TOPOLOGICAL UMBRAL MOONSHINE

THEO JOHNSON-FREYD

Caveat lector: These notes are unedited and sometimes elliptical. They start off well, and slowly peter.

Contents

These lectures were written for the workshop "Topological Moonshine" at the University of Illinois, Urbana-Champaign, July 17-21, 2023. My goal for these lectures is to explain some thoughts and ideas about how to express the statements of (umbral) moonshine in the language of of topological modular forms. These are not fully worked out by any means, and I invite you to refine (or reject) this story further. I thank the organizers Dan Berwick-Evans, Emily Cliff, Meng Guo, and Arnav Tripathy, for forcing me to organize my thoughts well enough to give a lecture series.

I thank Dan Berwick-Evans, John Duncan, and Davide Gaiotto for many clarifying conversations about this material. The workshop "Topological Moonshine" was supported by a grant from the National Science Foundation. The author's work is supported by an NSERC Discovery Grant and by the Simons Collaboration on Global Categorical Symmetries. Research at the Perimeter Institute is supported by the Government of Canada through Industry Canada and by the Province of Ontario through the Ministry of Economic Development and Innovation. The Perimeter Institute is in the Haldimand Tract, land promised to the Six Nations. Dalhousie University is in Mi'kma'ki, the unceded territory of the Mi'kmaq. The University of Illinois Urbana-Champaign is on the lands of the Peoria, Kaskaskia, Piankashaw, Wea, Miami, Mascoutin, Odawa, Sauk, Mesquaki, Kickapoo, Potawatomi, Ojibwe, and Chickasaw Nations. These lands were the traditional territory of these Native Nations prior to their forced removal; these lands continue to carry the stories of these Nations and their struggles for survival and identity.

Perimeter Institute for Theoretical Physics, Waterloo, Ontario, theojf@pitp.ca.
Department of Mathematics, Dalhousie University, Halifax, Nova Scotia, theojf@dal.ca.

1. Is equivariant elliptic cohomology the light that Moonshine reflects?

Summary: Yes and no. Certainly Moonshine selects equivariant elliptic cohomology classes, and some of the Moonshine statements are natural from this language. However, Moonshine also involves three aspects that go beyond plain equivariant elliptic cohomology:

(1) Moonshine involves certain *genus-zero phenomena*, or equivalently *optimal growth conditions.*
(2) Moonshine involves *Atkih–Lehner involutions* and other symmetries that go beyond $\mathrm{SL}(2,\mathbb{Z})$.
(3) Moonshine involves *positivity*, analogous to asking that a K-theory class be not just the class of a virtual representation, but actually be represented by an actual representation.

We will handle these aspects by:

(1) designing a bespoke version of equivariant Tmf that incorporates optimal growth;
(2) restricting to cases where Atkih–Lehner involutions don't occur;
(3) simply ignoring all aspects of positivity.

1.1. Classical modular forms.

I first review some standard vocabulary. Probably I won't say any of this in my lectures.

Let $\mathcal{M} = \mathcal{M}(\mathbb{C})$ denote the moduli stack of $\mathbb{C}$-analytic elliptic curves. In other words, $\mathcal{M} = \mathfrak{h}/\Gamma$, where:

- $\mathfrak{h} = \{\tau \in \mathbb{C} : \Im(\tau) > 0\}$ is the upper half plane
- $\Gamma = \mathrm{SL}(2,\mathbb{Z})$
- $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ acts by

$$\gamma \cdot \tau = \frac{a\tau + b}{c\tau + d}$$

Note that this action factors through $\mathrm{PSL}(2,\mathbb{Z})$. However, the elliptic curve E_τ has an involution $z \mapsto -z$, and so the correct moduli stack uses $\mathrm{SL}(2,\mathbb{Z})$ and not $\mathrm{PSL}(2,\mathbb{Z})$. By convention, an elliptic curve comes with a basepoint, undoing the stackiness by E_τ itself.

Actually, later I'll use a slightly different group. Namely, I will most likely want to talk about *spin* elliptic curves, i.e. elliptic curves equipped with a spin structure. There are two topological types of a spin elliptic curve: the spin structure might bound or it might not. The non-bounding spin structure on T^2 is (naïvely) invariant under all of $\mathrm{SL}(2,\mathbb{Z})$. More precisely, it is invariant under the double cover $\mathrm{Mp}(2,\mathbb{Z})$. The bounding spin structure is invariant under a certain double of $\Gamma_0(2)$. So the moduli stack of non-bounding spin elliptic curves is $\mathfrak{h}/\mathrm{Mp}(2,\mathbb{Z})$.

By construction, $\mathcal{M}$ is naturally a noncompact complex curve. A holomorphic function on $\mathcal{M}$ is called a *weak modular function*. By construction, this is a holomorphic function $f : \mathfrak{h} \to \mathbb{C}$ such that for each $\gamma \in \Gamma$, we have $f(\gamma \cdot \tau) = f(\tau)$. It is conventional to write the (right) action of Γ on $\mathcal{O}(\mathfrak{h})$ with a slash:

$$(f|\gamma)(\tau) := f(\gamma \cdot \tau).$$

Generalizing the modular functions, a *(weak) modular form of weight w* is a function $f(\tau)$ such that $f(\tau)(\mathrm{d}\tau)^{w/2}$ descends to $\mathcal{M}$. Note that

$$\mathrm{d}(\gamma \cdot \tau) = \mathrm{d}\left(\frac{a\tau + b}{c\tau + d}\right) = \frac{(a\,\mathrm{d}\tau)(c\tau + d) - (a\tau + b)(c\,\mathrm{d}\tau)}{(c\tau + d)^2} = \frac{(ad - bc)\,\mathrm{d}\tau}{(c\tau + d)^2}$$

and that $ad - bc = \det(\gamma) = 1$. So $f(\tau)$ is a weight-w modular form if it is invariant under the action

$$f \mapsto (f|_w\gamma)(\tau) := (c\tau + d)^{-w} f(\gamma \cdot \tau).$$

Another generalization are vector valued modular forms. Pick a representation V of Γ. A *vector valued (weak) modular form* is a holomorphic function $f : \mathfrak{h} \to V$ which is equivariant for the Γ-action. Note that this is *more* than a section of a vector bundle over $\mathcal{M}$. It is a section of a vector bundle over $\mathcal{M}$, together with a trivialization of that vector bundle when pulled back along

$\mathfrak{h} \to \mathcal{M}$. If you pick V, then you can talk about "V-valued modular forms of weight w" by adjusting the action by the weight-w action.

An example is the following. Let me write "Pic" for the invertible vector bundles, so that $\mathrm{Pic}(G) = \{\text{1D representations of } G\} = \hom(G, \mathbb{C}^\times)$. Recall that

$$\mathrm{Pic}(\mathrm{PSL}(2,\mathbb{Z})) = \mathbb{Z}/6, \qquad \mathrm{Pic}(\mathrm{SL}(2\mathbb{Z})) = \mathbb{Z}/12, \qquad \mathrm{Pic}(\mathrm{Mp}(2,\mathbb{Z})) = \mathbb{Z}/24.$$

Actually, these are really easy calculations if you are willing to quote the following presentation:

$$\mathrm{SL}(2,\mathbb{Z}) = \langle S, T : S^2 = (ST)^3 \text{ is central of order } 2\rangle.$$

If you change that "2" to a "1" or a "4", you get the other two groups. The generators are:

$$S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

So $\mathrm{SL}(2,\mathbb{Z})$ is almost a free product. It is the pushout (in Groups)

$$\mathrm{SL}(2,\mathbb{Z}) = \mathbb{Z}/4 \underset{\mathbb{Z}/2}{*} \mathbb{Z}/6$$

Cohomology turns pushouts into pullbacks (in Sp), and also you can compute cohomology prime-by-prime. So it's easy to read off things like $\mathrm{H}^1(G;\mathbb{C}^\times) = \mathbb{Z}/4 \underset{\mathbb{Z}/2}{\times} \mathbb{Z}/6 = \mathbb{Z}/12$. The generator is the representation where $S \mapsto e(1/4)$ and $T \mapsto e(-1/12)$, where $e(x) := \exp(2\pi i x)$.

Anyway, a V-valued modular form with V invertible is called a *modular form with multiplier*, since obviously it is like a modular form, except instead of being invariant under the action of Γ, it is an eigenform with some prescribed eigenvalues. Now, there is a line bundle over $\mathcal{M}$ whose sections are weight-w modular forms. This line bundle is isomorphic to the line bundle you get by taking $V = [w \mod 12] \in \mathbb{Z}/12 = \mathrm{Pic}(\mathrm{SL}(2,\mathbb{Z}))$. But modular forms of weight w are not the same as modular functions with multiplier $[w \mod 12]$. They almost are, but the trivializations are different. The difference is a factor of the function $\eta(\tau)^{2w}$ where

$$\eta(\tau) = q^{1/24} \prod_{m\geq 1}(1-q^m), \qquad q = e(\tau).$$

These differences matter for the following reason. So far, I haven't put any analytic control over the behaviour of $f(\tau)$ as τ approaches the boundary of $\mathfrak{h}$. Let's put that control in now. $\mathcal{M}$ has a Deligne-Mumford compactification $\overline{\mathcal{M}}$, constructed as follows. The boundary of $\mathfrak{h}$ is a copy of $P^1(\mathbb{R})$. The subset $P^1(\mathbb{Q}) \subset P^1(\mathbb{R})$ is stable under the $\mathrm{SL}(2,\mathbb{Z})$-action — it is the orbit of $\tau = i\infty$. By definition,

$$\overline{\mathcal{M}} = (\mathfrak{h} \cup P^1(\mathbb{Q}))/\Gamma,$$

with $\mathfrak{h} \cup P^1(\mathbb{Q})$ topologized as a subset of the closed upper half plane. As a set, $\overline{\mathcal{M}} = \mathcal{M} \cup \{i\infty\}$. The added point is called the *cusp*. Its stabilizer under $\mathrm{SL}(2,\mathbb{Z})$ is the stabilizer of a point in $P^1(\mathbb{Q})$, i.e. it is the upper Borel subgroup. This group is a copy of $\mathbb{Z} \times \mathbb{Z}/2 = \langle T, S^2\rangle$. A weak modular form f will extend over this cusp only if $f(\tau)$ stays bounded as $\tau \to i\infty$, in which case f is called a *holomorphic modular form*. If f blows up no worse than q^{-k}, $k < \infty$, as $q \to 0$, then f is called *weakly holomorphic*. If $f(q \to 0) = 0$, then f is called *cuspidal*. By construction, $\mathcal{M}$ has the homotopy type of a $\mathrm{BSL}(2,\mathbb{Z})$, and so $\mathrm{Pic}(\mathcal{M}) = \mathbb{Z}/12$. But $\mathrm{Pic}(\overline{\mathcal{M}}) = \mathbb{Z}$: each line bundle on $\mathcal{M}$ has refinements over $\overline{\mathcal{M}}$ with different orders of vanishing or pole at the cusp.

Number theorists like to add yet another generalization. The action $\gamma \cdot \tau := \frac{a\tau+b}{c\tau+d}$ makes sense for any $\gamma \in \mathrm{PSL}(2,\mathbb{R})$. Indeed, this is the full symmetry group of $\mathfrak{h}$. Rather than using $\mathrm{SL}(2,\mathbb{Z})$, you could use some other subgroup $\Gamma \to \mathrm{PSL}(2,\mathbb{R})$. (Er, "subgroup" in quotes.) Let's generate an equivalence relation on "sub"groups $\Gamma \to \mathrm{PSL}(2,\mathbb{R})$, called *virtual sameness*, by declaring that Γ and Γ' are virtually the same as soon as $\Gamma \to \mathrm{PSL}(2,\mathbb{R})$ factors through Γ' and the map $\Gamma \to \Gamma'$ has finite index (=cokernel) and coindex (=kernel). So $\mathrm{Mp}(2,\mathbb{Z}) \to \mathrm{SL}(2,\mathbb{Z}) \to \mathrm{PSL}(2,\mathbb{Z})$ are all virtually the same. Number theorists typically, but not always, focus on modular groups Γ that are virtually the same as as $\mathrm{SL}(2,\mathbb{Z})$.

[This is part of standard group theory terminology, where you say a group is "virtually ADJECTIVE" if it is ADJECTIVE up to finite extensions. Another example: a semisimple Lie group is virtually the same as its connected simply-connected form. Note that if $\rho : \Gamma \to \Gamma'$ is a homomorphism, then the homotopy fibre of $\mathrm{B}\Gamma \to \mathrm{B}\Gamma'$ has $\pi_0 = \Gamma/\rho(\Gamma')$ and $\pi_1 = \ker(\rho)$ and no other homotopy groups. So working virtually is working modulo π-finite spaces.]

A function or form that's invariant under Γ is said to be *a modular form of level* Γ. If $\rho : \Gamma \to \mathrm{SL}(2,\mathbb{Z})$ is a finite-index subgroup, then modular forms of level Γ are easily described in terms of vector-valued modular forms: you use the permutation representation $\mathrm{SL}(2,\mathbb{Z})/\Gamma$ (also called the result of *induction* of the trivial representation of Γ along ρ). The main examples are:

- $\Gamma(N) := \ker(\mathrm{SL}(2,\mathbb{Z}) \to \mathrm{SL}(2,\mathbb{Z}/N))$. The permutation representation is, of course, the image of $\mathrm{SL}(2,\mathbb{Z}) \to \mathrm{SL}(2,\mathbb{Z}/N)$. In other words,
$$\Gamma(N) = \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}(2,\mathbb{Z}) : a \equiv d \equiv 1 \mod N, b \equiv c \equiv 0 \mod N \right\}.$$
- $\Gamma_1(N) \supset \Gamma(N) :=$ stabilizer of $(1,0) \in (\mathbb{Z}/N)^2$ (under the standard action of $\mathrm{SL}(2,\mathbb{Z})$ via $\mathrm{SL}(2,\mathbb{Z}/N)$). The permutation action is, of course, the orbit of $(1,0) \in (\mathbb{Z}/N)^2$ under this action. In other words,
$$\Gamma_1(N) = \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}(2,\mathbb{Z}) : a \equiv d \equiv 1 \mod N, c \equiv 0 \mod N \right\}.$$
- $\Gamma_0(N) \supset \Gamma_1(N) :=$ stabilizer of $[1,0] \in P^1(\mathbb{Z}/N)$ (under the standard action of $\mathrm{SL}(2,\mathbb{Z})$ via $\mathrm{PSL}(2,\mathbb{Z}/N)$). The permutation acton is, of course, the orbit of $[1,0] \in P^1(\mathbb{Z}/N)$ under this action. In other words,
$$\Gamma_0(N) = \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}(2,\mathbb{Z}) : c \equiv 0 \mod N \right\}.$$

These are called *congruence groups*. Normally, when people say that they have "a modular form of level N," they mean its level is one of these three groups. More generally, people call a subgroup of $\mathrm{SL}(2,\mathbb{Z})$ a "congruence subgroup" if it contains $\Gamma(N)$ for some N. (Not every finite-index subgroup of $\mathrm{SL}(2,\mathbb{Z})$ is congruence.)

