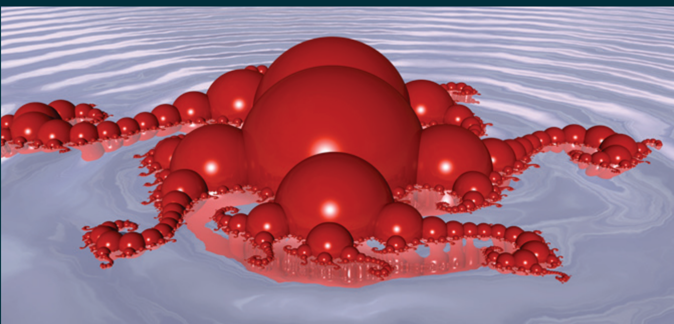
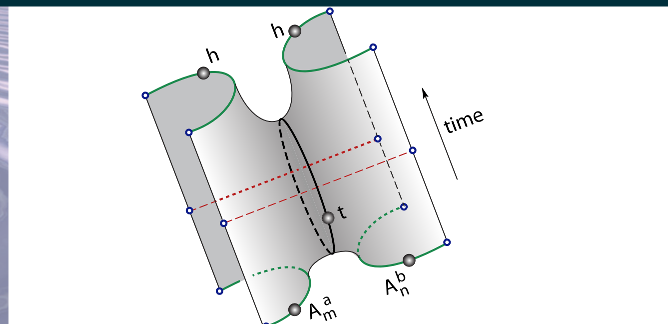
American Mathematical Society

CURRENT EVENTS BULLETIN

Friday, January 6, 2012, 1:00 PM to 4:45 PM
Room 200, Hynes Convention Center
Joint Mathematics Meetings, Boston, MA





1:00 PM Jeffrey F. Brock

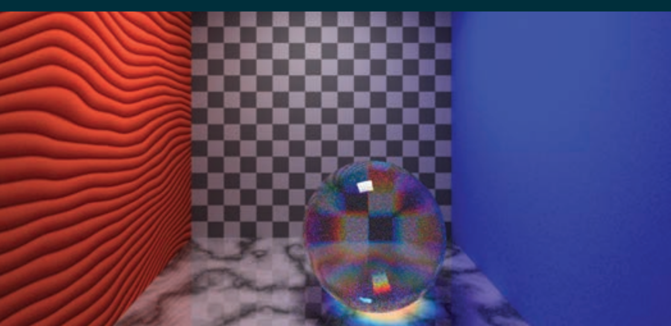
Assembling surfaces from random pants: mixing, matching and correcting in the proofs of the surface-subgroup and Ehrenpreis conjectures

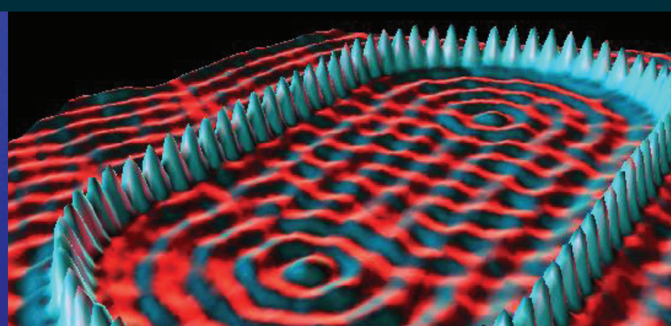
The revolution in low-dimensional topology precipitated by Thurston continues -- we will learn about two of the new breakthroughs descended from it.

2:00 PM Daniel S. Freed

The cobordism hypothesis: quantum field theory + homotopy invariance = higher algebra

Lurie's spectacular work on the cobordism hypothesis is one of the latest demonstrations of the unreasonable effectiveness of physical theory in shaping recent mathematical thought.





3:00 PM Gigliola Staffilani

Dispersive equations and their role beyond PDE

Everyone has heard of the Schrödinger equation, but few understand its surprising interplay with other fields.

4:00 PM Umesh Vazirani

How does quantum mechanics scale?

Quantum mechanics, bizarre as it is, has been tested to exquisite precision. But one aspect of it has not— it is too complex. Here is a computer science view of that complexity, and how the difficulties could shape future developments.

Organized by David Eisenbud, University of California, Berkeley

Introduction to the Current Events Bulletin

Will the Riemann Hypothesis be proved this week? What is the Geometric Langlands Conjecture about? How could you best exploit a stream of data flowing by too fast to capture? I love the idea of having an expert explain such things to me in a brief, accessible way. I think we mathematicians are provoked to ask such questions by our sense that underneath the vastness of mathematics is a fundamental unity allowing us to look into many different corners -- though we couldn't possibly work in all of them. And I, like most of us, love common-room gossip.

The Current Events Bulletin Session at the Joint Mathematics Meetings, begun in 2003, is an event where the speakers do not report on their own work, but survey some of the most interesting current developments in mathematics, pure and applied. The wonderful tradition of the Bourbaki Seminar is an inspiration, but we aim for more accessible treatments and a wider range of subjects. I've been the organizer of these sessions since they started, but a broadly constituted advisory committee helps select the topics and speakers. Excellence in exposition is a prime consideration.

A written exposition greatly increases the number of people who can enjoy the product of the sessions, so speakers are asked to do the hard work of producing such articles. These are made into a booklet distributed at the meeting. Speakers are then invited to submit papers based on them to the *Bulletin of the AMS*, and this has led to many fine publications.

I hope you'll enjoy the papers produced from these sessions, but there's nothing like being at the talks -- don't miss them!

David Eisenbud, Organizer
University of California, Berkeley
de@msri.org

For PDF files of talks given in prior years, see
http://www.ams.org/ams/current-events-bulletin.html.
The list of speakers/titles from prior years may be found at the end of this booklet.

Assembling surfaces from random pants: mixing, matching and correcting in the proofs of the surface-subgroup and Ehrenpreis conjectures

Jeffrey F. Brock *

December 7, 2011

1 Introduction

In his revolutionary work on 3-manifolds, William Thurston saw that many topological questions might be fruitfully pursued with geometric methods. The tools he developed frequently harness hyperbolic geometry to develop sufficient structure to arrive at a topological or algebraic conclusion.

The reader may take as a simple example an element in the fundamental group of a manifold is homotopy class containing an uncountable collection of closed loops, in a manifold with a constant negative curvature metric we can find a *unique* geodesic representative of each homotopy class. More refined structure emerges from dynamical considerations, such as the statistics of the geodesic flow $\phi_t : T^1M \to T^1M$ on the unit tangent bundle of manifold of constant negative curvature.

In the recent work of Jeremy Kahn and Vladimir Markovic [KM1], a beautiful example emerges of the power of Thurston's perspective in the answer to the following conjecture of Waldhausen:

Theorem 1.1 (Kahn-Markovic)**.** THE SURFACE SUBGROUP THEOREM — *Let M be a closed irreducible 3-manifold. Then $\pi_1(M)$ contains a subgroup Γ isomorphic to $\pi_1(S)$ for a closed surface S.*

*Research partially supported by the NSF.

On its face a purely algebraic statement, the Theorem in this generality relies on Perelman's resolution of the geometrization conjecture to ensure that each such 3-manifold admits a geometric decomposition and uses the geometry and dynamics of geodesics in hyperbolic 3-manifolds to arrive at a construction of an immersed, π_1-injective surface in the manifold. (In 3-manifolds with the other, non-hyperbolic geometries the existence had been established previously).

Remarkably, the three-dimensional perspective motivates a construction that well adapted to questions in purely two-dimensional geometry. Specifically, Leon Ehrenpreis asked whether given two hyperbolic Riemann surfaces one can find finite covers that are nearly isometric. (Though he phrased his question in the language of quasi-conformal mappings, the geometric formulation is equivalent). Using a two-dimensional version of the construction of Theorem 1.1, Kahn and Markovic have answered in the affirmative this long-standing *Ehrenpreis Conjecture*:

Theorem 1.2 (Kahn-Markovic). THE EHRENPREIS CONJECTURE — *Let X and Y be hyperbolic Riemann surfaces. Then for each* $\varepsilon > 0$ *there are finite covers* $\widehat{X}$ *and* $\widehat{Y}$ *of X and Y and a* $(1+\varepsilon)$*-bi-Lipschitz diffeomorphism* $f\colon \widehat{X} \to \widehat{Y}$.

The proof recently appeared in the preprint [KM2]. What could the proofs of these two apparently disparate statements have in common? Each relies on a construction of a specific hyperbolic Riemann surface made up from "pairs of pants" of a specific geometric type, with boundary curves comprised of geodesics. These pairs of pants are glued together along their geodesic boundary curves as they sit either in the hyperbolic 3-manifold M, in the first case, or abstractly to form the covering space $\widehat{X}$ in the second case. But in each a central technical effort is required to show that the pairs of pants in the construction can be matched along boundary curves without losing the geometric properties needed. To do this *matching*, one employs the (exponential) *mixing* of the geodesic flow on the unit tangent bundle of the hyperbolic 3-manifold M in the first case, and on the base hyperbolic surface X in the second.[1]

In this article, we will try to illuminate some key insights required to make the argument work. Though the details are beyond the scope of our treatment, the basic outline of proof in each case can be made readily understandable.

Acknowledgement. The author is grateful to Jeremy Kahn for discussions and

[1] A preprint of Lewis Bowen [Bow] observed the connection between the surface subgroup conjecture and the Ehrenpreis conjecture, proving weak forms of each in which the surface or covering space has bounded injectivity radius but may not be closed.

explanations of the proofs of Theorems 1.1 and 1.2. We borrow extensively from lecture notes of Kahn in our discussion, terminology, and examples.

2 Immersions and Covers

The essential elements of the approach of Kahn and Markovic to Theorems 1.1 and 1.2 involves how to build up mappings, be they immersions into a higher dimensional manifold or covering maps, from immersions of a standard piece.

As a test of the concept, let us first consider the case of 1-manifolds. Let C be the circle of radius 1 and consider a family of isometric immersions of the interval of length $2\pi/3$ into C with endpoints at the cube roots of unity. By choosing immersions whose endpoints match up at $\theta = 0$, $\theta = 2\pi/3$ and $\theta = 4\pi/3$, we can glue these immersions together to form a covering space of C. This first case is perhaps deceptive, as the covering space can be made to have degree 1.

Consider, alternatively, isometric immersions of the interval of length 3π for which the endpoints lie at $\theta = 0$ and $\theta = \pi$. Such immersions will not be injective, but choosing (a pair) of immersions and gluing them along their endpoints, we can form a degree-3 cover of the circle.

Finally, consider the set of all isometric immersions of the unit interval into C. The problem now becomes more subtle, as there is no way to build a locally isometric cover out of such intervals since the length of the circle is not a multiple of an integer. If, however, we are willing to take a very large number of intervals, we stretch our intervals slightly by allowing our maps to be *almost isometric* immersions, we may build a locally *almost* isometric cover. As the number of intervals we allow grows, we can construct a cover with covering map that is more and more nearly isometric.

In general, it is a homological problem whether a given collection intervals $\sigma_i \colon I \to C$ can be assembled into a cover: if

$$\sum_i n_i \sigma_i$$

represents a formal collection immersions of oriented intervals with multiplicity $n_i \in \mathbb{Z}$, then provided we have

$$\sum_i n_i \partial \sigma_i = 0,$$

where $\partial \sigma_i$ represents $\sigma_i(1) - \sigma_i(0)$ as a 0-chain, the maps σ_i may be glued together to form a covering space.

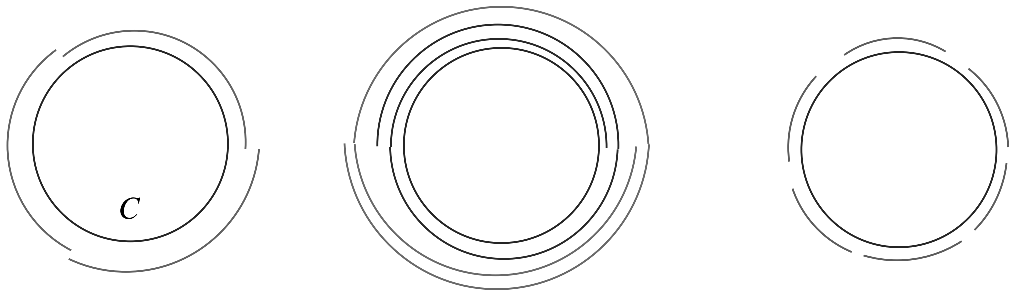


Figure 1. Building covers of the circle from intervals.

We may argue similarly to construct immersed 1-manifolds in a 2-manifold. Given a graph G with geodesic edges in a hyperbolic surface X, it is not entirely clear whether one can piece together the edges end-to-end to obtain an immersion of a 1-manifold (it depends on the structure of the graph). If however, we are willing to consider two copies of each edge, with opposite orientations, then the endpoints may once again be paired to form a map of a (closed) 1-manifold into X.

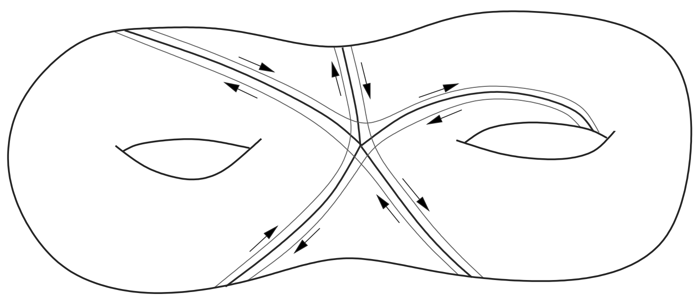


Figure 2. Constructing an immersed 1-manifold from a graph with geodesic edges.

What about the question of how isometric this immersion may be? Surely if there is a vertex in the graph that is a dead-end it is hopeless: the 1-manifold resulting from the above construction will double back on itself. If, however, the edges of G emanate from each vertex of G in an ε-dense and equidistributed set of directions, we can pair up edges in such a way that the resulting path has very small angle at each vertex. A piecewise geodesic in a hyperbolic surface that is

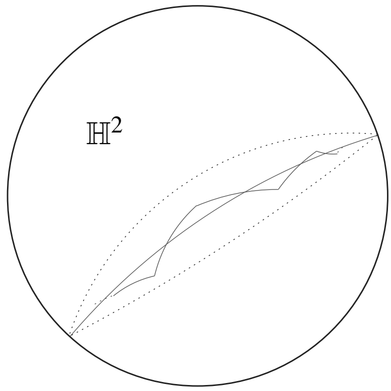


Figure 3. A piecewise geodesic path in the hyperbolic plane with small angles is close to a unique geodesic.

made up of segments of definite length meeting in very small angles lies very close to a unique geodesic in the surface, and is homotopic to this geodesic by a very short homotopy. In particular, if the geodesic is closed, the map is itself very close to an isometric immersion and the immersion is *homotopically non-trivial,* or π_1-*injective*.

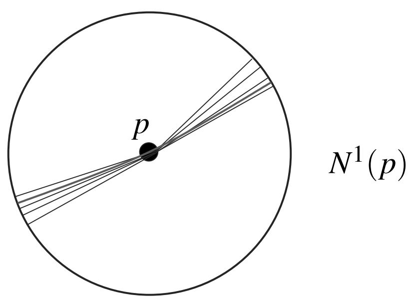


Figure 4. Equidistributed directions allow for nearly isometric immersions.

This circle of ideas motivates a similar discussion concerning constructions of nearly-isometric covers of surfaces and nearly isometric immersions of surfaces into 3-manifolds. Before working out the analogies we introduce some important geometric notions in the hyperbolic geometry of surfaces.

3 Hyperbolic Pairs of Pants

Any surface of negative Euler characteristic can be built out of "pairs of pants," namely, surfaces that are topologically the complement of three embedded open disks in the sphere, glued together along their boundary curves, or "cuffs."

A *hyperbolic pair of pants* is a compact surface with geodesic boundary that is topologically a sphere with three open disks removed. Each pair of cuffs (boundary components of pants) can be connected by a unique shortest *orthogeodesic*, namely, a geodesic joining the cuffs that is orthogonal to each at its endpoints.

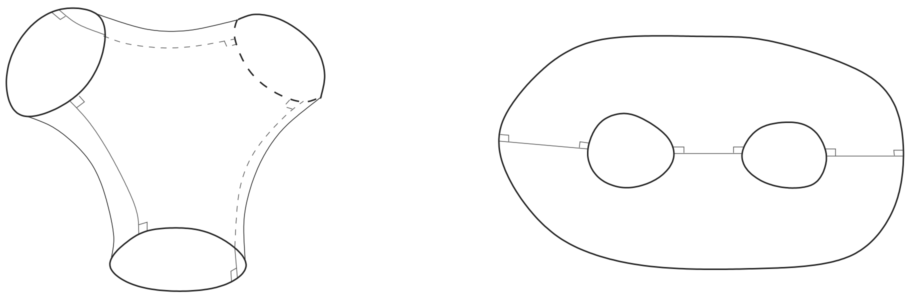


Figure 5. Different perspectives on pairs of pants and their orthogeodesics.

Cutting the hyperbolic pair of pants along the three orthogeodesics joining the boundary components, we obtain a pair of *right-angled hyperbolic hexagons*, namely, hexagons in the hyperbolic plane all of whose interior angles are $\pi/2$. An elementary exercise in hyperbolic geometry shows that three side-lengths of a right-angled hyperbolic hexagon determine its structure uniquely up to isometry. Thus, doubling a right-angled hyperbolic hexagon along three alternating edges produces a pair of pants whose structure is uniquely determined by the geodesic lengths ℓ_1, ℓ_2 and ℓ_3 of the remaining sides of the hexagon. The resulting pair of pants has boundary lengths $2\ell_i$, and we call ℓ_i the *half-lengths* of the pair of pants.

We may glue hyperbolic pairs of pants together to form a hyperbolic Rimeann surface provided we make choose pants whose boundary lengths match up pairwise. The structure of the result appears at first only to depend on the lengths of the *pants curves*, the curves involved in the gluing, but an additional parameter is involved, coming from the displacement of the feet of the orthogeodesics that lie on the given boundary curve. This "shear" parameter, together with the lengths of the pants curves produce Fenchel-Nielsen coordinates for the Teichmüller space, the parameter space of hyperbolic structures on a surface S up to isotopy. We will not need explicit use of these coordinates, but it is useful to keep in mind the fact that the lengths and shears along the pants curves determine the structure on the surface up to isometry.

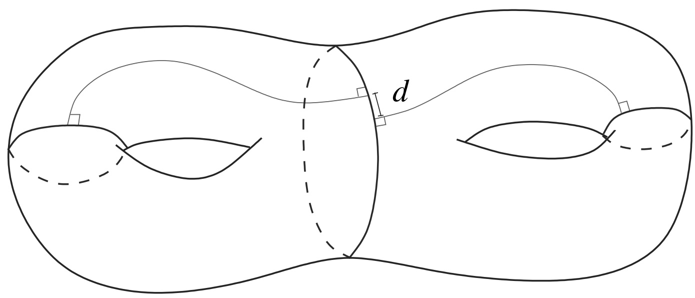


Figure 6. A surface obtained by gluing hyperbolic pants.

4 Perfect Surfaces and Merely Good Surfaces

A central element of the construction of Kahn and Markovic is to find a “perfect” model surface for their covering spaces. Given a (large) constant R, we consider the (unique) hyperbolic pair of pants P all of whose boundary half-lengths are exactly R. Then an *R-perfect surface* is obtained by gluing together copies of P so that the displacement of the feet of the orthogeodesics is precisely unit distance to the left (since the surface is oriented, displacement “to the left” makes sense from either direction).

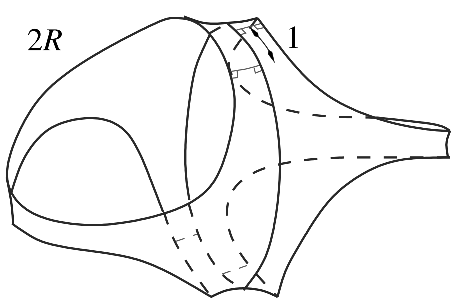


Figure 7. Gluing good pants with a good shear.

Kahn and Markovic show the following

Proposition 4.1. *Any two closed perfect surfaces have a common finite cover.*

The idea of proof is somewhat the reverse: they show that an R-perfect surface admits a branched cover to a universal *R-perfect orbifold O_R*, obtained by taking the quotient of an R-perfect surface by a maximal set of symmetries. Then one can find a common finite cover by intersecting the images of their fundamental groups in the orbifold fundamental group of O_R.

The transition from *perfect* surfaces to those that are merely *good* involves relaxing the conditions on half-lengths to allow an ε error, where $\varepsilon > 0$ and for the gluing displacement to differ from 1 by ε/R. More precisely, while a perfect surface admits a pants decomposition P so that each curve $\gamma \in P$ has half-length $hl(\gamma) = R$ and pants determined by P are glued with displacement $s(\gamma) = 1$, a *good* surface has

$$|hl(\gamma) - R| < \varepsilon \quad \text{and} \quad |s(\gamma) - 1| < \varepsilon/R$$

for each $\gamma \in P$.

Theorem 4.2. *For all $\varepsilon < \varepsilon_0$, and $R > R_0$, any (ε, R)-good surface is $10^{12}\varepsilon$-close to an R-perfect surface in the Teichmüller metric.*

The *Teichmüller metric* is defined in terms of quasiconformal distortion of minimal distortion quasiconformal maps in a given isotopy class. For the purposes of our discussion here, let us say that the surfaces are nearly isometric by an diffeomorphism that sends corresponding pants curves to pants curves and sends feet nearby to corresponding feet.

Provided, then, that we can find a good cover $\widehat{X}$ of an arbitrary hyperbolic Riemann surface X, this cover is close to a perfect surface $\widehat{X}'$. If Y is another hyperbolic surface and $\widehat{Y}$ is its good cover, the nearby perfect surface $\widehat{Y}'$ has a finite cover in common with the perfect surface $\widehat{X}'$. Since the distance in the Teichmüller metric is invariant under passing to covers, it follows that X and Y have finite covers that are close to the common perfect cover of $\widehat{X}'$ and $\widehat{Y}'$. This proves the Ehrenpreis conjecture modulo the task of finding a good cover of an arbitrary hyperbolic Riemann surface.

5 Good Covers

Though we have up to now considered surfaces built from pairs of pants by gluing along boundary components, a pair of pants decomposition of a finite cover of a Riemann surface X will not in general descend to a pants decomposition of the base except in very symmetric situations. The covering map will restrict to an immersion on each pair of pants, and the boundary curves will project to closed geodesics that need not be simple: their projections may have many self-intersections. But given an isometric immersion of a pair of pants, how can we see that it came from a cover of X?

The challenge of constructing a good cover can be reformulated into a challenge of matching up isometric immersions of *good pants*, pants whose half lengths are within ε of R, in such a way that they match along a closed geodesic in X with a shear within ε/R of 1.

In other words, given two good pants P_1 and P_2 immersed isometrically into X and a closed geodesic γ in X so that $\gamma_1 \subset \partial P_1$ and $\gamma_2 \subset \partial P_2$ and γ_1 and γ_2 each map isometrically to γ, the only obstruction to gluing these immersions along γ_1 and γ_2 is that the immersions must send P_1 and P_2 to the opposite side of γ.

Much like the situation in figure 1, given a finite collection of isometrically immersed pants in X, we may glue these pants into a finite cover of X provided only that for each closed geodesic γ in X, the number of pants mapping a boundary curve to γ that lie to one side of γ is the same as the number mapping a boundary curve to γ that lie to the other side.

But if we are in search of *good* covers, we would like the geodesics in this consideration to be of length close to R, and we would like additional shearing constraints on the gluings along such curves. To this end, we consider the sets $\mathscr{G}(X)$ of all closed geodesics in X and $\mathscr{P}(X)$ of all isometrically immersed hyperbolic pairs of pants. Then we may attempt to build good covers out of good pants: let

$$\Gamma_{\varepsilon,R} = \{\gamma \mid \gamma \in \mathscr{G}(X), |hl(\gamma) - R| < \varepsilon\}$$

and

$$\Pi_{\varepsilon,R} = \{P \mid P \in \mathscr{P}(X), \text{ and } \partial P \subset \Gamma_{\varepsilon,R}\}$$

the (ε,R)-good pants in X. To glue (ε,R)-good pants into an (ε,R)-good cover, we consider the *feet* of the orthogeodesics in P on each boundary curve of P. As these feet divide the cuff into two geodesic segments of equal length, we may think of the position of the feet as determining a single point in the unit normal bundle of the "square root" of the geodesic γ in X, denoted $N^1(\sqrt{\gamma})$, by a slight abuse of notation.

