsto\; \frac{\tau(U)}{-2p^2 \times \tau(u) + 1}\,, \tag{16}$$

which corresponds to the monodromy matrix

$$M_{\mathrm{mag}}(p) \;=\; \begin{pmatrix} 1 & 0 \\ -2p^2 & 1 \end{pmatrix}. \tag{17}$$

This monodromy leaves the purely magnetic charges $(0, m)$ unaffected, but the electrically charged particles change their magnetic charges by $\Delta m = -2p^2 \times n$ when they circle the singularity.

General dyons: Finally, suppose a pair of dyons with some generic (electric,magnetic) charges $\pm(q,p)$ becomes massless at some point U_0. The monodromy of a path enclosing this point is

$$\tau_{\rm dual}(U) \;\mapsto\; \tau_{\rm dual}(U) \;+\; 2r^2. \tag{18}$$

where $r = \gcd(q,p)$ and

$$\tau_{\rm dual}(U) \;=\; \frac{a\tau(U)+b}{c\tau(U)+d} \tag{19}$$

for the same matrix

$$K \;=\; \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,Z) \tag{20}$$

that turns the dyonic charges (q,p) into purely electric,

$$K\begin{pmatrix} q \\ -p \end{pmatrix} = \begin{pmatrix} r \\ 0 \end{pmatrix}. \tag{21}$$

Translating the monodromy (18) to the language of the original coupling $\tau(U)$ gives us the monodromy matrix

$$M(q,p) \;=\; K^{-1} \times \begin{pmatrix} 1 & 2r^2 \\ 0 & 1 \end{pmatrix} \times K \;=\; \begin{pmatrix} 1+2r^2cd & +2r^2c^2 \\ -2r^2d^2 & 1-2r^2cd \end{pmatrix}. \tag{22}$$

The second equality here uses the explicit form (20) of the duality matrix K. Note that only two of K's matrix elements are relevant in eq. (22), and we may easily find those elements from the matrix equation (21). Specifically, we have

$$cq \,-\, dp \;=\; 0 \quad\Longrightarrow\quad c \,=\, \frac{p}{r}\,, \quad d \;=\; \frac{q}{r}\,, \tag{23}$$

hence

$$M(q,p) \;=\; \begin{pmatrix} 1+2pq & +2q^2 \\ -2p^2 & 1-2pq \end{pmatrix} \tag{24}$$

Note that for any (q,p) charges of the massless dyons, the monodromy matrices have two curious properties. First, $M(q,p) \equiv 1_{2\times2} \pmod 2$ — the diagonal matrix elements are odd while the off-diagonal elements are even. This property defines a subgroup of $SL(2,\mathbf{Z})$ called Γ_2. The reason all monodromies are restricted to this subgroup is that out model has only integer electric charges (in units of $e = g$). Allowing for half-integer charges (coming from half-integer isospin multiplets of $SU(2)$) would produce monodromies generating the entire $SL(2,\mathbf{Z})$ S–duality group.

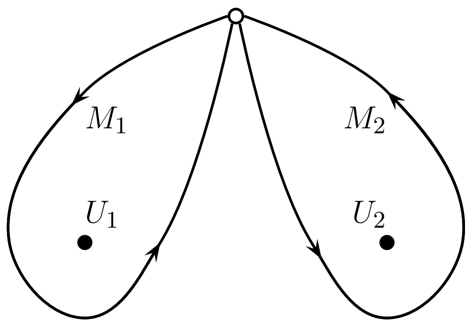
Second, all matrices (24) have $\mathrm{tr}(M) = 2$. This property is not preserved by matrix multiplication, so it does not restrict us to a subgroup. Instead, it's a characteristic feature of monodromies of path inside which only a single type of (q,p) charges become massless. If a path encloses multiple singularities with different types of massless charges, the monodromy of this path generally has $\mathrm{tr}(M) \neq 2$.

Group theory of monodromies.

Closed paths in the moduli space — considered modulo deformations which do not cross any singularities — form the *fundamental group* of the moduli space. The monodromies provide a map from this group into the S–duality group $SL(2,\mathbf{Z})$, and this map preserves the group law: The monodromy of a combined path P_2P_1 — which first follows one closed path P_1 and then another closed path P_2 — is a matrix product

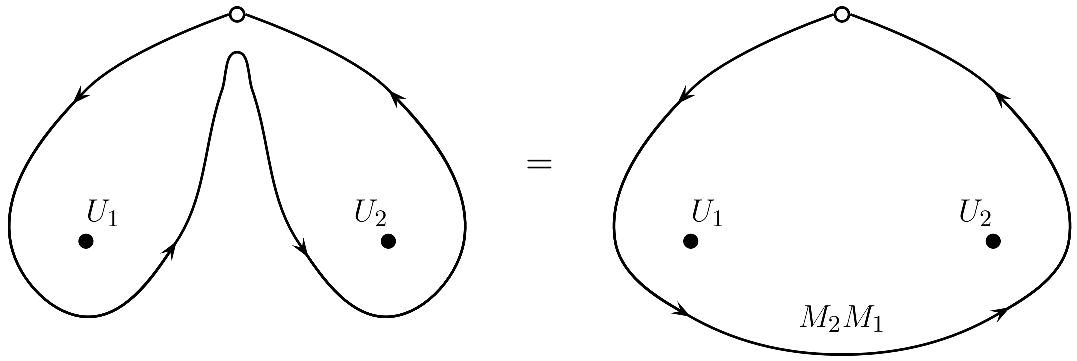
$$M(P_2P_1) \;=\; M(P_2)\times M(P_1). \tag{25}$$

For example, suppose one path circles a singularity at U_1 and has monodromy M_1 while another path circles a different singularity U_2 and has monodromy M_2,



(26)