However, note that to talk about virtual sameness, we didn't demand that the group be a subgroup of $\mathrm{SL}(2,\mathbb{Z})$. You could pick some $\Gamma \subset \mathrm{SL}(2,\mathbb{Z})$ of finite index, and then extend it, with finite index, to some larger group $\Gamma' \subset \mathrm{SL}(2,\mathbb{R})$ not contained in $\mathrm{SL}(2,\mathbb{Z})$. Number theorists almost always use $\Gamma' \supset \Gamma(N)$ for some N, as then a Γ'-modular form is in particular $\Gamma(N)$-modular and so of "level N."

The most important example is the following. Look at the matrix
$$W_N := \frac{1}{\sqrt{N}} \begin{pmatrix} 0 & 1 \\ -N & 0 \end{pmatrix} \in \mathrm{SL}(2,\mathbb{R}).$$
Then $W_N^2 = -1 \in \mathrm{SL}(2,\mathbb{R})$, so W_N acts on $\mathfrak{h}$ by an involution. Moreover, W_N normalizes $\Gamma_0(N)$: there is an index-2 extension of $\Gamma_0(N)$ to a group called "$\Gamma_0(N) + N$," which is not contained in $\mathrm{SL}(2,\mathbb{Z})$. More generally, a *Hall divisor* of N is a divisor $e|N$ such that e and $\frac{N}{e}$ are relatively prime; in other words, e is some subset of the distinct prime factors of N. The Hall divisors of N have a natural group structure, given by xor on subsets: in terms of divisors, it is $e * f = ef/\gcd(e,f)^2$. The unit is $1|N$, and this group is isomorphic to $(\mathbb{Z}/2)^{\text{distinct prime factors of } N}$. Since e and $\frac{N}{e}$ are relatively prime, we can find α, β such that $\alpha e - \beta \frac{N}{e} = 1$, and so the set of matrices of the form
$$\mathcal{W}_e := \left\{ \frac{1}{\sqrt{e}} \begin{pmatrix} ae & b \\ cN & de \end{pmatrix} \in \mathrm{SL}(2,\mathbb{R}) : a,b,c,d \in \mathbb{Z} \right\}$$
is nonempty. This set is a coset of $\Gamma_0(N) \subset \mathrm{SL}(2,\mathbb{R})$, and these cosets are closed under multiplication: $\mathcal{W}_e \mathcal{W}_f = \mathcal{W}_{e*f}$. Any element $W_e \in \mathcal{W}_e$ is called an *Atkin–Lehner involution*. There is an

interesting group

$$\Gamma_0(N)^+ := \bigcup_{\text{Hall divisors } e \text{ of } N} \mathcal{W}_e \subset \mathrm{SL}(2,\mathbb{R}).$$

It obviously contains $\Gamma_0(N)$ as a normal subgroup, and in fact $\Gamma_0(N)^+$ is the normalizer of $\Gamma_0(N)$ inside $\mathrm{SL}(2,\mathbb{R})$. Groups between $\Gamma_0(N)$ and $\Gamma_0(N)^+$ are given names like "$\Gamma_0(N) + e, f, \dots,$" where $e, f, \dots$ is some list of Hall divisors. Such a group is called *Fricke* or *non-Fricke* according to whether $N \in e, f, \dots$.

Suppose you have a finite-index subgroup $\Gamma \subset \mathrm{SL}(2,\mathbb{Z})$. As before, we can compactify $\mathfrak{h}/\Gamma =: \mathcal{M}(\Gamma)$ to

$$\overline{\mathcal{M}(\Gamma)} = (\mathfrak{h} \cup P^1(\mathbb{Q}))/\Gamma.$$

Now the *cusps* are the orbits of the action of Γ on $P^1(\mathbb{Q})$. Let's focus on $\Gamma = \Gamma_0(N)$. Consider the point $\frac{x}{y} \in P^1(\mathbb{Q})$, written in irreducible form. Define $g := \gcd(y, N)$. If $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(N)$, then $\gamma \cdot \frac{x}{y} = \frac{ax+by}{cx+dy}$ will define the same g, since $c \equiv 0 \pmod N$ and d is invertible $\pmod N$. So this g is a well-defined number for the cusp. A little bit more number theory (omitted — exercise, or see Diamond and Shurman) shows that the cusps for g are a torsor for $(\mathbb{Z}/\gcd(g, N/g))^\times$. So the total number of cusps for $\Gamma_0(N)$ is:

$$\#\text{cusps} = \sum_{g|N} \phi(\gcd(g, N/g)).$$

(This $\phi(-)$ is Euler's totient function, defined by $\phi(m) = \#(\mathbb{Z}/m)^\times$.) For example, of e is a Hall divisor, then there is only one cusp for e. Now let e be a Hall divisor, $\frac{x}{y} \in P^1(\mathbb{Q})$ a representative of a cusp, and $\gamma = \frac{1}{\sqrt{e}}\left(\begin{smallmatrix} ae & b \\ cN & de \end{smallmatrix}\right) \in \mathcal{W}_e$. Then $\gamma \cdot \frac{x}{y} = \frac{aex+by}{cNx+dey}$. Simplifying the fraction (which depends on $\gcd(y, e)$) and computing, we see that the new cusp is assigned a new divisor g, which is the "e-inverse" of g: there is a natural action of the group of Hall divisors on the set of all divisors, extending the permutation action, and this is how the cusps are permuted by $\Gamma_0(N)^+$. So for example if N is square-free, so that all divisors are Hall, then $\Gamma_0(N)^+$ has only one cusp.

1.2. ***G*-equivariant modular forms.** Let G be a finite group. Let $\mathcal{M}^G = \mathcal{M}^G(\mathbb{C})$ denote the moduli space of pairs consisting of an elliptic curve E_τ and a principle G-bundle $P : E_\tau \to \mathrm{B}G$. Note that $\pi_1 E_\tau = \mathrm{H}_1 E_\tau \cong \mathbb{Z}^2$, and a choice of $\tau \in \mathfrak{h}$ selects a basis for this group. So we can present $\mathcal{M}^G$ as a quotient space as follows:

$$\mathcal{M}^G = \frac{\mathfrak{h} \times \hom(\mathbb{Z}^2, G)}{\mathrm{SL}(2,\mathbb{Z}) \times G}$$

with action given by

$$\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}, g \right) \cdot \left(\tau, \begin{pmatrix} h \\ k \end{pmatrix} \right) = \left(\frac{a\tau + b}{c\tau + d}, \begin{pmatrix} gh^a k^b g^{-1} \\ gh^c k^d g^{-1} \end{pmatrix} \right),$$

where $h, k \in G$ commute (and so represent a map $\mathbb{Z}^2 \to G$) and $g \in G$ is arbitrary. [I am rapidly running out of letters!]

Let's call a holomorphic function on $\mathcal{M}^G$ a *G-equivariant modular function*. In other words, it is a function $f\left(\begin{smallmatrix} h \\ k \end{smallmatrix}\right)(\tau)$, where $\tau \in \mathfrak{h}$ and $h, k \in G$ commute, which is invariant under the action by $\mathrm{SL}(2,\mathbb{Z}) \times G$. More generally, a *$G$-equivariant modular form of weight w* is a function f on $\mathfrak{h} \times \hom(\mathbb{Z}^2, G)$ such that $f\left(\begin{smallmatrix} h \\ k \end{smallmatrix}\right)(\tau)\, (\mathrm{d}\tau)^{w/2}$ is $(\mathrm{SL}(2,\mathbb{Z}) \times G)$-invariant. There is an interesting generalization of this that incorporates a multiplier, explained in a paper by N. Ganter. Namely, pick $\alpha \in \mathrm{H}^3(G; \mathbb{C}^\times)$. Then α transgresses to a holomorphic line bundle $\mathcal{L}^\alpha$ over $\mathcal{M}^G$. A choice of cocycle for α will select a trivialization of $\mathcal{L}^\alpha$ over $\mathfrak{h} \times \hom(\mathbb{Z}^2, G)$. Sections of $\mathcal{L}^\alpha$ are *G-equivariant modular functions with anomaly α*. The ring of G-equivariant modular forms with anomaly α is $\mathrm{MF}^\alpha(\mathrm{B}G)$. [I'll write $\mathrm{B}G$ for the stack quotient of G. It knows G as a group.]

Let me briefly describe $\mathcal{L}^\alpha$. Take the trivial line bundle over $\mathfrak{h} \times \hom(\mathbb{Z}^2, G)$. To descend it to $\mathcal{M}^G$, we need to choose, for each $(\tau, \binom{h}{k}) \in \mathfrak{h} \times \hom(\mathbb{Z}^2, G)$ and each $(\gamma, g) \in \mathrm{SL}(2,\mathbb{Z}) \times G$, an isomorphism between the line over $(\tau, \binom{h}{k})$ and the line over $(\gamma \cdot \tau, g\gamma\binom{h}{k}g^{-1})$. An isomorphism of trivial lines is just a $\mathbb{C}^\times$-number. The number to use is computed as follows. Build the mapping cylinder for the isomorphism $\gamma : T^2 \cong E_\tau \cong E_{\gamma\cdot\tau} \cong T^2$. Give this cylinder a G-bundle with monodromies $\binom{h}{k}$ along the A- and B-cycles in E_τ, $g\gamma\binom{h}{k}g^{-1}$ along the A- and B-cycles in $E_{\gamma\cdot\tau}$, and monodromy g along the length of the cylinder. A 3-manifold M^3 with a G-bundle P is a thing to which (a cycle for) α assigns a number: the number is $\int_M P^*\alpha$. If M is closed, then this number depends only on the cohomology class of α; otherwise, under $\alpha \rightsquigarrow \alpha + \mathrm{d}\beta$, it changes by a boundary term $\int_{\partial M} P^*\beta$. In our case, the boundary of M is just some 2-tori. So the boundary term corresponds to changing the trivialization by $\int_{E_\tau} P^*\beta$, where β is a 2-cochain on G and P is the bundle on E_τ with monodromies $\binom{h}{k}$. For the usual bar complex for group cohomology, a 2-cochain is $\beta : G \times G \to \mathbb{C}^\times$, and $\int_{E_\tau} P^*\beta = \frac{\beta(h,k)}{\beta(k,h)}$. So for example if you instead use the normalized bar complex, then the trivialization of $\mathcal{L}^\alpha$, when restricted to the subspace where one of $k, h = 1$, does not depend on the cocycle.

Let's pick up a G-equivariant modular function $f \in \mathrm{MF}(\mathrm{B}G)$ and try to unpack it in non-equivariant language. Of course, f consists of some massive list of functions $f\binom{h}{k}$ indexed by $\binom{h}{k} : \mathbb{Z}^2 \to G$, each with some modularity, and also with lots of redundancy. First, obviously $f\binom{h}{k} = f\binom{ghg^{-1}}{gkg^{-1}}$, so that actually we care only about the G-conjugacy class of the map $\binom{h}{k} : \mathbb{Z}^2 \to G$. Second, changing $\binom{h}{k}$ by $\gamma \in \mathrm{SL}(2,\mathbb{Z})$ corresponds to acting on f by modular transformation γ, so in some sense you only care about the image of $\binom{h}{k} : \mathbb{Z}^2 \to G$, and not the choice of generators. But the generators are there to be able to talk about the actual function f: they are part of the trivialization over $\mathfrak{h}$, and modular forms are trivialized over $\mathfrak{h}$. So we'll keep them around.

Let's look first at $f\binom{h}{1}$, where we have set $k = 1$. This is a function of $\tau \in \mathfrak{h}$. It depends only on the conjugacy class of h, and moreover $f\binom{h}{1} = f\binom{h^{-1}}{1}$, since $\left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ acts trivially on τ. Let $N = \mathrm{order}(h)$. Pick $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. Then

$$f\binom{h}{1}(\tau) = f\binom{h^a}{h^c}(\gamma \cdot \tau).$$

In particular, if $c \equiv 0 \pmod N$ and $a \equiv 1 \pmod N$, then γ takes $f\binom{h}{1} \mapsto f\binom{h}{1}$; in other words, the function $f\binom{h}{1} : \mathfrak{h} \to \mathbb{C}$ is invariant under the group

$$\Gamma_1(N) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z}) \text{ s.t. } a \equiv 1 \pmod N \text{ and } c \equiv 0 \pmod N \right\}.$$

Note that, since $\det\gamma = 1 \equiv 1 \pmod N$, these conditions imply $d \equiv 1 \pmod N$ as well. The class of $b \mod N$ is unconstrained. So $f\binom{h}{1}$ is a modular form of level N. Since $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right) \in \Gamma_1(N)$, we do have

$$f\binom{h}{1}(\tau) = f\binom{h}{1}(\tau + 1).$$

In particular, $f\binom{h}{1}$ has a Fourier expansion in integer powers of q:

$$f\binom{h}{1}(\tau) = \sum_{n\in\mathbb{Z}} a_n(h)\, q^n.$$

Like the whole function, the coefficients $a_n(h)$ depend only on the conjugacy class of h, and moreover are invariant under $h \leftrightarrow h^{-1}$. In other words, for each n, the coefficients a_n define an element of

$$a_n \in \mathrm{RO}(G) \otimes \mathbb{C},$$

where $\mathrm{RO}(-)$ means the real representation ring

$$\mathrm{RO}(G) = \mathrm{K}_0\mathbf{Rep}_\mathbb{R}(G).$$

This is the end of the story if G is a cyclic group, but you can get a bit more modularity if G is nonabelian. Suppose that h is conjugate to h^a for some a. Then a is invertible mod N. Let $\gamma \in \mathrm{SL}(2,\mathbb{Z})$ be a lift of $\left(\begin{smallmatrix} a & 0 \\ 0 & a^{-1}\end{smallmatrix}\right) \in \mathrm{SL}(2,\mathbb{Z}/N)$. Then γ would take $f\left(\begin{smallmatrix} h \\ 1\end{smallmatrix}\right)$ to $f\left(\begin{smallmatrix} h^a \\ 1\end{smallmatrix}\right)$, but that's equal to $f\left(\begin{smallmatrix} h \\ 1\end{smallmatrix}\right)$ since h and h^a are by assumption conjugate. In particular, if h is conjugate to all of its primitive powers, then $f\left(\begin{smallmatrix} h \\ 1\end{smallmatrix}\right)$ will be modular for the group

$$\Gamma_0(N) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z}) \text{ s.t. } c \equiv 0 \pmod N \right\}.$$

In general, two elements $h_1, h_2 \in G$ are called *algebraic-conjugate* if they are conjugate up to primitive powers, i.e. if h_1 and gh_2g^{-1} generate the same cyclic group for some $g \in G$. A finite group is called *rational* when every element is conjugate to all of its primitive powers: when algebraic conjugacy implies conjugacy. The set of conjugacy classes in G is $G/G = \hom(\mathbb{Z}, G)/G$, and functions on this set are

$$\mathcal{O}(\{\text{conjugacy classes}\}) \cong \mathrm{R}(G) \otimes \mathbb{C}, \qquad \mathrm{R}(G) := \mathrm{K}_0\mathbf{Rep}_{\mathbb{C}}(G).$$

Since G is finite, the restriction map $\hom(\mathbb{Z}, G) \leftarrow \hom(\widehat{\mathbb{Z}}, G))$ is an iso, where $\widehat{\mathbb{Z}} \cong \prod_p \mathbb{Z}_p$ is the ring of profinite integers. The algebraic conjugacy classes are $\hom(\mathbb{Z}, G)/(\widehat{\mathbb{Z}}^\times \times G)$, and functions on this set are

$$\mathcal{O}(\{\text{algebraic conjugacy classes}\}) \cong \mathrm{R}\mathbb{Q}(G) \otimes \mathbb{C}, \qquad \mathrm{R}\mathbb{Q}(G) := \mathrm{K}_0\mathbf{Rep}_{\mathbb{Q}}(G).$$

So a rational group is one for which $\mathrm{R}\mathbb{Q}(G) \to \mathrm{R}(G)$ becomes an isomorphism after tensoring with $\mathbb{C}$, or equivalently after tensoring with $\mathbb{Q}$. This is equivalent in turn to asking that every character of G take rational values.