Applying the exponential mixing of the geodesic flow on X, Kahn and Markovic establish the following equidistribution theorem, which is a central technical tool in the construction.

Theorem 5.1 (Kahn-Markovic)**.** EQUIDISTRIBUTION THEOREM 1 — *For every good curve γ, the set of feet* $\mathrm{feet}_\gamma(P)$ *of for $P \in \Pi_{\varepsilon,R}$ and $\gamma \subset \partial P$ is evenly distributed in $N^1(\sqrt{\gamma})$ to the scale e^{-qR} for q depending only on the topology of X and the choice of ε.*

Here we say the set of feet are *evenly distributed to the scale ε* if for any two ε-intervals on the unit normal bundle the ratio of the number of feet in these intervals lies in the interval $(1-\varepsilon, 1+\varepsilon)$.

All that is lacking to piece together the pants in $\Pi_{\varepsilon,R}$ into a good cover is to know that there are an equal number of pants on either side of each geodesic $\gamma \in \Gamma_{\varepsilon,R}$. Indeed the equidistribution guarantees that the sets of feet that correspond to pants on each side of γ are almost the same, but an imbalance could in principle propogate through the whole construction. Unfortunately, the "doubling trick" of figure 2 does not directly work here, since doubling each pair of pants will just double whatever imbalance may exist across the normal bundle. This has ultimately to do with the fact that the normal bundle is *disconnected*; the key to making the doubling trick work previously was the connecteness of the normal bundle for codimension-2 submanifolds.

This suggests returning to the setting of 3-manifolds to reassess what the foregoing notions may imply.

6 Nearly Geodesic Immersed Surfaces in 3-Manifolds

As the normal bundle of a closed geodesic in a hyperbolic 3-manifold is connected, we may attempt to use the doubling trick of figure 2 to produce a nearly totally geodesic immersed surface in a hyperbolic 3-manifold. Recall we argued that a nearly geodesically immersed closed loop in a hyperbolic surface is homotopically essential – given a nearly geodesic immersion of a hyperbolic surface X into a hyperbolic 3-manifold, each element of the fundamental group of X is itself nearly geodesically immersed in M. As in the 2-dimensional case, a piecewise geodesic path made up of segments of definite length and with small angle is once again near a unique geodesic. It follows that such a surface immerses in M in a π_1-injective manner.

But what of the equidistribution theorem? The normal bundle to a closed geodesic γ in M is a torus, but we can give this torus an explicit representation in terms of the geodesic γ: γ is the quotient of the axis of an element $\varphi_\gamma \in \mathrm{PSL}_2(\mathbb{C})$, in the isometries of hyperbolic 3-space $\mathbb{H}^3$, acting *loxodromically*. Indeed, the element φ_γ is conjugate in $\mathrm{PSL}_2(\mathbb{C})$ to an element of the form $z \mapsto e^{\lambda(\gamma)}z$ where $\lambda(\gamma) = \ell(\gamma) + i\theta(\gamma)$ is the *complex length* of γ. The torus $\mathbb{C}^*/\langle\varphi_\gamma\rangle$ represents the unit-normal bundle of the geodesic γ.

A *skew pair of pants* in M is an immersion of a hyperbolic pair of pants P into M that sends geodesic boundary components to closed geodesics representatives in M, and sends orthogeodesics to geodesics orthogonal to the images of the boundary components. These orthogeodesics then determine points in the unit-normal bundle to γ, the corresponding boundary curve.

Theorem 6.1 (Kahn-Markovic)**.** EQUIDISTRIBUTION THEOREM 2 — *The feet*

$$\{\mathrm{feet}_\gamma(P) \mid \gamma \in \partial P, P \in \Pi_{\varepsilon,R}\}$$

are e^{-qR}-evenly distributed as points on $N^1(\sqrt{\gamma})$.

The equidistribution theorem guarantees that if we consider the set

$$A_\gamma = \{\mathrm{feet}_\gamma(P) \mid \gamma \in \partial P, P \in \Pi_{\varepsilon,R}\}$$

we can find a permutation $\sigma\colon A_\gamma \to A_\gamma$ that moves every point by a translation exactly $i\pi + 1$ up to error ε/R. In other words, the feet that correspond under the permutation σ are nearly opposed and are offset by nearly 1 along the geodesic γ. By doubling to obtain two copies A_γ^+ and A_γ^- of A_γ, the permutation and its inverse give a perfect matching of the union $A_\gamma^+ \sqcup A_\gamma^-$ with itself. Because the feet are matched in such a way that the "incoming" skew pants and the "outgoing" skew pants are nearly opposed in the normal bundle and sheared by 1 (using Hall's Marriage Theorem) a geodesic traversing the assembled pants will cross pants curves with small bending and do so in at worst segments of definite length – the shearing condition ensures that the pants curves do not "pile up." The result is a nearly Fuchsian, *quasi-Fuchsian* subgroup of the Kleinian group Γ uniformizing M.

In particular, Kahn and Markovic prove the following:

Theorem 6.2 (Kahn-Markovic)**.** GOOD SURFACE ALMOST FUCHSIAN — *There exist R_0, K_0, and $\varepsilon_0 > 0$ so that if $\rho\colon \pi_1(S) \to \mathrm{PSL}_2(\mathbb{C})$ is a representation induced from the inclusion of an (R,ε)-good panted surface with $R > R_0$ and $\varepsilon < \varepsilon_0$, we may find an R-perfect hyperbolic surface $\mathbb{H}/\Gamma$, with $\Gamma < \mathrm{PSL}_2(\mathbb{R}) < \mathrm{PSL}_2(\mathbb{C})$, and an equivariant mapping $h\colon \mathbb{H}^3 \to \mathbb{H}^3$ that extends to a $K_0\varepsilon$-quasiconformal map on $\partial_\infty\mathbb{H}^3 = \widehat{\mathbb{C}}$.*

In particular, ρ is a faithful, discrete, quasi-Fuchsian representation.

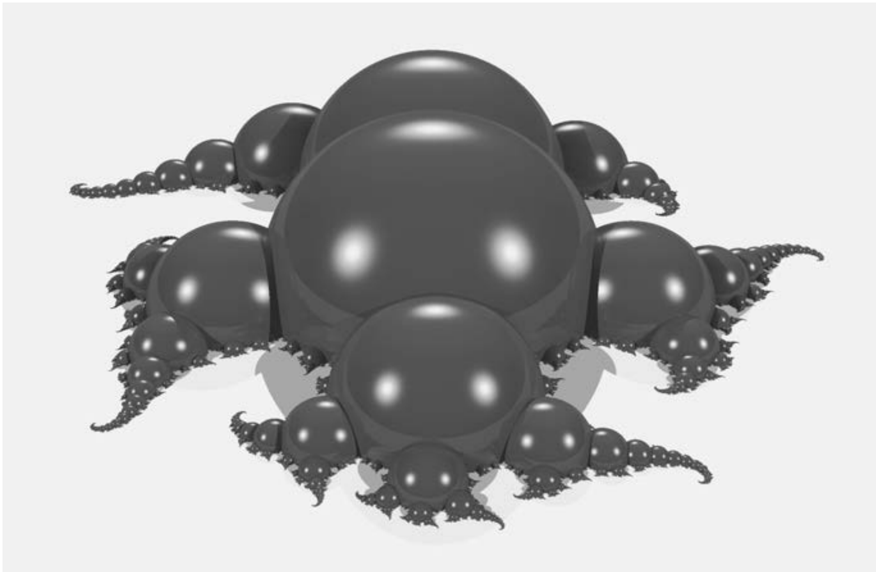


Figure 8. The universal cover of a quasi-Fuchsian surface obtained from gluing totally geodesic pairs of pants along their boundary components.

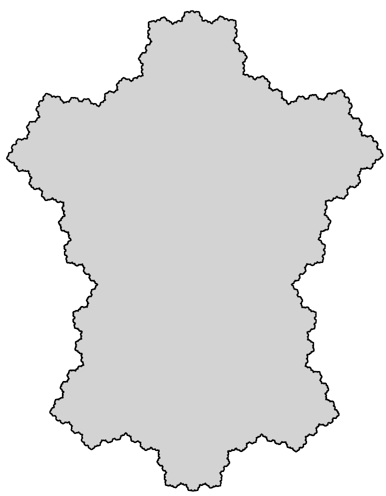


Figure 9. An almost Fuchsian Quasi-Fuchsian limit set.

The mapping h sends the circle $\mathbb{R}\cup\infty$ to a *quasi-circle* in $\widehat{\mathbb{C}}$, the limit set of the quasi-Fuchsian image group (see figure 9).

This suffices to verify Waldhausen's surface subgroup conjecture in the hyperbolic case, which was the remaining case that was unknown, after Perelman's proof of the geometrization conjecture. We will say more about implications of Theorem 1.1 after we briefly sketch the necessary modifications to the argument that address the problem presented by the disconnected normal bundle in the Ehrenpreis case.

7 Self-Correction and the Ehrenpreis Conjecture

As noted above in the case of the circle, when we are gluing 1-manifolds to form a cover of the circle, a homological condition should be satisfied: the formal sum of these 1-manifolds should be a cycle in homology, guaranteeing that their boundary points cancel in pairs.

While the case of connected normal bundle outlined above allows for a doubling argument, gluing pants in a surface along a geodesic presents the possibility of an imbalance among all the good pants that contain a given curve in their boundary. Equidistribution, however, guarantees that the pants that lie to each side are *almost* balanced, and we find that removing some of these pants from the collection gives a balanced collection that is still sufficiently equidistributed to produce an (ε,R)-good surface.

To proceed formally, Kahn and Markovic introduce a homology theory of *good pants* to show the desired "closing up" can be carried out with a good surface. We will omit this level of detail but we describe the following additional structure: let $\Gamma_{\varepsilon,R}$ be the oriented closed geodesics on X of length within ε of R, the so-called *good curves*. Then if $\mathbb{Z}\Gamma_{\varepsilon,R}$ is the formal sums of elements in $\Gamma_{\varepsilon,R}$, we may take the obvious boundary map

$$\partial\colon \Pi_{\varepsilon,R} \to \mathbb{Z}\Gamma_{\varepsilon,R}$$

from the collection of good pants to the formal sums of good curves.

Kahn and Markovic argue that when R is large, depending on the topology of X and on ε, there is a correction function

$$q\colon \mathbb{Q}\Gamma_{\varepsilon,R} \to \mathbb{Q}\Pi_{\varepsilon,R}$$

satisfying

1. $\partial q(\partial P) = \partial P$ if $P \in \mathbb{Q}\Pi_{\varepsilon,R}$ and
2. $\|q(\alpha)\|_\infty \leq e^{-R}P(R)\|\alpha\|_\infty$ for any $\alpha \in \mathbb{Q}\Gamma_{\varepsilon,R}$, where $P(R)$ is a polynomial in R.

The second item reflects, in some sense, that *most* of the pants can be matched if they are equidistributed. This is again a consequence of the exponential mixing of the geodesic flow.

Then letting

$$\pi = \sum_{P \in \Pi_{\varepsilon,R}} P$$

be the sum of the good pants, we can balance this collection by considering

$$\pi' = \pi - q(\partial\pi).$$

We note that $\partial\pi' = \partial\pi - \partial q(\partial\pi) = 0$ by (1) so π' is a collection of pants that is balanced. Furthermore, for large R $q(\partial\pi)$ is much smaller than π by (2) so π' is still very well distributed. It follows that the matching can take place in such a manner that yields a $(K\varepsilon, R)$-surface, where K is a universal constant.

This is what was required to find the good cover of X needed to prove the Ehrenpreis conjecture.

8 Epilogue

Though the proof of the Ehrenpreis conjecture represents a major breakthrough in the study of Teichmüller theory, particularly the behavior of Teichmüller geometry in the passage to covers, the work of Kahn and Markovic on π_1-injective immersed surfaces in hyperbolic 3-manifolds is part of a compelling larger story in the recent history of 3-manifold topology.

While Perelman's proof of the geometrization conjecture is a central part of the proof of Theorem 1.1 as stated, the existence of a π_1-injective surface in a hyperbolic 3-manifold had been a central question popularized by the work of Thurston. Thurston's original hyperbolization theorem applied to *Haken* 3-manifolds, namely, those 3-manifolds M that admit an *embedded* π_1-injective surface (see, e.g., [Th2], [Th3], [Ot], [Kap]).

Thurston asked as a kind of challenge problem whether a compact, orientable, irreducible, atoroidal 3-manifold M with infinite fundamental group admits a finite cover that *fibers over the circle* (he also remarked at the time that "this dubious sounding question seems to have a definite chance for a positive answer" [Th1]). This would produce, via Thurston's original theorem, a geometric structure for a finite cover which would then descend to the original manifold. Such manifolds are now known to be hyperbolic by Perelman.

But efforts toward the conjecture have generally taken the geometric structure as a hypothesis. The proof of Kahn-Markovic of Theorem 1.1 is a key ingredient, in that the existence of an injective surface subgroup is a necessary condition for either conjecture. Much work is underway to "separate" the Kahn-Markovic surface – were one to be able to pass to a finite cover of M that eliminated self intersections of the Kahn-Markovic surface one would have a proof of the virtually Haken conjecture (recent results of Agol together with announced work of Wise imply that this suffices to prove the virtually fibered conjecture as well [Ag]).

Even without this "virtually fibered conjecture" the Kahn-Markovic Surface Subgroup Theorem provides the possibility of deeper insight into the nature of the geometric structure on a closed 3-manifold M. Indeed, *the Geometrization Theorem tells one very little about a hyperbolic structure on M other than that it is indeed hyperbolic.*

The Kahn-Markovic Surface Subgroup Theorem represents an compelling instance of an strain in research begun by Thurston in 3-manifold topology directed at the following question:

Question 8.1. *How does the topology of a hyperbolic* 3*-manifold determine its geometry?*

The beautiful role of dynamics in the proof of Theorem 1.1 suggests a new tool for approaching Question 8.1 and other questions in the study of hyperbolic 3-manifolds. The techniques of Kahn and Markovic are as compelling as their results, and this body of work will undoubtedly inspire many more threads of inquiry than it ties off.

References

[Ag] I. Agol. Criteria for virtual fibering. *Journal of Topology* **1**(2008), 269–284.

[Bow] L. Bowen. Weak Forms of the Ehrenpreis Conjecture and the Surface Subgropu Conjecture. *Preprint, arXiv:math/0411662*.

[Br] J. Brock. Weil-Petersson translation distance and volumes of mapping tori. *Comm. Anal. Geom.* **11**(2003), 987–999.

[BCM] J. Brock, R. Canary, and Y. Minsky. The classification of Kleinian surface groups, II: the ending lamination conjecture. *Submitted (2004).*

[BD] J. Brock and N. Dunfield. Injevtivity radii of hyperbolic integer homology 3-spheres. *in preparation*.

[KM1] J. Kahn and V. Markovic. Immersing almost geodesic surfaces in a closed hyperbolic three manifold. *To appear, Annals of Math., arXiv:0910.5501* (2009).

[KM2] J. Kahn and V. Markovic. The good pants homology and a proof of the Ehrenpreis conjecture. *Preprint, arXiv:1101.1330* (2011).

[Kap] M. Kapovich. *Hyperbolic manifolds and discrete groups*. Birkhäuser Boston Inc., Boston, MA, 2001.

[Mos] D. Mostow. *Strong rigidity of locally symmetric spaces*. Annals of Math Studies 78, Princeton University Press, 1972.

[NS] H. Namazi and J. Souto. Heegaard splittings and pseudo-Anosov maps. *Geom. and Funct. Anal.* **19**(2009), 1195–1228.

[Ot] J. P. Otal. *Le théorème d'hyperbolisation pour les variétés fibrées de dimension trois*. Astérisque, 1996.

[Th1] W. P. Thurston. Three-dimensional manifolds, Kleinian groups and hyperbolic geometry. *Bull. AMS* **6**(1982), 357–381.

[Th2] W. P. Thurston. Hyperbolic structures on 3-manifolds I: Deformations of acylindrical manifolds. *Annals of Math.* **124**(1986), 203–246.

[Th3] W. P. Thurston. Hyperbolic structures on 3-manifolds II: Surface groups and 3-manifolds which fiber over the circle. *Preprint,* `arXiv:math.GT/9801045` (1986).

THE COBORDISM HYPOTHESIS

DANIEL S. FREED

Abstract. In this expository paper we introduce extended topological quantum field theories and the cobordism hypothesis.

1. Introduction

The *cobordism hypothesis* was conjectured by Baez-Dolan [BD] in the mid 1990s. It has now been proved by Hopkins-Lurie in dimension two and by Lurie in higher dimensions. There are many complicated foundational issues which lie behind the definitions and the proof, and only a detailed sketch [L1] has appeared so far.[1] The history of the Baez-Dolan conjecture goes most directly through quantum field theory and its adaptation to low-dimensional topology. Yet in retrospect it is a theorem about the structure of manifolds in all dimensions, and at the core of the proof lies Morse theory. Hence there are two routes to the cobordism hypothesis: algebraic topology and quantum field theory.

Consider the abelian group Ω_0^{SO} generated by compact oriented 0-dimensional manifolds, that is, finite sets Y of points each labeled with $+$ or $-$. The group operation is disjoint union. We deem Y_0 equivalent to Y_1 if there is a finite union X of compact oriented 1-manifolds with oriented boundary $Y_1 \amalg -Y_0$. Then a basic theorem in differential topology [Mi1, Appendix] asserts that Ω_0^{SO} is the free abelian group with a single generator, the positively oriented point pt_+.[2] This result is the cornerstone of smooth intersection theory. From the point of view of algebraic topology the cobordism hypothesis is a similar statement about a more ornate structure built from smooth

Received by the editors November 15, 2011.

The work of D.S.F. is supported by the National Science Foundation under grant DMS-0603964.

[1]Nonetheless, we use ‘theorem’ and its synonyms in this manuscript.

[2]Two important remarks: (1) we can replace orientations with framings; (2) for unoriented manifolds the group Ω_0^{O} is not free on one generator, but rather there is a relation and $\Omega_0^{O} \cong \mathbb{Z}/2\mathbb{Z}$.

manifolds. The simplest version is for framed manifolds. The language is off-putting if unfamiliar, and it will be explained in due course.

Theorem 1.1 (Cobordism hypothesis: heuristic algebro-topological version)**.** *For $n \geq 1$, $\mathrm{Bord}_n^{\mathrm{fr}}$ is the free symmetric monoidal (∞, n)-category with duals generated by pt_+.*

The 'Bord' in $\mathrm{Bord}_n^{\mathrm{fr}}$ stands for 'bordism',[3] and pt_+ is now the point with the standard framing. $\mathrm{Bord}_n^{\mathrm{fr}}$ is an elaborate algebraic gadget which encodes n-framed manifolds with corners of dimensions $\leq n$ and tracks gluings and disjoint unions. One of our goals is to motivate this elaborate algebraic structure.

An *extended topological field theory* is a representation of the bordism category, i.e., a homomorphism $F\colon \mathrm{Bord}_n^{\mathrm{fr}} \to \mathcal{C}$. The codomain $\mathcal{C}$ is a symmetric monoidal (∞, n)-category, typically linear in nature. In important examples F assigns a complex number to every closed n-manifold and a complex vector space to every closed $(n-1)$-manifold.

Theorem 1.2 (Cobordism hypothesis: weak quantum field theory version)**.** *A homomorphism $F\colon \mathrm{Bord}_n^{\mathrm{fr}} \to \mathcal{C}$ is determined by $F(\mathrm{pt}_+)$.*

The object $F(\mathrm{pt}_+) \in \mathcal{C}$ satisfies stringent finiteness conditions expressed in terms of dualities, and the real power of the cobordism hypothesis is an existence statement: if $x \in \mathcal{C}$ is *n-dualizable*, then there exists a topological field theory F with $F(\mathrm{pt}_+) = x$.

Our plan is to build up gradually to the categorical complexities inherent in extended field theories and the cobordism hypothesis. So in the next two sections we take strolls along the two routes to the cobordism hypothesis: algebraic topology (§2) and quantum field theory (§3). Section 4 is an extended introduction to non-extended topological field theory. The simple examples discussed there only hint at the power of this circle of ideas. In §5 we turn to extended field theories and so also to higher categories. The cobordism hypothesis is the subject of §6, where we state a complete version in Theorem 6.8. The cobordism hypothesis connects in exciting ways to other parts of topology, geometry, and representation theory as well as to some contemporary ideas in quantum field theory. A few of these are highlighted in §7.

The manuscript [L1] has leisurely introductions to higher categorical ideas and to the setting of the cobordism hypothesis, in addition to a detailed

[3]'Bordism' replaces the older 'cobordism', as bordism is part of homology whereas cobordism is part of cohomology [A1].

sketch of the proof and applications. The original paper [BD] is another excellent source of expository material. We have endeavored to complement these expositions rather than duplicate them. I warmly thank David Ben-Zvi, Andrew Blumberg, and Tim Perutz for their comments and suggestions.

2. Algebraic topology

The most basic maneuvers in algebraic topology extract algebra from spaces. For example, to a topological space X we associate a sequence of abelian groups $H_q(X)$. There are several constructions of these *homology groups*, but for nice spaces they are all equivalent [Sp]. The homology construction begins to have teeth only when we tell how homology varies with X. One elementary assertion is that if $X \simeq Y$ are homeomorphic spaces, then the homology groups are isomorphic. Thus numerical invariants of homology groups, such as the rank, are homeomorphism invariants of topological spaces: Betti numbers. But it is much more powerful to remember the isomorphisms of homology groups associated to homeomorphisms, and indeed the homomorphisms associated to arbitrary continuous maps. This is naturally encoded in the algebraic structure of a *category*. Here is an informal definition; see standard texts (e.g. [Mc]) for details.

Definition 2.1. *A* category $\mathcal{C}$ *consists of a set*[4] C_0 *of objects* $\{x\}$*, a set* C_1 *of morphisms* $\{f\colon x \to y\}$*, identity elements* $\{id_x\colon x \to x\}$*, and an associative composition law* $f, g \longmapsto g \circ f$ *for morphisms* $x \xrightarrow{f} y$ *and* $y \xrightarrow{g} z$*. If* $\mathcal{C}, \mathcal{D}$ *are categories then a* homomorphism[5] $F\colon \mathcal{C} \to \mathcal{D}$ *is a pair* (F_0, F_1) *of maps of sets* $F_i\colon \mathcal{C}_i \to \mathcal{D}_i$ *which preserve identity maps and compositions.*

More formally, there are source and target maps $C_1 \to C_0$, identity elements are defined by a map $C_0 \to C_1$, and composition is a map from a subset of $C_1 \times C_1$ to C_1—the subset consists of pairs of morphisms for which the target of the first equals the source of the second. Topological spaces comprise the objects of a category Top whose morphisms are continuous maps; abelian groups comprise the objects of a category Ab whose morphisms are group homomorphisms. Some basic properties of homology groups are summarized

[4] We do not worry about technicalities of set theory in this expository paper.

[5] The word ‘functor’ is usually employed here, but ’homomorphism’ is more consistent with standard usage elsewhere in algebra.

by the statement that

$$H_q\colon (\mathrm{Top}, \amalg) \longrightarrow (\mathrm{Ab}, \oplus) \tag{2.2}$$

is a homomorphism. We explain the '$\amalg$' and '$\oplus$' in the next paragraph.

The homomorphism property does not nearly characterize homology, and we can encode many more properties via extra structure on Top and Ab. We single out one here, an additional operation on objects and morphisms. If X_1, X_2 are topological spaces there is a new space $X_1 \amalg X_2$, the *disjoint union*. The operation $X_1, X_2 \mapsto X_1 \amalg X_2$ has properties analogous to a commutative, associative composition law on a set. For example, the empty set $\emptyset$ is an identity for disjoint union in the sense that $\emptyset \amalg X$ is canonically identified with X for all topological spaces X. Furthermore, if $f_i\colon X_i \to Y_i,\ i = 1, 2$ are continuous maps, there is an induced continuous map $f_1 \amalg f_2\colon X_1 \amalg X_2 \to Y_1 \amalg Y_2$ on the disjoint union. An operation on a category with these properties is called a *symmetric monoidal structure*, in this case on the category Top. Similarly, the category Ab of abelian groups has a symmetric monoidal structure given by direct sum: $A_1, A_2 \to A_1 \oplus A_2$. The homology maps (2.2) are homomorphisms of symmetric monoidal categories: there is a canonical identification of $H_q(X_1 \amalg X_2)$ with $H_q(X_1) \oplus H_q(X_2)$.