Then the combined path — which circles first U_1 and then U_2 —



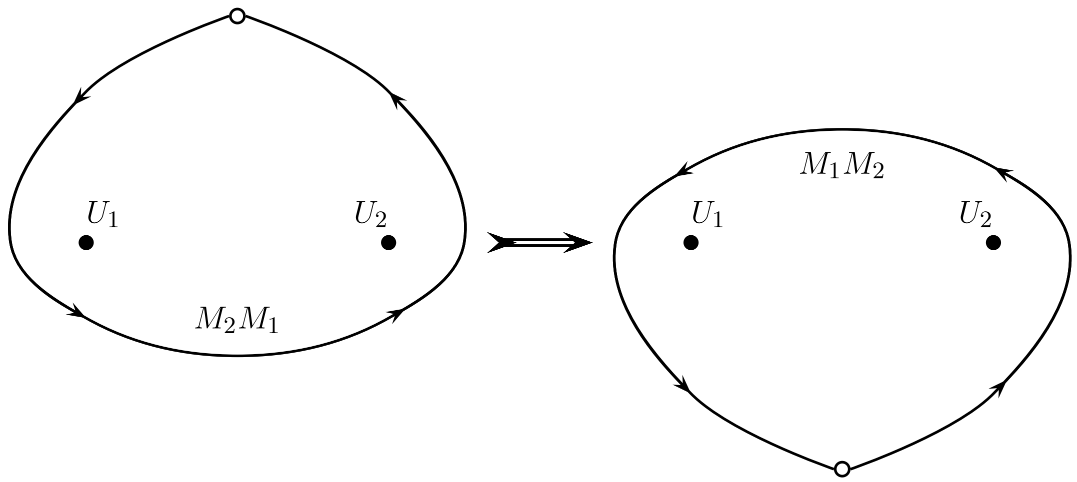
(27)

has monodromy M_2M_1.

The small open circle in figures (26) and (27) denotes the reference point where all closed loops begin and end. If we move this point, the monodromies may suffer a similarity transform,

$$M(P) \;\mapsto\; U^{-1} \times M(P) \times U \tag{28}$$

for some common $U \in SL(2,\mathbf{Z})$ for all paths P. For example, suppose we move down the reference point on figure (27):



(29)

The new path also circles both singularities U_1 and U_2 but in the opposite order — first U_2 then U_1 — so its monodromy is M_1M_2 instead of M_2M_1. The monodromy matrices (24) generally do not commute, so $M_1M_2 \neq M_2M_1$, but the two matrix products are always similar. Indeed, let $U = M_2^{-1}$, then

$$U^{-1}(M_1M_2)U \;=\; M_2M_1M_2M_2^{-1} \;=\; M_2M_1\,. \tag{30}$$

Monodromies of the Seiberg–Witten model:

We saw earlier in class that the Seiberg–Witten model has two singularities in the U–modulus plane, at $U_1 = -\Lambda^2$ and $U_2 = +\Lambda^2$. There are no singularities at any other *finite* values of U, but for $U \to \infty$ we have

$$\tau(U) \;=\; \frac{i}{2\pi}\log\frac{U^2}{\Lambda^4} \;+\; \sum_{n=1}^{\infty} C_n\left(\frac{\Lambda^4}{U^2}\right)^n \tag{31}$$

As we circle the infinity — *i.e.*, keep $|U|$ fixed and large, $|U| \gg |\Lambda|^2$ and change phase(U) by $+2\pi$, the instanton terms in eq. (31) come back to themselves, but the logarithm in the one-loop term changes by $+4\pi i$, which leads to $\tau \mapsto \tau - 2$ and hence monodromy matrix

$$M_\infty \;=\; \pm\begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}. \tag{32}$$

Note that the overall sign of this matrix affects the electric and magnetic charges of all the particles, but it does not affect the coupling τ, that's why we have left it open for a moment.

To determine the sign, consider the Higgs VEV

$$\langle\Phi^a\rangle \;=\; \bigl(0, 0, \sqrt{U}\bigr). \tag{33}$$

As the U modulus circles the ∞ once, this VEV changes sign — which is equivalent to a 180° isospin rotation that reverses the direction of the un-broken $U(1)$ inside the $SU(2)$. In terms of this reversed $U(1)$, the abelian gauge fields $\vec{E}$ and $\vec{B}$ change signs, and this requires all the electric and magnetic charges to change their signs too. This charge reversal corresponds to the '$-$' sign in eq. (32), thus the correct monodromy matrix for circling the infinity is

$$M_\infty = \begin{pmatrix} -1 & +2 \\ 0 & -1 \end{pmatrix}. \tag{34}$$

Topologically, a path circling the ∞ — *i.e.*, a very large counterclockwise circle — is equivalent to the path (27) circling both singularities at $U = \mp\Lambda^2$. Consequently, both path should have the same monodromies,

$$M_2 \times M_1 = M_\infty . \tag{35}$$

This equation allows us to determine the electric and magnetic charges of the particles that become massless at the singularities $U_{1,2} = \mp\Lambda^2$. Using Occam's razor we assume that there is only one pair of massless charges $\pm(q_i, p_i)$ at each singularity U_i $(i = 1, 2)$, so the monodromies M_i have form (24),

$$M_i = M(q_i, p_i) = \begin{pmatrix} 1 + 2p_i q_i & +2q_i^2 \\ -2p_i^2 & 1 - 2p_i q_i \end{pmatrix} \tag{36}$$

All we need to do is to plug these matrices into eq. (35) and solve for the charges (q_1, p_1) and (q_2, p_2).

Let me skip the algebra and simply state the result: All solutions have charges

$$p_1 = p_2 = 1, \quad q_2 = q_1 + 1, \quad \text{any } q_1 \in \mathbf{Z}. \tag{37}$$

This looks like an infinite family of solutions, but all solutions with the same q_1 mod 2 are equivalent to each other by monodromies around the infinity. Indeed, the M_∞ matrix (35) (combined with the charge conjugation $\pm(q, p) \mapsto \mp(q, p)$) changes the electric charges of dyons with $p = 1$ by 2 units, so applying it several times can change any even q_1 to $q_1 = 0$ (and hence $q_2 = +1$) and any odd q_1 to $q_1 = -1$ (and hence $q_2 = 0$). Either way, this gives *massless magnetic monopoles (with no electric charges) at one singularity and massless dyons (with $q = \pm 1$) at the other singularity.* We still has to choose which singularity has the monopoles, but that's equivalent to choosing which branch of $\sqrt{\Lambda^4}$ is $+\Lambda^2$ and which is $-\Lambda^2$ — it's a distinction without any physical difference. The usual choice is to have massless monopoles at $U = +\Lambda^2$ and massless dyons at $U = -\Lambda^2$, but that's simply a convention.