Very few groups are rational. I mean, symmetric groups are, but rationality is pretty rare, and the Monster $\mathbb{M}$ is not rational.

But we don't need it to be. Since $-1 \in \mathrm{SL}(2,\mathbb{Z})$ acts trivially on $\mathfrak{h}$, to get modularity for $\Gamma_0(N)$ rather than $\Gamma_1(N)$, it suffices if every primitive power of every $g \in G$ is conjugate to either g or g^{-1} (or both). A finite group is called *inverse semi-rational* [an ugly name] if it has this property. The same property is also called *cut* [an unintuitive name]. Conjugacy classes up to inversion are $\hom(\mathbb{Z}, G)/(\mathbb{Z}^\times \times G)$, and functions on this set are

$$\mathcal{O}(\{\text{conjugacy-up-to-inversion classes}\}) \cong \mathrm{RO}(G) \otimes \mathbb{C}, \qquad \mathrm{RO}(G) := \mathrm{K}_0\mathbf{Rep}_{\mathbb{R}}(G).$$

So G is inverse semi-rational when $\mathrm{R}\mathbb{Q}(G) \to \mathrm{RO}(G)$ is an isomorphism after tensoring with $\mathbb{Q}$; equivalently, if every real character is rational.

Inverse semi-rationality is slightly more common than rationality, but they are still rare. Using the classification of finite simple groups, in 2021 by Bächle, Caicedo, Jespers, Maheshwary gave a complete classification of which finite simple groups are inverse semi-rational. Their complete list is:

$$\begin{gathered} \mathrm{C}_2, \quad \mathrm{C}_3, \\ \mathrm{A}_7, \quad \mathrm{A}_8, \quad \mathrm{A}_9, \quad \mathrm{A}_{12}, \\ \mathrm{L}_2(7), \quad \mathrm{U}_3(3), \quad \mathrm{U}_3(5), \quad \mathrm{U}_4(3), \quad \mathrm{U}_5(2), \quad \mathrm{U}_6(2), \quad \mathrm{S}_4(3), \quad \mathrm{S}_6(2), \quad \mathrm{O}_8^+(2), \\ \mathrm{M}_{11}, \quad \mathrm{M}_{12}, \quad \mathrm{M}_{22}, \quad \mathrm{M}_{23}, \quad \mathrm{M}_{24}, \quad \mathrm{Co}_1, \quad \mathrm{Co}_2, \quad \mathrm{Co}_3, \quad \mathrm{HS}, \quad \mathrm{McL}, \quad \mathrm{Th}, \quad \mathbb{M}. \end{gathered}$$

Moonshiners will recognize some of their favourite friends. Of course, almost-simple groups can be inverse semi-rational even if their socles are not (e.g. symmetric groups), and moonshiners also like almost-simple groups, perfect abelian extensions of simple groups, etc. So this is not quite a complete list of moonshine-friendly groups. But still.

Anyway, for these groups G, the $\left(\begin{smallmatrix} h \\ 1\end{smallmatrix}\right)$-components of a G-equivariant modular are automatically $\Gamma_0(\mathrm{order}(h))$-modular, but for other groups they are not. I suspect that, if you really wanted to get a good theory, you should try to build some version of equivariant modular forms with Γ_0-modularity built in: some "rational subring" of $\mathrm{MF}(\mathrm{B}G)$. Maybe you should call these "algebraic modular forms" in analogy with algebraic K-theory, since $\mathrm{R}\mathbb{Q}(G)$ is the 0th algebraic K-group of the group algebra $\mathbb{Q}[G]$. Comparing to the presentation of $\mathcal{M}^G$ as a quotient, this "algebraic MF" should involve upgrading the modular group from $\mathrm{SL}(2,\mathbb{Z})$ to $\mathrm{SL}(2,\widehat{\mathbb{Z}})$. Comparing to the lectures of Ganter in this workshop, it probably involves something about Adams operations, since those act on $\mathrm{R}(G) = \mathrm{K}(\mathrm{B}G)$ by power operations on the character table, equivalently by the Galois group of $\mathbb{Q}$ (which acts through its abelianization, which is of course $\widehat{\mathbb{Z}}^\times$).

There is one final enhancement of modularity that appears in Moonshine, and which I do not know how to describe in terms of $\mathrm{MF}(\mathrm{B}G)$. Namely, it often, but not always, happens that $f\binom{h}{1}$ is invariant under certain *Atkin–Lehner involutions*, which are matrices in $\mathrm{PGL}(2,\mathbb{Q})$ but not in $\mathrm{PSL}(2,\mathbb{Z})$.

[Aside: $\mathrm{PGL}(2,\mathbb{R})$ does not quite act on the upper half plane $\mathfrak{h}$ — it has two components, and the identity component is $\mathrm{PSL}(2,\mathbb{R})$ and does act on $\mathfrak{h}$, whereas the nonidentity component reverses the sign of $\Im(\tau)$. In general, there is a natural map $\mathrm{PGL}(n,\mathbb{K}) \to \mathrm{H}^1(\mathbb{K};\mu_n) = \mathbb{K}^\times/(\mathbb{K}^\times)^n$, given by the class of the determinant. For $n=2$ and $\mathbb{K}=\mathbb{R}$, there is an isomorphism $\mathbb{R}^\times/(\mathbb{R}^\times)^2 = \{+,-\}$. The Atkin–Lehner involutions live in $\mathrm{PGL}^+(2,\mathbb{Q}) = \mathrm{PGL}(2,\mathbb{Q}) \times_{\mathrm{PGL}(2,\mathbb{R})} \mathrm{PSL}(2,\mathbb{R})$.]

Specifically, for each divisor $e|N$, consider the set of matrices

$$\mathcal{W}_e = \left\{ \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{ s.t. } b \in \mathbb{Z}, a, d \in e\mathbb{Z}, c \in N\mathbb{Z}, \det\gamma = e \right\}.$$

This set is nonempty if and only if e is a *Hall divisor*, i.e. $\gcd(e, \frac{N}{e}) = 1$. The Hall divisors are in obvious bijection with subsets of the prime factors of N. $\mathcal{W}_e$ is coset of $\Gamma_0(N)$, and it normalizes, and $\mathcal{W}_e\mathcal{W}_{e'} = \mathcal{W}_{e \operatorname{xor} e'}$, where the group operation $e \operatorname{xor} e'$ is exclusive-or of subsets. Note that $e \operatorname{xor} e = 1$, so, modulo $\Gamma_0 N$, elements of $\mathcal{W}_e$ are involutions. This explains the name *Atkin–Lehner involution*. Given some (generators for a) subgroup of the Hall divisors $e, e', \dots$, people write "$\Gamma_0(N) + e, e', \dots$" for the extension of $\Gamma_0(N)$ through these cosets. The original Monstrous Moonshine involves some complicated rules that tell you, for each h, which Hall divisors $e, e', \dots$ to use so that $f\binom{h}{1}$ is modular for $\Gamma_0(\mathrm{order}(h)) + e, e', \dots$, and then it turns out that there is no larger modularity for these functions. When all AL involutions are included, you call the group $\Gamma_0(N)^+$.

Work of Paquette, Persson, and Volpato hints that there is some connection between these Atkin–Lehner involutions and some extension of $G = \mathbb{M}$ by "noninvertible symmetries," specifically "duality defects of Tambara–Yamagami type." But that extension is definitely not the end of the story: we know myriad examples of things with TY-type duality defects, and most of them do not have extended modularity.

For the remainder of these lectures, I'm going to largely ignore these Atkin–Lehner involutions, and, other than a few comments here or there, pretend as if Moonshine was just about modularity for groups explainable by $\mathrm{MF}(\mathrm{B}G)$. The special case $e = N$ of an Atkin–Lehner involution is called the *Fricke involution*, and a group $\Gamma_0(N) + e, e', \dots$ is called *non-Fricke* if $N \notin e, e', \dots$. Correspondingly, an element $h \in \mathbb{M}$ is called *(non-)Fricke* if its assigned group $\Gamma_0(\mathrm{order}(h)) + e, e', \dots$ is (non-)Fricke. The rule for assigning AL involutions to h involves various powers of h, and the only way that there are no AL involutions is if all powers of h are non-Fricke. So the remaining lectures will largely focus the *entirely non-Fricke* part of Moonshine.

1.3. **Lunarity and optimal growth.** What makes moonshine *moonshine* is the genus-zero phenomenon. The classical story goes as follows. Suppose you have a (nonconstant) holomorphic function $f : \mathfrak{h} \to \mathbb{C}$; let's assume it is, at least, a modular form of level N for some large N. Let $\Gamma(f) \supset \Gamma(N)$ be the stabilizer of f under the $\mathrm{SL}(2,\mathbb{R})$-action. By definition, f descends to a function $\mathfrak{h}/\Gamma(f) \to \mathbb{C}$. Since f was at least $\Gamma(N)$-modular, $\mathfrak{h}/\Gamma(f)$ has a compactification

$\overline{\mathfrak{h}/\Gamma(f)} = (\mathfrak{h} \cup P^1(\mathbb{Q}))/\Gamma(f)$. A modular function f is called *weakly holomorphic* if it is merely meromorphic near the infinite points in this compactification. In this case, f defines a regular function $\overline{\mathfrak{h}/\Gamma(f)} \to P^1(\mathbb{C})$. The function f is called a *hauptmodul* when this $\overline{\mathfrak{h}/\Gamma(f)} \to P^1(\mathbb{C})$ is an isomorphism. Note that this implies, but is slightly stronger than, asking that the punctured curve $\mathfrak{h}/\Gamma(f)$ have genus one. [Following the German, the plural of "hauptmodul" is "hauptmoduln": "ox"⇝"oxen," "brother"⇝"brethren," "child"⇝"children," "sow"⇝"swine," ...] Conversely, a subgroup $\Gamma \subset \mathrm{SL}(2, \mathbb{R})$ is called *genus zero* if $\mathfrak{h}/\Gamma$ has genus one, i.e. if $\Gamma = \Gamma(f)$ for some hauptmodul.

There are only finitely many hauptmoduln, up to the obvious actions like rescaling and adding constants. For example, Ogg showed that a prime p is *supersingular* (in the sense that every supersingular elliptic curve in characteristic p can be defined over $\mathbb{F}_p$) if and only if $\Gamma_0(p)^+ = \Gamma_0(p) + p$ is genus zero. He also knew (showed?) that this occurs exactly for the primes

$$p = 2,\ 3,\ 5,\ 7,\ 11,\ 13,\ 17,\ 19,\ 23,\ 29,\ 31,\ 41,\ 47,\ 59,\ 71.$$

Ogg was at a lecture by Tits, in which the latter wrote up the prime factorization of $\mathbb{M}$. Ogg immediately recognized his list, and this is what started the whole Moonshine game. Ogg's *Jack Daniels Problem*, only partially solved, is to explain why $p | \#\mathbb{M}$ if and only if $\Gamma_0(p)^+$ is genus zero. The "only if" direction is fundamental to the statement of Monstrous Moonshine, but the "if" direction remains largely unexplained (although Duncan and Ono have some partial story). If, as I intend to do, we throw away Atkin–Lehner involutions, then genus-zero groups are even harder to come by: $\Gamma_0(N) + e, e', \dots$ has an easier time be being genus-zero than plain $\Gamma_0(N)$ does, since it involves a further quotient of the corresponding modular curve. The list of N for which $\Gamma_0(N)$ has genus zero is:

$$N = 1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7,\ 8,\ 9,\ 10,\ 12,\ 13,\ 16,\ 18,\ 25.$$

Suppose that f is $\Gamma_0(N)$-modular. When is f a hauptmodul with $\Gamma_0(N)$? Following Cheng and Duncan, we can translate the question into one about growth rates. Indeed: by Riemann–Roch, $f : \overline{\mathfrak{h}/\Gamma_0(N)} \to P^1(\mathbb{C})$ is an isomorphism exactly when f has a single simple pole. Let's assume that f does grow near $q = 0$. Then that must be the pole. So f is an isomorphism if and only if the pole at $q = 0$ is simple, and f is bounded at all other cusps of $\Gamma_0(N)$.

This growth condition makes sense for other weights, modularities, etc. In general, Cheng and Duncan say that a weakly holomorphic Γ-modular form has *optimal growth* if its growth is as slow as possible given its weight and modularity. This turns out to be a good translation of the genus-zero property to other types of moonshine, so it's the one I'll focus on.

To find a good version of optimal growth for G-equivariant modular forms, we should first work out the cusps of $\mathcal{M}^G$. When an elliptic curve E degenerates towards a cuspidal curve, one of its cycles gets very small. If E carries a principle G-bundle P, then this vanishing cycle will carry some monodromy k, well defined up to conjugation and inversion — in other words, $k \in G/(G \times \mathbb{Z}^\times)$. Let $C(k) \subset G$ denote the centralizer of $k \in G$. The other monodromy of P will be some element $h \in C(k)$, well-defined up to conjugatation by $C(k)$ and up to multiplication by k. In other words, the cusps are $\left(\begin{smallmatrix} h \\ k \end{smallmatrix}\right)$ where $k \in G/(G \times \mathbb{Z}^\times)$ and $h \in (C(k)/\langle k \rangle)/(C(k)/\langle k \rangle)$.

There is a more stacky way to say all this. The cusps of $\mathcal{M}^G$ are the fibre of $\overline{\mathcal{M}^G} \to \overline{\mathcal{M}}$ over $\tau = i\infty$. Selecting this τ breaks the $\mathrm{SL}(2, \mathbb{Z})$-symmetry to a stackiness of $\langle T, \pm 1 \rangle \cong \mathbb{Z} \times \mathbb{Z}/2$. So the cusps are classified by

$$\hom(\mathbb{Z}^2, G)/(\langle T, \pm 1 \rangle \times G).$$

Note that the choice $\left(\begin{smallmatrix} h \\ k \end{smallmatrix}\right)$ breaks the stackiness further: in particular, it breaks the stackiness by $\langle T \rangle$ down to $\langle T^{\mathrm{order}\, k} \rangle$.

A bit of language. The monodromy k around the vanishing cycle is called the *twisting*. The other monodromy h is called the *twining*.