Remark 2.3. Homology is *classical* in that disjoint unions map to *direct sums*. We will see that a characteristic property of *quantum* systems is that disjoint unions map to *tensor products*. The passage from classical to quantum is therefore a kind of exponentiation.

Our interest here is not all topological spaces, but rather smooth manifolds. Fix a positive integer n.

Definition 2.4. *Let Y_0, Y_1 be smooth compact $(n-1)$-dimensional manifolds without boundary. A* bordism from Y_0 to Y_1 *is a compact n-dimensional manifold X with boundary, a decomposition $\partial X = \partial X_0 \amalg \partial X_1$, and diffeomorphisms $Y_i \to \partial X_i$, $i = 1, 2$.*

Figure 1 depicts an example which emphasizes that manifolds need not be connected. The empty set $\emptyset$ is a manifold of any dimension. So a *closed* n-manifold—that is, a compact manifold without boundary—is a bordism from $\emptyset^{n-1}$ to $\emptyset^{n-1}$. Note also that the disjoint union of smooth manifolds is a smooth manifold, and the disjoint union of bordisms is a bordism.

To turn bordism into algebra we observe that bordism defines an equivalence relation: closed $(n-1)$-manifolds Y_0, Y_1 are bordant if there exists a

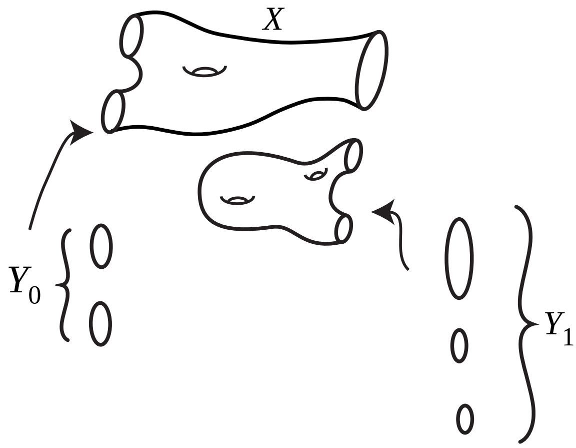


FIGURE 1. A bordism $X\colon Y_0 \to Y_1$

bordism from Y_0 to Y_1. (Observe that to prove transitivity it is convenient to modify Definition 2.4 so that boundary identifications are between the open manifold $(-\epsilon, \epsilon)\times Y_i$ and an open 2-sided collar neighborhood of ∂X_i: smooth functions glue nicely on open sets.) Disjoint union defines an abelian group structure on the set Ω^O_{n-1} of equivalence classes. For example, $\Omega^O_0 \cong \mathbb{Z}/2\mathbb{Z}$ is generated by a single point. Twice a point is the disjoint union of two points, and as two points bound a closed interval, two points are bordant to the empty 0-manifold. Life is more interesting when we consider manifolds with extra topological structure. For example, there are bordism groups Ω^{SO}_q of oriented manifolds. An orientation on a 0-manifold consisting of a single point is a choice of $+$ or $-$. Then $\Omega^{SO}_0 \cong \mathbb{Z}$ by the map which sends a finite set of oriented points to the number of positive points minus the number of negative points. This is a foundational result in differential topology which enables oriented counts in intersection theory [Mi1]. Another interesting structure is a *stable framing*. It arises in the Pontrjagin-Thom construction. Let $f\colon S^{q+N} \to S^N$ be a smooth map. By Sard's theorem there is a regular value $p \in S^N$, whence $M := f^{-1}(p) \subset S^{q+N}$ is a smooth q-dimensional submanifold. Also, a basis of T_pS^N pulls back under f to a global framing of the normal bundle to M in S^N. If we deform p to another regular value, then the framed manifold M undergoes a bordism. The same is true if f deforms to a smoothly homotopic map. The precise correspondence works in the stable limit $N \to \infty$: the stably framed bordism group Ω^{fr}_q is isomorphic to the stable homotopy group of the sphere $\lim\limits_{N\to\infty} \pi_{q+N}(S^N)$. This is the most basic link between bordism and homotopy theory.

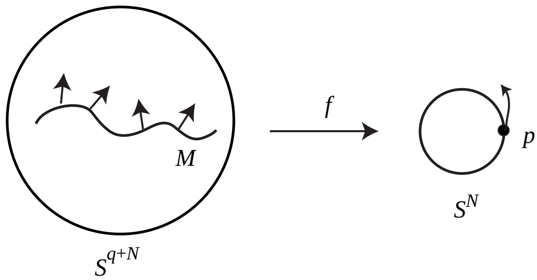


FIGURE 2. The Pontrjagin-Thom construction

Bordism has a long history in algebraic topology. By 1950 it appears[6] that Pontrjagin had defined abelian groups based on the notion of a bordism, though it was Thom [T] who made the first systematic computations of bordism groups using homotopy theory. There are many variations according to the type of manifold: oriented, spin, framed, etc. Theory and computation of bordism groups were an important part of algebraic topology in the 1950s and 1960s, and they found applications in other parts of topology and geometry. For example, Hirzebruch's 1954 proof of the Riemann-Roch theorem was based on bordism computations, as was the first proof of the Atiyah-Singer index theorem [Pa] in 1963.

The bordism group of d-dimensional manifolds arises when $(d+1)$-dimensional bordisms are used to define an equivalence relation. Disjoint union of d-manifolds gives the abelian group structure. One lesson from classical algebraic topology is that the passage from Betti numbers to homology groups is very fruitful. The analog here is to *track* bordisms between closed manifold, not merely to observe their existence—in our "categorified" world we *encode* the bordism as a map. Segal [Se2] introduced a bordism category of Riemann surfaces in his axiomatization of 2-dimensional conformal field theory, which inspired Atiyah [A2] to axiomatize topological field theories in any dimensions using bordism categories of smooth manifolds with no continuous geometric structure (such as a metric or conformal structure). Tillmann [Til1, Til2] observed that the classifying space of the bordism category, which has the abelian group-like operation of disjoint union, is a *spectrum* in the sense of stable homotopy theory. Together with Madsen [MT] they conjecturally identify the classifying spectrum of an enriched bordism category—a step towards the ∞-categories we meet in §5—and show that

[6] According to [May, §6] a 1950 Russian paper of Pontrjagin contains bordism groups; see [P] for a later account. Thom [T] also cites work of Rohlin relevant to computations of bordism in low dimensions, but I do not know if Rohlin phrased them in terms of bordism groups.

their conjecture implies Mumford's conjecture [Mu] about the rational cohomology of the mapping class group. The Madsen-Tillmann conjecture was subsequently proved in [MW] and is now known as the Madsen-Weiss theorem. The relation with the spectra Thom used to compute bordism groups is elucidated in [GMTW, §3], where another proof is given.

For now we restrict to manifolds with boundary—no corners—and so organize closed $(n-1)$-manifolds into a symmetric monoidal category which refines the abelian group Ω_{n-1}.

Definition 2.5. $\mathrm{Bord}_{\langle n-1,n\rangle}$ *is the symmetric monoidal category whose objects are compact* $(n-1)$*-manifolds and in which a morphism* $X\colon Y_0 \to Y_1$ *is a bordism from* Y_0 *to* Y_1*, up to diffeomorphism. The monoidal structure is disjoint union.*

So now a bordism is a map—a morphism in a category—and so it carries an "arrow of time', at least near the boundary so as to distinguish incoming boundary components from outgoing boundary components. Composition (Figure 3) is defined by gluing bordisms. We identify diffeomorphic

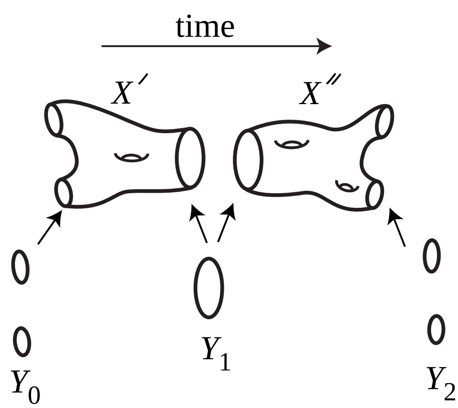


FIGURE 3. Composition of bordisms

bordisms—the diffeomorphism must commute with the boundary identifications—in order to obtain an associative composition law. The identity morphism $Y \to Y$ is the cylinder $[0,1] \times Y$ with obvious boundary identifications. There are variants $\mathrm{Bord}^{SO}_{\langle n-1,n\rangle}$ and $\mathrm{Bord}^{\mathrm{fr}}_{\langle n-1,n\rangle}$ for oriented and framed manifolds, but with one important change: in $\mathrm{Bord}^{\mathrm{fr}}_{\langle n-1,n\rangle}$ the morphisms X carry framings of the *tangent* bundle (not stabilized) and the objects Y carry framings of $TY \oplus (1)$, where '(1)' here denotes the trivial real line bundle of rank one.

By analogy to the homology homomorphism (2.2) we are led to the following definition.

Definition 2.6 ([A2])**.** *An* n-dimensional topological field theory *is a homomorphism*

$$F\colon \mathrm{Bord}_{\langle n-1,n\rangle} \longrightarrow (\mathrm{Ab}, \otimes) \tag{2.7}$$

of symmetric monoidal categories.

As telegraphed in Remark 2.3 in a quantum field theory disjoint unions map to *tensor products*, not direct sums. There are many variations on this definition. The domain can be a bordism category of smooth manifolds with extra structure, or even of singular manifolds. The codomain may be replaced by any symmetric monoidal category, algebraic or not. We introduce a more drastic variant of Definition 2.6 in §5. A typical choice for the codomain is $(\mathrm{Vect}_{\mathbb{C}}, \otimes)$, the category of complex vector spaces under tensor product. A topological field theory with values in $\mathrm{Vect}_{\mathbb{C}}$ is a linearization—a linear representation—of manifolds.

We have been led naturally to Definition 2.6 by combining basic ideas in homology and bordism. But this is hardly the historical path! For that we turn in the next section to notions in quantum field theory. Before leaving bordism, though, we pause to remind the reader of the connection with Morse theory.

Intuitively, a Morse function refines the arrow of time to a particular time function. Let $X\colon Y_0 \to Y_1$ be an n-dimensional bordism. A function $f\colon X \to \mathbb{R}$ is compatible with the bordism structure if there exist $t_0 < t_1$ such that t_0, t_1 are regular values of f and $Y_i = f^{-1}(t_i)$. Furthermore, f is a *Morse function* if it has finitely many isolated *nondegenerate* critical points. The main theorems in Morse theory [Mi2] assert that slices $f^{-1}(t)$ and $f^{-1}(t')$ are diffeomorphic if there are no critical values between t and t', and at an isolated critical point there is a topology change which is described by a standard surgery. For example, in Figure 4 the local slice evolves from

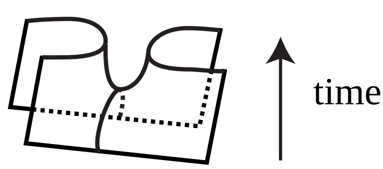


FIGURE 4. An elementary bordism

the two parallel line segments at the bottom to the two curves at the top; the saddle depicts the *elementary bordism* which connects the two local slices. Figure 5 displays the standard example of a Morse function on the torus—

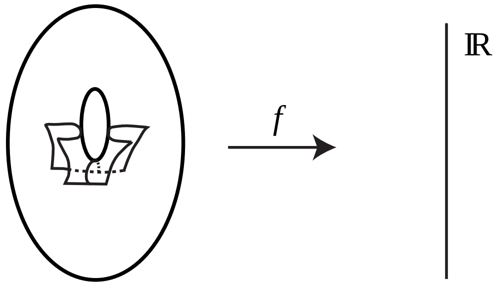


FIGURE 5. A Morse function

the height function—and embeds the elementary bordism of Figure 4 into a neighborhood of one of the critical points of index 1.

Remark 2.8. The local description of the topology change at a critical point uses a manifold with *corners*, as in Figure 4. Manifolds with boundary and no corners do not suffice. The additional locality afforded by admitting corners—and eventually higher codimensional corners—is a crucial idea for the cobordism hypothesis; see §5.

Morse functions exist, as a consequence of Sard's theorem. This means that any bordism can be decomposed as a composition of elementary bordisms, one for each critical point. Manipulations with Morse functions are a key ingredient in Milnor's presentation [Mi3] of Smale's h-cobordism theorem [Sm]. The space of Morse functions on a fixed bordism has many components: Morse functions in different components induce qualitatively different decompositions into elementary bordisms. Cerf [C] relaxed the Morse condition to construct a *connected* space of functions. This enables a systematic study of transitions between decompositions. For example, Cerf theory is the basis for Kirby calculus [K], which describes links in 3-manifolds and 4-manifolds. As we shall see it is also a crucial tool for constructing topological field theories.

An elementary illustrative example of a Cerf transition is the family of functions

$$f_t(x) = \frac{x^3}{3} - tx, \qquad x, t \in \mathbb{R}. \tag{2.9}$$

For $t > 0$ this is a Morse function with nondegenerate critical points at $x = \pm\sqrt{t}$. For $t < 0$ it is a Morse function with no critical points. At $t = 0$

the function fails to be Morse: $x = 0$ is a degenerate critical point. So as t increases from negative to positive two critical points are born on the x-line, and they separate at birth. In the other direction, as t decreases from positive to negative the two critical points collide and annihilate. This simple "birth-death transition" is all that is needed to connect different components of Morse functions.

3. Quantum field theory

For much of its history quantum field theory was tied to four spacetime dimensions and a handful of physically realistic examples. As opposed to quantum mechanics, where the underlying theory of Hilbert spaces and operator theory has been fully developed, the analytic underpinnings of quantum field theory remain unsettled. Still, there has been a huge transformation over the past three decades. Quantum field theorists now study a large set of examples in a variety of dimensions, not all of which are meant to be physically relevant. A deeper engagement with mathematicians and mathematics has led physicists to study models whose consequences are more relevant to geometry than to accelerators. Topological and algebraic aspects of quantum field theories have come to the fore. From another direction string theory has illuminated the subject, and there are new ties to condensed matter theory as well.

In this section we briefly sketch how Definition 2.6 of a topological quantum field theory emerges from physics. Our exposition is purely formal, extracting the structural elements which most directly lead to our goal. Let's begin with quantum mechanics, which is a 1-dimensional quantum field theory. (The dimension of a theory refers to spacetime, and at least in mainstream theories there is a single time dimension. Thus a 1-dimensional theory only has time; space is treated externally.) The basic ingredients are a complex separable Hilbert space $\mathcal{H}$ and for each time interval of length t a unitary operator

$$U_t = e^{-itH/\hbar}. \tag{3.1}$$

Here H is the self-adjoint *Hamiltonian* which describes the quantum system, and $\hbar$ is Planck's constant. The states of the system are vectors (really complex lines of vectors) in $\mathcal{H}$, and the unitary operators (3.1) describe the evolution of a state in time. Self-adjoint operators $\mathcal{O}$ on $\mathcal{H}$ act on the system—they are the *observables*—and the physics is encoded in expectation values

$$\langle \Omega, U_{t_n}\mathcal{O}_n \cdots U_{t_2}\mathcal{O}_2 U_{t_1}\mathcal{O}_1 U_{t_0}\Omega\rangle. \tag{3.2}$$

In this expression the state Ω evolves for time t_0, is acted on by the operator $\mathcal{O}_1$, then evolves for time t_1, then is acted on by the operator $\mathcal{O}_2$, etc. See Figure 6 for a pictorial representation. We recommend [Ma, Fa]

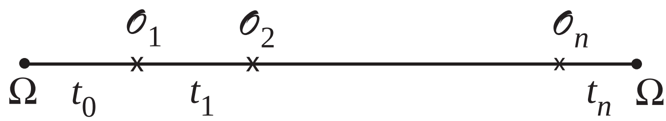


FIGURE 6. Vacuum expectation value in quantum mechanics

for structural expositions of mechanics which elucidate the pairing of states and observables.

It is convenient and powerful to analytically continue the time t from the real line to the complex line, and we restrict to $\operatorname{Im} t < 0$. Real times are now at the boundary of allowed complex times. If the Hamiltonian H is nonnegative, and $\operatorname{Im} t < 0$, then the evolution operator $e^{-itH/\hbar}$ is a contracting operator. *Wick rotation* to imaginary time is then the restriction to purely imaginary $t = \tau/\sqrt{-1}$, where the *Euclidean time* τ is strictly positive. We associate the Euclidean contracting evolution $F_\tau = e^{-\tau H/\hbar}$ to an interval of length τ, that is, to a compact, connected Riemannian 1-manifold with boundary whose total length is τ. The evolution obeys a semigroup law

$$F_{\tau_2+\tau_1} = F_{\tau_2} \circ F_{\tau_1}, \tag{3.3}$$

as illustrated in Figure 7. This is already reminiscent of bordism. We

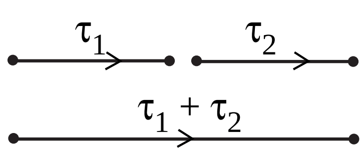


FIGURE 7. Composition of 1-dimensional bordisms

can imagine a bordism category $\mathrm{Bord}^{\mathrm{Riem}}_{\langle 0,1\rangle}$ whose objects are compact oriented 0-manifolds and whose morphisms are compact Riemannian oriented 1-manifolds with boundary. The semigroup law for the evolution evolution of a quantum mechanical system is encoded in the statement that

$$F\colon \mathrm{Bord}^{\mathrm{Riem}}_{\langle 0,1\rangle} \longrightarrow \mathrm{Hilb} \tag{3.4}$$

is a homomorphism to the category of Hilbert spaces and contracting linear maps. Notice that F encodes more than evolution. For example, we demand that F be a homomorphism of *symmetric monoidal* categories mapping disjoint unions to tensor products, which encodes the idea that the state space of the union of quantum mechanical systems is a tensor product. Exotic "evolutions" are now possible; see Figure 8. In a more careful axiomatization [Se1] one takes the codomain to be a category of topological vector spaces; then the Hilbert space structure emerges more organically from the geometry, as do the operator insertions in (3.2).

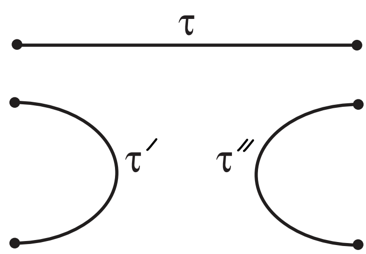


FIGURE 8. Exotic evolutions in quantum mechanics

It is a small step now to pass from the formal description (3.4) of a quantum mechanical system to the assertion that an n-dimensional quantum field theory is a homomorphism

$$F\colon \mathrm{Bord}^{\mathrm{Riem}}_{\langle n-1,n\rangle} \longrightarrow \mathrm{Hilb} \tag{3.5}$$

from the bordism category of Riemannian n-dimensional bordisms ("Riemannian spacetimes") to the category of Hilbert spaces (better: topological vector spaces). If X is such a bordism, and $x \in X$ a point not on the boundary, then the boundary sphere of the geodesic ball of sufficiently small radius r maps under F to a vector space $\mathcal{H}_r$, and the limit as $r \to 0$ is a vector space of operators associated to the point x. We can approximate it by the vector space at some small finite radius r_0. Remove an open ball

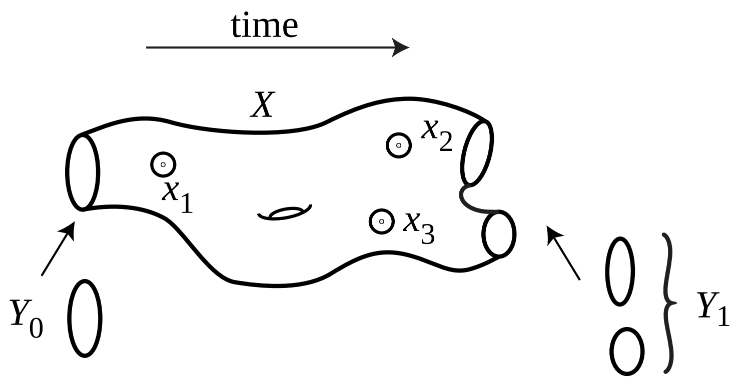


FIGURE 9. Operator insertions

of radius r_0 about x. Choose the arrow of time so that the new boundary component—the sphere of radius r_0 about x—is incoming. For example, the bordism in Figure 9 has incoming boundary Y_0 union the spheres about x_1, x_2, and x_3 and outgoing boundary Y_1. A field theory F determines vector spaces $F(Y_0), F(Y_1)$ for the boundary components and then vector spaces V_1, V_2, V_3 associated to the points x_1, x_2, x_3. The bordism X goes over to a linear map

$$F(X)\colon V_1 \otimes V_2 \otimes V_3 \longrightarrow \mathrm{Hom}\big(F(Y_0), F(Y_1)\big). \tag{3.6}$$

This is the sense in which the vector spaces V_i attach a space of operators to x_i, analogously to the operators which appear in (3.2) as illustrated in Figure 6. In case $Y_0 = Y_1 = \emptyset^{n-1}$, then $F(X)$ is called a *correlation function* between "operators" at the points x_i. If in addition there are no points x_i, then $F(X)$ is a complex number, the *partition function* of the closed manifold X.

This geometric formulation of quantum field theory developed in the 1980s out of the interaction between mathematicians and physicists centered around 2-dimensional conformal field theory. Graeme Segal's samizdat manuscript *The definition of conformal field theory*, now published [Se2], was widely distributed and very influential among both mathematicians and physicists. Segal's recent series of lectures [Se1] explores and expands on these ideas in the context of general quantum field theories. More traditional mathematical treatments of quantum field theory [SW], [H], [GJ] are set in four-dimensional Minkowski spacetime and focus on analytic aspects. The geometric formulation set the stage for the advent of *topological* field theories. In 1988 Witten [W1] introduced twistings of supersymmetric quantum field theories on Minkowski spacetime which allow them to be formulated on arbitrary oriented Riemannian manifolds. Special correlation functions in twisted theories are topological invariants. Witten's first application was to a supersymmetric gauge theory in four dimensions—a theory whose principal field is a connection on a principal bundle—where he showed that Donaldson's polynomial invariants of 4-manifolds [D] are correlation functions in that twisted supersymmetric gauge theory. Two-dimensional supersymmetric σ-models—whose principal field is a map $\Sigma \to M$ from a 2-manifold into a Riemannian target manifold—also admit topological twistings in case there is enough supersymmetry (which constrains the target manifold to be Kähler in the basic case). These 2-dimensional topological field theories [W2] have had profound consequences for algebraic geometry in the form of Gromov-Witten invariants and mirror symmetry. By late 1988 Witten realized [W3] that the Jones polynomials of knots and links in S^3 are encoded in a 3-dimensional field theory—called *Chern-Simons theory* after the classical action functional of connections which defines it—and he used it to introduce new invariants of 3-manifolds. This theory, as opposed to the topologically twisted supersymmetric models, is topological at the classical level and has an immediate connection to combinatorially accessible invariants. For many mathematicians it served as an accessible entrée into quantum field theory. In early 1989 Atiyah [A2] introduced a set of axioms for topological quantum field theory which essentially amount to Definition 2.6.

4. Topological quantum field theory

1-dimensional theories

Let's begin our exploration Definition 2.6 with a 1-dimensional topological field theory of oriented manifolds. Recall that the domain of such a theory is the bordism category $\mathrm{Bord}^{SO}_{\langle 0,1\rangle}$ in which an object is a compact oriented 0-manifold—a finite set of points each with a '+' of '−' attached—and a morphism is an oriented 1-dimensional bordism. There are two basic objects, the + point and the − point, and any other object is a tensor product (disjoint union) of these. Some basic morphisms are illustrated in Figure 10. The arrow of time points to the right, whereas the orientation is notated by an arrow on each component of the bordism. Notice that there is a correlation

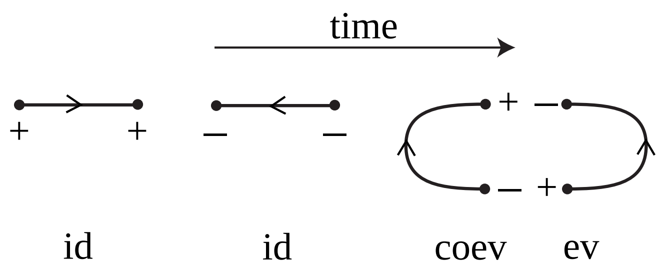


FIGURE 10. Elementary oriented 1-dimensional bordisms

between the orientation, the arrow of time, and the boundary orientation. The first two morphisms are identities. The third is called *coevaluation* and the fourth *evaluation*. Notice that the second bordism is obtained from the first by reversing the arrow of time, and the same holds for the third and fourth bordisms. *Time reversal* is a *duality* operation. Thus the − point is the dual of the + point and the evaluation is dual to the coevaluation.[7] The coevaluation and evaluation are evolutions in 1-dimensional topological field theory which go beyond the standard evolutions in quantum mechanics (Figure 8). Also, in quantum mechanics the closed intervals are Riemannian, so have a length τ, whereas in the topological theory all closed intervals are diffeomorphic and lead to the identity evolution. Comparison with (3.1) shows that the Hamiltonian vanishes in a topological field theory. There is no local evolution: all of the non-identity behavior comes from topology.