Confinement in the Massive Theory

Thus far, we have focused on the massless theory with $\mathcal{N} = 2$ SUSY and exactly degenerate moduli space of SUSY vacua. Now let's break the extra SUSY (down to $\mathcal{N} = 1$) by giving the Φ^a superfields a small mass m,

$$W_{\text{tree}} \;=\; \frac{m}{2}\,\Phi^a\Phi^a. \tag{38}$$

Earlier in class we saw that this mass lifts the degeneracy of the moduli space and gives us two discrete SUSY vacua with $U = \pm\Lambda^2$. But as long as $m \ll \Lambda$, we may obtain the effective low-energy theory (for $E \ll |\Lambda|$) by starting with the EFT of the massless theory for U near $\pm\Lambda^2$ and then perturbing it by the superpotential (38).

Let's start with the EFT for U near $+\Lambda^2$. This EFT comprises the U, the massless magnetic monopoles $\mathcal{M}^\pm$, and the abelian vector superfield V, or rather its magnetic dual $\widetilde{V}$. The effective Wilsonian Lagrangian is

$$\mathcal{L}_w \;=\; \int\! d^4\theta\, K_w \;+\; \int\! d^2\theta \left(\frac{\tau_w^{\text{dual}}(U)}{8\pi i}\, \widetilde{W}^\alpha \widetilde{W}_\alpha \;+\; W(U,\mathcal{M}^\pm) \right) \;+\; \text{H.c.} \tag{39}$$

where

$$K_w \;=\; \hat{K}_w(U,\bar{U}) \;+\; Z_{\mathcal{M}} \left(\overline{\mathcal{M}}^+ \exp(+2\widetilde{V}) \mathcal{M}^+ \;+\; \overline{\mathcal{M}}^- \exp(-2\widetilde{V}) \mathcal{M}^- \right) \;+\; \cdots \tag{40}$$

and

$$W \;=\; \mathcal{M}^+\mathcal{M}^- \times \Big(\lambda(U-\Lambda^2) \;+\; O\big((U-\Lambda^2)^2\big) \Big) \;+\; \frac{m}{2} \times U. \tag{41}$$

Note that all of the Wilsonian couplings — including the $\tau_w^{\text{dual}}(U)$ and the Kähler metric $g_{U\bar{U}} = \partial^2 \hat{K}_w / \partial U \partial \bar{U}$ — are smooth functions of the modulus U. The τ and the Kähler metric become singular only after we integrate out the monopole superfields $\mathcal{M}^\pm$.

The mass term (38) for the Φ^a fields becomes an O'Raifeartaigh term for the U modulus. Thanks to the Yukawa coupling $\lambda = O(1/\Lambda)$ of the U to the monopole superfields $\mathcal{M}^\pm$, this O'Raifeartaigh term does not break SUSY; instead, it pushes the scalar components of $\mathcal{M}^\pm$ to non-zero values

$$\langle\mathcal{M}^+\rangle \;=\; \langle\mathcal{M}^-\rangle \;=\; \sqrt{\frac{-m}{2\lambda}\,+\,O(m^2)} \;=\; O(\sqrt{m\Lambda}) \tag{42}$$

and fixes the modulus scalar at $\langle U\rangle = +\Lambda^2$. Consequently, we have a SUSY vacuum with a mass gap, *i.e.* no massless particles. Indeed, from the *dual* vector $\widetilde{V}$ point of view, the $\mathcal{M}^\pm$ fields have *electric* charges $\pm e^{\text{dual}}$, so their VEVs (42) cause the Higgs effect: The vector multiplet $\widetilde{V}$ and the chiral multiplet $\delta\mathcal{M}^+ - \delta\mathcal{M}^-$ get masses

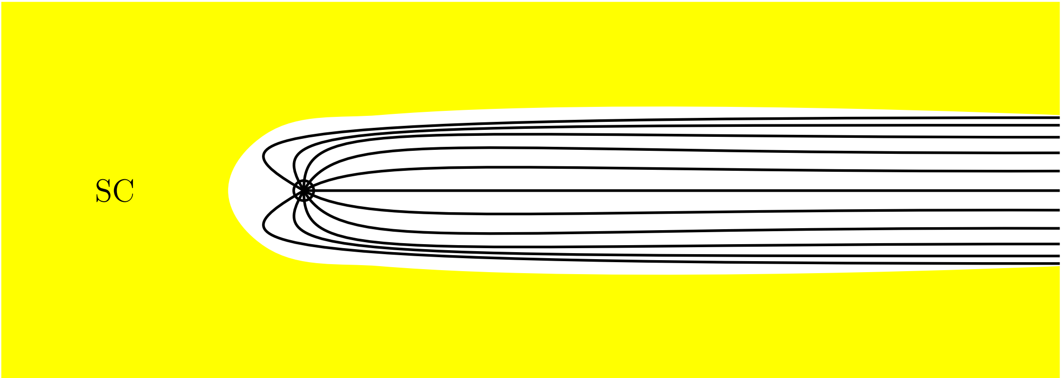
$$m_v \;=\; 2e^{\text{dual}}\sqrt{Z_{\mathcal{M}}}\times|\langle\mathcal{M}\rangle| \;=\; O(\sqrt{m\Lambda}). \tag{43}$$

The remaining chiral multiplets δU and $\delta\mathcal{M}^+ + \delta\mathcal{M}^-$ get comparable masses

$$m_s \;=\; \frac{\sqrt{2}\lambda\,\langle\mathcal{M}\rangle}{\sqrt{Z_{\mathcal{M}}g_{U\bar{U}}}} \;=\; O(\sqrt{m\Lambda}) \tag{44}$$

from their Yukawa couplings to the VEVs (42).

The ordinary Higgs mechanism (involving electrically-charged scalar VEVs) is a relativistic version of superconductivity. In a superconductor some electrically-charged superfluid short-circuits static electric fields and screens the electric charges of all particles. It also repels the magnetic fields in the Meissner effect or constricts them into thin flux tubes, and that causes problems for magnetic monopoles. Indeed, if you put a monopole in the middle of a bulk superconductor, the magnetic charge is not screened, so the magnetic field has a fixed flux μ and it has to go somewhere. But instead of spreading out uniformly in all directions, the magnetic field is forced into a flux tube of fixed diameter $\sim 1/m_v$.