Recall that a function on G/G is precisely a class in $\mathrm{R}(G)\otimes\mathbb{C} = \mathrm{K}^0(\mathrm{B}G)\otimes\mathbb{C}$. A function on $\hom(\mathbb{Z}^2,G)/G$ is a class in

$$\mathrm{KU}^0(L\mathrm{B}G)\otimes\mathbb{C}$$

where of course $L\mathrm{B}G = G/G$ in the stacky sense. The further quotient by ±1 replaces KU with KR, the "capital-R Real" K-theory. Of course, the Fourier expansion of a modular form near a cusp is not just a single number, but a q-series. In terms of loop spaces, q encodes an action of S^1 rotating the loop. All together, Fourier expansion, aka q-expansion, for G-equivariant modular functions is almost a map

$$\mathrm{MF}(\mathrm{B}G) \to \mathrm{KR}(L\mathrm{B}G/S^1)\otimes\mathbb{C} \approx \mathrm{KR}(L\mathrm{B}G)\otimes\mathbb{C}((q)).$$

This is not quite correct, but almost is: the rotation/S^1 action gets mixed into the the rest of the structure a bit. Very explicitly, the RHS is the following. Let me first describe the version with KR replaced by KU, i.e. where we just care about the S^1 aka T-symmetry and not the ±1 symmetry. Let $k\in G$ select a point $[k]\in L\mathrm{B}G$; suppose $\mathrm{order}(k)=N$. Its connected component is a $\mathrm{B}C(k)$. A class in $\mathrm{KU}(\mathrm{B}C(k))\otimes\mathbb{C}$ restricts to a class in $\mathrm{KU}(\mathrm{B}\langle k\rangle)\otimes\mathbb{C} = \mathbb{C}[t]/(t^N=1)$. We want those classes in $\mathrm{KU}(\mathrm{B}C(k))[q^{1/N}]\otimes\mathbb{C}$ whose restriction to $\langle k\rangle$ lives in $\mathbb{C}[tq^{1/N}]\subset\mathbb{C}[t,q^{1/N}]/(t^N=1)$. Now take a direct sum over conjugacy classes k to get the KU version of the RHS. To get the actual RHS, also quotient from KU to KR.

As indicated in the previous section, I would actually prefer to conflate conjugacy-up-to-inversion with algebraic-conjugacy, and when these are different, I'd prefer to use the latter. So I would prefer if the codomain of this q-expansion map were not built from KR, but rather the following. Since G is finite,

$$L\mathrm{B}G = \hom(\mathrm{B}\mathbb{Z},\mathrm{B}G)\cong\hom(\mathrm{B}\widehat{\mathbb{Z}},\mathrm{B}G) = \widehat{L}\mathrm{B}G$$

where you might think of $\mathrm{B}\widehat{\mathbb{Z}}$ as an object of étale homotopy rather than homotopy. Thus this space carries a natural action by $\widehat{Z}^\times$. This also acts on the characters of all finite-group representations, and so there should be a "KQ" that imposes equivariance for this action — it should be to the algebraic K-theory of $\mathbb{Q}$ what KR is to KO. Then the codomain of q-expansion should look more like "$\mathrm{KQ}(\widehat{L}\mathrm{B}G/\mathrm{B}\widehat{\mathbb{Z}})\otimes\mathbb{C}$," with a little bit of adjustment in the q-degree.

The Monstrous Moonshine example tells us that we want functions $f\binom{h}{k}(\tau)$ such that $f\binom{h}{1}(\tau)\sim q^{-1}+0+O(q)$ as $\tau\to i\infty$. When $k\neq1$, it depends a bit on h, and the AL involutions get in the way, but it is simple in the entirely non-Fricke case: $f\binom{h}{k}(\tau)\sim O(1)$ if $k\neq1$.

All together, we see the following. There is a well-defined map

$$\{\text{cusps of }\mathcal{M}^G\}\to L\mathrm{B}G/\widehat{\mathbb{Z}}^\times, \qquad \binom{h}{k}\mapsto[k].$$

The fibre over $\mathrm{B}G\to L\mathrm{B}G$ are the *infinite cusps* and the rest of the cusps are *finite*: in other words, a cusp represented by $\binom{h}{k}$ is infinite or finite according where $k=1$ or $k\neq1$. A G-equivariant modular function $f:\mathcal{M}^G\to\mathbb{C}$ is *lunar* if it grows like $O(q^{-1})$ at the infinite cusps and like $O(1)$ at the finite cusps, and if moreover the residue at the infinite cusps $\binom{h}{1}$ is independent of h.

I suggest that G-equivariant lunar forms are what Moonshine is about.

1.4. **Topological lunar forms.** Topological modular forms TMF are a derived integral refinement of MF. Here are the main ingredients of the definition. There is a derived moduli stack $\mathcal{M}^{\mathrm{der}}$ of "elliptic curves over the sphere spectrum." It carries a sheaf $\mathcal{O}^{\mathrm{der}}$ of spectra whose fibre at an elliptic curve E is E-elliptic cohomology, and TMF is by definition the spectrum of global sections of $\mathcal{O}^{\mathrm{der}}$. If X is a space, then you are supposed to consider the derived stack $\mathcal{M}^{X,\mathrm{der}}$ of pairs consisting of an elliptic curve E together with a map $E\to X$ factoring through a formal neighbourhood of X. Restrict $\mathcal{O}^{\mathrm{der}}$ along $\mathcal{M}^{X,\mathrm{der}}\to\mathcal{M}^{\mathrm{der}}$ and take its global sections; this produces $\mathrm{TMF}(X)$.

The homotopy groups of $\mathcal{O}^{\mathrm{der}}$ are: $\pi_{\mathrm{odd}}\mathcal{O}^{\mathrm{der}}=0$; $\pi_{2n}\mathcal{O}^{\mathrm{der}}|_E = \mathrm{Lie}(E)^{\otimes n} =: \omega^n$. So a section of $\pi_\bullet\mathcal{O}^{\mathrm{der}}$ is a modular form of weight $\bullet/2$. This supplies an elliptic spectral sequence $\mathrm{H}^\bullet(\mathcal{M}_{\mathbb{Z}},\omega^{\bullet/2})\Rightarrow\pi_\bullet\mathrm{TMF}$. In particular, $\mathrm{TMF}\otimes\mathbb{C}=\mathrm{MF}$ (with homological degree $\approx$ weight).

The equivariant version of TMF is defined roughly as follows. Fix a group G, which for me will always be finite. Consider the moduli stack $\mathcal{M}^{\mathrm{B}G,\mathrm{der}}$ of pairs (E,P) where $E \in \mathcal{M}^{\mathrm{der}}$ is a derived elliptic curve and $P : \mathrm{H}_1\, E \to G$, where the homology theory is such that $\mathrm{H}_1\, E \cong \mathbb{Z}^2$ no matter who E is. The point is that fundamental groups and G-bundles and so on for derived elliptic curves, and indeed for (singular) elliptic curves over other bases, are complicated. So we use maps $\mathrm{H}_1\, E \to G$ in place of G-bundles. Over a field of characteristic 0, these are the same except for something about the Galois actions that I never wrapped my head around. I publically hope that sometime this week one of the experts in the audience will explain this to me. I secretly hope that this will also deal with the thing about KR versus KQ that I spoke about in the previous sections, but I'm not optimistic on that front. Pull $\mathcal{O}^{\mathrm{der}}$ back along $\mathcal{M}^{\mathrm{der}} \to \mathcal{M}^{G,\mathrm{der}}$ and take its sections. This gives $\mathrm{TMF}_G = \mathrm{TMF}(\mathrm{B}G)$. If X is a space with a G action, then you are supposed to think about $\mathrm{TMF}_G(X) = \mathrm{TMF}(X/G)$ as having to do with elliptic curves which explore the stack X/G, but can only be "macroscopic" in the stacky/G directions — the elliptic curves have to be "microscopiuc" in the X directions. This is so that you do indeed get something that depends locally in X, and if you do it right you will end up with a genuinely equivariant cohomology theory.

The upshot is that TMF lets us drop those distracting $\otimes\mathbb{C}$'s from the previous discussion. Specifically, there is a q-series expansion

$$\mathrm{TMF} \to \mathrm{KO}((q))$$

of nonequivariant cohomology theories which enhances to a map

$$\mathrm{TMF}(\mathrm{B}G) \to \mathrm{KR}_{S^1}(L\mathrm{B}G),$$

where the codomain is slightly twisted in a way I tried to sketch above, but that someone really should spell out. Either through some magic about the group G or through modifying the domain, I'd really like this to end up in "KQ" — some version of algebraic K-theory — rather than KR, so let's just talk about "algebraic TMF" as if it exists and has such a map. As briefly mentioned above, maybe "algebraic TMF" means something about Hecke eigenfunctions.

Derived algebraic geometry is far enough along that one can talk about divisors, poles, etc. The compactifications $\overline{\mathcal{M}}$ and $\overline{\mathcal{M}^G}$ admit derived refinements, and $\mathcal{O}^{\mathrm{der}}$ extends over them — at the cuspidal point $\tau = i\infty \in \overline{\mathcal{M}}$, the fibre of $\mathcal{O}^{\mathrm{der}}$ is KU, which reduces to KO when you take into account the $\mathbb{Z}/2$-automorphism of the cusp. (For "algebraic TMF", the cusp should reduce to KQ, because the stackiness should be $\widehat{\mathbb{Z}}^\times$ rather than $\mathbb{Z}^\times$.) I think the cusps of $\overline{\mathcal{M}^{G,\mathrm{der}}}$ are a copy of $L\mathrm{B}G$, and I think $\mathrm{B}G \hookrightarrow L\mathrm{B}G$ is a divisor. So I think you can define:

Given a finite group G, the *Topological lunar forms* Tlf_G are the sections of $\mathcal{O}^{\mathrm{der}}(-(\mathrm{B}G))$ over $\overline{\mathcal{M}^{G,\mathrm{der}}}$. In other words, they are the topological modular forms with (at most) a simple pole along the divisor $\mathrm{B}G \hookrightarrow L\mathrm{B}G$. More generally, if X is a space with a G-action, then $\mathrm{Tlf}_G(X)$ looks like $\mathrm{Tmf}(X/G)$ — functions of pairs $(E, f : E \to X/G)$ which treat E as "microscopic" compared to X — except with a simple pole allowed along those curves (E,f) in which the stackiness trivializes along one cycle.

2. Topological mock modularity

Summary: A generalized mock modular form with shadow s is precisely a nullhomotopy, in a certain cochain model for TMF, of a class constructed from s. This gives a derived/homotopical origin of mock modularity. Number theorists and harmonic analysts care about "pure" mock modular forms which is when the shadow s is purely antiholomorphic. Mock modular moonshine is specifically about pure mock modular forms of weight $w = 1/2$. The (Segal–)Stolz–Teichner proposal says that each nice enough $N = (0,1)$ 2D SQFT supplies a TMF class; "null" means that supersymmetry spontaneously breaks. Pure mock modularity of weight $1/2$ corresponds to having an $N = (1,1)$ SQFT which is nullhomotopic if you forget the left-moving susy. The Bunke–Naumann "secondary Witten genus" makes sense for all SQFTs, not just sigma models. Its recipe is: compute the 1-point function s of the supercurrent; find a (generlized) mock modular form with shadow s; take the class of that mock modular form modulo $\mathrm{MF} + \mathbb{Z}((q))$.

2.1. **Classical mock modular forms.** I first review the standard story. Probably I won't say any of this in my lectures. This story is more or less due to Zwegers.

A *modular function* was a holomorphic function on $\mathcal{M} = \mathcal{M}(\mathbb{C}) = \mathfrak{h}/\mathrm{SL}(2,\mathbb{Z})$, satisfying some analytic condition near $\tau = i\infty$ ($q = 0$). As a space, $\mathcal{M}$ is a (finitely stacky) complex curve, so a real surface. We can of course consider C^∞ or analytic (C^ω) functions on $\mathcal{M}$. People normally assume analyticity from the get go, so I will too. Unsurprisingly, such functions are called *real-analytic modular functions*. More generally, a *real-analytic modular form* of weight $(w,\bar{w})$ if a function $f(\tau,\bar\tau)$ on $\mathfrak{h}$ such that $f(\tau,\bar\tau)(\mathrm{d}\tau)^{w/2}(\mathrm{d}\bar\tau)^{w/2}$ descends to $\mathcal{M}$. In other words, f should transform as

$$f(\gamma\cdot\tau) = (c\tau+d)^w(c\bar\tau+d)^{\bar w}f(\tau), \qquad \gamma = \begin{pmatrix} a & b \\ c & d\end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z}).$$

An important, and fun, example is $\Im(\tau) := (\tau-\bar\tau)/2i$. Indeed:

$$\begin{aligned}\Im(\gamma\cdot\tau) &= \frac{1}{2i}\left(\frac{a\tau+b}{c\bar\tau+d} - \frac{a\bar\tau+b}{c\tau+d}\right) = \frac{1}{2i}\frac{(a\tau+b)(c\bar\tau+d)-(a\bar\tau+b)(c\tau+d)}{(c\tau+d)(c\bar\tau+d)} \\ &= \frac{1}{2i}(c\tau+d)^{-1}(c\bar\tau+d)^{-1}\big((ac\tau\bar\tau+bc\bar\tau+ad\tau+bd)-(ac\tau\bar\tau+bc\tau+ad\bar\tau+bd)\big) \\ &= \frac{1}{2i}(c\tau+d)^{-1}(c\bar\tau+d)^{-1}(ad-bc)(\tau-\bar\tau) = (c\tau+d)^{-1}(c\bar\tau+d)^{-1}\Im(\tau)\end{aligned}$$

since $ad-bc=1$. In other words, $\Im(\tau)$ transforms as a C^∞ modular form of weight $(-1,-1)$. As such, we can move weight between the holomorphic and antiholomorphic sides by multiplying by powers of $\Im(\tau)$ — the controlling factor is just $w-\bar w$. Note that this adjusts the cuspidal behaviour, but not by much: as a function of $q = \exp(2\pi i\tau)$, we have $\Im(\tau) = \frac{1}{2\pi}\log(|q|)$.

Suppose you have a C^ω modular form f. You can ask: is it holomorphic? If not, by how much is it off? The answer, of course, is measured by derivative $\partial_{\bar\tau}f := \frac{\partial f}{\partial\bar\tau}$. To make it sound like I'm doing physics, I'll call this the *holomorphic anomaly* of f. If f had modular weight $(w,\bar w)$, then $\partial_{\bar\tau}f$ will be modular of weight $(w,\bar w+2)$.

A function f is *harmonic* if $\partial_\tau\partial_{\bar\tau}f = 0$, i.e. if $\partial_{\bar\tau}f$ is purely antiholomorphic. Suppose that f is harmonic and transforms as a modular function (weight $(0,0)$). Then the complex conjugate $\overline{\partial_{\bar\tau}f}$ will be a (weakly) holomorphic modular form of weight 2. More generally, the modular form folks like to move all the weight to the holomorphic side, and don't really like antiholomorphic things. So the convention is to say: start with f of weight $(w,0)$; then $\partial_{\bar\tau}f$ has weight $(w,2)$; so $\overline{\partial_{\bar\tau}f}$ has weight $(2,w)$; so $\Im(\tau)^w\overline{\partial_{\bar\tau}f}$ has weight $(2-w,0)$. We'll call f a *harmonic Maass form* if this $\Im(\tau)^w\overline{\partial_{\bar\tau}f}$ is holomorphic, i.e. if f is annihilated by the "Laplace" operator $\partial_\tau\Im(\tau)^w\partial_{\bar\tau}$. (Maass was more generally interested in eigenvectors for this Laplacian.) This modified holomorphic anomaly $\Im(\tau)^w\overline{\partial_{\bar\tau}f}$ is called the *shadow* of the harmonic Maass form f. Even if it's not holomorphic (so that f is not harmonic) I'll still call it the *shadow*.