Now suppose F is a 1-dimensional oriented topological field theory (2.7) with values in complex vector spaces:

$$F\colon \bigl(\mathrm{Bord}^{SO}_{\langle 0,1\rangle}, \amalg\bigr) \longrightarrow \bigl(\mathrm{Vect}_{\mathbb{C}}, \otimes\bigr). \tag{4.1}$$

The notation recalls that F is a homomorphism of symmetric monoidal categories, so maps disjoint unions to tensor products. The homomorphism F assigns a vector space $F(\mathrm{pt}_+) = V_+$ to the + point and a vector space $F(\mathrm{pt}_-) = V_-$ to the − point. This determines the value of F on all compact oriented 0-manifolds as they are disjoint unions of + and − points. Also, since the empty 0-manifold $\emptyset^0$ is the tensor unit for disjoint union, it maps under the homomorphism F to the tensor unit for complex vector spaces under tensor product, which is the complex line $\mathbb{C}$. Next, consider F evaluated on the bordisms in Figure 10. As F is a homomorphism it sends identities

[7] The latter statement is not true for all geometric structures. For example, if we use 2-framings—framings of the tangent bundle made 2-dimensional by adding on a trivial bundle—then the adjoints of coevaluation differ from evaluation by a change of framing.

to identities, so the first two bordisms map to id_{V_+} and id_{V_-}, respectively. The last two bordisms map under F to linear maps

$$c\colon \mathbb{C} \longrightarrow \begin{matrix} V_+ \\ \otimes \\ V_- \end{matrix} \qquad\qquad e\colon \begin{matrix} V_- \\ \otimes \\ V_+ \end{matrix} \longrightarrow \mathbb{C} \tag{4.2}$$

where we have written the tensor product vertically to match the figure. The sense in which coevaluation and evaluation give rise to duality is illus-

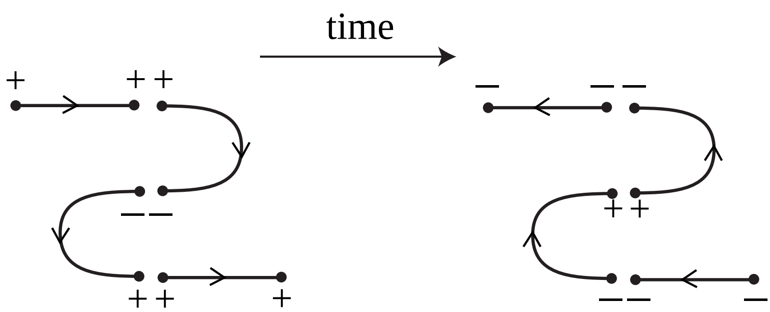


FIGURE 11. The S-diagrams

trated in Figure 11. The left figure is the composition of two 1-dimensional bordisms, each with two components. The first maps a single + point to the tensor product (disjoint union) of 3 points: $+, -, +$. The second maps these 3 points back to the + point. The composition is computed by gluing at the 3 points in the middle. The result is diffeomorphic to the identity map on the + point. Recall that morphisms in $\mathrm{Bord}^{SO}_{\langle 0,1\rangle}$ are 1-dimensional bordisms *up to diffeomorphisms* which preserve the boundary identifications. Comparing the first composition in Figure 11 with the first bordism in Figure 10 we see that the composition is the identity. To see the relation to duality we apply the homomorphism F. Now the homomorphism property has two consequences: (1) F sends a disjoint union of bordisms to the tensor product of the corresponding linear maps, and (2) F sends a composition of bordisms to the corresponding composition of linear maps. Using these rules we see that F sends the compositions in Figure 11 to compositions of linear maps

$$V_+ \xrightarrow{\begin{matrix}\mathrm{id}_{V_+} \\ \otimes \\ c\end{matrix}} \begin{matrix} V_+ \\ \otimes \\ V_- \\ \otimes \\ V_+ \end{matrix} \xrightarrow{\begin{matrix} e \\ \otimes \\ \mathrm{id}_{V_+}\end{matrix}} V_+ \qquad\qquad V_- \xrightarrow{\begin{matrix}\mathrm{id}_{V_-} \\ \otimes \\ c\end{matrix}} \begin{matrix} V_- \\ \otimes \\ V_+ \\ \otimes \\ V_- \end{matrix} \xrightarrow{\begin{matrix} e \\ \otimes \\ \mathrm{id}_{V_+}\end{matrix}} V_- \tag{4.3}$$

(Note that we have used the symmetry in the first diagram to exchange the order of the tensor product in the maps c, e from (4.2).)

Lemma 4.4. *If the compositions* (4.3) *are identity maps, then V_+, V_- are finite dimensional vector spaces and e is a nondegenerate duality pairing.*

Proof. Set $c(1) = \sum_{i=1}^{N} v_+^i \otimes v_-^i$ for some $v_\pm^i \in V_\pm$ and some positive integer N. Then the first composition in (4.3) is the map $\xi \mapsto \sum e(v_-^i, \xi)v_+^i$. Since this is the identity map, it follows that $\{v_+^i\}_{i=1}^N$ spans V_+, whence V_+ is finite dimensional. The same argument with the second composition proves that V_- is finite dimensional. If $\xi \in V_+$ satisfies $e(v_-, \xi) = 0$ for all $v_- \in V_-$, then $\xi = \sum e(v_-^i, \xi)v_+^i = 0$. The same argument with the second composition in (4.3) proves that if $\eta \in V_-$ satisfies $e(\eta, v_+) = 0$ for all $v_+ \in V_+$, then $\eta = 0$. Hence e is a nondegenerate pairing. □

Remark 4.5. A similar argument for a field theory $F\colon \big(\mathrm{Bord}^{SO}_{\langle 0,1\rangle}, \amalg\big) \longrightarrow \big(\mathrm{Ab}, \otimes\big)$ with values in abelian groups proves that $F(\mathrm{pt}_+)$ is finitely generated and free.

Lemma 4.4 illustrates an important *finiteness* principle in topological field theories: the vector space attached to an $(n-1)$-manifold in an n-dimensional topological field theory with values in $\mathrm{Vect}_{\mathbb{C}}$ is finite dimensional. We derived this finiteness from duality: the + point and − point are duals, and that duality is expressed by the existence of coevaluation and evaluation maps. Notice that any vector space V has a dual space, defined algebraically as the space of linear maps $V \to \mathbb{C}$, which comes with a canonical evaluation map. However, the coevaluation map exists if and only if V is finite dimensional.

This notion of finiteness generalizes to any symmetric monoidal category.

Definition 4.6. *Let $\mathcal{C}$ be a symmetric monoidal category and $x \in \mathcal{C}$. Then* duality data *for x is a triple (x', c, e) consisting of an object $x' \in \mathcal{C}$, a coevaluation $c\colon 1 \to x \otimes x'$, and an evaluation $e\colon x' \otimes x \to 1$ such that the compositions*

$$(4.7)\quad x \xrightarrow{c\otimes \mathrm{id}_x} x \otimes x' \otimes x \xrightarrow{\mathrm{id}_x \otimes e} x \qquad\qquad x' \xrightarrow{\mathrm{id}_{x'} \otimes c} x' \otimes x \otimes x' \xrightarrow{e\otimes \mathrm{id}_{x'}} x'$$

are identity maps. We say x is dualizable *if there exists duality data for x.*

The argument in Lemma 4.4 with the S-diagrams in Figure 11 apply in an n-dimensional field theory—take the Cartesian product of the S-diagrams with a fixed $(n-1)$-manifold—which shows that objects in the image of a field theory F are always dualizable. In the next section we define an extension of the notion of a field theory and there is a corresponding extension of dualizability, which we take up in §6.

At this point we can state and prove a very simple special case of the cobordism hypothesis.

Theorem 4.8. *Let V be a finite dimensional complex vector space. Then there is a homomorphism F as in* (4.1) *such that $F(\mathrm{pt}_+) = V$.*

Proof. If Y is an oriented compact 0-manifold set

$$F(Y) = \bigotimes_{y\in Y:y=\mathrm{pt}_+} V \quad\otimes\quad \bigotimes_{y\in Y:y=\mathrm{pt}_-} V^*. \tag{4.9}$$

Referring to the third and fourth bordisms in Figure 10 define $F(\mathrm{coev})$ as the inverse duality pairing $\mathbb{C}\to V\otimes V^*$ and $F(\mathrm{ev})$ as the duality pairing $V^*\otimes V\to\mathbb{C}$. A Morse function on a 1-dimensional bordism decomposes it as a composition of the elementary bordisms coev and ev: a nondegenerate critical point of a real-valued function on a 1-manifold is either a local maximum or a local minimum. The only Cerf move (Figure 12) cancels a local maximum against a local minimum, and the proof that this does not change the value of F is the statement that the S-diagrams in Figure 11 map to the identity. $\square$

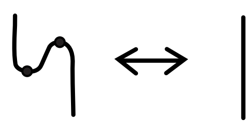


FIGURE 12. Cerf move in dimension one

2-dimensional theories

Next, consider a 2-dimensional oriented topological field theory

$$F\colon \big(\mathrm{Bord}^{SO}_{\langle 1,2\rangle}, \amalg\big) \longrightarrow \big(\mathrm{Vect}_{\mathbb{C}}, \otimes\big). \tag{4.10}$$

There is only one compact connected oriented 1-manifold up to diffeomorphism: a circle has orientation-reversing diffeomorphisms (reflection). Let

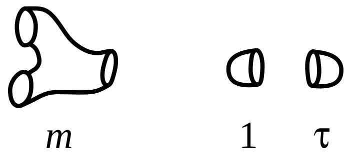


FIGURE 13. Some elementary oriented 2-dimensional bordisms

$V = F(S^1)$. Elementary 2-dimensional bordisms, as depicted in Figure 13, give extra structure on V, namely linear maps

$$\begin{aligned} m\colon &\quad V \otimes V \longrightarrow V \\ 1\colon &\quad \mathbb{C} \longrightarrow V \\ \tau\colon &\quad V \longrightarrow \mathbb{C} \end{aligned} \tag{4.11}$$

The multiplication m gives V an algebra structure with respect to which the image of $1 \in \mathbb{C}$ under the linear map 1 is an identity element. The linear map τ is a trace on V.[8] Standard arguments with oriented surfaces and their diffeomorphisms prove that m is associative and commutative and that the trace is nondegenerate in the sense that the pairing $v_1, v_2 \mapsto (\tau \circ m)(v_1, v_2)$ is a nondegenerate pairing on V. For example, the composition of the bordisms labeled m and τ in Figure 13 is the product of the circle with the bordism labeled ev in Figure 10; then the argument of Lemma 4.4 with the S-diagram proves that the pairing $\tau \circ m$ is nondegenerate. Thus an oriented 2-dimensional topological field theory determines a *commutative Frobenius algebra*, a commutative algebra with a nondegenerate trace. The converse is also true.

Theorem 4.12. *Let V be a commutative Frobenius algebra. Then there is a homomorphism*

$$F\colon \bigl(\mathrm{Bord}^{SO}_{\langle 1,2\rangle}, \amalg\bigr) \longrightarrow \bigl(\mathrm{Vect}_{\mathbb{C}}, \otimes\bigr) \tag{4.13}$$

with $F(S^1) = V$.

This is one of the oldest theorems in the subject. In the physics literature the statement dates at least to Dijkgraaf's thesis [Di]. There are several proofs in

[8] Note that the bordism τ is the time-reversal of 1. There is also a time-reversal of m, which may be expressed as a composition of the maps in (4.11) together with the inverse of $\tau \circ m$.

the mathematics literature, for example in [Ab, Ko]. The appendix to [MS] contains a proof of Theorem 4.12 as well as several important variations. As in the proof of Theorem 4.8 we first extend F to all closed oriented 1-manifolds via tensor products. The data (4.11) which defines the Frobenius structure on V tells what to attach to elementary 2-dimensional bordisms arising from critical points of a Morse function of index 1,0,2. It remains to verify that different Morse functions lead to the same linear map. That check, for which we refer to the reader to [MS], uses the basic properties of a commutative Frobenius algebra.

These explicit arguments with Morse functions quickly become tedious. The situation simplifies for *extended* field theories (§5) which are more local. They are the province of the cobordism hypothesis. The cobordism hypothesis is proved using on the one hand more powerful results about space of Morse functions and on the other more sophisticated algebra to organize the argument.

One source of examples of commutative Frobenius algebras is the cohomology algebra $H^\bullet(M;\mathbb{C})$ of a compact oriented n-manifold M. The trace is pairing with the fundamental class $[M]\in H_n(M)$. If there is odd cohomology, then it is commutative in the graded sense because of signs in the commutation rule for cup products. For example, if $M=S^2$ then we obtain the truncated polynomial algebra $\mathbb{C}[x]/(x^2)$. The corresponding field theory plays a role in the construction of Khovanov homology for links [Kh, B-N]. If the Frobenius algebra V is *semisimple*, then we can simultaneously diagonalize the multiplication operators $M_a(b)=ab,\ a,b\in V$ and so find a basis of commuting idempotents $e_1,e_2,\dots,e_n\in V$: thus $e_ie_i=e_i$ and $e_ie_j=0$ if $i\neq j$. The Frobenius algebra is determined up to isomorphism by nonzero complex numbers $\lambda_1,\lambda_2,\dots,\lambda_n$ defined by $\tau(e_i)=\lambda_i$. In this case everything in the field theory F with $F(S^1)=V$ is easily computed in terms of the basis $\{e_i\}$ and the numbers λ_i. For example the 2-holed torus in Figure 14 maps to the endomorphism $e_i\mapsto\lambda_i^{-1}e_i$ of V and a closed surface X_g of genus g maps to the complex number

$$F(X_g)=\sum\lambda_i^{1-g}. \tag{4.14}$$

These computations are made by chopping the surfaces into the elementary bordisms in Figure 13 and their time-reversals. Let G be a finite group and $A=\operatorname{Map}(G,\mathbb{C})$ the vector space of complex-valued functions on G. Then

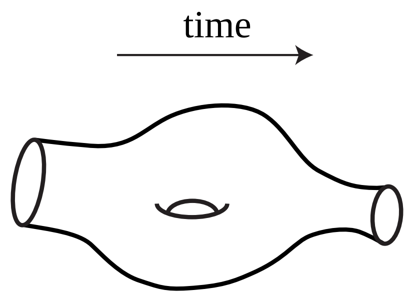


FIGURE 14. Torus with incoming and outgoing boundary circles

A is an associative algebra under the convolution product

$$(f_1 * f_2)(g) = \sum_{g_1 g_2 = g} f_1(g_1) f_2(g_2), \qquad g, g_1, g_2 \in G, \quad f_1, f_2 \colon G \to \mathbb{C}. \tag{4.15}$$

We also define the trace

$$\tau(f) = \frac{f(e)}{\#G}, \tag{4.16}$$

where $e \in G$ is the identity element. The product is not commutative if G is not abelian. Let V be the center of A, the space of *class functions* on G; it is a commutative Frobenius algebra which can be identified with the complexification $R(G) \otimes \mathbb{C}$ of the representation ring of G. Let F_G denote the 2-dimensional oriented topological field theory with $F_G(S^1) = V$ guaranteed by Theorem 4.12. The complexified representation ring is semisimple. Classical orthogonality formulas of Schur show that the characters χ_i of the irreducible complex representations of G are, up to scale, the commuting idempotents $e_i = \bigl(\chi_i(1)/\#G\bigr)\chi_i$. Then we easily compute that $\lambda_i = \sum \chi_i(1)^2/\#G$ and from (4.14) the partition function of a closed connected oriented surface is

$$F_G(X) = \sum_{\substack{\chi \text{ irreducible} \\ \text{character of } G}} \left(\frac{\chi(1)}{\#G}\right)^{\mathrm{Euler}(X)}, \tag{4.17}$$

where $\mathrm{Euler}(X)$ is the Euler characteristic of X.

The construction of F_G which relies on Theorem 4.12 takes as input the complexified representation ring and uses Morse theory to produce a topological field theory. There is also a direct geometric construction of this simple finite theory. For any manifold M let $\mathcal{F}_M$ denote the collection of principal G-bundles $P \to M$. So P is a manifold with a free right G-action and quotient M. In other terms $P \to M$ is a covering space which is Galois, or regular, but note that P need not be connected. For example, if $M = S^1$ and $G = \mathbb{Z}/n\mathbb{Z}$ for some positive integer n, then there are n distinct isomorphism classes of principal G-bundles over M; the connectivity of the total space of a cover depends on the prime factorization of n. For any manifold $\mathcal{F}_M$ is a category: a morphism $(P' \to M) \longrightarrow (P \to M)$ is a smooth map $\varphi \colon P' \to P$ which commutes with the G-action and covers the identity map of M. This category is a *groupoid* since all morphisms are invertible. For $M = \mathrm{pt}$ there is only one G-bundle up to isomorphism, the trivial bundle $P = G$ with G acting by right multiplication, and the group of automorphisms is G acting by left multiplication on P. Figure 15 depicts an groupoid equivalent to $\mathcal{F}_{\mathrm{pt}}$. There is a single object, the set of arrows

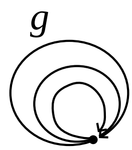


FIGURE 15. G-bundles over pt

is G, and composition of arrows is given by the group law. For $M = S^1$ if we introduce a basepoint $p \in P$ on a G-bundle $P \to S^1$, then we can compute the holonomy, or monodromy, around the circle (after choosing an orientation), which is an element of G. The bundle with basepoint is rigid: any automorphism which fixes the basepoint is the identity. The group G acts simply transitively on the set of basepoints over a fixed point of S^1, and it conjugates the holonomy. In this way we see that $\mathcal{F}_{S^1}$ is equivalent to the groupoid $G/\!/G$ of G acting on itself by conjugation. It is depicted in Figure 16. The set of isomorphism classes $\pi_0(\mathcal{F}_{S^1})$ is the set of conjugacy classes in G and the automorphism group $\pi_1(\mathcal{F}_{S^1}, P)$ at a G-bundle with

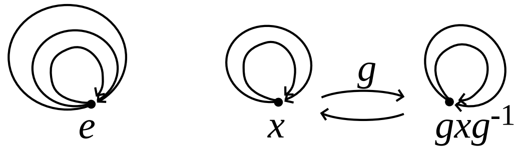


FIGURE 16. G-bundles over S^1

holonomy x is the centralizer group of x in G.

Principal G-bundles are local and contravariant. Consider a bordism, as in Figure 1 with the arrow of time pointing to the right. The inclusions of the incoming and outgoing boundary induce restriction maps of bundles, which are homomorphisms of groupoids:

$$\begin{tikzcd} & \mathcal{F}_X \arrow[dl, "s"'] \arrow[dr, "t"] & \\ \mathcal{F}_{Y_0} & & \mathcal{F}_{Y_1} \end{tikzcd} \tag{4.18}$$

A diagram of the form (4.18) is a *correspondence*, which is a generalization of a homomorphism from $\mathcal{F}_{Y_0}$ to $\mathcal{F}_{Y_1}$. Namely, if s is invertible, then $s \times t$ embeds $\mathcal{F}_X$ into $\mathcal{F}_{Y_0} \times \mathcal{F}_{Y_1}$ as the graph of $t \circ s^{-1}$. A composition of bordisms (Figure 3) induces a composition of correspondences

$$\begin{tikzcd} & & \mathcal{F}_{X'' \circ X'} \arrow[dl, "r'"'] \arrow[dr, "r''"] & & \\ & \mathcal{F}_{X'} \arrow[dl, "s'"'] \arrow[dr, "t'"] & & \mathcal{F}_{X''} \arrow[dl, "s''"'] \arrow[dr, "t''"] & \\ \mathcal{F}_{Y_0} & & \mathcal{F}_{Y_1} & & \mathcal{F}_{Y_2} \end{tikzcd} \tag{4.19}$$

The locality of principal G-bundles is hidden in this statement: the groupoid $\mathcal{F}_{X'' \circ X'}$ of G-bundles on the composition $X'' \circ X'$ is the fiber product of t' and s''; that is, a G-bundle $P \to X'' \circ X'$ is a triple (P', P'', θ) consisting of G-bundles $P' \to X'$, $P'' \to X''$, and an isomorphism $\theta \colon P'\big|_{Y_1} \to P''\big|_{Y_1}$ of their restrictions to Y_1.

Correspondence diagrams can often be "linearized" into honest maps. For the field theory F_G we use closed oriented 1-manifolds Y and compact oriented 2-dimensional bordisms X. On 1-manifolds we define

$$F_G(Y) = \operatorname{Hom}\bigl(\mathcal{F}_Y, \mathbb{C}\bigr). \tag{4.20}$$

Here we view $\mathbb{C}$ as a groupoid with only identity morphisms. Then homomorphisms $\mathcal{F}_Y \to \mathbb{C}$ assign complex numbers to objects in $\mathcal{F}_Y$ so that the numbers at each end of a morphism are equal. In other words, $\operatorname{Hom}(\mathcal{F}_Y, \mathbb{C})$ is the vector space of invariant functions on $\mathcal{F}_Y$, so can be identified with

$\mathrm{Map}\bigl(\pi_0(\mathcal{F}_Y),\mathbb{C}\bigr)$, the space of functions on equivalence classes of G-bundles. Then to a correspondence (4.18) we define

$$F_G(X) = t_*\circ s^*\colon F_G(Y_0)\longrightarrow F_G(Y_1) \tag{4.21}$$

as pullback followed by pushforward. The fibers of t are (equivalent to) groupoids with finitely many objects, each with a finite stabilizer group. The pushforward t_* of a function ϕ on $\mathcal{F}_X$ is the sum

$$t_*(\phi)(y) = \sum_x \frac{\phi(x)}{\#\,\mathrm{Aut}(x)},\qquad y\in\mathcal{F}_{Y_1}, \tag{4.22}$$

over the equivalence classes x in the fiber $t^{-1}(y)$ of the value of ϕ divided by the order of the automorphism group. (This formula makes clear that F_G may be defined on rational vector spaces.) Key point: the fact that (4.19) is a fiber product implies that the push-pull construction takes compositions of bordisms to compositions of linear maps. In other words, there is an *a priori* proof that the push-pull construction produces a homomorphism $F_G\colon \mathrm{Bord}^{SO}_{\langle 1,2\rangle}\to\mathrm{Vect}_{\mathbb{C}}$ of symmetric monoidal categories. The enterprising reader can now *compute* that $F_G(S^1)$ is the vector space of central functions on G, and that the basic bordisms in Figure 13 map to the convolution product, the character of the identity representation, and the trace (4.16).

Now suppose X is a closed oriented 2-manifold. It is interpreted as a bordism $X\colon\emptyset^1\to\emptyset^1$. In grand Bourbaki style the groupoid of G-bundles $\mathcal{F}_{\emptyset^1}$ has a single object with only the identity object. (After all, $\mathcal{F}$ maps disjoint unions to Cartesian products, and $\emptyset^1$ is the tensor unit for disjoint union.) In this case (4.21) specializes to the sum of the constant function 1 over $\mathcal{F}_X$: it counts (with automorphisms) the G-bundles over X. If X is connected then that count of bundles is

$$F_G(X) = \frac{\#\,\mathrm{Hom}\bigl(\pi_1(X,x),G\bigr)}{\#G}; \tag{4.23}$$

the numerator counts G-bundles with a basepoint over x and the group G acts simply transitively on the basepoints.