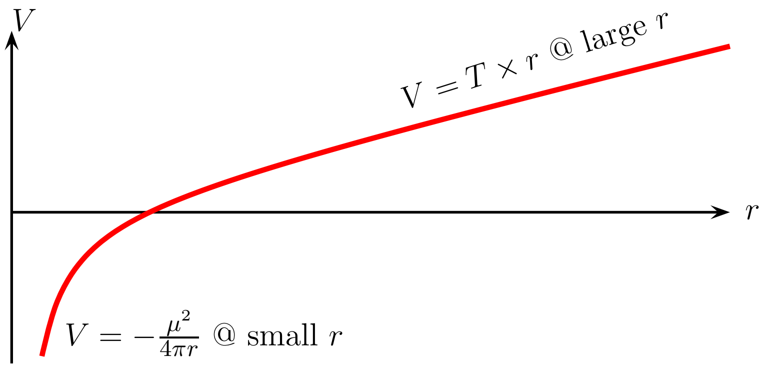
(45)

At one end of the tube is the monopole which produces the magnetic flux. At the other end, we need either another monopole with an opposite magnetic charges to soak up that flux, or else the tube needs to reach the superconductor's surface beyond which the magnetic field can spread out. Since in the Higgs mechanism the whole space becomes superconducting, the flux tube has to connect two monopoles with opposite charges.

Consider a pair of monopoles connected by a long flux tube. The tube has a finite *tension*

$$T = \frac{E_{\text{tube}}}{L_{\text{tube}}}\,, \tag{46}$$

so it likes to form a straight line, and then it pulls the monopoles towards each other with a distance-independent force T. Consequently, instead of a Coulomb potential between the two monopoles, at large distances we get a *confining potential*



(47)

Thus, *in the electric superconductor, magnetic monopoles are confined:* no matter how much energy you give a monopole-antimonopole pair, you can never break it into two separate particles.

Now consider a *magnetic superconductor* in which the superfluid is made of bosons with magnetic rather then electric charges. Using the electric-magnetic duality as a mirror, we immediately see that in a magnetic superconductor, the magnetic fields are short-circuited while the electric fields are repelled in a Meissner-like effect or restricted to electric flux tubes. Consequently, the magnetic charges are screened while the electric charges are confined.

Gerard 't Hooft had argued back in the 1970-s that this is the origin of quark confinement in QCD: The QCD vacuum should be some kind of a chromo-magnetic superconductor, so the chromo-electric charges of the quarks become confined. Of course, the chromo-magnetic monopoles in an un-broken $SU(3)$ gauge theory are somewhat different from the ordinary magnetic monopoles in $U(1)$ or $SU(2)$ broken to $U(1)$, but that's a technical issue. If you are interested, it's explained in chapter 5 of 't Hooft's lecture notes (`arXiv:hep-th/0010225`) I've assigned in homeworks 9 and 11. The real problem is explaining why those monopoles form a Bose–Einstein condensate (or some other kind of a superfluid).

To this day, there is no rigorous proof of monopole condensation in QCD, but the Seiberg–Witten theory provides a useful model how it could happen. Indeed, in the massive Seiberg–Witten theory we have non-zero VEVs (42) of scalar fields $\mathcal{M}^{\pm}$ with magnetic charges. Those VEVs make the vacuum into a *magnetic superconductor* in which all electric charges are confined. This includes the quarks (if we add them to the theory), the $W^{\pm}$ vector bosons, or even the dyons. All isolated particles must have zero *net* electric charges; a net magnetic charge is OK as it would screened by the monopole condensate.

Oblique confinement: The massive Seiberg–Witten theory has two vacua with $\langle U\rangle = \pm\Lambda^2$. Both vacua are confining, but while the $\langle U\rangle = +\Lambda^2$ vacuum confines the electric charges, the $\langle U\rangle = -\Lambda^2$ vacuum confines the $n - m$ combination of electric and magnetic charges; such confinement is called *oblique.*

To see how this works, let's recall that the massless theory with $U = -\Lambda^2$ has massless dyons $\mathcal{D}^{\pm}$ with (electric,magnetic) charges $\pm(1,1)$. Consequently, for small $m \ll \Lambda$ we have an effective theory of U, $\mathcal{D}^{+}$, and $\mathcal{D}^{-}$ chiral superfields and an abelian vector superfield $\widetilde{V}$; the latter is dual to V but the duality here is not electric-magnetic as before but rather electric-dyonic, thus

$$\tau^{\rm dual} \;=\; \frac{\tau}{-\tau+1} \tag{48}$$

instead of $-1/\tau$. The effective Wilsonian Lagrangian is similar to (39)–(41); in particular, the superpotential is

$$W \;=\; \mathcal{D}^+\mathcal{D}^- \times \Big(\lambda(U+\Lambda^2) \;+\; O\big((U+\Lambda^2)^2\big)\Big) \;+\; \frac{m}{2}\times U. \tag{49}$$

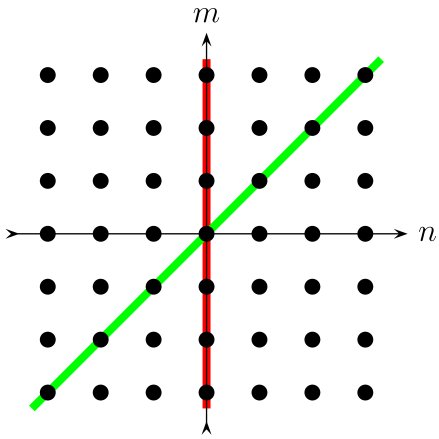
Thanks to the O'Raifeartaigh term, the dyon fields have non-zero VEVs

$$\left\langle\mathcal{D}^+\right\rangle \;=\; \left\langle\mathcal{D}^-\right\rangle \;=\; \sqrt{\frac{-m}{2\lambda}\;+\;O(m^2)} \;=\; O(\sqrt{m\Lambda}), \tag{50}$$

the U modulus is fixed at $\langle U\rangle = -\Lambda^2$, and the theory has a mass gap due to (dual) Higgs mechanism.