Remark: This Laplacian seemed a bit ad hoc. Here's another description. Under some growth rate condition, the space of C^ω modular forms of weight $(w,0)$ is (pre)Hilbert, with Hilbert pairing

$$\langle f_1,f_2\rangle = \int_{\mathcal{M}} \overline{f_1}f_2\,\Im(\tau)^{w-2}\,\mathrm{d}\tau\mathrm{d}\bar\tau$$

Indeed, $\overline{f_1}f_2$ has weight (w,w), whereas volume forms on $\mathcal{M}$ have weight $(2,2)$, so you need to adjust. In particular, you could consider the pairing

$$\langle\partial_{\bar\tau}f_1,\partial_{\bar\tau}f_2\rangle = \int_{\mathcal{M}}(\partial_\tau\overline{f_1})(\partial_{\bar\tau}f_2)\Im(\tau)^w\,\mathrm{d}\tau\mathrm{d}\bar\tau.$$

Now integrate by parts: this is equal up to a sign to

$$\langle f_1, (\partial_{\bar\tau} \Im(\tau)^w \partial_{\bar\tau}) f_2 \rangle.$$

So this is where Maass's Laplacian comes from.

Suppose you really really like holomorphic functions. Perhaps you like them so much that, when handed a harmonic function, you will just immediately try to work with its holomorphic part. In other words, you want to subtract off the $\bar\tau$-dependency. Well, you know that f solves an inhomogeneous differential equation

$$\partial_{\bar\tau} f(\tau, \bar\tau) = \Im(\tau)^{-w} \bar s(\bar\tau, \tau).$$

You'd really like to work with solutions to the corresponding homogeneous equation. Of course, these are the same if you can select a reference inhomogeneous solution. And you can if you can choose a distinguished basepoint for $\bar\tau$. We do have one: $\bar\tau = \overline{i\infty} = -i\infty$. So we can try to modify f to

$$f_{\text{holo}}(\tau) = f(\tau, \bar\tau) - (2i)^w \int_{\bar z = -i\infty}^{\bar\tau} (\tau - \bar z)^{-w}\, \bar s(\bar z, \tau)\, \mathrm{d}\bar z$$

which looks marginally cleaner if s is holomorphic, so that $\bar g(\bar z, \tau) = \overline{s(z)}$ does not depend on depend on τ. Formally, you should think that what you've done is to analytically continue f to the $\tau, \bar\tau$ plane, and then take

$$f_{\text{holo}}(\tau) = \text{“} \lim_{\bar\tau \to -i\infty} f(\tau, \bar\tau). \text{”}$$

Of course, whether the integral (or, for that matter, the limit) makes sense depends on some assumptions about growth rates: we need the combination $(\tau - \bar z)^{-w}\, \bar s(\bar z, \tau)$ to decay sufficiently quickly as $\bar z \to -i\infty$. From undergraduate calculus, we know that "inverse quadratic" is sufficiently quick. So we need something like $s(z, \bar\tau) = O(z^{w-2})$ as $z \to i\infty$. You might also worry that about a pole at $\bar z = \tau$, since the value of the integral would depend on which side you go around it, but $\bar z$ is in the lower half plane and τ is in the upper half plane so there is no issue.

By choosing the cusp $(-)i\infty$ as the starting point for our integral, we've broken the $\mathrm{SL}(2, \mathbb{Z})$-equivariance of f to a manifest equivariance just for the upper Borel $\langle T \rangle \times \{\pm 1\}$. But we can measure how far off f_{holo} is from being modular. Of course, f is modular of weight $(w, 0)$, so the failure of f_{holo} will be, up to a constant like $\pm(2i)^w$, the failure of $\int_{\bar z = -i\infty}^{\bar\tau} (\tau - \bar z)^{-w}\, \bar s(\bar z, \tau)\, \mathrm{d}\bar z$ to be modular. Oh, you still like holomorphic things? Then its complex conjugate, i.e.

$$\int_{z = i\infty}^{\tau} (z - \bar\tau)^{-w}\, s(z)\, \mathrm{d}z$$

where perhaps $g(z)$ actually is $g(z, \bar\tau)$. The integrand transforms really well under modular transformations (g has weight $2 - w$, so the total integrand is well defined on $\mathcal{M}$). The only issue is the cycle of integration, which under a modular transformation transforms as

$$\int_{i\infty}^{\tau} \mapsto \int_{\gamma \cdot i\infty}^{\gamma \cdot \tau}$$

whereas modularity would require that it instead transform as

$$\int_{i\infty}^{\tau} \mapsto \int_{i\infty}^{\gamma \cdot \tau}.$$

These differ by the cycle

$$\int_{i\infty}^{\gamma \cdot i\infty} = \int_{i\infty}^{a/c}$$

Dropping the subscript, people say that f is a *mock modular form of weight w with shadow s* if (s is modular of weight $2-w$ and) f is invariant under the transformation

$$f \mapsto (f|_{w;s}\gamma)(\tau) := (c\tau+d)^{-w} f(\gamma\cdot\tau) - \int_{z=-i\infty}^{a/c} \bar{s}(z)\,(z-\tau)^{-w}\mathrm{d}z$$

up to some w-dependent constant that I've dropped. Any mock modular form has a *modular completion* which just restores the integral term. The modular completion is more natural from the perspective of harmonic analysis. As far as I can tell, the reason number theorists like mock modular forms is that they have q-series (since we still have manifest T-invariance). A question I don't know: can you recover s from these functions $\int_{-i\infty}^{a/c} \bar{s}(z)\,(z-\tau)^{-w}\mathrm{d}z$?

Recall that from the harmonic analysis perspective, s being holomorphic is special, but not that special. People say that a *pure mock modular form* is one where the shadow s is holomorphic. A mock modular form is *mixed* if s is the dot product of a vector-valued holomorphic modular form with a vector-valued antiholomorphic modular form (of course in the conjugate representation). A mock modular form is *generalized* if s is just some random C^ω function. In other words, a generalized mock modular form is just the same data as a C^ω modular form with some specific rules about the cuspidal behaviour, whereas pure mock modular forms are much closer to being holomorphic.

2.2. **Mock modularity from derived modularity.** We are interested in the ring(-spectrum)s TMF and TMF $\otimes\,\mathbb{C}$, and their variants with various cuspidal behaviour. The latter is simply the space of (weakly) holomorphic modular forms, which are essentially just holomorphic functions on $\mathcal{M}$. More precisely, as Dan Berwick-Evans has explained quite well, TMF $\otimes\,\mathbb{C}$ is the *derived* space of holomorphic modular forms.

The most naïve model of this is a Dolbeault model. Let's just talk about functions. A C^∞ function on a complex variety X is holomorphic when it is in the kernel of the Dolbeault operator $\bar\partial$. The *Dolbeault complex* is $C^\infty(X) \overset{\bar\partial}{\to} \Omega^{0,1}(X) \overset{\bar\partial}{\to} \Omega^{0,2}(X) \overset{\bar\partial}{\to} \dots$. Just the same way, you could look at a Dolbeault model for the ring of modular forms. It will be a double complex, with only a horizontal differential $\bar\partial$; the other grading is twice (!) the modular form degree. Remembering that we have a favourite coordinate τ on a cover $\mathfrak{h}$ of $\mathcal{M}$, and the stackiness of this cover, while infinite, is rationally acyclic. Thus we looking at the $\mathrm{SL}(2,\mathbb{Z})$ fixed points for a complex like

$$\dots \to C^\infty(\mathfrak{h}) \overset{\partial/\partial\bar\tau}{\to} C^\infty(\mathfrak{h}) \overset{0}{\to} C^\infty(\mathfrak{h}) \overset{\partial/\partial\bar\tau}{\to} C^\infty(\mathfrak{h}) \to \dots$$

In cohomological degree $-2w$, the action of $\mathrm{SL}(2,\mathbb{Z})$ is the weight-$(w,0)$ action. In degree $-2w+1$, it's the weight-$(w,2)$ action. Dan has explained how to incorporate a target space: if M is a manifold, then $\mathrm{TMF}(M)\otimes\mathbb{C}$ has a model like this that also involves to the de Rham cohomology of M.

For Tmf $\otimes\,\mathbb{C}$, there is a similar description, where you have to put in the correct cuspidal behaviour. I ran out of time when preparing these notes to work out exactly what that behaviour is.

Anyway, the point is: up to issues of growth rates, a generalized mock modular form of weight w is just (the same data as) a C^ω modular form of weight $(w,0)$ is just a cochain in degree $-2w$. Genuine (holomorphic) modularity is that this cochain is a cocycle.

Cochains are a model-dependent notion. But nullhomotopies are less so. Let's suppose fixed a shadow s, made into a cochain, automatically a cocycle, of degree $-2w+1$. A mock modular form with shadow s is then precisely a realization of this cocycle as a coboundary.

This is still not an entirely coherent thing, since I asked you to fix a cochain representative for s, but then immediately made s nullhomotopic and so derivèdly-zero. Where this notion really shines is when you have a morphism of complexes, perhaps an inclusion, with some homotopical kernel.

Indeed: pick a class $[s] \neq [0] \in \text{hoker}$; then its image is derivedly-zero, and so you can pick a class of a nullhomotopy.

In the case at hand, the stack $\mathcal{M}$ of smooth elliptic curves is affine, up to finite stackiness that $\mathbb{C}$-valued functions don't see. So line bundles on it have no cohomology in degrees other than 0. And indeed $\pi_\bullet(\mathrm{TMF} \otimes \mathbb{C})$ is just the ring MF of weakly-holomorphic modular forms, entirely in even cohomological degree (not only that — entirely in even weights, i.e. twice-even degree). On the other hand, the compactified moduli stack $\overline{\mathcal{M}}$ is not affine — it's a stacky refinedment of $\mathbb{P}^1(\mathbb{C})$. So sufficiently ample lines on it only have degree-0 cohomology, but their duals only have cohomology in degree 1.

Over $\mathbb{C}$, these computations are not too hard. What you find is that $\pi_{-21}(\mathrm{Tmf} \otimes \mathbb{C}) \cong \mathbb{C}$ is nonzero: there is a holomorphic derived modular form σ of weight $-21/2$. I'll come back to who it is later. But since $\pi_{\mathrm{odd}}(\mathrm{TMF} \otimes \mathbb{C}) = 0$, there is some weakly holomorphic (generalized) mock modular form of weight $-20/2$ with its as the shadow.

Recall that there is an interesting function $\Delta(\tau) = q \prod_{n=1}^{\infty}(1 - q^n)^{24}$. It is a holomorphic modular form of weight 12, i.e. it is class in $\pi_{24}(\mathrm{TMF} \otimes \mathbb{C})$. (It does not refine to TMF, but it does up to a scalar multiple.) Actually, I can say more: it is a *cusp form* in the sense that $\Delta(\tau = i\infty) = 0$. The derived version of cusp forms is called, not surprisingly, *topological cusp forms* Tcf: $\Delta \in \pi_{24}(\mathrm{Tcf} \otimes \mathbb{C})$. Anyway, multiplying or dividing by Δ is a good way to adjust cuspidal behaviour, trading it for modular degree.

So in particular we have an interesting class $\Delta\sigma \neq 0 \in \pi_3(\mathrm{Tcf} \otimes \mathbb{C})$. But $\pi_3(\mathrm{Tmf} \otimes \mathbb{C}) = 0$. So there is a (generalized) mock modular form φ of weight $w = 2$ (homotopical degree 4) with shadow the cusp form $\Delta\sigma$, and this f is bounded at the cusp!

Dividing back by Δ, we could equivalently say that σ is bounded at the cusp, whereas $\phi = \varphi/\Delta$ has a simple pole at the cusp. In other words, ϕ is what I called *lunar* before, albeit without any equivariance.

2.3. **Supersymmetry.** So far in my narrative, there's no reason for these φ or ϕ to be "pure" as a mock modular form. Indeed, there's no good way so far to even formulate pureness from the perspective of a chain model for $\mathrm{TMF} \otimes \mathbb{C}$. What would pureness mean? It would mean that the shadow σ or $\Delta\sigma$ is somehow "antiholomorphic." But I've written it as a derived modular form — it is "holomorphic" in the derived sense. If it is also somehow antiholomorphic, then it should be somehow "constant." What should this mean?

Mayuko Yamashita has already discussed at this workshop the *(Segal–)Stolz–Teichner proposal.* Stolz and Teichner have made a precise conjecture that implements this proposal, but I will prefer to speak about the meaning rather than the technical details (because my perspective is that probably there are many implementations). What they propose is summarized as saying that: TMF *has a cocycle model whose cocycles are compact unitary 1+1D quantum field theories, not necessarily conformal, equipped with an $N = (0, 1)$ supersymmetry.*

Let me elaborate on this briefly. A *quantum field theory* in $d + 1$D is some sort of object that lives on Lorentzian spacetime manifolds of dimension $d + 1$. "Lives on" should mean that it enjoys some sort of sheaflike locality condition, and the manifolds might be required to have some background fields beyond the metric, and furthermore the metric might be required to have some special holonomy, depending on the QFT in question. A unitary QFT can be Wick-rotated into Euclidean signature, where it becomes a reflection-positive QFT. (More generally, as well-explained by Kontsevich–Segal, unitary QFTs should be formulated on manifolds with complex-valued metric, whose real part is positive-definite; the Lorentzian metrics live on the boundary of this space.) A priori, QFT is about physics aka dynamics: it explains how states evolve; evolution, infinitesimally, is captured by spacetime manifolds that are germs of spacial slices, and longer evolutions are captured by cylinders, but in physics of course sometimes the macroscopic evolution has trouble:

waves crash. That said, in a Wick-rotated theory, you can certainly *ask* if the evolution is well-defined on all macroscopic spacetimes — if the path integral over those manifolds "converges" (absolute convergence, not conditional). For example, it does for sigma models with compact target (the high-energy oscillations are suppressed by a Gaussian), and so these QFTs are called *compact*.

A priori, a quantum field theory can have an *anomaly*. In a path integral model, this happens when the path integral doesn't quite resolve as a number, but rather as a section of some line bundle. From the perspective of operators and expectation values, this is because the value of the path integral itself is not physical: only ratios are. The upshot of this is that two QFTs describe the same physics whenever they differ by stacking by an invertible theory, and so if you want to work with the space of "physicses described by QFTs", you should quotient some space by this action. Well, if you can make this quotient precisely, then there will certainly a map $\{\text{physicses equipped with a path integral description}\}/\{\text{invertible things}\} \to \mathrm{B}\{\text{invertible things}\}$. If "QFTs" means the domain of this map, then this map itself records the anomaly. More precisely, people have taken to distinguishing "(possibly-)anomalous QFTs" from "nonanomalous aka absolute QFTs," the former being the collection of actual physicses, and the latter being the homotopy kernel of the anomaly map (i.e. equipped with anomaly cancellation data).