Theorem 4.24. *Let X be a compact oriented connected 2-manifold and G a finite group. Then*

$$\#\,\mathrm{Hom}\bigl(\pi_1(X,x),G\bigr) = (\#G)\sum_{\substack{\chi\text{ irreducible}\\ \text{character of }G}}\left(\frac{\chi(1)}{\#G}\right)^{\mathrm{Euler}(X)}. \tag{4.25}$$

The theorem follows immediately by comparing (4.23) and (4.17). The proof is representative of how topological field theory is used in more complicated situations. The invariant on the left hand side of (4.25), initially defined for closed 2-manifolds, is extended to an invariant for compact 2-manifolds with boundary which obeys a gluing law. So it is computed by chopping X into elementary pieces (as in Figure 13 together with the time-reversal of m).

Remark 4.26. The appearance of the Euler characteristic in (4.25) suggests an extension of F_G which includes 0-manifolds. They would appear as corners of 2-manifolds and boundaries of 1-manifolds. Then in a triangulation of X, the count of vertices, edges, and triangles in the triangulation should combine to give Euler(X) and a new proof of (4.25). In such an *extended* field theory we have more locality, so more decompositions and hence more computational flexibility. We take up extended theories in §5 and pursue this idea in Example 5.7.

Remark 4.27. There is a variation on (4.21) in which $\mathcal{F}_X$ in (4.18) carries an integral kernel. In that case the pull-push formula (4.21) is modified to pull-multiply-push. The integral kernel must be local in that it multiplies in the fiber product (4.19). In this 2-dimensional theory we can obtain such an integral kernel by starting with a cocycle for a class in the group cohomology $H^2(G; \mathbb{C}/\mathbb{Z})$.

The theory F_G was introduced by Dijkgraaf and Witten [DW]. See [FQ],[F] for more details about defining F_G by counting principal G-bundles. The lecture notes [Q] contain elaborations and many more examples.

The push-pull construction is a finite version of the Feynman functional integral in quantum field theory. The groupoid $\mathcal{F}_M$ consists of *gauge fields* for a finite group G; if G is a Lie group, then gauge fields form the groupoid of G-connections on M. The integral kernel described in Remark 4.27 is the exponential of the *classical action* of the field theory. The pushforward t_* is the *Feynman integral* or *functional integral* or *path integral* over the space of fields (with fixed boundary condition). In almost all physically interesting examples the space, or stack, of fields is not finite, but rather is infinite dimensional. One way to define pushforward t_* on functions is via integration theory, which of course requires a measure on the space of fields. (There are alternatives, at least for some topological theories; see [FHT] for one example.) Furthermore, the measures must be consistent with the fiber product (4.19) under composition of bordisms. Such measures have not been constructed rigorously in most examples of physical interest. The example of finite gauge theories, while it nicely illustrates many topological and algebraic aspects, misses completely many of the central analytical issues in quantum field theory.

5. n-categories and extended topological quantum field theory

In this section we extend the definition of an n-dimensional topological field theories in two directions: (i) to invariants of manifolds of all dimensions $\leq n$ and (ii) to invariants of families of manifolds. These extensions go beyond what was traditionally done in quantum field theory.

Standard topological field theories, as in Definition 2.6, are *local* in that invariants of n-manifolds are computed by cutting along closed codimension 1 submanifolds. We saw after Theorem 4.24 that it might be desirable to go further and cut along codimension 2 submanifolds as well, so have n-manifolds with corners. Once we take that plunge we may as well continue cutting in higher and higher codimension until we are cutting along 0-manifolds. In other words, we end up considering n-manifolds with corners of all codimension: the local model for the maximal corner is a corner in real affine space, $\{(x^1, x^2, \dots, x^n) \in \mathbb{A}^n : x^i \geq 0\}$ near $(0, 0, \dots, 0)$.

In a bottom up view, rather than a top down view, we build higher dimensional manifolds by time evolution of lower dimensional manifolds. This is illustrated in Figure 10 by the time evolution of 0-manifolds to produce 1-manifolds. Now we evolve again, introducing a second time as in Figure 17. Let $t_1, t_2 \in [0, 1]$ denote the times, so the space of times is the

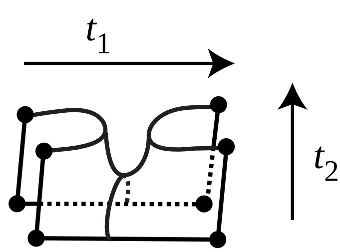


FIGURE 17. Two-time evolution of two points

square $[0, 1] \times [0, 1]$. At each of the four corners $t_1, t_2 \in \{0, 1\}$ lies the 0-manifold Y consisting of two points. At time $t_2 = 0$ they evolve in t_1 via the identity bordism, whereas at time $t_2 = 1$ they evolve as the evaluation followed by the coevaluation. (These 1-dimensional bordisms are pictured in Figure 10). The evolution in t_2 is a 2-dimensional bordism W between these two 1-dimensional bordisms X_0, X_1. As a manifold it is a 2-dimensional manifold with corners, but as a bordism we remember the time evolutions. Morally, as in §2 it is only the arrows of time which matter—and these only near the boundaries and corners—but it is convenient both heuristically and technically to think in terms of actual time functions. An algebraic representation of this two-time evolution is:

$$\begin{tikzcd} Y \arrow[rr, bend left, "X_1"] \arrow[rr, bend right, "X_0"'] \arrow[rr, phantom, "\Uparrow W"] & & Y \end{tikzcd} \tag{5.1}$$

The algebraic structure which includes (5.1) is a *2-category*. In addition to objects x and morphisms $f, g\colon x \to y$ mapping between them, there are now 2-morphisms $\eta\colon f \Rightarrow g$ which map between morphisms. For clarity 'morphisms' are now termed '1-morphisms'. In the 2-category $\mathrm{Bord}_{\langle 0,1,2\rangle}$ the objects are compact 0-manifolds, the 1-morphisms are 1-dimensional bordisms, and the 2-morphisms are 2-dimensional bordisms. A 2-category has two associative composition laws, easily seen pictorially in $\mathrm{Bord}_{\langle 0,1,2\rangle}$. Namely, we can compose horizontally in the first time t_1 or vertically in the second time t_2. Disjoint union is an extra algebraic structure—still called a symmetric monoidal structure—and the empty manifolds are identity elements for disjoint union. So, for example, a closed 2-manifold W is interpreted as a 2-morphism $W\colon \emptyset^1 \Rightarrow \emptyset^1$ in $\mathrm{Bord}_{\langle 0,1,2\rangle}$. For now we leave unspecified what sort of extra topological data (orientation, framing, ...) we assume present.

The saddle in Figure 17 is the elementary bordism in Morse theory depicted in Figure 4. In other words, it is the 2-manifold $D^1 \times D^1$ which implements the surgery beginning with $S^0 \times D^1$ and ending with $D^1 \times S^0$. Here D^1 is the standard closed 1-ball. The general surgery

$$D^p \times D^q\colon S^{p-1} \times D^q \longrightarrow D^p \times S^{q-1}, \tag{5.2}$$

can be written algebraically in a diagram similar to (5.1) with $Y = S^{p-1} \times S^{q-1}$. Morse theory tells that a manifold has a handlebody decomposition into elementary bordisms (5.2). We might conclude that 2-categories go far enough, and that nothing is to be gained by chopping further. We could, after all, make a 2-category whose objects are closed $(n-2)$-manifolds and with 1-morphisms and 2-morphisms as their time evolutions. But the structure simplifies if we don't stop there and rather go all the way down to points.

Therefore, to study manifolds of dimension $\le n$, or equivalently to study topological field theories of dimension n, we are led to the *n-category* $\mathrm{Bord}_{\langle 0\cdots n\rangle}$ whose objects are compact 0-manifolds and whose k-morphisms $(1\le k\le n)$ are k-time evolutions of objects. There are k composition laws for k-morphisms, and they satisfy various compatibilities. Disjoint union gives a symmetric monoidal structure. It is a complicated combinatorial problem to track all of this data. The relevance of higher categories to topological field theory was understood in the early 1990s, but at that time rigorous foundations were not available. In the intervening years several approaches and definitions have been advanced. We will not attempt a formal definition here, but refer the reader to [BD, L1] for more detailed exposition and references.

The n-category $\mathrm{Bord}_{\langle 0\cdots n\rangle}$ is the first extension we envisioned at the beginning of this section. The second is to families of manifolds. It turns out that this can be encoded by extending the n-category $\mathrm{Bord}_{\langle 0\cdots n\rangle}$ higher up: we adjoin $(n+1)$-morphisms, $(n+2)$-morphisms, etc. Namely, if W_0, W_1 are n-dimensional bordisms we define an $(n+1)$-morphism $\varphi\colon W_0\to W_1$ to be a diffeomorphism which preserves all of the "boundary data". An n-morphism is a map between two $(n-1)$-morphisms, each of which is a map between two $(n-2)$-morphisms, and on down. The diffeomorphism φ must preserve the implicit identifications. In terms of the n-time evolution, φ must be compatible with the data at each of the 2^n extreme times $t_i\in\{0,1\}$. Since φ is a diffeomorphism, it is invertible. We continue and define an $(n+2)$-morphism $\varphi_0\to\varphi_1$ to be an isotopy between the diffeomorphisms φ_0 and φ_1, again preserving the boundary data. Isotopies are also invertible, up to a higher isotopy. Continuing in this way we have k-morphisms for all k, so an ∞-category. But it has the property that every k-morphism for $k>n$ is invertible.

Remark 5.3. A higher category in which *every* morphism is invertible—i.e., an $(\infty,0)$-category—is a combinatorial model for a space. Since every morphism is invertible, this is also called an *∞-groupoid.* So whereas an

n-category has *sets* of n-morphisms, an (∞, n)-category has *spaces* of n-morphisms. An n-category may be extended to a *discrete* (∞, n)-category in which all k-morphisms for $k > n$ are identity maps.

Definition 5.4. *Let $n \in \mathbb{Z}^{>0}$. An (∞, n)-category is an ∞-category in which every k-morphism is invertible for $k > n$.*

'Definition' is not really appropriate as we have not defined ∞-categories! There are complete definitions for (∞, n)-categories, in fact several [Ba, R, Be], and some ongoing work [BS] which axiomatizes the notion of (∞, n)-category, studies the collection of all such, and compares the existing models.

Definition 5.5. Bord_n *is the (∞, n)-category whose objects are compact 0-manifolds, k-morphisms for $1 \leq k \leq n$ are k-time evolutions of objects, and k-morphisms for $k > n$ are $(k - n - 1)$-fold iterated isotopies of diffeomorphisms. It is symmetric monoidal under disjoint union.*

Again this is only a descriptive definition.

The manifolds in Bord_n typically carry extra data. For example, there is an (∞, n)-category Bord_n^{SO} of oriented bordisms. There is also a bordism category of bordisms with *framing*, but in an unstable[9] sense. Namely, an *n-framing* on a k-bordism W in $\mathrm{Bord}_n^{\mathrm{fr}}$ is a trivialization of $TW \oplus (n - k)$, where $(n-k)$ is the trivial bundle of the indicated rank. The (∞, n)-category of unoriented manifolds is denoted Bord_n^O. We use 'Bord_n' generically to denote any of these and many other similar possibilities.

Analogous to Definition 2.6 we consider representations of Bord_n. We allow an arbitrary codomain.

Definition 5.6. *Let $\mathcal{C}$ be a symmetric monoidal (∞, n)-category. An* extended topological field theory *with values in $\mathcal{C}$ is a homomorphism $F\colon \mathrm{Bord}_n \to \mathcal{C}$.*

The homomorphism property means that F respects the n composition laws as well as the symmetric monoidal structures. The cobordism hypothesis, which we take up in the next section, determines the space of homomorphisms F in terms of $\mathcal{C}$.

For the remainder of this section we indicate some examples which illuminate the idea of an extended field theory and the flexibility of Definition 5.6.

Example 5.7. Let G be a finite group. Recall from §4 the 2-dimensional topological field theory $F_G\colon \mathrm{Bord}_{\langle 1,2\rangle}^{SO} \to \mathrm{Vect}_{\mathbb{C}}$. In (4.15) we introduced the

[9] Framings on manifolds used to define framed bordism groups—isomorphic to stable homotopy groups of spheres—are stable.

algebra $A = \mathrm{Map}(G, \mathbb{C})$ of functions on G under convolution, but only its center made an appearance in F_G—as $F_G(S^1)$. There is an extended field theory $\widehat{F}_G$ of 0-, 1-, and 2-manifolds which has $\widehat{F}_G(\mathrm{pt}_+) = A$. The codomain $(\infty, 2)$-category $\mathcal{C}$ of any extension has the property that the $(\infty, 1)$-category $\mathrm{Hom}_{\mathcal{C}}(1, 1)$ of endomorphisms of the tensor unit 1 is identified with $\mathrm{Vect}_{\mathbb{C}}$. In fact, $\mathrm{Vect}_{\mathbb{C}}$ is discrete: objects are complex vector spaces, 1-morphisms are linear maps, and there are no non-identity higher morphisms. So we might hope that $\mathcal{C}$ is also discrete, an ordinary 2-category. Furthermore, if $\widehat{F}_G(\mathrm{pt}_+)$ is to be A, then objects of $\mathcal{C}$ are algebras. Thus let $\mathcal{C} = \mathit{Alg}_{\mathbb{C}}$ be the 2-category whose objects are complex algebras. If $A_0, A_1 \in \mathit{Alg}_{\mathbb{C}}$, then we define a 1-morphism $B\colon A_0 \to A_1$ to be an (A_1, A_0)-bimodule B, a complex vector space B with a left action of A_1 and a right action of A_0. Composition is by tensor product over algebras: if $B\colon A_0 \to A_1$ and $B'\colon A_1 \to A_2$, then $B' \circ B\colon A_0 \to A_2$ is the (A_2, A_0)-bimodule $B' \otimes_{A_1} B$. The symmetric monoidal structure is given by tensor product over $\mathbb{C}$. The algebra $\mathbb{C}$ is the tensor unit 1 and $\mathrm{Hom}_{\mathit{Alg}_{\mathbb{C}}}(1, 1)$ is the collection of $(\mathbb{C}, \mathbb{C})$-bimodules, which is canonically $\mathrm{Vect}_{\mathbb{C}}$, as desired. A 2-morphism between bimodules is a linear map which intertwines the algebra actions.

We pause to remark that we have climbed to the next categorical level—from 1-categories to 2-categories—by endowing objects in a 1-category with an associative unital composition law. Complex vector spaces form a 1-category, whereas complex vector spaces which are algebras form a 2-category. This is an important general idea, which can be implemented at all categorical levels and also can be iterated. For example, if we consider complex vector spaces with 2 composition laws we obtain a 3-category (of commutative algebras). We will meet more examples below. We can embed $\mathit{Alg}_{\mathbb{C}}$ into the more familiar (?) 2-category of $\mathbb{C}$-linear categories $\mathrm{Cat}_{\mathbb{C}}$: an algebra A maps to the linear category of left A-modules. It is usually easier to scale categorical heights via algebra structures than by introducing more elaborate objects. To put this construction in context, we observe that an *isomorphism* in the 2-category of algebras is a *Morita equivalence* of algebras.

Returning to

$$\widehat{F}_G\colon \mathrm{Bord}_2^{SO} \longrightarrow \mathit{Alg}_{\mathbb{C}}, \tag{5.8}$$

once we posit $\widehat{F}_G(\mathrm{pt}_+) = A = \mathrm{Map}(G, \mathbb{C})$, we can compute $\widehat{F}_G(S^1)$ as follows. We know that $\widehat{F}_G(\mathrm{pt}_-)$ is the dual to $\widehat{F}_G(\mathrm{pt}_+)$, since pt_+ and pt_- are dual in Bord_2^{SO}, and it turns out that the dual algebra is the *opposite* algebra A^o. The coevaluation in Figure 10 is the left $(A \otimes A^o)$-module A, and the evaluation is the right $(A^o \otimes A)$-module A. After permuting the two boundary points of the evaluation, we compose coevaluation and evaluation to compute

$$\widehat{F}_G(S^1) = A \otimes_{A \otimes A^o} A. \tag{5.9}$$

This tensor product is the *Hochschild homology* of the algebra A. We can easily compute it explicitly. Tensoring over A gives the tensor product $A \otimes_A A$ of the right A-module A with the left A-module A, which is canonically A by multiplication. Then the A^o-action is by left and right multiplication, so letting $[A, A] \subset A$ denote the subspace spanned by elements of the form $a_1a_2 - a_2a_1,\ a_1, a_2 \in A$, we conclude $\widehat{F}_G(S^1) = A/[A, A]$. This is *not* the center of A, which is what we expect from the text after (4.22). To identify the vector space $A/[A, A]$ with the center of A we need one more piece of data, a nondegenerate *trace* $\tau\colon A \to \mathbb{C}$ on A. Nondegeneracy means that $a_1, a_2 \mapsto \tau(a_1a_2)$ is a nondegenerate pairing, and then we identify the quotient $A/[A, A]$ with the orthogonal subspace $[A, A]^\perp \subset A$, which is easily identified with the center of A. The pair (A, τ) is a *Frobenius algebra.* For $A = \mathrm{Map}(G, \mathbb{C})$ we use the trace (4.16).

The cobordism hypothesis, stated for framed manifolds in Theorem 1.2, asserts that $\widehat{F}_G$ is determined by its value on pt_+. This is true here, but 'value on pt_+' must be interpreted as the pair (A, τ). The extra datum τ is necessary as $\widehat{F}_G$ is an oriented theory, not simply a framed theory; see Theorem 6.11 and Example 6.13.

In §4 we described an approach to the non-extended theory F_G using a finite version of the path integral in physics, which amounts to counting principal G-bundles. The finite path integral extends to give an *a priori* construction of $\widehat{F}_G$ in which $\widehat{F}_G(\mathrm{pt}_+) = A$ is the result of a computation; see [F, FHLT] for details.

Example 5.10. 3-dimensional Chern-Simons theory [W3] was the example which most pointed the way towards extended topological field theories. The approach of Reshetikhin-Turaev [RT1, RT2] to the resulting invariants of 3-manifolds and links begins with a quantum group, in the form of a complex linear category with extra structure, a *modular tensor category* [MSei]. By contrast, Witten begins with the Chern-Simons functional and uses the path integral. The relationship between the approaches, worked out in [F] for finite gauge groups, is that Chern-Simons is a (partially) extended theory of 1-, 2-, and 3-manifolds whose value on S^1 is the modular tensor category. A complete construction of this 1-2-3 theory beginning from quantum group data was given in [Tu]; see also [Wa]. There is current work, for example [BDH], to construct a fully extended 0-1-2-3 theory.

Example 5.11. The previous two examples are discrete: there are no interesting invariants for families of manifolds beyond those for single manifolds. That an extension of Definition 2.6 to families would be fruitful emerged in the 1990s from 2-dimensional field theories. Segal promoted the idea of a cochain-valued topological field theory [Se3], and there were several mathematical works which pointed towards invariants for families of manifolds; a quirky sample is [LZ, G, KM, BC]. The most definitive work in this direction is by Kevin Costello [Co], who constructed a theory of "open-closed" topological 2-dimensional field theories in families from Calabi-Yau categories. These are closely related to fully extended 2-dimensional theories; see [L1, §4.2].

Example 5.12. Another motivating example for the cobordism hypothesis which includes invariants for families of manifolds is *string topology*, which defines invariants of compact manifolds using its loop space and Riemann surfaces. It was introduced by Chas-Sullivan [CS], and there is a large literature which follows. See [L1, §4.2] for the relation with the cobordism hypothesis.

6. The cobordism hypothesis

Recall from §4 that objects in the image of a non-extended topological field theory obey a finiteness condition, expressed in categorical terms by dualizability (Definition 4.6). There is an analogous finiteness condition called *adjointability* for k-morphisms, $1 \le k \le n-1$, in an extended n-dimensional field theory. We give the definition for 1-morphisms, which specializes to the traditional notion of *adjoint functors* in category theory [Ka] for the 2-category of categories.

Definition 6.1. *Let $\mathcal{C}$ be a 2-category; $x, y \in \mathcal{C}$ objects in $\mathcal{C}$; and suppose $f\colon x \to y$, $g\colon y \to x$ are 1-morphisms. Then f is a* left adjoint *to g if there exist 2-morphisms $u\colon \mathrm{id}_x \Rightarrow g \circ f$ and $c\colon f \circ g \Rightarrow \mathrm{id}_y$ such that the* *compositions*

$$f = f \circ \mathrm{id}_x \xRightarrow{\mathrm{id}\times u} f \circ g \circ f \xRightarrow{c\times \mathrm{id}_y} \mathrm{id}_y \circ f = f \tag{6.2}$$

and

$$g = g \circ \mathrm{id}_y \xRightarrow{\mathrm{id}\times u} g \circ f \circ g \xRightarrow{c\times \mathrm{id}_x} \mathrm{id}_x \circ g = g \tag{6.3}$$

are identity 2-morphisms.

We then say that g is a right adjoint to f, and u, c are the unit and counit of an adjunction. The compositions (6.2) and (6.3) are the 2-morphism version of the S-diagram compositions (4.7). The corresponding definition for (∞, n)-categories and higher morphisms is similar, but the compositions are only the identity maps up to higher morphisms, or equivalently are identity maps in a homotopy category which remembers higher morphisms only up to equivalence. *Invertible* maps have adjoints—the inverse *is* an adjoint—so adjointability is weaker than invertibility.

Remark 6.4. If an n-morphism in an n-category, or (∞, n)-category, is adjointable then it is invertible. This follows since the unit and counit of an adjunction, which are $(n+1)$-morphisms, are invertible.

Let $\mathcal{C}$ be a symmetric monoidal (∞, n)-category and $F\colon \mathrm{Bord}_n \to \mathcal{C}$ an extended field theory. Then just as $F(\mathrm{pt})$ is dualizable, so too is $F(M)$ adjointable for every k-dimensional bordism M with $1 \le k \le n-1$. This is an extended finiteness condition satisfied by an extended topological field theory. We extract from $\mathcal{C}$ all objects which have duals, and whose duality data have adjoints, which in turn have adjoints, etc.

Lemma 6.5. [L1, §2.3] *Let $\mathcal{C}$ be a symmetric monoidal (∞, n)-category. There is an (∞, n)-category $\mathcal{C}^{\mathrm{fd}}$ and a homomorphism $i\colon \mathcal{C}^{\mathrm{fd}} \to \mathcal{C}$ so that (i) every object in $\mathcal{C}^{\mathrm{fd}}$ is dualizable and every k-morphism, $1 \le k \le n-1$, is adjointable, and (ii) $i\colon \mathcal{C}^{\mathrm{fd}} \to \mathcal{C}$ is universal with respect to (i).*

Here 'fd' stands either for 'fully dualizable' or 'finite dimensional'. An (∞, n)-category which satisfies (i) is said to "have duals", as in the statement of Theorem 1.1. The finiteness condition on a topological field theory $F\colon \mathrm{Bord}_n \to \mathcal{C}$ may be summarized by

$$\begin{tikzcd} & \mathcal{C}^{\mathrm{fd}} \arrow[d, "i"] \\ \mathrm{Bord}_n \arrow[ur, dashed] \arrow[r, "F"'] & \mathcal{C} \end{tikzcd} \tag{6.6}$$

In other words, F factors through $\mathcal{C}^{\mathrm{fd}}$.

Extended topological field theories $F\colon \mathrm{Bord}_n \to \mathcal{C}$ are the objects of an (∞, n)-category we denote $\mathrm{Hom}(\mathrm{Bord}_n, \mathcal{C})$. A 1-morphism $\eta\colon F_0 \to F_1$ between two homomorphisms assigns a $(k+1)$-dimensional bordism $\eta(M)\colon F_0(M) \to F_1(M)$ to each k-dimensional bordism M. The fact that adjointable n-morphisms are invertible (Remark 6.4) implies that any 1-morphism η is in fact an isomorphism. The same applies to higher morphisms. It follows that $\mathrm{Hom}(\mathrm{Bord}_n, \mathcal{C})$ is an $(\infty, 0)$-category—all morphisms are invertible—so according to Remark 5.3 can be viewed as a *space*. In other words, the collection of extended topological field theories with values in $\mathcal{C}$ is a space.

The cobordism hypothesis identifies the space $\mathrm{Hom}(\mathrm{Bord}_n, \mathcal{C})$ with a space constructed directly from $\mathcal{C}$ by combining Lemma 6.5 with another universal construction.