Any kind of a Higgs mechanism gives us a superconducting vacuum, but this time the superfluid is made of dyons with charges $\pm(1,1)$ rather then purely magnetic monopoles. This superfluid short-circuits a particular combination $\vec{E}+\alpha\vec{B}$ of the electric and magnetic fields, while all other combinations are restricted to flux tubes. Consequently, there is screening of $n=m$ combinations of electric and magnetic charges, but all other combinations of electric and magnetic charges are confined.

The purely-electric charges of quarks and $W^\pm$ 'gluons' do not feel the difference between the 'straight' and the 'oblique' confinements. But for magnetically charged particles, the difference become important: for the straight confinement, all isolated particles must have zero net electric charge, regardless of the magnetic charge, while for the oblique confinement, the net electric charge of an isolated particle should be equal to its net magnetic charge. On the charge lattice



(51)

the isolated particles lie on the red vertical line for the straight confinement versus the green diagonal line for the oblique confinement.

Elliptic Curves

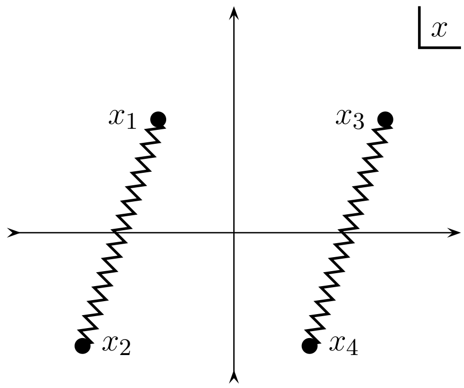
Let's go back to the massless theory and consider the holomorphic gauge coupling τ as a function of the modulus U. Now that we know all the singularities of the holomorphic function $\tau(U)$, we would like to write it down in an analytic form. Unfortunately, this is rather difficult because this function is not only multi-valued but has non-abelian monodromies, which require a very complicated Riemann surface with non-commuting branch cuts. To avoid this problem, we are going to construct an *elliptic curve* which uniquely determines τ modulo an $SL(2,\mathbf{Z})$ S–duality, and then we are going to write down a formula for that elliptic curve as a function of the modulus U.

The name *'elliptic curve'* is rather misleading to non-mathematicians. It is not an ellipse. It is not even a curve in the usual sense of the word but a Riemann surface of two real dimensions; it's a *complex curve* because it can be parametrized by one *complex* coordinate. Specifically, the elliptic curve is the Riemann surface of an elliptic integral

$$\int \frac{dx\ \mathrm{Polynomial}(x)}{\sqrt{(x-x_1)(x-x_2)(x-x_3)(x-x_4)}} \tag{52}$$

— integrals like these are used for calculating the lengths of elliptic arcs, hence the name. And that's the only connection between the ellipses and the elliptic curves.

The integrand of the elliptic integral (52) has two square-root branch cuts,



(53)

so its Riemann surface has two sheets. Introducing another coordinate y to distinguish between the sheets, we can describe the elliptic curve as a complex curve in $\mathbf{C}^2$ defined by the polynomial equation

$$y^2 \;=\; P_4(x) \;\equiv\; (x-x_1)(x-x_2)(x-x_3)(x-x_4). \tag{54}$$

And if we want to describe a whole family of elliptic curves, we simply make the roots $x_1(U),\ldots,x_4(U)$ into functions of the modulus U. However, we should remember that two seemingly different equations

$$y^2 \;=\; \prod_{i=1}^{4}(x-x_i) \qquad \text{and} \qquad y'^2 \;=\; \prod_{i=1}^{4}(x'-x'_i) \tag{55}$$

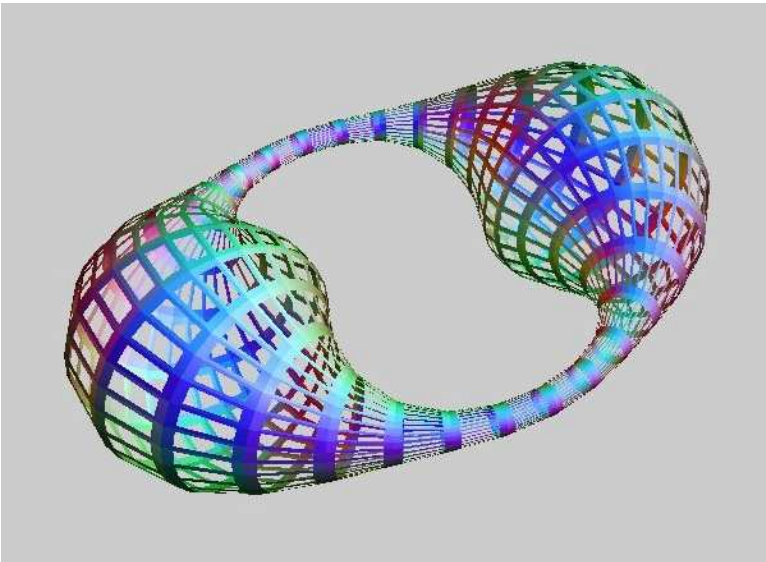
may describe the same elliptic curve in different coordinates related by an $SL(2,\mathbf{C})$ conformal symmetry of the complex sphere,

$$x'_i \;=\; \frac{\alpha x_i+\beta}{\gamma x_i+\delta}\,, \qquad x' \;=\; \frac{\alpha x+\beta}{\gamma x+\delta}\,, \qquad y' \;=\; \frac{y}{(\gamma x+\delta)^2}\times\prod_{i=1}^{4}(\gamma x_i+\delta)^{-1/2}. \tag{56}$$

Using such symmetries we may fix 3 out of four roots of the curve, for example $x_2\equiv 0$, $x_3\equiv 1$, $x_4\equiv\infty$, and only the x_1 varies from curve to curve. However, for our purposes it's convenient to fix only one root, $x_4 \equiv \infty$, but let the other three roots vary with the modulus U, thus

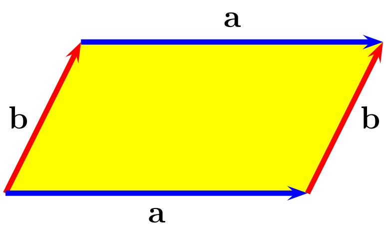
$$y^2 \;=\; (x - x_1(U))(x - x_2(U))(x - x_3(U)) \;=\; x^3 \;+\; a(U) \times x^2 \;+\; b(U) \times x \;+\; c(U). \tag{57}$$

Topologically, the elliptic curve is a torus. Indeed, each sheet of the Riemann surface is a sphere, while the two branch cuts provide two handles connecting the two spheres.