Freed has explained this quite well, especially in his arXiv note this week. An invertible QFT is someone who takes invertible values and lives on manifolds: for example, its Hilbert spaces are lines, and its partition functions are invertible numbers. Who is $\mathrm{B}\{\text{invertible things}\}$? The B makes it categorical: the anomaly of a $d+1$D QFT is someone who assigns a line to a spacetime manifold, or a gerbe to a spatial slice.

A priori, this is all an anomaly needs to define. That said, the *anomaly inflow mechanism* is the statement that this anomaly is in fact the $\leq d{+}1$D data of an invertible $d{+}2$D QFT. Whether inflow occurs depends on details of the model, and in examples seems to require a unitary hypothesis that in general has not been formally formulated. This is especially true in the nontopological case, and especially especial in the supersymmetric case.

Well, we care only about things up to deformation, since we just want a cocycle model for a topological object (TMF). So we only care about the anomaly up to deformation. One expects (if unitary) anomalies to be classified by a certain "differential Anderson dual" of the "differential cobordism spectrum" for you the background fields and cetera. Up to deformation, you should be able to drop the word "differential." We care about supersymmetric field theories: our tangential structure is Spin. So one expects the anomalies to be classified by $\Sigma^4 \mathrm{I}_{\mathbb{Z}}\mathrm{MSpin}$. There is a distinguished map $\mathrm{KO} = \Sigma^4\mathrm{I}_{\mathbb{Z}}\mathrm{KO} \to \Sigma^4\mathrm{I}_{\mathbb{Z}}\mathrm{MSpin}$ dual to the Spin orientation of KO: this map is called *the massive free fermion*, because given an element of $\pi_0\mathrm{KO}$, it writes down the corresponding massive free fermion in 3D. Recall that {possibly-anomalous 1+1D SQFTs} was the total space of a bundle over $\Sigma^4\mathrm{I}_{\mathbb{Z}}\mathrm{MSpin}$. Pulling this bundle back to KO gives a space spread over KO and so it is plausible at least that this can be made into an orthogonal spectrum.

Which, at a physical level of rigour, it can. I call this spectrum $\mathrm{SQFT}^\bullet$. The Stolz–Teichner proposal is that $\mathrm{SQFT}^\bullet = \mathrm{TMF}^\bullet$.

Here's how you should think of this. A supersymmetry is like a differential — a "curved" differential since it squares to a translation operator rather than to 0. In particular, a $(0,1)$-supersymmetry is an odd operator $\hat{G}$ which squares to $\bar{\partial}$. So $\bar{\partial}$ is "superexact" — physicists would say it's a "superdescendent" — which of course is just saying it's derivedly zero. So $(0,1)$ SQFTs are "derived holomorphic." The fact that we're working with the full SQFT and not just its partition function is what supplies the extra integrality.

In particular, any (0,1) SQFT $\mathcal{Q}$ has a (Ramond) partition function $Z(\mathcal{Q})$. There is a well-known physics argument that $Z(\mathcal{Q})$ is a (weakly) holomorphic modular function with a multiplier, and a deformation invariant of $\mathcal{Q}$; the argument is a manifestation of the "derived holomorphicity." The multiplier is $\bullet$ mod 24, if the SQFT was in cohomological degree $\bullet$. This is the same as the multiplier of $\eta(\tau)^\bullet$, where

$$\eta(\tau) = q^{1/24}\prod_{n=1}^{\infty}(1-q^n) = \sqrt[24]{\Delta}$$

is Dedekind's modular form (with multiplier) of weight 1/2. So we can correct the multiplier and simultaneously record the degree by working with

$$Z(\mathcal{Q})(\tau) \times \eta(\tau)^{-\bullet} \in \mathrm{MF}_{-\bullet/2}.$$

I call this combination the *Witten genus* of $\mathcal{Q}$. So this supplies a map

$$\mathrm{SQFT}^\bullet \to \mathrm{MF}_{-\bullet/2}.$$

When $\bullet = -m < 0$, the correction by η has the following interpretation. Take your SQFT. Now stack on m-many free fermions. The result $\mathcal{Q} \times \mathrm{Fer}(m)$ is now nonanomalous. Its partition function vanishes, but because I have m distinguished fermions $\psi_1, \dots, \psi_m$, I have a distinguished local operator $:\!\psi_1 \cdots \psi_m\!:$. Standard calculation: up to some conventions about how to handle odd self-adjoint operators, we have the 1-point function

$$\langle :\!\psi_1 \cdots \psi_m\!: \rangle_{\mathrm{Fer}(m)} = \big(\langle \psi \rangle_{\mathrm{Fer}(1)}\big)^m = \eta(\tau)^m$$

in the Ramond sector. So the upshot is that

$$Z(\mathcal{Q})(\tau) \times \eta(\tau)^m = \langle 1 \rangle_{\mathcal{Q}} \times \langle :\!\psi_1 \cdots \psi_m\!: \rangle_{\mathrm{Fer}(m)} = \langle :\!\psi_1 \cdots \psi_m\!: \rangle_{\mathcal{Q}\times\mathrm{Fer}(m)}.$$

Now suppose that $\mathcal{Q}$ enjoyed also a left-moving supersymmetry, i.e. it has $N = (1,1)$ susy. Then it is also derived-antiholomorphic, and in particular $Z(\mathcal{Q})$ is both holomorphic and antiholomorphic i.e. just a constant. The Witten genus will not be constant because of this conventional factor of η, but that's the only reason it won't be.

So, in this SQFT model, I propose the following. A *topological pure mock modular form* is a compact $N = (1,1)$ SQFT $\mathcal{S}$, called the shadow, together with a nullhomotopy among compact $N = (0,1)$ SQFTs.

In work with Gaiotto, we argue the following. Physics does not mandate that SQFTs be compact, but if they are not, then their partition functions will not be well-defined, and in particular will not be holomorphic or modular. There is a mild type of noncompactness, that we call having *tubular ends*: it is the SQFT version of a Riemannian manifold M^n that near its end approximates $\mathbb{R} \times N^{n-1}$. An SQFT with tubular ends does have a partition function, but the path integral converges conditionally, not absolutely. So it actually has a couple different functions that deserve the name "the" partition function. One of the reasonable interpretations is manifestly modular, and the cost of the noncompactness is that the "well-known physics argument" establishing holomorphicity fails.

Rather, what we find is that, if $\mathcal{Q}$ has tubular ends $\mathbb{R} \times \mathcal{S}$, then the Witten genus $Z(\mathcal{Q}) \times \eta(\tau)^{-\bullet}$ satisfies the following holomorphic anomaly equation:

$$\sqrt{-8}(\Im(\tau))^{1/2}\frac{\partial}{\partial\bar\tau}\big[Z(\mathcal{Q})(\tau,\bar\tau) \times \eta(\tau)^{-\bullet}\big] = \langle \hat G \rangle_{\mathcal{S}} \times \eta(\tau)^{-(\bullet+1)}.$$

The factors of η are there for my degree conventions. The factor of $\sqrt{-8}$ is just something about conventions for supersymmetry operators. The factor of $\sqrt{\Im(\tau)}$ looks like the factor you'd use for a mock modular form of weight 1/2.

What's really going on is that there is an *super* moduli space $\mathcal{M}^{(0,1)}$ of "super elliptic curves" which is $2|1$ real-dimensional. An $(0,1)$ SCFT can be placed on any elliptic curve in this stack. If you use the curve with moduli $(\tau, \bar\tau, \theta)$, then the partition function is

$$Z(\mathcal{Q})(\tau,\bar\tau,\theta) = Z(\mathcal{Q})(\tau,\bar\tau) + \theta\langle \hat G \rangle$$

up to some convention factor about the units of θ. This starts getting close to the picture that Dan Berwick-Evans has developed. The point of the holomorphic anomaly equation is that $\langle \hat G \rangle$ is not a deformation invariant, but its class modulo the image of the holomorphic anomaly equation is.

In my holomorphic anomaly equation, $\hat{G}$ was the generator of the right-moving susy in $\mathcal{S}$. If $\mathcal{S}$ has $N = (1,1)$ susy, then the well-known argument tells you that $\langle \hat{G} \rangle_{\mathcal{S}}$ should be antiholomorphic: in general, if an operator commutes with a right- (left-) moving susy, then it (anti)holomorphic. So the strictly-correct statement is that $Z(\mathcal{Q}) \times \eta(\tau)$ is [the modular completion of] a pure mock modular form of weight 1/2. If you want some other degree w of pure mock modularity, you'd need $\langle \hat{G} \rangle_{\mathcal{S}} (\Im(\tau)\eta(\tau)^2)^{w-1/2}$ to be antiholomorphic. I don't know a physical process which would make this to happen.

2.4. **Motivating example.** Consider the (0,1) sigma model with target a round S^3. So the dynamical fields are: a full boson exploring S^3; its superpartner, a right-moving fermion. Because of that mismatched fermion, this theory requires a String structure on S^3 to be well-defined — this String structure manifests as a term in the volume form for the path integral over the fermi field. Orientations are a torsor for $\mathbb{Z}/2$; Spin structures are a torsor for $\mathbb{Z}/2$-bundles; String structures are a torsor for $\mathrm{H}^3(-;\mathbb{Z})$. Typicallly they really do live just in a torsor. But for S^3, there is a distinguished String structure: the unique one that extends over D^4, equivalently the unique one which is isomorphic to its image under a reflection automorphism of S^3. Any other one is $k \in \mathbb{Z}$ units of String structure away from that "zero" one.

Use that $S^3 = \mathrm{SU}(2)$ as a Riemannian manifold. It is understood that 1+1D sigma models with group manifold target exist quantumly and flow to WZW models in the deep IR. Specifically, we'll flow to a $(0,1)$ WZW model. The left-moving chiral algebra will have bosonic fields J_1, J_2, J_3 and OPE

$$J_a(z)J_b(0) \sim \frac{\kappa\delta_{ab}}{z^2} + \frac{i\eta_{abc}J_c(0)}{z}.$$

This η is the totally-antisymmetric tensor on $\mathbb{R}^3$, i.e. the Lie bracket on $\mathfrak{su}(2)$. The number $\kappa \in \mathbb{N}$ is called the *level* of the left-moving sector. There are also some twist fields that are exponentials in the Js. The right-moving chiral algebra will have conjugate fields $\bar{J}_a$ and also their superpartners $\bar{\psi}_a$. The bosons will OPE with themselves, and they will act on the fermions as the adjoint representation, and the fermions will have an OPE like

$$\psi_a(z)\psi_b(0) \simeq \frac{\delta_{ab}}{z^2}.$$

The question is to work out κ. Here's how to do it. Considering the left- and right-multiplication actions by SU(2). These have to flow to some actions in the deep IR: surely they will become the left and right Noether currents $J, \bar{J}$, although the two handednesses are a priori uncorrelated. The String structure k introduces an 't Hooft anomaly for those actions. To anomaly match, you find that $\kappa = |k| - 1$ depends on the absolute value of k. Changing the sign of k stacks the theory with an Arf invertible phase, which is not seen by VOA. When $k = 1$, the theory spontaneously breaks supersymmetry (it flows to zero).

Suppose $k > 0$. The supercurrent is

$$\hat{G} = \sqrt{-1}\sqrt{\frac{2}{k+1}}\bar{\psi}_1\bar{\psi}_2\bar{\psi}_3 + (\text{proportionality factor}) \sum_a \bar{\psi}_a \bar{J}_a.$$

The second term has vanishing 1-point function. The 1-point of a free fermion is η, and then you have the partition function of the remaining fields. So, dropping a power of $\sqrt{i}$, we have

$$\langle \hat{G} \rangle = \sqrt{\frac{2}{k+1}}\eta(\bar{\tau})^3 Z_{k-1}^{\mathrm{WZW}}(\tau,\bar{\tau})$$

where Z_{k-1}^{WZW} is the partition function of the bosonic WZW model. This is a RCFT: its partition function is the norm of the characters of the modules for the chiral algebra V; these transform as a vector-valued modular form $\xi_j(\tau)$, where j ranges over $K_0\mathbf{Rep}(V) \otimes \mathbb{C}$.

These characters are pretty well known. We are working with SU(2) at level $k-1$, so there are k representations, naturally indexed by their spins (hence "j") $0, \frac{1}{2}, 1, \dots, \frac{k-1}{2}$. Define the weight 3/2 theta function

$$\Theta_{k+1,\ell}(\tau) = \sum_{m \in \mathbb{Z} + \frac{\ell}{2(k+1)}} m q^{(k+1)m^2}.$$

Then

$$\chi_j(\tau) = 2(k+1)\Theta_{k+1,2j+1}(\tau)\eta(\tau)^{-3}.$$

So our shadow is

$$\langle \hat{G} \rangle = 2(2(k+1))^{3/2}\eta(\bar{\tau})^3 \sum_{2j+1=1}^{k+2} |\Theta_{k+1,2j+1}(\tau)\eta(\tau)^{-3}|^2.$$

I won't go through the derivation of a mock modular form with this shadow: it uses the standard type of hard-to-motivate calculational trick that the number theorists and string theorists are good at. I'll just record that its holomorphic part is explicitly computable and looks like

$$q^{-1/8}\left(-\frac{k}{12} + 2\mathbb{Z}((q))\right).$$

When $k=1$, we have a bosonic WZW model at level 0. This is just the vacuum (massive aka gapped) theory. In terms of these formulas, the point is that the unique Theta series is precisely $\eta(\tau)^3$, so $\chi_0 = 1$ in this case. Thus, exactly in this case, $\langle \hat{G} \rangle$ is antiholomorphic. The physical explanation for this is the following. Exactly when $k=1$ (or $k=-1$ or $k=0$), the left-moving sector is completely trivial: the SCFT is entirely right-moving. So the left handed stress-energy tensor is 0. A left-moving susy is a square root of this. So we can decide to supply the model with a left-moving supersymmetry equal to 0!

When S^3 has 1 unit of String structure, it is the Lie group framing on SU(2). Every group has a framing, and every framed manifold represents a class in the sphere spectrum Mfr = \$ and so in any ring spectrum. The class represented by SU(2) is always called "ν". What I'm saying is that, after restoring powers of η, we have:

$$\nu = 0 + \theta \|\eta(\tau)\|^3.$$

2.5. **Integrality and KOMF.** Suppose that you have a $(0,1)$-SQFT $\mathcal{Q}$ with tubular ends $\mathcal{S}$. You can compactify on an S^1 (nonbounding spin structure) to get an $N=1$ quantum mechanics model. This model has an S^1 symmetry by rotating the circle of compactification, and so the Hilbert space decomposes as a sum of subspaces. There is also a time-reversing symmetry, whose precise behaviour depends on the anomaly of $\mathcal{Q}$.

When $\mathcal{Q}$ is compact, then $\int_{S^1}\mathcal{Q}$ is S^1-equivaraintly compact: its Hamiltonian has discrete spectrum on each S^1 eigenspace. So its supersymmetry $\bar{G}_0 : \mathcal{H}_{\text{ev}} \to \mathcal{H}_{\text{odd}}$ is Fredholm, and defines a KO$((q))$. The graded dimension in $\mathbb{Z}((q))$ of this object is easy to compute: compactify on another circle. But then you see that this is just the q-series expansion of the modular function $Z(\mathcal{Q})$ (up to normalization conventions).