Lemma 6.7. [L1, §2.4] *Let $\mathcal{D}$ be an (∞, n)-category. There is an ∞-groupoid $\mathcal{D}^\sim$ and a homomorphism $j\colon \mathcal{D}^\sim \to \mathcal{D}$ so that (i) every k-morphism, $k > 0$, in $\mathcal{D}^\sim$ is invertible, and (ii) $j\colon \mathcal{D}^\sim \to \mathcal{D}$ is universal with respect to (i).*

The ∞-groupoid $\mathcal{D}^\sim$, which is an ∞-category in which every morphism is invertible, may be constructed from $\mathcal{D}$ by removing all noninvertible morphisms.

Finally, we can state a precise version of the cobordism hypothesis, first for n-framed manifolds.

Theorem 6.8 (Cobordism hypothesis: framed version)**.** *Let $\mathcal{C}$ be a symmetric monoidal (∞, n)-category. Then the map*

$$\begin{aligned} \mathrm{Hom}(\mathrm{Bord}_n^{\mathrm{fr}}, \mathcal{C}) &\longrightarrow (\mathcal{C}^{\mathrm{fd}})^\sim \\ F &\longmapsto F(\mathrm{pt}_+) \end{aligned} \tag{6.9}$$

is a homotopy equivalence of spaces.

At this point the reader should refer back to the heuristic versions stated in §1 as well as the discrete 1-dimensional version in Theorem 4.8. In particular, the cobordism hypothesis is a theorem about smooth manifolds and their diffeomorphism groups, which is reflected by the method of proof.

Suppose M is a bordism of dimension $k \le n$ which is n-framed. Recall that the n-framing is an isomorphism $\underline{\mathbb{R}^n} \to TM \times \underline{\mathbb{R}^{n-k}}$, where $\underline{\mathbb{R}^j}$ denotes the trivial real vector bundle of rank j over M. The orthogonal group[10] $O(n)$ acts on framings by precomposition with constant orthogonal maps $(n) \to (n)$. This induces an action of $O(n)$ on the space $\mathrm{Hom}(\mathrm{Bord}_n^{\mathrm{fr}}, \mathcal{C})$.

Corollary 6.10. *The orthogonal group $O(n)$ acts on $(\mathcal{C}^{\mathrm{fd}})^{\sim}$.*

Let G be a Lie group equipped with a homomorphism $\rho\colon G \to O(n)$. A *G-structure* on a bordism M is a reduction of structure group of its tangent bundle to G along ρ. More precisely, choose a Riemannian metric on M (this is a contractible choice). Then a G-structure is a principal G-bundle $P \to M$ together with an isomorphism of the associated G-bundle $\rho(P)$ with the bundle of orthonormal frames of $TM \oplus (n-k)$. For example, for $G = \{e\}$ a G-structure is an n-framing, and for $G = SO(n)$ it is an orientation. There is a bordism category Bord_n^G of manifolds with G-structure.

Theorem 6.11 (Cobordism hypothesis: G-structure version)**.** *The map*

$$\begin{aligned} \mathrm{Hom}(\mathrm{Bord}_n^G, \mathcal{C}) &\longrightarrow \big((\mathcal{C}^{\mathrm{fd}})^{\sim}\big)^{hG} \\ F &\longmapsto F(\mathrm{pt}_+) \end{aligned} \tag{6.12}$$

is a homotopy equivalence between the space of extended topological field theories on G-manifolds and the homotopy fixed point space of the G-action on $(\mathcal{C}^{\mathrm{fd}})^{\sim}$.

Here G acts through the homomorphism $\rho\colon G \to O(n)$ and the $O(n)$-action given in Corollary 6.10.

Example 6.13. For $n = 2$ an oriented 2-dimensional theory is determined by the value on pt_+, but in the fixed point space. Consider $\mathcal{C} = Alg_{\mathbb{C}}$, as in Example 5.7. First, the 2-category $Alg_{\mathbb{C}}^{\mathrm{fd}}$ of fully dualizable complex algebras has objects finite dimensional semisimple algebras, i.e., finite products of matrix complex algebras. (A careful proof may be found in [Da, §3.2].) A point in the *homotopy* fixed point space of the $SO(2)$-action includes extra *data*—in this case being a fixed point is not a *condition*—and the extra data here is the nondegenerate trace τ discussed in Example 5.7; see [FHLT, Example 2.8] for details.

The exhausted reader... and author... may be relieved to know that we are not going to attempt to summarize the proof sketched in [L1] in any

[10] It is perhaps more natural to use the full general linear group $GL(n;\mathbb{R})$, but all of the topological information is carried by the maximal compact subgroup $O(n) \subset GL(n;\mathbb{R})$.

detail. Rather, we give a very rough intuition for why the cobordism hypothesis might be true. Our exposition in §4, in particular the proof of Theorem 4.8, emphasizes the role of Morse theory. The existence of Morse functions allows the decomposition of a bordism into a composition of elementary bordisms (5.2). These elementary bordisms encode the evaluations and coevaluations, or units and counits, of duality and adjointness data. That is clear in the proof of Theorem 4.8. As another example, Figure 17 may be read as a counit for the adjunction between the two 1-morphisms coev, ev in Figure 10. So if $x \in \mathcal{C}$ is fully dualizable, a choice of duality data—duals and adjoints all the way up—defines F on elementary bordisms. As arbitrary bordisms are compositions of elementary bordisms, F can be extended to arbitrary bordisms. In other words a Morse function gives, in principle, a way to evaluate $F(M)$ for every bordism M. The issue is whether $F(M)$ is well-defined. The duality data involves choices, and we must be sure that those choices can be made coherently. This is expressed via contractibility statements. The first is that the space of duality data for a dualizable object x is contractible. The second generalizes the connectivity statement at the heart of Cerf theory [C]. Lurie uses a higher connectivity theorem of Kiyoshi Igusa [I] for the space of *generalized framed Morse functions*. Such functions relax the nondegeneracy condition at a critical point to allow a single degeneracy, as in (2.9), and also include a framing of the negative definite subspace at a critical point. Igusa proves that on a k-dimensional manifold this space is k-connected.[11] These contractibility statements are central to the proof, but it is a highly nontrivial problem to organize the higher categorical data to apply these theorems. The solution to that problem, described in detail in [L1], is equally central to the proof.

7. Implications, extensions, and applications

Some brief vignettes illustrate the scope of the extended topological field theory and the cobordism hypothesis.

[11] It is in fact a consequence of the cobordism hypothesis that this space of functions is weakly contractible. This has been proved independently of the cobordism hypothesis in [EM] and also in unpublished work of Galatius.

Invertible theories and Madsen-Tillmann spectra

Recall from Lemma 6.7 that any (∞, n)-category $\mathcal{D}$ has an underlying ∞-groupoid $\mathcal{D}^{\sim}$, which may be identified with a space. There is a quotient construction as well.

Lemma 7.1. *Let $\mathcal{D}$ be an (∞, n)-category. There is an ∞-groupoid $|\mathcal{D}|$ and a homomorphism $q\colon \mathcal{D} \to |\mathcal{D}|$ so that (i) every k-morphism, $k > 0$, in $|\mathcal{D}|$ is invertible, and (ii) $q\colon \mathcal{D} \to |\mathcal{D}|$ is universal with respect to (i).*

These constructions are relevant to *invertible topological field theories.*

Definition 7.2. *A topological field theory $\alpha\colon \mathrm{Bord}_n \to \mathcal{C}$ is* invertible *if $\alpha(M)$ is invertible for all objects and morphisms M.*

It follows from the cobordism hypothesis that α is invertible if and only if $\alpha(\mathrm{pt}_+)$ is invertible. An invertible field theory $\alpha\colon \mathrm{Bord}_n \to \mathcal{C}$ factors through $|\,\mathrm{Bord}_n\,|$ and $(\mathcal{C}^{\mathrm{fd}})^{\sim}$:

(7.3)
$$\begin{tikzcd} \mathrm{Bord}_n \arrow[r, "\alpha"] \arrow[d, "q"'] & \mathcal{C} \\ |\,\mathrm{Bord}_n\,| \arrow[r, "\tilde{\alpha}"] & (\mathcal{C}^{\mathrm{fd}})^{\sim} \arrow[u, "j"'] \end{tikzcd}$$

Since Bord_n and $\mathcal{C}$ are symmetric monoidal, so too are $|\,\mathrm{Bord}_n\,|$ and $(\mathcal{C}^{\mathrm{fd}})^{\sim}$. An $infty$-groupoid is equivalent to a space (Remark 5.3), and a symmetric monoidal ∞-groupoid is equivalent to the 0-space of a spectrum. Furthermore, $\tilde{\alpha}$ is an infinite loop space map. This reduces the study of invertible topological field theories to a problem in stable homotopy theory.

Remark 7.4. Invertible field theories play a role in ordinary quantum field theory, for example as *anomalies.*

A corollary of the cobordism hypothesis [L1, §2.5] determines the homotopy type of the spectrum $|\,\mathrm{Bord}_n\,|$. Consider first the bordism (∞, n)-category $\mathrm{Bord}_n^{\mathrm{fr}}$ of n-framed manifolds. The cobordism hypothesis, in the heuristic form Theorem 1.1, asserts that $\mathrm{Bord}_n^{\mathrm{fr}}$ is free on one generator. It follows that so too is $|\,\mathrm{Bord}_n^{\mathrm{fr}}\,|$. The latter is a spectrum, and the free spectrum on one generator is the sphere spectrum. For the bordism (∞, n)-category of G-manifolds Bord_n^G the cobordism hypothesis in the form Theorem 6.11 implies that $|\,\mathrm{Bord}_n^G\,|$ is the n^{th} suspension of a Madsen-Tillmann spectrum. (These spectra are mentioned in §2 before Definition 2.5.)

An ∞-groupoid—or $(\infty, 0)$-category—is a model for a space. We may view an (∞, n)-category as a generalization of a space which allows noninvertibility. From that perspective the cobordism hypothesis is a generalization of the Madsen-Tillmann conjecture.

Variations on the cobordism hypothesis

For another approach to extended topological field theories, see [MoW].

In [L1, §4] Lurie describes several applications and extensions of the cobordism hypothesis. One important extension is to manifolds with singularities, though there are many special cases which do not in fact involve singularities. To illustrate, in Example 5.7 we described a 2-dimensional oriented field theory F associated to a Frobenius algebra A. Now suppose that M is a left A-module. Recall that M determines a 1-morphism $M\colon 1 \to A$ in the Morita 2-category of algebras, where the tensor unit 1 is the trivial algebra $\mathbb{C}$. We might ask what sort of field theory we can associate to the pair (A, M), assuming sufficient finiteness.. A physicist might describe M as giving a *boundary condition* for F, and so extend F to a field theory $\widetilde{F}$ in which some boundaries are "colored" with the boundary condition M. For example, a closed interval with one endpoint colored is associated to M as a left A-module; the closed interval with both endpoints colored is associated to M as a vector space. The coloring represents a coning off of a point, which is viewed as a manifold with singularities. This is just the tip of the iceberg of possibilities opened up by the cobordism hypothesis with singularities.

From the point of view of algebra, given that $\mathrm{Bord}_n^{\mathrm{fr}}$ is the free symmetric monoidal (∞, n)-category with duals on one generator, we might ask how to describe more general symmetric monoidal (∞, n)-categories specified by generators and relations. Roughly speaking, the cobordism hypothesis with singularities identifies these as bordism categories of manifolds with singularities.

Applications to topology

We indicated briefly in Example 5.10 the important role that Chern-Simons theory played in the development of extended topological quantum field theories. That theory encodes invariants of 3-manifolds and links. Newer invariants of links and low dimensional manifolds were in part inspired by notions in extended field theory. Crane and Frenkel [CF] suggested that "categorification" of the 3-dimensional invariants would lead to new invariants, potentially related to Donaldson invariants. Later Khovanov [Kh] introduced such a categorification of the Jones polynomial. This now has a proposed derivation from quantum field theory [GSV, W4].

There is current research in many directions which will potentially take advantage of more powerful aspects of extended field theories and the cobordism hypothesis in contexts which are not discrete and semisimple. For example, the cobordism hypothesis illuminates string topology invariants and topological versions of Hochschild homology and its cousins [BCT]. It also appears in several discussions of the 2-dimensional extended topological field theories relevant for mirror symmetry: the "A-model" and the "B-model". There is an enormous literature on this subject; see [Te] for one recent example which uses ideas around the cobordism hypothesis.

Applications to algebra

Now we shift focus from topology and bordism categories to the codomain $\mathcal{C}$. Quite generally a homomorphism in algebra organizes the codomain according to the structure of the domain. This principle is often applied in the context of group actions on sets, for example: the structure of orbits and stabilizers illuminates the situation at hand. Here if $F\colon \mathrm{Bord}_n \to \mathcal{C}$ is a homomorphism, and $F(\mathrm{pt}_+) = x$ then we can study x using smooth manifolds and their gluings.

One application is to E_k*-algebras*, which are objects in a symmetric monoidal category which have k associative composition laws. We met E_1-algebras (ordinary associative algebras) in the category $\mathrm{Vect}_{\mathbb{C}}$ of complex vector spaces in Example 5.7. An E_2-algebra in $\mathrm{Vect}_{\mathbb{C}}$ is a commutative algebra and there is nothing higher up: an E_k-algebra for $k > 2$ is also a commutative algebra. More interesting examples are obtained if we look in other symmetric monoidal categories, for example the ∞-category of chain complexes. In [L1, §4.1] Lurie describes some relationships between the cobordism hypothesis and E_k-algebras in (∞, n)-categories. In particular, an E_k-algebra A in an (∞, n)-category $\mathcal{C}$ is automatically k-dualizable, so determines a homomorphism $F\colon \mathrm{Bord}_k^{\mathrm{fr}} \to E_k(\mathcal{C})$, where $E_k(\mathcal{C})$ is the $(\infty, n+k)$-category whose objects are E_k-algebras in $\mathcal{C}$. Thus E_k-algebras may be studied with smooth manifolds. For example, if A is an ordinary algebra (E_1-algebra), then in the associated field theory $F(S^1)$ is the Hochschild homology of A (see (5.9) for a simple example). Since the circle is an E_2-algebra in the bordism category, so too is the Hochschild homology $F(S^1)$. This is a version of the Deligne conjecture, and precise versions of this argument and generalizations appear in many works, for example [Co, KS, L2, BFN]. (We remark that there are several other proofs of the Deligne conjecture.)

As another application of the cobordism hypothesis to algebra, we mention ongoing work [DHS] which proves that a *fusion category* [ENO] is 3-dualizable. A fusion category is a special type of *tensor category*, and a tensor category is an E_1-algebra in the 2-category of linear categories. So tensor categories form a 3-category, and it is in that 3-category that fusion categories are fully dualizable. The associated 3-dimensional framed field theory can be brought to bear on the study of fusion categories. We remark that simple topological diagrams involving 0- and 1-dimensional manifolds are usually used to study fusion categories and their cousins. The cobordism hypothesis opens up the possibility of using the more powerful topology of 3-dimensional manifolds. In related ongoing work of the author and Teleman, we consider E_2-algebras in the 2-category of linear categories; they comprise the 4-category of *braided tensor categories*. We prove that *modular* tensor categories are invertible, which now gives a 4-dimensional perspective on quantum groups.

Applications to representation theory

In §4 and in Example 5.7 we illustrated a very simple, discrete 2-dimensional field theory associated to a finite group G. There is also a 3-dimensional field theory with values in the 3-category of tensor categories; it attaches the tensor category of vector bundles over G under convolution to pt. (The theory is unoriented—as is the 2-dimensional theory—so we have an unframed unoriented unadorned point.) That theory may be viewed as the simplest case of 3-dimensional Chern-Simons theory (Example 5.10). Ben-Zvi and Nadler [BN] study the analogous theory for a reductive complex group G. Discrete categories are futile here; the full force of ∞-categories comes into play. One would like a 3-dimensional theory which generalizes that of a finite group, and now attaches the symmetric monoidal ∞-category of $\mathcal{D}$-modules on G to a point. However, the necessary finiteness conditions are not satisfied. Instead, they construct a related 2-dimensional field theory, the *character theory*, which assigns to a point the *Hecke category* associated to G. Then one computes that the category of Lusztig's *character sheaves* is attached to S^1. The character theory may be viewed as a dimensional reduction of a 4-dimensional field theory [KW] related to the geometric

Langlands program. It seems likely that the topological field theory perspective, and the cobordism hypothesis, will shed light on old questions in the representation theory of semisimple Lie groups.

Echos in quantum field theory

As mentioned earlier, quantum field theorists traditionally only studied *2-tier* theories: correlation functions on n-manifolds and Hilbert spaces attached to $(n-1)$-manifolds. In recent years the ideas mathematicians have developed around extended field theories, including the cobordism hypothesis, have seeped into physics. In 2-dimensional conformal field theory there is a category of boundary conditions, called *D-branes*, and in topological versions this is understood to be part of an extended field theory. Higher dimensional analogs are now common; see [Kap] for a recent review. For example, Kapustin-Witten [KW] study a topological twist of the 4-dimensional $N = 4$ supersymmetric Yang-Mills theory. Going beyond the traditional two tiers, this theory attaches a category to every closed 2-manifold. Kapustin-Witten relate that to a category which appears in the geometric Langlands program. The story is richer: there is a family of theories parametrized by $\mathbb{CP}^1$ and *S-duality* acts as an involution on the theories. This suggests an equivalence between two different categories attached to a 2-manifold, which is a topological version of the basic conjecture in the geometric Langlands program.

The maximally supersymmetric $N = 4$ Yang-Mills theory is the dimensional reduction of a 6-dimensional supersymmetric field theory called Theory $\mathcal{X}$ which has superconformal invariance. This theory has no classical description. It is predicted to exist from limiting arguments in string theory. Its mysterious nature justifies the appellation 'Theory $\mathcal{X}$'. A few properties can be predicted from string theory, and these can be used to study dimensional reductions. Among the many protagonists here we mention Gaiotto [Ga] and Gaiotto-Moore-Neitzke [GMN]. One important idea—which is clearly inspired by extended field theory and the activity surrounding the cobordism hypothesis—is to study dimensional reductions of the 6-dimensional theory as a function of the compactifying manifold. This is formalized as follows. Suppose $F\colon \mathrm{Bord}_6 \to \mathcal{C}$ is a 6-dimensional extended topological theory. Then for any closed 2-manifold N we obtain a 4-dimensional theory $F_N\colon \mathrm{Bord}_4 \to \mathrm{Hom}_{\mathcal{C}}\big(F(N), F(N)\big)$ defined using

Cartesian product:[12] $F_N(M) = F(N \times M)$. Now view F_N as a function of N. Then we obtain a 2-dimensional extended field theory with values in the $(\infty, 4)$-category of 4-dimensional field theories! The flexibility in Definition 5.6 which allows arbitrary codomains is heavily used here. One can get other field theories by composing with homomorphism out of 4-dimensional theories. A recent paper [MoT] implements this idea in a physics context, and predicts the existence of certain holomorphic symplectic manifolds.

Finally, the renewed interest in E_n-algebras and their role in extended topological field theories may bring some fresh perspectives to quantum field theories which are not topological. One axiomatic approach to quantum field theory [H] assigns operator algebras to open sets and describes how they fit together. This idea was imported in an algebro-geometric framework in certain mathematical approaches to 2-dimensional conformal field theory, in vertex operator algebras [Bo] and chiral algebras [BeDr]. These ideas are circling back to general quantum field theories [CG] with potential to shed new light on their structure.

These are only a few examples of the potential that extended topological field theories and the cobordism hypothesis hold in both mathematics and physics.

References

[A1] M. F. Atiyah, *Bordism and cobordism*, Proc. Cambridge Philos. Soc. **57** (1961), 200–208.

[A2] ———, *Topological quantum field theories*, Inst. Hautes Études Sci. Publ. Math. (1988), no. 68, 175–186 (1989).

[Ab] Lowell Abrams, *Two-dimensional topological quantum field theories and Frobenius algebras*, J. Knot Theory Ramifications **5** (1996), no. 5, 569–587.

[B-N] Dror Bar-Natan, *On Khovanov's categorification of the Jones polynomial*, Algebr. Geom. Topol. **2** (2002), 337–370 (electronic).

[Ba] Clark Barwick, *(infinity, n)-Cat as a closed model category*, ProQuest LLC, Ann Arbor, MI, 2005. Thesis (Ph.D.)–University of Pennsylvania.

[BC] Martin Betz and Ralph L. Cohen, *Graph moduli spaces and cohomology operations*, Turkish J. Math. **18** (1994), no. 1, 23–41.

[BCT] Andrew J. Blumberg, Ralph L. Cohen, and Constantin Teleman, *Open-closed field theories, string topology, and Hochschild homology*, `0906.5198`.

[12] The bordism groups of Pontrjagin and Thom are *rings* with multiplication given by Cartesian product. Our discussion of topological field theory has not used this ring structure until now.

[BD] John C. Baez and James Dolan, *Higher-dimensional algebra and topological quantum field theory*, J. Math. Phys. **36** (1995), no. 11, 6073–6105.

[BDH] Arthur Bartels, Christopher L. Douglas, and André G. Henriques, *Conformal nets and local field theory*, `0912.5307`.

[Be] Julia E. Bergner, *A survey of* $(\infty, 1)$*-categories*, Towards higher categories, IMA Vol. Math. Appl., vol. 152, Springer, New York, 2010, pp. 69–83.

[BeDr] Alexander Beilinson and Vladimir Drinfeld, *Chiral algebras*, American Mathematical Society Colloquium Publications, vol. 51, American Mathematical Society, Providence, RI, 2004.

[BFN] David Ben-Zvi, John Francis, and David Nadler, *Integral transforms and Drinfeld centers in derived algebraic geometry*, J. Amer. Math. Soc. **23** (2010), no. 4, 909–966.

[BN] David Ben-Zvi and David Nadler, *The Character Theory of a Complex Group*, `0904.1247`. preprint.

[Bo] Richard E. Borcherds, *Vertex algebras, Kac-Moody algebras, and the Monster*, Proc. Nat. Acad. Sci. U.S.A. **83** (1986), no. 10, 3068–3071.

[BS] C. Barwick and C. Schommer-Pries, in preparation.

[C] Jean Cerf, *La stratification naturelle des espaces de fonctions différentiables réelles et le théorème de la pseudo-isotopie*, Inst. Hautes Études Sci. Publ. Math. (1970), no. 39, 5–173.

[CF] Louis Crane and Igor B. Frenkel, *Four-dimensional topological quantum field theory, Hopf categories, and the canonical bases*, J. Math. Phys. **35** (1994), no. 10, 5136–5154. Topology and physics.

[CG] K. Costello and O. Gwilliam, *Factorization algebras in perturbative quantum field theory*. `http://math.northwestern.edu/~costello/factorization_public.html`. in preparation.

[Co] Kevin Costello, *Topological conformal field theories and Calabi-Yau categories*, Adv. Math. **210** (2007), no. 1, 165–214.

[CS] Moira Chas and Dennis Sullivan, *String Topology*, `math/9911159`.

[D] S. K. Donaldson, *Polynomial invariants for smooth four-manifolds*, Topology **29** (1990), no. 3, 257–315.

[Da] O. Davidovich, *State sums in 2-dimensional fully extended topological field theories*, Ph.D. thesis, University of Texas at Austin, 2011. `http://repositories.lib.utexas.edu/handle/2152/ETD-UT-2011-05-3139`.

[DHS] C. Schommer-Pries C. Douglas, A. Henriques. in preparation.

[Di] R. Dijkgraaf, *A geometrical approach to two-dimensional conformal field theory*, Ph.D. thesis, University of Utrecht, 1989. `http://igitur-archive.library.uu.nl/dissertations/2011-0929-200347/UUindex.html`.

[DW] Robbert Dijkgraaf and Edward Witten, *Topological gauge theories and group cohomology*, Comm. Math. Phys. **129** (1990), no. 2, 393–429.

[EM] Y. M. Eliashberg and N. M. Mishachev, *The space of framed functions is contractible*, `arXiv:1108.1000`.

[ENO] Pavel Etingof, Dmitri Nikshych, and Viktor Ostrik, *On fusion categories*, Ann. of Math. (2) **162** (2005), no. 2, 581–642.

[F] Daniel S. Freed, *Higher algebraic structures and quantization*, Comm. Math. Phys. **159** (1994), no. 2, 343–398, `arXiv:hep-th/9212115`.

[Fa] L. Faddeev, *Elementary introduction to quantum field theory*, Quantum fields and strings: a course for mathematicians. Vol. 1 (P. Deligne et. al., ed.), American Mathematical Society, 1999, pp. 513–550. Notes by L. Jeffrey.