(58)

Despite its funny appearance, this torus is conformally flat, so it has a flat complex coordinate z — which has a holomorphic but very complicated relation to x and y — with constant metric $ds^2 = \text{const} \times dz\, dz^*$. This flat coordinate spans a parallelogram whose opposite sides glued together to make a torus,

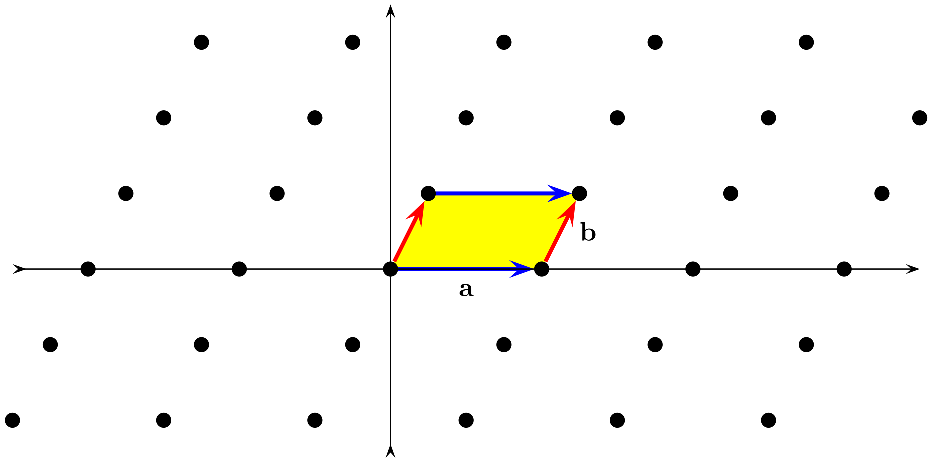


(59)

Equivalently, we may treat z as a multi-valued coordinate that spans the whole complex plane, but whose values are taken modulo a discrete lattice,

$$z \equiv z + n\mathbf{a} + m\mathbf{b} \qquad \forall n, m \in \mathbf{Z}. \tag{60}$$

In this language, the parallelogram (59) occupies a unit cell of the lattice with complex periods **a** and **b**.



(61)

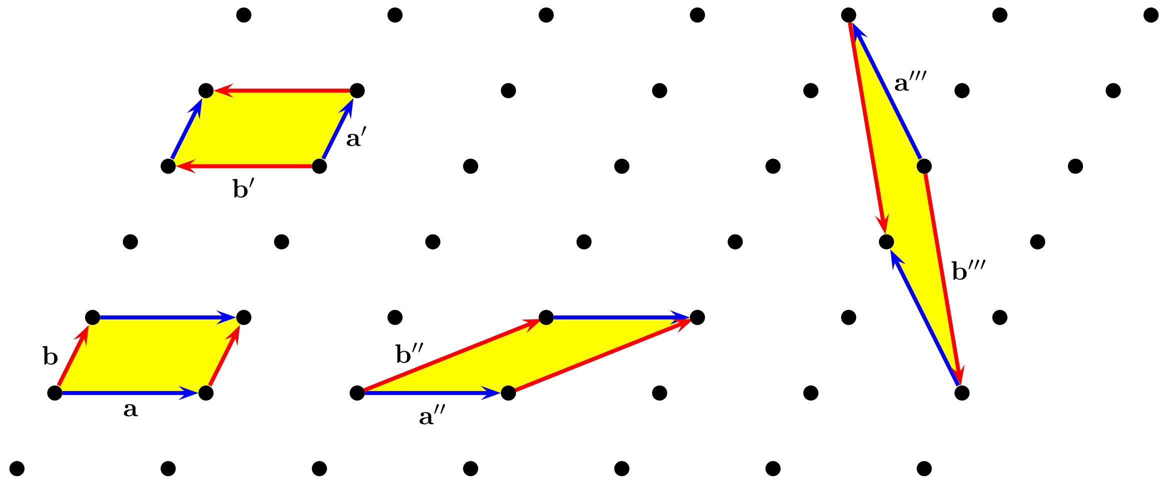
Note that defining z as the flat coordinate of the torus determines it only up to $z \to cz + d$ for some complex constants c and d. Consequently, we are free to multiply the two periods of the lattice by the same factor, $\mathbf{a} \to c\mathbf{a}$, $\mathbf{b} \to c\mathbf{b}$. On the other hand, the ratio

$$\tau = \frac{\mathbf{b}}{\mathbf{a}} \tag{62}$$

which controls the aspect ratio and the skew angle of the torus is completely determined by the complex structure of the variable z and cannot be changes without a non-holomorphic redefinition such as $z \to \alpha z + \gamma z^*$. Thus, as long as z is holomorphically related to the complex coordinates x and y of the elliptic curve (57), the curve's geometry controls the period ratio (62) of the torus.

Or rather, the complex structure of the elliptic curve determines the complex structure of the lattice (61), but we are free to choose any unit cell we like to define the periods **a** and

b. Here are a few examples:



(63)

When you glue together the opposite sides of any such cell, we always get exactly the same torus geometry, but different cells provide different homology bases for the circles wrapping the torus. Or in terms of the lattice, different unit cells correspond to different bases for the same lattice. And of course, all such bases are related to each other by the $SL(2,\mathbf{Z})$ symmetries,

$$\begin{pmatrix} \mathbf{b}' \\ \mathbf{a}' \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} \mathbf{b} \\ \mathbf{a} \end{pmatrix} \tag{64}$$

so the period ratios for different bases are related according to

$$\tau' = \frac{\alpha\tau+\beta}{\gamma\tau+\delta}, \qquad \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in SL(2,\mathbf{Z}). \tag{65}$$

The bottom line is, *the complex structure of a complex curve determines the period ratio* τ *of a complex lattice modulo* $SL(2,\mathbf{Z})$.