When $\mathcal{Q}$ is noncompact with tubular ends $\mathcal{S}$, it often still happens that $\int_{S^1}\mathcal{Q}$ is compact. It can fail in a mild way, by a sort of "$\mathbb{Z}/2$-index" of $\mathcal{S}$, which can adjust some integrality in a controlled way. If you think about the compactifications, what it really is doing is evaluating $Z(\mathcal{Q})$ at $(\tau,\bar{\tau}) = (\tau, -i\infty)$. This is the type of formula that does not parse unless $Z(\mathcal{Q})$ is analytic (and so can be analytically continued away from the real locus where τ and $\bar{\tau}$ are conjugate) and also has an entire analytic continuation. If $\mathcal{S}$ is sufficiently nice, then $Z(\mathcal{Q})$ will solve an ODE like $\partial_{\bar{\tau}} Z(\mathcal{Q}) =$ (nice function), and so $Z(\mathcal{Q})$ will also be nice. Anyway, the point is that this q-series is precisely the holomorphic part of the Maass form $Z(\mathcal{Q})$.

Fix $\mathcal{S}$, and imagine that you have a filling $\mathcal{Q}$. There are many fillings, of course, and so $Z(\mathcal{Q})$ is not determined by $\mathcal{S}$, but the holomorphic anomaly equation already determines $[Z(\mathcal{Q})] \in$

$\{C^\omega$ modular functions$\}/\{$holomorphic modular functions$\}$. The integrality anomaly is also determined. The upshot is that $\mathcal{S}$ determines a well-defined class in $\mathbb{C}((q))/(\mathrm{MF}+\mathbb{Z}((q)))$. Moreover, although the ingredients (for example, $\langle \bar{G}\rangle$) of this class turn out not to be deformation invariants of $\mathcal{S}$, the total class in. And it vanishes if $\mathcal{Q}$ exists.

Reversing the logic, this class can obstruct $\mathcal{S}$ from being nullhomotopic. I told you the class for $\mathcal{S} = S^3_k$: it was

$$\left[-\frac{k}{12} + 2\mathbb{Z}((q))\right] \in \frac{\mathbb{C}((q))}{\mathrm{MF} + 2\mathbb{Z}((q))}.$$

I snuck in a factor of 2: it is because actually in this case the anomaly of any putative filling would make $\int_{S^1} \mathcal{Q}$ into a quaternionic module, and so its dimensions are even. I also dropped the $\eta(\tau)^{-3}$ factor for convention reasons.

This confirms what you know from geometry: S^3_k is String-nullcobordant iff $k = 0 \pmod{24}$. What we're seeing is that also its sigma model is quantum-nullhomotopic only under the same condition.

Dan Berwick-Evans introduced a cohomology theory KOMF that is the home for invariants like this that play integrality against modularity. The specific invariant I've described is a version of the Bunke–Naumann secondary Witten genus, and Dan's KOMF is exactly designed to cohere that invariant.

3. Umbral and categorical symmetries

3.1. **Neimeier lattices.** I first tell a quite classical story. Probably I will say any of this in my lectures.

A (positive-definite) *lattice* of rank r is a full-rank discrete subgroup $L \subset \mathbb{R}^r$. [More precisely, it is a full-rank discrete subgroup of some orthogonal vector space over $\mathbb{R}$ with positive-definite inner product. I do not need to trivialize that vector space.] The subgroup of $\mathrm{O}(n)$ that preserves L is $\mathrm{O}(L)$. The lattice L is *integral* if $\langle,\rangle : L \times L \to \mathbb{R}$ takes values in $\mathbb{Z}$, and *even* if $\|\ell\|^2 = \langle \ell, \ell\rangle$ takes values in $2\mathbb{Z}$ for all $\ell \in L$ (evenness implies integrality by polarization). For any lattice, define its *dual lattice* $L^\vee := \{x \in \mathbb{R}^n : \langle x, \ell\rangle \in \mathbb{Z}\ \forall \ell \in L\}$. Integrality says that $L \subset L^\vee$. The lattice is *self-dual* if $L = L^\vee$. More generally, if L is integral, then $L^\vee/L$ is a finite abelian group with a nondegenerate $\mathbb{R}/L$-valued inner product, called the *discriminant group* $\Delta(L)$. If L is even, then the metric on $\Delta(L)$ refines canonically to a quadratic form.

Suppose that L is an even lattice. A *root* in L is a vector with $\|\ell\|^2 = 2$, the smallest possible length. A lattice is a *root lattice* if it is spanned by its roots. Suppose α is a root of L. Then the reflection map that sends $\alpha \mapsto -\alpha$ but fixes vectors orthogonal to α is $\ell \mapsto \ell - \langle \alpha, \ell\rangle \alpha$. So this takes $L \to L$. The subgroup of $\mathrm{O}(L)$ generated by this reflections is the *Weyl group* $\mathrm{Weyl}(L)$. It is a normal subgroup, and the quotient is the group $\mathrm{Out}(L)$ of *outer automorphisms of* L.

You know these things in examples. Suppose I draw a simply-laced Dynkin diagram X of rank r. Then you can construct a lattice of rank r as follows: you put a basis vector α for each of the roots; you make $\|\alpha\|^2 = 2$; you make $\langle \alpha, \beta\rangle = 0, -1$ according to whether α and β are connected by an edge or not. Now take the span of these roots, call it L_X. Fact: $\mathrm{Out}(L_X) = \mathrm{Aut}(X)$. Another fact: if L is a root lattice (as I've defined it), then it is L_X for a unique Dynkin diagram X.

Another fact. Suppose the simply-laced Dynkin diagram X is connected. Then the ratio

$$h(X) = \frac{\#\{\text{roots in } L_X\}}{\text{rank of } X}$$

is an integer, called the *Coxeter number* of X.

Evenness forces the lattice points to be rather far apart, whereas self-duality forces them close together. Low-dimensional intuition makes these seem in conflict. But it leads you somewhat astray: the volume of a sphere relative to its bounding box goes down. It turns out that there do exist self-dual even lattices, and you know one of them: the E_8 root lattice. It turns out that self-dual even lattices can live only in rank $r = 8$. There are various famous proofs. One uses Poisson resummation to show that the Theta series of an even self-dual lattice has good modularity, and then rules out other ranks from considering MF. Another converts an even self-dual lattice into an invertible element in KO.

The only even self-dual lattice of rank 8 is E_8 root lattice. In rank 16, there are two self-dual even lattices. One is $E_8 \times E_8$. The other is not a root lattice. It is constructed as follows. You take D_{16}. Now inspect the highest weight of the fundamental half-spin representation of Spin(32). This weight σ is not in the root lattice, but it is a weight, so it has integral pairing with the roots, and it is even: for D_n in general, $\|\sigma\|^2 = n/4$. So the union

$$D_{16}^+ := D_{16} \cup (D_{16} + \sigma)$$

is an even lattice. Now $\#\Delta(D_{16}) = 4$, so $\#\Delta(D_{16}^+) = 4/(\#(D_{16}^+/D_{16})) = 4/2^2 = 1$. So D_{16}^+ is self-dual. Remark: $\mathrm{Out}(D_{16}) = \mathbb{Z}/2$, but $\mathrm{Out}(D_{16}^+) = *$, because we broke the outer automorphism by selecting "the" fundamental half-spin representation.

Niemeier, famously, classified the self-dual even lattices of rank 24. The main steps go as follows. Leech had already shown that if $\mathrm{rk}(L) = 24$ and L has no roots, then $L \cong$ Leech is the lattice Leech had discovered. Niemeier shows that if L contains a root, then its root system is full rank. Moreover, he shows that each connected component of the Dynkin diagram of this root system has the same Coxeter number. Now just do some combinatorics: find all Dynkin diagrams of rank 24 such that all components have the same Coxeter number; there turn out to be 23 of them. For example, E_8 and D_{16} both have Coxeter number 30, so $E_8 \times D_{16}$ is on your list. Finally, for each of these 23 cases, study the ways to extend that root lattice to a self-dual lattice, like $D_{16} \rightsquigarrow D_{16}^+$, and show that there is a unique way up to outer automorphisms.

Here's how this works in general. Suppose you have an even lattice L. Even extensions of this lattice are classified by subgroups $I \subset \Delta(L)$ which are *isotropic* in the sense that the quadratic form trivializes on I. Then $\Delta(L + I) = I^\vee/I$, where $I^\vee \subset \Delta(L) = \Delta(L)^\vee$ is the dual subgroup. (Orthogonal subgroup for the quadratic form on $\Delta(L)$.) It could happen that $L + I$ has more roots — for example, $D_8^+ = D_8 + \{0, \sigma\} = E_8$, because in rank 8, the half-spinor has length^2 $8/4 = 2$.

If L is a root lattice, then $\Delta(L)$ is just a product of groups indexed by the components: $\Delta(A_n) = n+1$; $\Delta(D_n) = 4$ or 2^2 according to the parity of n; $E_{6,7,8} = 3, 2, 1$. Specifically, these numbers are the numbers of "1"s appearing in the markings on the affine Dynkin diagram. There is a classical notion: a *code* over q is a subgroup of $\mathbb{F}_q^n$. So subgroups of $\Delta(L)$ are also some types of codes. The condition that $I \subset \Delta(L)$ is isotropic for the quadratic form generalizes being an *doubly even code* (over $\mathbb{F}_2$). If you rule out that $L + I$ should have new roots, then that's like ruling out codewords of Hamming weight 4. In classical coding theory, a *word* is a tuple of symbols, perhaps drawn from a finite field, and a *code* is a set, perhaps a group, of words. The *Hamming weight* of a word is the number of nonzero letters in the word. The idea (if your code is a group) is that when you transmit a message, perhaps some of the letters get changed because of noise in the transmission. If that number is less than half the minimal Hamming weight of your code, then you can definitely correct them back to being in your code. So codes with high minimal Hamming weight have good error correction properties. The subgroups of $\Delta(L)$ that we care about are basically the self-dual codes of high minimal Hamming weight.

Given a Dynkin diagram X, I'll write $\mathrm{Niem}(X)$ for the corresponding Niemeier lattice. From the remarks above, you see that $\mathrm{Out}(\mathrm{Niem}(X)) \subset \mathrm{Aut}(X)$. For example, $\mathrm{Aut}(A_2^{12}) = 2 \wr \Sigma_{12}$. It turns out that in this case the code is Golay's ternary code, and $\mathrm{Out}(\mathrm{Niem}(X)) = 2M_{12}$. Indeed, $\Delta(A_2) = \mathbb{Z}/3$, so $\Delta(A_2^{12}) = \mathbb{F}_3^{12}$, and $2M_{12}$ is its automorphism group. Similarly, $\mathrm{Out}(\mathrm{Niem}(A_1^{24})) = M_{24}$ because the code in question is the binary Golay code inside $\Delta(A_1^{24}) = \mathbb{F}_2^{24}$.

The groups $\mathrm{Out}(\mathrm{Niem}(X))$ were rediscovered by the moonshiners when they were thinking about mock modular forms. Because mock modular forms have a "shadow", they called their new moonshine "umbral". So they called these the *umbral groups*: $\mathrm{Umb}(X) := \mathrm{Out}(\mathrm{Niem}(X))$. So for example $\mathrm{Umb}(A_1^{24}) = M_{24}$ and $\mathrm{Umb}(A_2^{12})) = 2M_{12}$.

3.2. **Leech lattice.** Take a root lattice L_X, of constant dual Coxeter number h. There is a vector $\rho \in L_X^\vee$ that pairs with all the simple roots to 1. In fact $2\rho \in L_X$: it is the sum of the positive roots. Using the evenness of the extension $L_X \to \mathrm{Niem}(X)$, in fact $\rho \in \mathrm{Niem}(X)$.

Since X has constant Coxeter number h, we can consider the function

$$\langle \rho, - \rangle \mod h : \mathrm{Niem}(X) \to \mathbb{Z}/h.$$

Its kernel is a sublattice $\mathrm{Niem}(X)_0 \subset \mathrm{Niem}(X)$ of index $\mathbb{Z}/h$. (The *index* of a subgroup is the quotient, thought of as a set [cardinality] or orbit or quotient group or whatever you have.) Thus $\Delta(\mathrm{Niem}(X)_0)$ is an extension of shape $(\mathbb{Z}/h).(\mathbb{Z}/h)$. Now you do a calculation, which holds universally provided the rank is divisible by 24: $\Delta(\mathrm{Niem}(X)_0) = (\mathbb{Z}/h)^2$ with the split quadratic form. There are various Lagrangian subgroups: one of them corresponds to the inclusion $\mathrm{Niem}(X)_0 \subset \mathrm{Niem}(X)$. There is a unique Lagrangian subgroup that intersects this one transversely.

(You should think of this $(\mathbb{Z}/h)^2$ with the split quadratic form as a "cotangent bundle" $T^*(\mathbb{Z}/h)$. The extension $\mathrm{Niem}(X) \supset \mathrm{Niem}(X)_0$ corresponds to the zero section. It turns out that the Lagrangian subgroups of $(\mathbb{Z}/h)^2$ are precisely the conormal bundles of subgroups of $\mathbb{Z}/h$. The unique transverse Lagrangian subgroup is the cotangent fibre at the origin.)

So there is one other distinguished extension $\mathrm{Niem}(X)_0 \subset L$ self-dual. It turns out that this L never has roots — you killed the roots you had when you passed to $\mathrm{Niem}(X)_0$, and what you have to calculate is that you don't get any new ones under the extension, and this again uses something about the code not having words of low Hamming weight. So, in the rk $= 24$ case, it is a Leech lattice. Note that the selection of ρ broke the $\mathrm{Weyl}(X)$-action, but it did not break the action by $\mathrm{Aut}(X)$. Of course, the Niemeier construction itself broke $\mathrm{Aut}(X) \rightsquigarrow \mathrm{Umb}(X)$.

These 23 constructions of the Leech lattice are what Conway and Sloane call the *hol(e)y construction*. The reason for the name is that it makes manifest a subgroup of Co_0 isomorphic to $\mathrm{Umb}(X)$.

Here $\mathrm{Co}_0 := \mathrm{O}(\mathrm{Leech})$. Since Leech has no roots, this is also $\mathrm{Out}(\mathrm{Leech})$. Obviously Co_0 is not simple: it contains a central ± 1. Conway showed that this is the only reason it isn't simple: $\mathrm{Co}_1 := \mathrm{Co}_0 / \pm 1$ is simple. This is somewhat remarkable. For example, Co_0 cannot contain any elements of determinant -1, as otherwise it would emit a nontrivial surjection (the determinant). So $\mathrm{Co}_0 \subset \mathrm{SO}(24)$.

The group Co_1 appears as the symmetries of a few things.

First, it is the (projective) automorphism group of Leech.