[FHLT] Daniel S. Freed, Michael J. Hopkins, Jacob Lurie, and Constantin Teleman, *Topological quantum field theories from compact Lie groups*, A celebration of the mathematical legacy of Raoul Bott, CRM Proc. Lecture Notes, vol. 50, Amer. Math. Soc., Providence, RI, 2010, pp. 367–403.

[FHT] Daniel S. Freed, Michael J. Hopkins, and Constantin Teleman, *Consistent orientation of moduli spaces*, The many facets of geometry, Oxford Univ. Press, Oxford, 2010, pp. 395–419.

[FQ] Daniel S. Freed and Frank Quinn, *Chern-Simons theory with finite gauge group*, Comm. Math. Phys. **156** (1993), no. 3, 435–472, `arXiv:hep-th/911100`.

[G] E. Getzler, *Batalin-Vilkovisky algebras and two-dimensional topological field theories*, Comm. Math. Phys. **159** (1994), no. 2, 265–285.

[Ga] Davide Gaiotto, *N=2 dualities*, `arXiv:0904.2715`.

[GJ] James Glimm and Arthur Jaffe, *Quantum physics*, second ed., Springer-Verlag, New York, 1987. A functional integral point of view.

[GMN] Davide Gaiotto, Gregory W. Moore, and Andrew Neitzke, *Wall-crossing, Hitchin Systems, and the WKB Approximation*, `0907.3987`.

[GMTW] Søren Galatius, Ulrike Tillmann, Ib Madsen, and Michael Weiss, *The homotopy type of the cobordism category*, Acta Math. **202** (2009), no. 2, 195–239, `arXiv:math/0605249`.

[GSV] Sergei Gukov, Albert Schwarz, and Cumrun Vafa, *Khovanov-Rozansky homology and topological strings*, Lett. Math. Phys. **74** (2005), no. 1, 53–74.

[H] Rudolf Haag, *Local quantum physics*, second ed., Texts and Monographs in Physics, Springer-Verlag, Berlin, 1996. Fields, particles, algebras.

[I] Kiyoshi Igusa, *The space of framed functions*, Trans. Amer. Math. Soc. **301** (1987), no. 2, 431–477.

[K] Robion Kirby, *A calculus for framed links in S^3*, Invent. Math. **45** (1978), no. 1, 35–56.

[Ka] Daniel M. Kan, *Adjoint functors*, Trans. Amer. Math. Soc. **87** (1958), 294–329.

[Kap] Anton Kapustin, *Topological field theory, higher categories, and their applications*, `arXiv:1004.2307`.

[Kh] Mikhail Khovanov, *A categorification of the Jones polynomial*, Duke Math. J. **101** (2000), no. 3, 359–426.

[KM] M. Kontsevich and Yu. Manin, *Gromov-Witten classes, quantum cohomology, and enumerative geometry*, Comm. Math. Phys. **164** (1994), no. 3, 525–562.

[Ko] Joachim Kock, *Frobenius algebras and 2D topological quantum field theories*, London Mathematical Society Student Texts, vol. 59, Cambridge University Press, Cambridge, 2004.

[KS] M. Kontsevich and Y. Soibelman, *Notes on A_∞-algebras, A_∞-categories and non-commutative geometry*, Homological mirror symmetry, Lecture Notes in Phys., vol. 757, Springer, Berlin, 2009, pp. 153–219.

[KW] Anton Kapustin and Edward Witten, *Electric-magnetic duality and the geometric Langlands program*, Commun. Number Theory Phys. **1** (2007), no. 1, 1–236, `arXiv:hep-th/0604151`.

[L1] Jacob Lurie, *On the Classification of Topological Field Theories*, `0905.0465`.

[L2] ———, *Higher Algebra*. `http://math.harvard.edu/ lurie/papers/higheralgebra.pdf`. draft version.

[LZ] Bong H. Lian and Gregg J. Zuckerman, *New perspectives on the BRST-algebraic structure of string theory*, Comm. Math. Phys. **154** (1993), no. 3, 613–646.

[Ma] George W. Mackey, *The mathematical foundations of quantum mechanics: A lecture-note volume*, W., A. Benjamin, Inc., New York-Amsterdam, 1963.

[May] J. P. May, *Stable algebraic topology, 1945–1966*, History of topology, North-Holland, Amsterdam, 1999, pp. 665–723.

[Mc] Saunders Mac Lane, *Categories for the working mathematician*, second ed., Graduate Texts in Mathematics, vol. 5, Springer-Verlag, New York, 1998.

[Mi1] John W. Milnor, *Topology from the differentiable viewpoint*, Based on notes by David W. Weaver, The University Press of Virginia, Charlottesville, Va., 1965.

[Mi2] ———, *Morse theory*, Based on lecture notes by M. Spivak and R. Wells. Annals of Mathematics Studies, No. 51, Princeton University Press, Princeton, N.J., 1963.

[Mi3] ———, *Lectures on the h-cobordism theorem*, Notes by L. Siebenmann and J. Sondow, Princeton University Press, Princeton, N.J., 1965.

[MoT] Gregory W. Moore and Yuji Tachikawa, *On 2d TQFTs whose values are holomorphic symplectic varieties*, `1106.5698`.

[MoW] Scott Morrison and Kevin Walker, *Higher categories, colimits, and the blob complex*, Proc. Natl. Acad. Sci. USA **108** (2011), no. 20, 8139–8145.

[MS] G. Segal G. W. Moore, *Dirichlet branes and mirror symmetry*, Clay Mathematics Monographs, vol. 4, ch. D-branes and K-theory in 2D conformal field theory, pp. x+681, American Mathematical Society, Providence, RI, 2009.

[MSei] Gregory Moore and Nathan Seiberg, *Classical and quantum conformal field theory*, Comm. Math. Phys. **123** (1989), no. 2, 177–254.

[MT] Ib Madsen and Ulrike Tillmann, *The stable mapping class group and $Q(\mathbb{CP}^\infty_+)$*, Invent. Math. **145** (2001), no. 3, 509–544.

[Mu] David Mumford, *Towards an enumerative geometry of the moduli space of curves*, Arithmetic and geometry, Vol. II, Progr. Math., vol. 36, Birkhäuser Boston, Boston, MA, 1983, pp. 271–328.

[MW] Ib Madsen and Michael Weiss, *The stable moduli space of Riemann surfaces: Mumford's conjecture*, Ann. of Math. (2) **165** (2007), no. 3, 843–941.

[P] L. S. Pontryagin, *Smooth manifolds and their applications in homotopy theory*, American Mathematical Society Translations, Ser. 2, Vol. 11, American Mathematical Society, Providence, R.I., 1959, pp. 1–114.

[Pa] Richard S. Palais, *Seminar on the Atiyah-Singer index theorem*, With contributions by M. F. Atiyah, A. Borel, E. E. Floyd, R. T. Seeley, W. Shih and R. Solovay. Annals of Mathematics Studies, No. 57, Princeton University Press, Princeton, N.J., 1965.

[Q] Frank Quinn, *Lectures on axiomatic topological quantum field theory*, Geometry and quantum field theory (Park City, UT, 1991), IAS/Park City Math. Ser., vol. 1, Amer. Math. Soc., Providence, RI, 1995, pp. 323–453.

[R] Charles Rezk, *A Cartesian presentation of weak n-categories*, Geom. Topol. **14** (2010), no. 1, 521–571.

[RT1] N. Yu. Reshetikhin and V. G. Turaev, *Ribbon graphs and their invariants derived from quantum groups*, Comm. Math. Phys. **127** (1990), no. 1, 1–26.

[RT2] ———, *Invariants of* 3*-manifolds via link polynomials and quantum groups*, Invent. Math. **103** (1991), no. 3, 547–597.

[Se1] G. B. Segal, *Felix Klein Lectures 2011.* `http://www.mpim-bonn.mpg.de/node/3372/abstracts`.

[Se2] ———, *The definition of conformal field theory*, Topology, geometry and quantum field theory, London Math. Soc. Lecture Note Ser., vol. 308, Cambridge Univ. Press, Cambridge, 2004, pp. 421–577.

[Se3] ———, *String algebras.* `http://www.cgtp.duke.edu/ITP99/segal/stanford/lect5.pdf`.

[Sm] S. Smale, *On the structure of manifolds*, Amer. J. Math. **84** (1962), 387–399.

[Sp] Edwin H. Spanier, *Algebraic topology*, Springer-Verlag, New York, 1981. Corrected reprint.

[SW] R. F. Streater and A. S. Wightman, *PCT, spin and statistics, and all that*, Princeton Landmarks in Physics, Princeton University Press, Princeton, NJ, 2000. Corrected third printing of the 1978 edition.

[T] René Thom, *Quelques propriétés globales des variétés différentiables*, Comment. Math. Helv. **28** (1954), 17–86.

[Te] C. Teleman, *Lie group actions in categories and Langlands duality*. `http://media.ma.utexas.edu/media/Teleman_Lecture1`. Perspectives in Geometry lectures at University of Texas, October, 2011.

[Til1] Ulrike Tillmann, *The classifying space of the* $(1+1)$*-dimensional cobordism category*, J. Reine Angew. Math. **479** (1996), 67–75.

[Til2] ———, *On the homotopy of the stable mapping class group*, Invent. Math. **130** (1997), no. 2, 257–275.

[Tu] V. G. Turaev, *Quantum invariants of knots and 3-manifolds*, de Gruyter Studies in Mathematics, vol. 18, Walter de Gruyter & Co., Berlin, 1994.

[W1] Edward Witten, *Topological quantum field theory*, Comm. Math. Phys. **117** (1988), no. 3, 353–386.

[W2] ———, *Topological sigma models*, Comm. Math. Phys. **118** (1988), no. 3, 411–449.

[W3] ———, *Quantum field theory and the Jones polynomial*, Comm. Math. Phys. **121** (1989), no. 3, 351–399.

[W4] ———, *Fivebranes and Knots*, `1101.3216`.

[Wa] K. Walker, *On Witten's 3-manifold invariants*. `http://canyon23.net/math/1991TQFTNotes.pdf`.

Department of Mathematics, University of Texas, 1 University Station C1200, Austin, TX 78712-0257

E-mail address: `dafr@math.utexas.edu`

DISPERSIVE EQUATIONS AND THEIR ROLE BEYOND PDE

GIGLIOLA STAFFILANI[1]

Abstract. Arguably the star in the family of dispersive equations is the Schrödinger equation. Among many mathematicians and physicists it is regarded as fundamental, in particular to understand complex phenomena in quantum mechanics.

But not many people may know that this equation, when defined on tori for example, has a very reach and more abstract structure that touches several fields of mathematics, among which analytic number theory, symplectic geometry, probability and dynamical systems.

In this talk I will illustrate in the simplest possible way how all these different aspects of a unique equation have a life of their own while interacting with each other to assemble a beautiful and subtle picture. This picture is not yet completely well understood and many questions and open problems are there ready to be solved by a new generation of mathematicians.

1. Introduction

In these notes I would like to collect some old and new results addressing very different mathematical aspects related to semilinear periodic Schrödinger equations in low dimensions. In doing so I will present some open problems that often go behind the field of partial differential equations and touches upon analytic number theory, probability, symplectic geometry and dynamical systems.

After the introduction in Section 1 I will set up the stage in Section 2. I will start Section 3 I will start with a (now classical) Strichartz inequality for the periodic linear Schrödinger equation in two dimensions due to Bourgain. I will continue with some results on local and global well-posedness for certain nonlinear Schrödinger equations.

In Section 4 I will elaborate on the growth in time of high order Sobolev norms for the global flow, whenever it exists. I will explain how the estimate of this growth could give some information on how the frequency profile of a certain wave solution could move from low to high frequencies while maintaining constant mass and energy (*forward cascade.*) I will present two results for the defocusing, cubic, periodic, two dimensional Schrödinger equation: the first is a polynomial upper bound in time for Sobolev norms of a global generic solution; the second is a weak growth result, namely that after fixing a small constant δ and a large one K, one can find a certain solution that at time zero is as small[1] as δ and at a certain time far in the future is as big as K.

In Section 5 I will use certain periodic Schrödinger equations as examples of infinite dimension Hamiltonian systems and for them I will present some old and recent results that are generalizations of fine dimensions ones. As a first example I will consider the cubic periodic defocusing NLS and I will recall the squeezing theorem due to Bourgain. Next I will introduce the concept of Gibbs measures associated to periodic semilinear Schrödinger equations in one dimension. These measure already proposed by Lebowitz, Rose and Speer

[1] GS is funded in part by NSF DMS-1068815.

[1]In terms of a fixed Sobolev norm.

were later proved to be invariant by Bourgain who also used this invariance to show global well-posedness at a level in which conservation laws are not available. Of course in this case global well-posedness should be understood as an almost sure result. I will then introduce the periodic derivative nonlinear Schrödinger (DNLS) equation. This is an integrable system, that also can be viewed as an Hamiltonian system. Proving that it is globally well-posed for rough data is very challenging. In fact in order to be able to use certain estimates one needs to apply a gauge transformation to the equation. Moreover even for the gauged equation local well-posedness can be obtained via a fixed point argument only on certain spaces that are of type l^p, not necessarily $p = 2$, with respect to frequency variables. Because of this when later one wants to introduce a Gibbs measure, which is in turn related to the Gaussian measure defined on Sobolev spaces $H^s, s < \frac{1}{2}$, one needs to generalize the definition and take advantage of the more abstract Wiener theory. In spite of several obstacles that one needs to overcome in order to apply a variant of Bourgain's argument, one still obtains for the gauged DNLS problem an almost surely global well-posedness result. Of course at the end one needs to "un-gauge" and I will show how a purely probabilistic argument will translate the almost surely global well-posedness for the gauged DNLS into a similar one for the original derivative nonlinear Schrödinger equation.

2. Setting up the stage

The objects of study in these notes is mainly the semilinear Schrödinger (NLS) initial value problems (IVP)

$$\begin{cases} iu_t + \frac{1}{2}\Delta u = \lambda |u|^{p-1}u, \\ u(x,0) = u_0(x) \end{cases} \tag{1}$$

where $p > 1$, $u : \mathbb{R} \times \mathbb{T}^n \to \mathbb{C}$, and $\mathbb{T}^n$ is a n-dimensional torus[2]. We observe right away that (1) admits two conservation laws

$$H(u(t)) = \frac{1}{2}\int |\nabla u|^2(x,t)\,dx + \frac{2\lambda}{p+1}\int |u(t,x)|^{p+1}\,dx = H(u_0) \tag{2}$$

called the Hamiltonian and

$$M(u(t)) = \int |u|^2(x,t)\,dx = M(u_0) \tag{3}$$

called the mass.

Schrödinger equations are classified as *dispersive* partial differential equations and the justification for this name comes from the fact that if no boundary conditions are imposed their solutions tend to be waves which spread out spatially. A simple and complete mathematical characterization of the word *dispersion* is given to us for example by R. Palais in [41].

It is probably common knowledge that dispersive equations are proposed as models of certain wave phenomena that occur in nature. But it turned out that some of these equations appear also in more abstract mathematical areas such as algebraic geometry [29] and they are found to possess surprisingly beautiful structures. Certainly I am not in the position to discuss this part of mathematics here, but nevertheless I hope I will be able to give a glimpse of various connections of these equations with other areas of mathematics.

The interesting aspect of dispersive equations, Schrödinger equations in particular, is that in later times their solutions do not acquire extra smoothness and neither remain

[2] Later we will distinguish between a rational and an irrational torus.

compact if the initial profiles were. In particular, since we will impose periodic boundary conditions, dispersion will be extremely weak. All this will make our analysis more difficult, but also more interesting.

Probably the most standard questions that one may want to ask about an IVP such as (1), since it does model physical phenomena, are existence of solutions, stability, time-asymptotic properties of solutions, blow up etc. Until recently these questions were addressed in a very deterministic way and I will report on some of these results in Sections 2, 3 and 4. In recent years there has been an increasing interest on addressing these questions using a natural probabilistic approach, this is some of the content of the remaining Section 5. The set up for this probabilistic approach is based on viewing (1) as an infinite dimensional Hamiltonian system. This is done by rewriting the equation as an Hamiltonian systems for the Fourier coefficients of the solutions to (1). Using this structure one can then formally define an invariant measure [34] acting on the infinite dimensional space given by the vectors of Fourier coefficients. This measure, proved to be invariant [5], is able to select data in rough spaces that can be evolved globally in time even when blow up may occur and in so doing gives what we call an almost surely global well-posedness.

The infinite dimensional Hamiltonian structure that we can recognize for some NLS equations, in some cases can be also equipped with a symplectic structure. Then the natural question is whether one may be able to extend fundamental concepts such a *capacity* or prove results such as Gromov's non-squeezing theorem in this infinite dimensional context, [6, 21, 32]. In these notes I will recall one of such results, see Theorem 5.13, but much more needs to be studied and discovered in this area.

It is clear by now that when possible, a strong deterministic and probabilistic approach to the study of an IVP such as (1) is certainly bound to generate not just some abstract and beautiful mathematics, but also a deeper understanding of the physical phenomena that semilinear Schrödinger equations represent.

2.1. **Notation.** Throughout these notes we use C to denote various constants. If C depends on other quantities as well, this will be indicated by explicit subscripting, e.g. $C_{\|u_0\|_2}$ will depend on $\|u_0\|_2$. We use $A \lesssim B$ to denote an estimate of the form $A \leq CB$, where C is an absolute constant. We use $a+$ and $a-$ to denote expressions of the form $a + \varepsilon$ and $a - \varepsilon$, for some $0 < \varepsilon \ll 1$.

Finally, since we will be making heavy use of Fourier transforms, we recall here that $\hat{f}$ will usually denote the Fourier transform of f with respect to the space variables and when there is no confusion we use the hat notation even when we take Fourier transform also with respect to the time variable. In general though we will use the notation $\tilde{u}$ if we want to emphasize that we take the Fourier transform of a function $u(t, x)$ in both space and time variables.

3. Periodic Strichartz estimates

Let's start with the classical result of existence, uniqueness and stability of solutions. for an IVP. It is not hard to understand that these results strongly depend on the regularity one asks for the solutions themselves and the given data. So we first have to decide how we "measure" the regularity of function. The most common way of doing so is by deciding where the weak derivatives of the function "live". Most of the times we assume that the data are in Sobolev spaces H^s. In more sophisticated instances one may need to replace Sobolev spaces with different ones, like Besov spaces, Hölder spaces, and so on.

Since we will be dealing with functions that have a time variable we will often need mixed norm spaces, so for example, we may need that $f \in L^p_x L^q_t$, that is $\|(\|f(x,t)\|_{L^q_t})\|_{L^p_x} < \infty$. Finally, for a fixed interval of time $[0,T]$ and a Banach space of functions Z, we denote with $C([0,T],Z)$ the space of continuous maps from $[0,T]$ to Z.

We are now ready to give the definition of well-posedness for the IVP (1). We start with the linear Schrödinger IVP

$$\begin{cases} iv_t + \Delta v = 0, \\ v(x,0) = u_0(x). \end{cases} \tag{4}$$

The solution $v(x,t) =: S(t)u_0(x)$ of this IVP will be studied below, for now we will use it to write the solution to (1).

Definition 3.1. *We say that the IVP* (1) *is* locally well-posed (l.w.p) *in $H^s(\mathbb{R}^n)$ if for any ball B in the space $H^s(\mathbb{R}^n)$ there exist a time T and a Banach space of functions $X \subset L^\infty([-T,T], H^s(\mathbb{R}^n))$ such that for each initial data $u_0 \in B$ there exists a unique solution $u \in X \cap C([-T,T], H^s(\mathbb{R}^n))$ for the integral equation*[3]

$$u(x,t) = S(t)u_0 + c\int_0^t S(t-t')|u|^{p-1}u(t'))\,dt'. \tag{5}$$

Furthermore the map $u_0 \to u$ is continuous as a map from H^s into $C([-T,T], H^s(\mathbb{R}^n))$. If uniqueness is obtained in $C([-T,T], H^s(\mathbb{R}^n))$, then we say that local well-posedness is "unconditional".

If Definition 3.1 holds for all $T > 0$ then we say that the IVP is *globally well-posed* (g.w.p).

Remark 3.2. *The intervals of time are symmetric about the origin because the problems that we study here are all time reversible (i.e. if $u(x,t)$ is a solution, then so is $-u(x,-t)$).*

Usually the way one proves well-posedness, at least locally, is by defining an operator

$$Lv = S(t)u_0 + c\int_0^t S(t-t')|v|^{p-1}v(t'))\,dt'$$

and then showing that in a certain space of functions X one has a fixed point and as a consequence a solution according to (5). The hard part is to decide what space X could work. The general idea is to show strong estimates[4] for the solution $S(t)u_0$ of the linear problem (4), identify the space X from these estimates and expect that the solution u also satisfies them at least when through (5) one can show that u is a perturbation of the linear problem. This kind of argument usually works in so called subcritical regimes[5] and for short times; for long times and critical regimes the situation could be much more complicated.

Remark 3.3. *Our notion of global well-posedness does not require that $\|u(t)\|_{H^s(\mathbb{R}^n)}$ remains uniformly bounded in time. In fact, unless $s = 0, 1$ and one can use the conservation of mass or energy, it is not a triviality to show such an uniform bound. This can be obtained as a consequence of scattering, when scattering is available. In general this is a question related to* weak turbulence theory *and we will address it more in details in Section 4.*

[3]Note that (1) is equivalent to (5) via the Duhamel principle when enough regularity is assumed.

[4]For example Strichartz estimates in Section 3.

[5]If we write $H(u(t)) = K(u(t)) + \lambda P(u(t))$, where $K(u(t)) = \frac{1}{2}\int |\nabla u|^2(x,t)\,dx$ is the kinetic energy and $P(u(t)) = \frac{2}{p+1}\int |u(t,x)|^{p+1}\,dx$ is the potential one, then the energy subcritical regime is when the kinetic energy is stronger than the potential one.

We are now ready to introduce some of the most important estimates relative to the solution $S(t)u_0$ to the linear Schrödinger IVP (4). This solution is easily computable by taking Fourier transform. In fact for each fixed frequency ξ problem (4) transforms into the ODE

$$\begin{cases} i\hat{v}_t(t,\xi) - |\xi|^2\hat{v}(t,\xi) = 0, \\ \hat{v}(\xi,0) = \hat{u}_0(\xi) \end{cases} \tag{6}$$

and we can write its solution as

$$\hat{v}(t,\xi) = e^{-i|\xi|^2 t}\hat{u}_0(\xi).$$

We observe that what we just did works both in $\mathbb{R}^n$ and $\mathbb{T}^n$.

3.1. **Strichartz estimates in $\mathbb{R}^n$.** If we define, in the distributional sense,

$$K_t(x) = \frac{1}{(\pi i t)^{n/2}} e^{i\frac{|x|^2}{2t}},$$

we then have

$$S(t)u_0(x) = e^{it\Delta}u_0(x) = u_0 \star K_t(x) = \frac{1}{(\pi i t)^{n/2}} \int e^{i\frac{|x-y|^2}{2t}} u_0(y)\, dy. \tag{7}$$

As mentioned already

$$\widehat{S(t)u_0}(\xi) = e^{-i\frac{1}{2}|\xi|^2 t}\hat{u}_0(\xi), \tag{8}$$

and from here $S(t)u_0(x)$ can be interpreted as the adjoint of the Fourier restriction operator on the paraboloid $P = \{(\xi, |\xi|^2) \ \text{ for } \xi \in \mathbb{R}^n\}$. This remark, strictly linked to (7) and (8), can be used to prove a variety of very deep estimates for $S(t)u_0$, see for example [15, 45]. From (7) we immediately have the so called *dispersive estimate*

$$\|S(t)u_0\|_{L^\infty} \lesssim \frac{1}{t^{n/2}}\|u_0\|_{L^1}. \tag{9}$$

From (8) instead we have the conservation of the homogeneous Sobolev norms[6]

$$\|S(t)u_0\|_{\dot{H}^s} = \|u_0\|_{\dot{H}^s}, \tag{10}$$

for all $s \in \mathbb{R}$. Interpolating (9) with (10) when $s = 0$ and using a so called TT^* argument one can prove the famous Strichartz estimates summarized in the following theorem:

Theorem 3.4. *[Strichartz Estimates in $\mathbb{R}^n$] Fix $n \geq 1$. We call a pair (q, r) of exponents admissible if $2 \leq q, r \leq \infty$, $\frac{2}{q} + \frac{n}{r} = \frac{n}{2}$ and $(q, r, n) \neq (2, \infty, 2)$. Then for any admissible exponents (q, r) and $(\tilde{q}, \tilde{r})$ we have the homogeneous Strichartz estimate*

$$\|S(t)u_0\|_{L^q_t L^r_x(\mathbb{R}\times\mathbb{R}^n)} \lesssim \|u_0\|_{L^2_x(\mathbb{R}^n)} \tag{11}$$

and the inhomogeneous Strichartz estimate

$$\left\| \int_0^t S(t-t')F(t')\, dt' \right\|_{L^q_t L^r_x(\mathbb{R}\times\mathbb{R}^n)} \lesssim \|F\|_{L^{\tilde{q}'}_t L^{\tilde{r}'}_x(\mathbb{R}\times\mathbb{R}^n)}, \tag{12}$$

where $\frac{1}{\tilde{q}} + \frac{1}{\tilde{q}'} = 1$ and $\frac{1}{\tilde{r}} + \frac{1}{\tilde{r}'} = 1$.