At this point, I can finally explain why did I introduce the elliptic curves in the first place. The holomorphic gauge coupling τ is the period ratio of the lattice of electric and magnetic charges

$$Q + i\mu = n\times e + m\times \tau e. \tag{66}$$

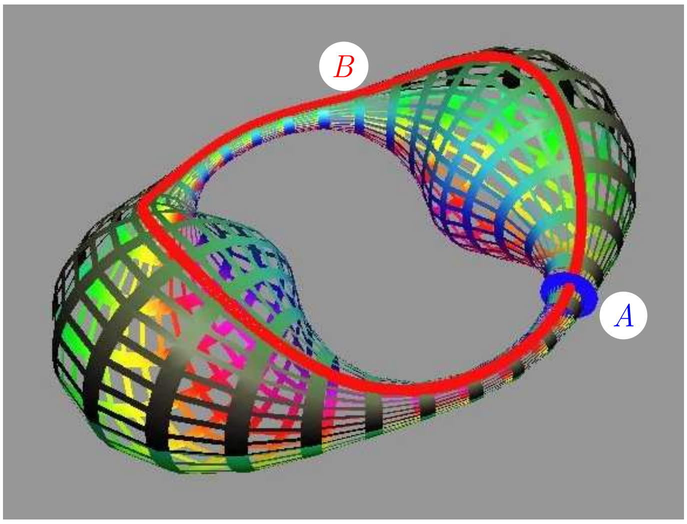
The electric-magnetic duality and its $SL(2,\mathbf{Z})$ cousins change the basis of this lattice, but they don't change the lattice itself. To encode $\tau(U)$ modulo $SL(2,\mathbf{Z})$, we need to specify how the charge lattice depends on U without fixating on a particular pair of periods $\mathbf{a}$ and $\mathbf{b}$. Equivalently, we want to specify how the complex structure of a torus depends on U without picking a particular flat coordinate on the torus, and we may do just that in terms of the elliptic curve family (57).

To make this approach practical, we need to do to things: (1) Learn how to translate between τ modulo $SL(2,\mathbf{Z})$ and the parameters of the elliptic curve, and (2) find out which family of the elliptic curves describes the Seiberg–Witten model. Let's start with the translation problem.

Getting τ form an elliptic curve:
We can write the two periods of the torus as contour integrals

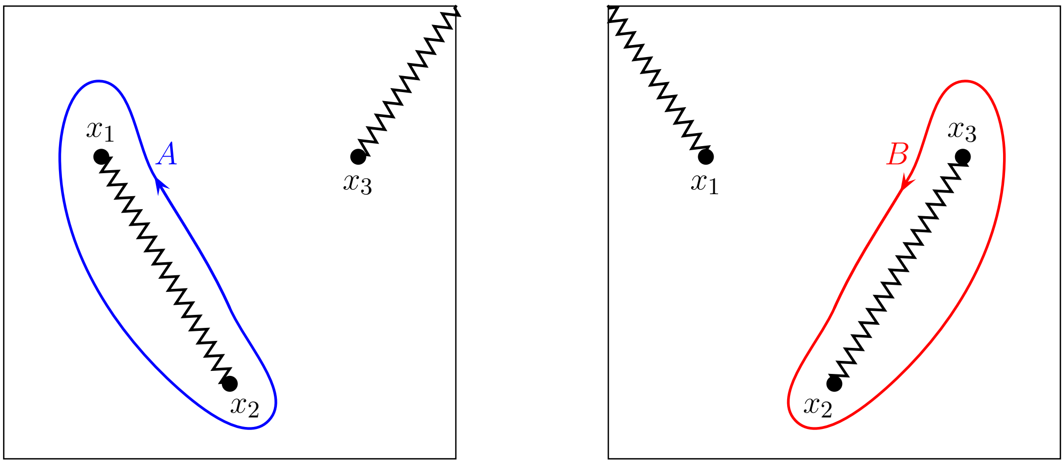
$$\mathbf{a} = \oint_A dz \qquad \mathbf{b} = \oint_B dz \tag{67}$$

over the minor and major circles of the torus:



(68)

In terms of the x coordinate of the elliptic surface, these cycles become contours surrounding different branch cuts,



(69)

What about the differential dz? Although the z coordinate of the torus is multivalued, the differential dz is single-valued, so it must translate into a single-valued differential of the elliptic curve, hence

$$dz = \frac{dx}{y} \times f(x) \tag{70}$$

where f is a single-valued function of x. Moreover, $f(x)$ may not have any singularities at finite x's — otherwise, the residue of that singularity would produce an extra ambiguity of z besides the lattice $n\mathbf{a} + m\mathbf{b}$. Likewise, a double loop around $x = \infty$ (this loop is closed on the curve) also should not have a residue, which requires

$$x \times \frac{f(x)}{y} \to 0 \quad \text{for } x \to \infty. \tag{71}$$

Since $y \sim x^{3/2}$ for large x, this limits the growth of $|f(x)|$ with $|x|$ to be slower then $\sqrt{|x|}$. For a holomorphic nowhere-singular function, this limit implies $f = \text{const}$, so without loss of generality we may take $f = 1$ and hence

$$dz = \frac{dx}{y}. \tag{72}$$

Consequently, we may obtain τ from an elliptic curve as a ratio of two contour integrals,

$$\tau = \oint_B \frac{dx}{y} \Big/ \oint_A \frac{dx}{y}. \tag{73}$$

Singularities:

When the U modulus of the Seiberg–Witten model approach the singularity at ∞, the gauge coupling becomes weak and $\operatorname{Im}\tau \to +\infty$, *cf.* eq. (31). When U approaches a finite singularity at $\pm\Lambda^2$, the *dual gauge coupling* becomes weak and $\operatorname{Im}\tau^{\text{dual}} \to +\infty$, *cf.* eq. (8). Either way, such a τ describes a long but skinny torus like a bicycle tire, and at the singularity itself the torus becomes infinitely long but infinitely thin.

In the elliptic curve language, singular tori correspond to degenerate roots $x_{1,2,3}$. Indeed, consider what happens when $x_1 = x_2 \neq x_3$ and

$$\frac{dx}{y} \;=\; \frac{dx}{(x-x_{12})\sqrt{x-x_3}} \tag{74}$$

The B contour normally encloses the x_2 but not the x_1, so when those roots coincide, the contour has to go right through the pole of (74) at $x = x_{12}$. This makes the integral diverge and the **b** period of the torus becomes infinite. On the other hand, the A contour does not go through the pole at x_{12} but circles it, so the integral — *i.e.*, the **a** period of the torus — remains finite. Consequently, $\tau = \mathbf{b}/\mathbf{a}$ is infinite and the torus indeed degenerates.