To tell you the second, I need to tell you a rather remarkable fact about the lattice $II_{25,1}$. By definition, $II_{m,n} \subset \mathbb{R}^{m,n}$ is the even unimodular lattice of indefinite signature (m, n). The word "the" is justified because, unlike in the definite-signature case, as long as $m, n > 0$, this lattice is unique (up to isomorphism) if it exists, and it exists if and only if $m - n \in 8\mathbb{Z}$. The roots in a definite-signature lattice are the integral points on a sphere; in the Lorentzian signature $(25, 1)$, they are the integral points on a 25-dimensional hyperboloid. Pick a hyperplane defining the notion of "positive", and look at the simple roots, i.e. those closest to this hyperplane. In the Euclidean signature case, these are points near (just barely north of) the equator in the sphere of roots. Now they are just north of a 24-dimensional hyperboloid. You are used to the idea that there is one simple root for each dimension, but in Lorentzian signature there can be more than that. You can still draw a sort of "Dynkin diagram" that encodes the angles between the simple roots. The rather remarkable fact is that in this case, this 24-dimensional Dynkin diagram is *the entire Leech lattice*: there is one simple root for each vertex in Leech. And the angles are related to the Leech metric. So $\mathrm{Co}_0 = \mathrm{Aut}(\text{Dynkin diagram}) = \mathrm{Out}(II_{25,1})$.

This second representation of Co_1 is intimately related to the first. There is a third representation that I don't know a satisfactory relationship. It was discovered by Duncan in his thesis. In their book, Frenkel–Lepowsky–Meurman developed a twisted orbifold construction that inputed the Leech lattice holomorphic CFT V_{Leech} and outputted the Monster CFT $V^\natural$. Duncan investigated what would happen if you ran the same construction, but inputed the *N=1 supersymmetric* holomorphic SCFT built from E_8. (I.e. $V_{E_8} \otimes \mathrm{Fer}(8)$, with supersymmetry that couples the free fermions to the 8 Cartan bosons.) What Duncan discovered was a $N = 1$ holomorphic SCFT $V^{f\natural}$ with no continuous symmetries (just as $\mathrm{Aut}(V^\natural) = \mathbb{M}$ has no continuous elements), and he showed that $V^{f\natural}$ was the unique such holomorphism SCFT with this property of central charge $c = 12$. (The corresponding uniqueness of $V^\natural$ is still open.) Finally, he showed that $\mathrm{Aut}_{N=1}(V^{f\natural}) = \mathrm{Co}_1$.

In some sense there is a one-way relationship. If you forget the susy, $V^{f\natural}$ is a very easy object: it is the lattice CFT of the (integral but not even) lattice D_{12}^+. So its susy-destroying automorphism group is $\mathrm{SO}^+(24)$, the image of $\mathrm{Spin}(24)$ in the half-spin representation. Cohomological calculation: $\mathrm{Co}_0 = \mathrm{O}(\mathrm{Leech}) \subset \mathrm{O}(24)$ has trivial Schur multiplier H_2 (and trivial abelianization H_1), and so lifts (canonically) to $\mathrm{Spin}(24)$. Actually I've been sloppy: since Co_0 does not contain reflections, although there is only one Leech up to $\mathrm{O}(24)$-action, there are two chiralities / oriented isomorphism classes. For one, but not the other, of these chiralities, the map $\mathrm{Co}_0 \to \mathrm{Spin}(24) \to \mathrm{Spin}^+(24)$ that you get factors through Co_1. Now another calculation: For that chirality, and not the other, there exists a unique spin-3/2 field in $V_{D_{12}^+}$. A final calculation: The conformal vector is the unique Co_0-fixed field of spin-24 (this is just because the 24-dimensional real representation is simple). These together imply that that unique spin-3/2 field is, up to a conventional scalar, an $N = 1$ superconformal vector, and with this vector you get $V^{f\natural}$. All together, these give a map $\mathrm{Co}_1 \to \mathrm{Aut}_{N=1}(V^{f\natural})$. Final ingredient is a nontrivial theorem: $\mathrm{Co}_1 \times \pm 1$ is a maximal subgroup of $\mathrm{SO}^+(24)$, and -1 does not fix the supersymmetry (nor do any of the continuous symmetries). So the map $\mathrm{Co}_1 \to \mathrm{Aut}_{N=1}(V^{f\natural})$ must be an iso.

What is missing from this story is how to go back. Can you use the geometry of $V^{f\natural}$ to construct the Leech lattice? I don't know. Not in general: the same discrete subgroup of $\mathrm{O}(n)$ can be $\mathrm{O}(L)$ for inequivalent L's. So you will need some other special facts about the problem.

3.3. **A conjecture.** Recall that, according to Stolz–Teichner, an $N = (0, 1)$ SCFT $\mathcal{Q}$ of anomaly $\bullet = 2(c_L - c_R)$ represents a TMF class of cohomological degree $\bullet$, whose image in $\mathrm{MF}_{-\bullet/2}$ is $Z(\mathcal{Q})\eta^{-\bullet}$. If the SCFT has $N = (1, 1)$ susy, then $Z(\mathcal{Q}) \in \mathbb{Z}$ is a constant. So the $(1, 1)$ SCFTs are basically the polynomials in η.

For example, $Z(V^{f\natural}) = 24$, and $(c_L, c_R) = (12, 0)$, and $V^\natural$ has a left-moving susy and its right-moving half is trivial so it is automatically $(1, 1)$. So $V^{f\natural}$ represents a class "$24\Delta^{-1}$" in TMF^{24}. Indeed, there is a unique such class. Integrally, Δ^{-1} itself is not a TMF class. One of the stronger pieces of evidence, I think, supporting the Stolz–Teichner proposal (at least in the weak form that SQFTs admit a TMF-valued deformation invariant, not a priori a complete invariant) is that, as far as we know, the subtle divisibility properties of TMF classes are satisfied by <u>all</u> SQFTs.

We also have the class ν, represented by S^3_1, or equivalently, in the deep IR, by $\overline{\mathrm{Fer}}(\mathfrak{su}(2))$ (since, for this amount of String structure, the bosons gap out). The product $24\Delta^{-1} \times \nu$ is then represented by $V^{f\natural} \times \overline{\mathrm{Fer}}(\mathfrak{su}(2))$. This product turns out to vanish in TMF. But there is a more to say. The class ν is a cusp form: it has a zero at $\tau = i\infty$. So $24\Delta^{-1}\nu$ is formally a regular function at $i\infty$. And indeed, by inspecting the elliptic spectral sequence, you discover that this is the class I earlier called "σ" that generates $\mathrm{Tmf}^{21} \cong \mathbb{Z}$.

If the Stolz–Teichner <u>picture</u> is correct, saying that the TMF-valued class in a complete invariant, then it says that $V^{f\natural} \times \overline{\mathrm{Fer}}(\mathfrak{su}(2))$ can be deformed in the space of compact (0,1)-SQFTs to one with spontaneous supersymmetry breaking.

Moreover, that deformation can be chosen so that its partition function [more precisely, the partition function of the corresponding SQFT with tubular ends] blows up with only a simple pole at $q = 0$. Moreover, after better understanding the topological meaning of "regular at the cusp," I think that deformation should be unique up to higher-order deformation, since I think [but haven't checked] that the next cohomology group vanishes. These are not at all obvious statements physically, and would be interesting to investigate.

Moreover, since $\mathrm{M}_{24} \subset \mathrm{Co}_1$, it acts on $V^{f\natural}$ and hence on $V^{f\natural}S^3_1$. **Conjecture [application of Stolz–Teichner proposal]:** $24\Delta^{-1} \in \mathrm{TMF}^{24}$ has a natural Co_1-equivariant refinement. $24\Delta^{-1}\nu \in \mathrm{Tmf}^{21}$ has a natural Co_1-equivariant refinement.

Topological Mathieu Moonshine Conjecture: (This M_{24}-equivariant refinement of) $24\Delta^{-1}\nu$ is nullhomotopic in M_{24}-equivariant topological lunar forms (i.e. allowing a pole only at the infinite cusp). Note: I think it is not nullhomotopic Co_1-equivariantly, even with much worse poles, because I think I can rule this out by calculating its image of Co_1-equivariant KOMF (which is just to say: by playing modularity against integrality).

To tell you the more general case, I have to tell you one more feature of umbral moonshine. Suppose your Niemeier lattice had Coxeter number $h > 2$. Then the umbral moonshine functions are not scalar-valued mock-modular forms. Rather, they are vector-valued. The vector space in question has a categorification: it is the SL_2-module associated to the modular tensor category associated to $\mathrm{SU}(2)$ at level $h-2$. This is precisely the level with which the left copy of $\mathrm{SU}(2)$ acts on S^3_{h-1}.

S^3_{h-1}, in the deep IR, is not purely antiholomoprhic (unless $h = 2$). But the holomorphic part is precisely a WZW algebra for $\mathrm{SU}(2)$ at level $h-2$. This means that if you work equivariantly with respect to this holomorphic $\mathrm{SU}(2)$, what you have left is naturally an antiholomorphic, and hence $N = (\infty, 1)$-supersymmetric, object. The only correction is that it is an object whose partition function transforms as an $\mathrm{SU}(2)_{h-2}$-character and not a scalar. Which is precisely what we want!

Physically, what's going on is the following. The modular tensor category for $\mathrm{SU}(2)_{h-2}$ selects a 3D topological quantum field theory. This antiholomorphic $N = 1$ theory is a boundary condition for the 3D TFT. For any MTC $\mathcal{C}$, one can consider the moduli space of its $N = (0, 1)$ boundary conditions. This is not naturally $\mathbb{Z}$-graded, but it is graded by the torsor $\mathbb{Z} + c$ where $c \in \mathbb{Q}/\mathbb{Z}$ is the central charge of the MTC (up to conventions). This space is formally a spectrum, and Stolz–Teichner would predict it is a TMF-module spectrum, which deserves the name "$\mathcal{C}$-equivariant TMF." It is still open to construct $\mathcal{C}$-equivariant TMF, but Henriques and Morrison have constructed "$\mathcal{C}$-equivariant $\mathrm{TMF} \otimes \mathbb{Q}$," which is to say they've correctly controlled the Galois on the (well-defined) $\mathcal{C}$-equivariant $\mathrm{TMF} \otimes \mathbb{C}$.

Given my narrative, you could reasonably think that my conjecture is that $V^{f\natural} \times S^3_k$ has a (hopefully unique) lunar nullhomotopy in an $\mathrm{SU}(2)_{h-2} \times \mathrm{Umb}(X)$-equivariant way. This is almost what I want. But it turns out to not quite reproduce the shadow functions of the original umbral moonshine: it produces shadows which are precisely the degree-24 permutation representation, but actually I need to take that one and subtract off some copy of the permutation representation on the components of the Dynkin diagram.

Recall that the $\mathrm{Umb}(X)$-action on Leech arose through the hol(e)y construction from $\mathrm{Niem}(X)$. This construction is reversable (to go the other way, you start with a deep hole in Leech), and is a sort of "gauging" construction for lattices, where the gauge group is a cyclic group. Indeed: in terms of lattice VOAs, the hol(e)y construction is exactly a gauging by $\mathbb{Z}/h$. I think of $V^{f\natural}$ as a sort of "quantum version" of Leech in a way I cannot make precise. There is a "quantum version" of $\mathbb{Z}/h$: it is the modular tensor category corresponding to $\mathrm{SU}(2)_{h-2}$. One way this occurs: this MTC has a $\mathbb{Z}$-form, whose mod-h reduction is the semisimplification of $\mathbf{Rep}_{\mathbb{F}_h}(\mathbb{Z}/h)$, at least when h is prime.

I think that each deep hole in Leech should correspond to an action of the $\mathrm{SU}(2)_{h-2}$ fusion category (forget the braiding) on $V^{f\natural}$. By definition, an action of a fusion category $\mathcal{C}$ on a holomorphic VOA V is a sub-VOA $W \subset V$ together with an equivalence $\mathbf{Rep}(W) \cong Z_{\mathrm{Drinfeld}}(\mathcal{C})$, where Z_{Drinfeld}

is the Drinfeld centre: you think of W as the fixed points of the $\mathcal{C}$-action. You think of the fusion rules of $\mathcal{C}$ as providing a hypergroup, and the associator of $\mathcal{C}$ as being the anomaly of the action.

Conjecture: For each Niemeier lattice, there is an action of the $\mathrm{SU}(2)_{h-2}$ fusion category on $V^{f\natural}$, commuting with the $\mathrm{Umb}(X)$-action.

I know how to construct the correct action for $h = 2$ (trivial), $h = 3$ ($X = A_2^{12}$), and $h = 4$ ($X = A_3^8$). The reason I can do this is that because in those cases the fusion category is an extension of cyclic groups. In all other cases the fusion category is inherently non-groupal in the sense that its simple composition facts have noninvertible elements, and we don't have good tools to do that representation theory. It's not impossible: Brandon Rayhoun has some techniques, and he has used them to find what look like the correct actions for the next couple Niemeier lattices. His methods seem to work when h in the single digits, but his methods break down when h gets too large.

There is also an action of this fusion category on the deep IR of S^3_{h-1} by Verlinde lines. You should think of the Verlinde lines as a sort of "quantum centre" of the SU(2) at this level: the classical centre is level-independent, but the quantum centre depends on the level. Neither of these actions is separately gaugeable, but, using that this fusion category is actually modular, there is a way to "gauge the diagonal action."

Topological umbral moonshine conjecture: The diagonal gauging $(V^{f\natural} \times S^3_{h-1}) /\!\!/ \mathrm{SU}(2)_{h-2}$ is nullhomotopic in $\mathrm{Umb}(X) \times \mathrm{SU}(2)_{h-2}$-equivariant topological lunar forms, and the space of nullhomotopies is connected. Equivalently, $V^{f\natural} \times S^3_{h-1}$ is $\mathrm{Umb}(X) \times \mathrm{SU}(2)_{h-2}$-equivariantly nullhomotopic, but where the (geometric) $\mathrm{SU}(2)_{h-2}$-action on S^3_{h-1} is skewed by a (quantum) fusion-categorical action of $\mathrm{SU}(2)_{h-2}$ on $V^{f\natural}$.

In other words, the conjecture is "just" a cohomology calculation. If everything is set up correctly, the nullhomotopy will necessarily reproduce the umbral moonshine functions, just because those are unique.

Some words of warning. There are good reasons to be suspicious of this conjecture. First, I am worried that it doesn't quite reproduce the growth rates of umbral moonshine. Second, Cheng et al calculated the 't Hooft anomalies of most of the umbral group actions on the putative umbral moonshine SQFTs. Separately, Treumann and I calculated the anomaly of the Co_1 action on $V^{f\natural}$, and the restriction of this class to the umbral groups. The anomalies don't match, and I don't think the as-yet-unknown gauging can meaningfully adjust them. The values calculated by Cheng et al don't seem to follow any rhyme or reason, and certainly there is not a known formula that interpolates them. So there is a missing piece of mathematics: either the Cheng et al calculations have an error, which is highly unlikely, or, more likely, there is some as-yet-undiscovered rule that assigns a cohomology class to each Niemeier lattice. Of course, there is a rule: Cheng et al did the calculation, so the rule can be "look at the lookup table". But that's a stupid rule: there should be an organic one. I think finding this rule (or, unlikely, finding an error in the Cheng et al calculations) would help a lot to clarify the umbral moonshine statements.