Now consider an almost-degenerate curve with two roots that are very close to each other but not exactly equal. Without loss of generality, let us take $x_1 = 0$, $x_2 = \epsilon \ll 1$, and $x_3 = 1$. For these roots, we can take the A cycle to be a circle of radius $|\epsilon| \ll 1$ with a center between x_1 and x_2. Along this circle we may approximate $x - x_3 \approx -1$, hence

$$y \;\approx\; i\sqrt{(x-x_1)(x-x_2)}, \tag{75}$$

and

$$\mathbf{a} \;=\; \oint\limits_A \frac{dx}{y} \;\approx\; -i\oint \frac{dx}{\sqrt{(x-x_1)(x-x_2)}} \;=\; 2\pi. \tag{76}$$

The B cycle surrounds the long branch cut of $\sqrt{(x-x_2)(x-x_3)}$, so we may convert it into a definite integral along one side of the branch cut,

$$\mathbf{b} \;=\; \oint\limits_B \frac{dx}{\sqrt{(x-x_1)(x-x_2)(x-x_3)}} \;=\; 2i\int\limits_{x_2}^{x^3} \frac{dx}{\sqrt{(x-x_1)(x-x_2)(x_3-x)}} \tag{77}$$

For $x_1 = 0$, $x_2 = \epsilon \ll 1$ and $x_3 = 1$ we may split the integration range into two pieces, from $x_2 = \epsilon$ to some $C \gg \epsilon$ — but $C \ll 1$ — and from C to $x_3 = 1$. In the first piece we approximate $1 - x \approx 1$ for $|x| \ll 1$, hence

$$2i\int\limits_{\epsilon}^{C} \frac{dx}{\sqrt{x(x-\epsilon)(1-x)}} \;\approx\; 2i\int\limits_{\epsilon}^{C} \frac{dx}{\sqrt{x(x-\epsilon)}} \;=\; 2i\,\mathrm{ar\cosh}(\sqrt{C/\epsilon}) \;=\; 2i\log\frac{4C}{\epsilon} \;+\; O(\epsilon/C). \tag{78}$$

In the second piece we have $|x| \geq |C| \gg |\epsilon|$, which allows us to approximate $x - \epsilon \approx x$ and consequently

$$2i\int\limits_{C}^{1} \frac{dx}{\sqrt{x(x-\epsilon)(1-x)}} \;\approx\; 2i\int\limits_{C}^{1} \frac{dx}{x\sqrt{1-x}} \;=\; 2i\log\frac{1+\sqrt{1-C}}{1-\sqrt{1-C}} \;=\; 2i\log\frac{4}{C} \;+\; O(C). \tag{79}$$

Altogether,

$$\mathbf{b} \;\approx\; 2i\log\frac{16}{\epsilon} \tag{80}$$

and

$$\tau \;\approx\; \frac{2i}{2\pi}\log\frac{16(x_3-x_1)}{x_2-x_1}\,. \tag{81}$$

More generally, when any two roots are much closer to each other than to the third root, we have

$$\tau \text{ or some } \tau^{\text{dual}} \;\approx\; \frac{2i}{2\pi}\log\frac{16\times\Delta x_{\text{large}}}{\Delta x_{\text{small}}}\,. \tag{82}$$

Seiberg–Witten curve:

Now let's compare eq. (82) for a near-degenerate elliptic curve with the known singularities of the Seiberg–Witten model. For $U \to \infty$ we have a weakly-coupled singularity

$$\tau(U) \;=\; \frac{2i}{2\pi}\log\frac{U}{\Lambda^2} \;+\; \mathrm{smooth}(U); \tag{83}$$

according to eq. (82) this calls for a curve with

$$\frac{\Delta x_{\text{large}}}{\Delta x_{\text{small}}} \;\propto\; \frac{U}{\Lambda^2} \qquad \text{for } U \to \infty. \tag{84}$$

We also have strongly-coupled singularities for $U \to \pm\Lambda^2$ where

$$\tau^{\text{dual}}(U) \;=\; \frac{2i}{2\pi}\log\frac{U\mp\Lambda^2}{\Lambda^2} \;+\; \text{smooth}(U); \tag{85}$$

according to eq. (82) this calls for

$$\frac{\Delta x_{\text{large}}}{\Delta x_{\text{small}}} \;\propto\; \frac{U\mp\Lambda^2}{\Lambda^2} \qquad \text{for } U\to\pm\Lambda^2. \tag{86}$$

At all other values of the modulus U the gauge coupling is non-singular, the three roots $x_{1,2,3}$ of the elliptic curve should stay non-degenerate and finite. All we need to do now is to find the holomorphic $x_1(U)$, $x_2(U)$, and $x_3(U)$ that satisfy these conditions, and the solution is obvious:

$$x_1 \;=\; U, \qquad x_2 \;=\; +\Lambda^2, \qquad x_3 \;=\; -\Lambda^2, \tag{87}$$

and the Seiberg–Witten elliptic curve is

$$y^2 \;=\; (x-U)\times(x^2-\Lambda^4). \tag{88}$$

Curiously, this curve has a very simple instanton expansion which stops at the one-instanton level,

$$y^2 \;=\; x^2(x-U) \;-\; \Lambda^4\times(x-U) \;+\; 0. \tag{89}$$

The gauge coupling itself is affected by all orders of the instanton expansion, but somehow the entire series

$$\tau(U) \;=\; \frac{i}{2\pi}\,\log\frac{U^2}{\Lambda^4} \;+\; \sum_{n=1}^{\infty} C_n\left(\frac{\Lambda^4}{U^2}\right)^n \tag{31}$$

can be summarized by a simple one-instanton formula (89). Similar simplifications happen in many other SUSY gauge theories with abelian Coulomb phases, including theories with multiple $U(1)$ factors and matrices of gauge couplings. Somehow, when all the couplings are encoded in a hyper-elliptic curve, the curve has a only classical and one-instanton terms while the couplings themselves have horribly complicated instanton expansions. I don't know why the curves are so simple, but I know they are.