See [30] and [47] for some concise proofs, and [15] for a complete list of authors who contributed to the final version of this theorem.

[6] We will see later that the L^2 norm is conserved also for the nonlinear problem (1).

3.2. **Strichartz estimates in $\mathbb{T}^n$.** In this section we will see how essential is the assumption that $\mathbb{T}^n$ is a rational torus[7] in order to be able to prove sharp Strichartz estimates. The conjecture is that for irrational tori one should be able to prove similar estimates, if not better in some cases, but for now the best available results are due to Bourgain in [9, 10]. In a sense irrational tori should generate some sort of weak dispersion since the reflections of the wave solutions through periodic boundary conditions, with periods irrational with respect to each other, should interact less in the nonlinearity. As for now there are no results of this type in the literature.

Assume that $c_i > 0$, $i = 1, ..., n$ are the periods with respect to each coordinate. In the periodic case one cannot expect the range of admissible pairs (q, r) as in Theorem 3.4. We concentrate on the pairs $q = r$, that is $q = \frac{2(n+2)}{n}$. There is the following conjecture:

Conjecture 3.1. *Assume that $\mathbb{T}^n$ is a rational torus and the support of $\hat{\phi}_N$ is in the ball $B_N(0) = \{|n| \lesssim N\}$. Write*

$$S(t)\phi_N(x) = \sum_{k\in\mathbb{Z}^n, |k|\sim N} a_k e^{i(\langle x,k\rangle - \gamma(k)t)},$$

where (a_k) are the Fourier coefficients of ϕ_N and

$$\gamma(k) = \sum_{i=1}^{n} c_i k_i^2. \tag{13}$$

If the torus is rational we can assume without loss of generality that $c_i \in \mathbb{N}$. Then

$$\|S(t)\phi_N\|_{L^q_t L^q_x([0,1]\times\mathbb{T}^n)} \lesssim C_q \|\phi_N\|_{L^2_x(\mathbb{T}^n)} \quad \text{if} \quad q < \frac{2(n+2)}{n} \tag{14}$$

$$\|S(t)\phi_N\|_{L^q_t L^q_x([0,1]\times\mathbb{T}^n)} \ll N^{\epsilon} \|\phi_N\|_{L^2_x(\mathbb{T}^n)} \quad \text{if} \quad q = \frac{2(n+2)}{n} \tag{15}$$

$$\|S(t)\phi_N\|_{L^q_t L^q_x([0,1]\times\mathbb{T}^n)} \lesssim C_q N^{\frac{n}{2} - \frac{n+2}{q}} \|\phi_N\|_{L^2_x(\mathbb{T}^n)} \quad \text{if} \quad q < \frac{2(n+2)}{n} \tag{16}$$

For a partial resolution of the conjecture see [4]. We present Bourgain's argument for $n = 2$, $q = 4$ below to show how the rationality of the torus comes into play.

Proof. In this proof we restrict further to the case when $c_i = 1$ for $i = 1, ..., n$. Then

$$\left\| \sum_{|k|\le N} a_n e^{i(\langle x,k\rangle - |n|^2 t)} \right\|^4_{L^4([0,1]\times\mathbb{T}^2)} = \left\| \left[\sum_{|k|\le N} a_k e^{i(\langle x,k\rangle - |k|^2 t)} \right]^2 \right\|^2_{L^2([0,1]\times\mathbb{T}^2)} = \sum_{k,m} |b_{k,m}|^2,$$

where

$$b_{k,m} = \sum_{k=k_1+k_2; m=|k_1|^2+|k_2|^2, |k_i|\le N, i=1,2} a_{k_1} a_{k_2}$$

since

$$\begin{aligned} \left[\sum_{|k|\le N} a_k e^{i(\langle x,k\rangle - |n|^2 t)} \right]^2 &= \sum_{|n_1|\le N, |n_2|\le N} a_{k_1} a_{k_2} e^{i(\langle x,(k_1+k_2)\rangle - (|k_1|^2+|k_2|^2)t)} \\ &= \sum_{k,m} b_{k,m} e^{i(\langle x,k\rangle + mt)}. \end{aligned}$$

---

[7] For us a torus is irrational if there are at least two coordinates for which the ratio of their periods is irrational.

Now it is easy to see that

$$\|S(t)\phi_N\|^4_{L^4_tL^4_x([0,1]\times\mathbb{T}^2)} \sim \sum_{k,m}|b_{k,m}|^2 \lesssim \sup_{|k|\lesssim N, |m|\lesssim N^2} \#M(k,m)\|(a_n)\|^4_{l^2}, \tag{17}$$

where

$$\#M(k,m) = \#\{(k_1\in\mathbb{Z}^2 \ / \ 2m-|k|^2 = |k-2k_1|^2\} = \#\{(z\in\mathbb{Z}^2 \ / \ 2m-|k|^2=|z|^2\}.$$

If $2m-|k|^2<0$ there are no points in $M(k,m)$, and if $R^2 := 2m-|k|^2 \geq 0$, there are at most $\exp C\frac{\log R}{\log\log R}$ many points on the circle of radius R [26], and since $R^2\leq N^2$, using (17) we obtain

$$\|S(t)\phi_N\|_{L^4_tL^4_x([0,1]\times\mathbb{T}^2)} \lesssim N^\epsilon\|\phi_N\|_{L^2}, \tag{18}$$

for all $\epsilon>0$. □

Remark 3.5. *Thanks to a very precise translation invariance in the frequency space for $S(t)u$, estimate* (18) *holds also when the support of ϕ_N is on a ball of radius N centered in an arbitrary point $z_0\in\mathbb{Z}^2$.*

In order to set up a fixed point theorem to prove well-posedness one defines $X^{s,b}$ spaces, introduced in this context by Bourgain [4]. The norms in these spaces are defined for $s,b\in\mathbb{R}$ as:

$$\|u\|_{X^{s,b}(\mathbb{T}^2\times\mathbb{R})} := \Big(\sum_{n\in\mathbb{Z}^2}\int_{\mathbb{R}}|\widetilde{u}(n,\tau)|^2\langle n\rangle^{2s}\langle\tau+|n|^2\rangle^{2b}d\tau\Big)^{\frac{1}{2}},$$

One can immediately see that these spaces are measuring the regularity of a function with respect to certain parabolic coordinates, this to reflect the fact that linear Schrödinger solutions live on parabolas. Having defined the spaces one wants to relate their norms to certain $L^q_tL^p_x$ norms that are typical of Strichartz estimates as proved above in a special case. A key estimate, proved in [4], is

$$\|u\|_{L^4_{t,x}} \lesssim \|u\|_{X^{0+,\frac{1}{2}+}}. \tag{19}$$

This is proved by viewing u as sum of components supported on paraboloids that are at distance one from each other, using (18) on each of them and then reassembling the estimates using the weight $\langle\tau+|n|^2\rangle^{2b}$. An additional estimate is:

$$\|u\|_{L^4_{t,x}} \lesssim \|u\|_{X^{\frac{1}{2}+,\frac{1}{4}+}}. \tag{20}$$

The estimate (20) is a consequence of the following lemma [8].

Lemma 3.6. *Suppose that Q is a ball in $\mathbb{Z}^2$ of radius N and center z_0. Suppose that u satisfies supp $\widehat{u}\subseteq Q$. Then*

$$\|u\|_{L^4_{t,x}} \lesssim N^{\frac{1}{2}}\|u\|_{X^{0,\frac{1}{4}+}}. \tag{21}$$

Lemma 3.6 is proved in [8] by using Hausdorff-Young and Hölder's inequalities. We omit the details. We can now interpolate between (19) and (20) to deduce:

Lemma 3.7. *Suppose that u is as in the assumptions of Lemma 3.6, and suppose that $b_1,s_1\in\mathbb{R}$ satisfy $\frac{1}{4}<b_1<\frac{1}{2}+, s_1>1-2b_1$. Then*

$$\|u\|_{L^4_{t,x}} \lesssim N^{s_1}\|u\|_{X^{0,b_1}}. \tag{22}$$

Lemma 3.7 can then be used to prove local well-posedness for the cubic NLS in $\mathbb{T}^2$ in H^s, $s > 0$. One in fact can set up a fixed point argument in the space $X^{s,b}$, $s > 0$, $b \sim \frac{1}{2}$. The key point is that the problem at hand has a cubic nonlinearity which by duality forces us to consider a product of four functions in L^1. This translates into estimating L^4 norms which via (22) are related back to the space $X^{s,b}$. In the proof one shows that the interval of time $[-T, T]$ suitable for a fixed point argument is such that

$$T \sim \|u_0\|_{H^s}^{-\alpha}, \tag{23}$$

for some $\alpha > 0$. As a consequence, the defocusing, cubic, periodic NLS problem (1) can be proved to be globally well-posed in $H^s, s \geq 1$ thanks to (23) and the conservation of the Hamiltonian (2). See [4, 8].

4. Growth of Sobolev norms and energy transfer to high frequencies

We consider the cubic, defocusing, periodic (rational) NLS initial value problem:

$$\begin{cases} iu_t + \Delta u = |u|^2 u, \ x \in \mathbb{T}^2 \\ u|_{t=0} = u_0 \in H^s(\mathbb{T}^2), \ s > 1. \end{cases} \tag{24}$$

From Section 3 we know that (24) is globally well-posed in $H^s, s \geq 1$. Hence, it makes sense to analyze the behavior of $\|u(t)\|_{H^s}$. But as we will discuss later this estimate is related to an important physical phenomenon: energy transfer to higher modes or forward cascade. We will elaborate more on this below.

Theorem 4.1 (Bound for the defocusing cubic NLS on $\mathbb{T}^2$ [42, 51])**.** *Let u be the global solution of* (24) *on $\mathbb{T}^2$. Then, there exists a function $C = C_{s,\|u_0\|_{H^1}}$ such that for all $t \in \mathbb{R}$:*

$$\|u(t)\|_{H^s(\mathbb{T}^2)} \leq C(1 + |t|)^{s+} \|u_0\|_{H^s(\mathbb{T}^2)}. \tag{25}$$

See also [8, 17].

Remark 4.2. *Let us note that, if we consider the spatial domain to be $\mathbb{R}^2$, one can obtain uniform bounds on $\|u(t)\|_{H^s}$ for solutions $u(t)$ of the defocusing cubic NLS by the recent scattering and highly non trivial result of Dodson* [23]*.*

The growth of high Sobolev norms has a physical interpretation in the context of the *low-to-high frequency cascade.* In other words, we see that $\|u(t)\|_{H^s}$ weighs the higher frequencies more as s becomes larger, and hence its growth gives us a quantitative estimate for how much of the support of $|\widehat{u}|^2$ has transferred from the low to the high frequencies. This sort of problem also goes under the name of *weak turbulence* [1, 49]. By local well-posedness theory discussed in Section 3, one can show that there exist $C, \tau_0 > 0$, depending only on the initial data u_0 such that for all t:

$$\|u(t + \tau_0)\|_{H^s} \leq C \|u(t)\|_{H^s}. \tag{26}$$

Iterating (26) yields the exponential bound:

$$\|u(t)\|_{H^s} \leq C_1 e^{C_2 |t|}, \tag{27}$$

where $C_1, C_2 > 0$ again depend only on u_0.

For a wide class of nonlinear dispersive equations, the analogue of (27) can be improved to a polynomial bound, as long as we take $s \in \mathbb{N}$, or if we consider sufficiently smooth initial data. This observation was first made in the work of Bourgain [7], and was continued in the work of Staffilani [43, 44].

The crucial step in the mentioned works was to improve the iteration bound (26) to:

$$\|u(t+\tau_0)\|_{H^s} \leq \|u(t)\|_{H^s} + C\|u(t)\|_{H^s}^{1-r}. \tag{28}$$

As before, $C, \tau_0 > 0$ depend only on u_0. In this bound, $r \in (0,1)$ satisfies $r \sim \frac{1}{s}$. One can show that (28) implies that for all $t \in \mathbb{R}$:

$$\|u(t)\|_{H^s} \leq C(1+|t|)^{\frac{1}{r}}. \tag{29}$$

In [7], (28) was obtained by using the *Fourier multiplier method*. In [43, 44], the iteration bound was obtained by using multilinear estimates in $X^{s,b}$-spaces due to Kenig-Ponce-Vega [31]. A slightly different approach, based on the analysis in the work of Burq-Gérard-Tzvetkov [11], is used to obtain (28) in the context of compact Riemannian manifolds in the work of Catoire-Wang [16], and Zhong [51].

The main idea in the proof of Theorem 4.1 in [42] is to introduce $\mathcal{D}$, an *upside-down I-operator*. This operator is defined as a Fourier multiplier operator. By construction, one is able to relate $\|u(t)\|_{H^s}$ to $\|\mathcal{D}u(t)\|_{L^2}$ and to consider the growth of the latter quantity. Following the ideas of the construction of the standard *I-operator*, as defined by Colliander, Keel, Staffilani, Takaoka, and Tao [18, 19, 20], the goal is to show that the quantity $\|\mathcal{D}u(t)\|_{L^2}^2$ is *slowly varying*. This is done by applying a Littlewood-Paley decomposition and summing an appropriate geometric series. A similar technique was applied in the low-regularity context in [19]. This first step though is not enough to prove Theorem 4.1. Instead one has to use *higher modified energies*, i.e. quantities obtained from $\|\mathcal{D}u(t)\|_{L^2}^2$ by adding an appropriate multilinear correction, again an idea introduced in [18, 19, 20]. In this way one obtains $E^2(u(t)) \sim \|\mathcal{D}u(t)\|_{L^2}^2$, which is even more slowly varying. Due to a complicated resonance phenomenon in two dimensions, the construction of E^2 is very involved and we do not present the details here.

4.1. **Example of energy transfer to high frequencies.** In this subsection we show that a very weak growth of Sobolev norms may indeed occur. More precisely we can prove

Theorem 4.3. *[Colliander-Keel-Staffilani-Takaoka-Tao,* [22]*] Let* $s > 1$, $K \gg 1$ *and* $0 < \sigma < 1$ *be given. Then there exist a global smooth solution* $u(x,t)$ *to the IVP* (24) *and* $T > 0$ *such that*

$$\|u_0\|_{H^s} \leq \sigma \quad \textit{and} \quad \|u(T)\|_{\dot{H}^s}^2 \geq K.$$

We start by listing the elements of the proof. The first is a reduction to a resonant problem that we will refer to as the RFNLS system, see (32). Then in Subsection 4.2 we introduce a special finite set Λ of frequencies and we reduce the RFNLS system to a finite-dimensional Toy Model ODE system, see (33). We study this Toy Model dynamically in Subsection 4.3 and we show some sort of "sliding property" for it, see Theorem 4.4. In Subsection 4.4 we introduce the approximation Lemma 4.5 together with a scaling argument and finally in Subsection 4.5 we sketch the proof of Theorem 4.3.

We consider the gauge transformation

$$v(t,x) = e^{-i2Gt}u(t,x),$$

for $G \in \mathbb{R}$. If u solves the NLS (24) above, then v solves the equation

$$(-i\partial_t + \Delta)v = (2G + v)|v|^2.$$

We make the ansatz

$$v(t,x) = \sum_{n \in \mathbb{Z}^2} a_n(t) e^{i(\langle n,x\rangle + |n|^2 t)}.$$

Now the dynamics is all recast trough $a_n(t)$:

$$i\partial_t a_n = 2Ga_n + \sum_{n_1-n_2+n_3=n} a_{n_1}\overline{a_{n_2}}a_{n_3}e^{i\omega_4 t}, \tag{30}$$

where $\omega_4 = |n_1|^2 - |n_2|^2 + |n_3|^2 - |n|^2$. By choosing

$$G = -\|v(t)\|_{L^2}^2 = -\sum_k |a_k(t)|^2$$

which is constant from the conservation of the mass, one can rewrite equation (30) as

$$i\partial_t a_n = -a_n|a_n|^2 + \sum_{n_1,n_2,n_3\in\Gamma(n)} a_{n_1}\overline{a_{n_2}}a_{n_3}e^{i\omega_4 t}, \tag{31}$$

where

$$\Gamma(n) = \{n_1, n_2, n_3 \in \mathbb{Z}^2 \;/\; n_1 - n_2 + n_3 = n; n_1 \neq n; n_3 \neq n\}.$$

From now on we will be referring to this system as the $FNLS$ system, with the obvious connection with the original NLS equation.

We define the resonant set

$$\Gamma_{res}(n) = \{n_1, n_2, n_3 \in \Gamma(n) \,/\, \omega_4 = 0\}.$$

The geometric interpretation for this set is as follows: If n_1, n_2, n_3 are in $\Gamma_{res}(n)$, then the four points (n_1, n_2, n_3, n) represent the vertices of a rectangle in $\mathbb{Z}^2$. We finally define the resonant truncation $RFNLS$ to be the system

$$-i\partial_t b_n = -b_n|b_n|^2 + \sum_{n_1,n_2,n_3\in\Gamma_{res}(n)} b_{n_1}\overline{b}_{n_2}b_{n_3}. \tag{32}$$

We now would like to restrict the dynamics to a finite set of frequencies and this set would need several important properties. The first one is closeness under resonance. A finite set $\Lambda \subset \mathbb{Z}^2$ is closed under resonant interactions if

$$n_1, n_2, n_3 \in \Gamma_{res}(n), n_1, n_2, n_3 \in \Lambda \implies n = n_1 - n_2 + n_3 \in \Lambda.$$

Hence a Λ-finite dimensional resonant truncation of $RFNLS$ is

$$-i\partial_t b_n = -b_n|b_n|^2 + \sum_{(n_1,n_2,n_3)\in\Gamma_{res}(n)\cap\Lambda^3} b_{n_1}\overline{b}_{n_2}b_{n_3}. \tag{33}$$

We will refer to this systems as the $RFNLS_\Lambda$ system.

4.2. **Λ: a very special set of frequencies.** We can construct a special Λ of frequencies with the following properties [22]

- **Generational set up:** $\Lambda = \Lambda_1 \cup \cdots \cup \Lambda_N$, N to be fixed later. A nuclear family is a rectangle (n_1, n_2, n_3, n_4) where the frequencies n_1, n_3 (the 'parents') live in generation Λ_j and n_2, n_4 ('children') live in generation Λ_{j+1}.
- **Existence and uniqueness of spouse and children:** $\forall\; 1 \leq j < N$ and $\forall\; n_1 \in \Lambda_j$ $\exists$ unique nuclear family such that $n_1, n_3 \in \Lambda_j$ are parents and $n_2, n_4 \in \Lambda_{j+1}$ are children.
- **Existence and uniqueness of siblings and parents:** $\forall\; 1 \leq j < N$ and $\forall\; n_2 \in \Lambda_{j+1}$ $\exists$ unique nuclear family such that $n_2, n_4 \in \Lambda_{j+1}$ are children and $n_1, n_3 \in \Lambda_j$ are parents.
- **Non degeneracy:** The sibling of a frequency is never its spouse.
- **Faithfulness:** Besides nuclear families, Λ contains no other rectangles.

- **Intergenerational Equality:** The function $n \longmapsto a_n(0)$ is constant on each generation Λ_j.
- **Multiplicative Structure:** If $N = N(\sigma, K)$ is large enough then Λ consists of $N \times 2^{N-1}$ disjoint frequencies n with $|n| > R = R(\sigma, K)$, the first frequency in Λ_1 is of size R and we call R the *inner radius* of Λ. Moreover for any $n \in \Lambda$, $|n| \leq C(N)R$.
- **Wide Spreading:** Given $\sigma \ll 1$ and $K \gg 1$, if N is large enough then $\Lambda = \Lambda_1 \cup .... \cup \Lambda_N$ as above and

$$\sum_{n\in\Lambda_N} |n|^{2s} \geq \frac{K^2}{\sigma^2} \sum_{n\in\Lambda_1} |n|^{2s}. \tag{34}$$

4.3. **The Toy Model.** The intergenerational equality hypothesis (that the function $n \longmapsto b_n(0)$ is constant on each generation Λ_j) persists under $RFNLS_\Lambda$ (33):

$$\forall\ m, n \in \Lambda_j,\ \ b_n(t) = b_m(t).$$

Also $RFNLS_\Lambda$ may be reindexed by generation index j and the recast dynamics is the Toy Model:

$$-i\partial_t b_j(t) = -b_j(t)|b_j(t)|^2 - 2b_{j-1}(t)^2\overline{b_j(t)} - 2b_{j+1}(t)^2\overline{b_j(t)}, \tag{35}$$

with boundary condition

$$b_0(t) = b_{N+1}(t) = 0. \tag{36}$$

Using direct calculation[8], we will prove[9] that our Toy Model evolution $b_j(0) \longmapsto b_j(t)$ is such that:

$$\begin{array}{rcl} (b_1(0), b_2(0), \ldots, b_N(0)) & \sim & (1, 0, \ldots, 0) \\ (b_1(t_2), b_2(t_2), \ldots, b_N(t_2)) & \sim & (0, 1, \ldots, 0) \\ & \cdot & \\ (b_1(t_N), b_2(t_N), \ldots, b_N(t_N)) & \sim & (0, 0, \ldots, 1) \end{array}$$

that is the bulk of conserved mass is transferred from Λ_1 to Λ_N and the weak transfer of energy from lower to higher frequencies follows from the *Wide Spreading* property (34) of Λ listed above.

We now make few observations that are simple, but they are nevertheless meant to show how nontrivial it is to move from Λ_1 to Λ_N. Global well-posedness for the Toy Model (35) is not an issue. Then we define

$$\Sigma = \{x \in \mathbb{C}^N\ /\ |x|^2 = 1\}\ \text{ and the flow map }\ W(t) : \Sigma \to \Sigma,$$

where $W(t)b(t_0) = b(t + t_0)$ for any solution $b(t)$ of (35). It is easy to see that for any $b(t)$ with $b(0) \in \Sigma$

$$\partial_t |b_j|^2 = 4\mathrm{Re}(i\bar{b}_j^{\,2}(b_{j-1}^2 + b_{j+1}^2)) \leq 4|b_j|^2.$$

So if

$$b_j(0) = 0 \implies b_j(t) = 0, \quad \text{for all }\ t \in [0, T]$$

and if we define the torus

$$\mathbb{T}_j = \{(b_1, ..., b_N) \in \Sigma\,/\,|b_j| = 1,\, b_k = 0,\, k \neq j\}$$

then

$$W(t)\mathbb{T}_j = \mathbb{T}_j \ \text{ for all } j = 1, ..., N$$

---

[8] Maybe dynamical systems methods are useful here?

[9] See Theorem 4.4.

hence $\mathbb{T}_j$ is invariant. This suggests that if we want to move from a torus $\mathbb{T}_j$ to $\mathbb{T}_i$ we cannot start from data on $\mathbb{T}_j$ and moreover we need to show that we can manage to avoid to hit any $\mathbb{T}_k$, $j < k < i$. This is in fact the content of the following instability-type theorem:

Theorem 4.4. *[Sliding Theorem] Let $N \geq 6$. Given $\epsilon > 0$ there exist x_3 within ϵ of $\mathbb{T}_3$ and x_{N-2} within ϵ of $\mathbb{T}_{N-2}$ and a time τ such that*

$$W(\tau)x_3 = x_{N-2}.$$

What the theorem says is that $W(t)x_3$ is a solution of total mass 1 arbitrarily concentrated near mode $j = 3$ at some time 0 and then gets moved so that it is concentrated near mode $j = N - 2$ at later time τ.

For the complete, and unfortunately lengthy proof of this theorem see [22]. Here we only give a motivation for it that should clarify the dynamics involved. Let us first observe that when $N = 2$ we can easily demonstrate that there is an orbit connecting $\mathbb{T}_1$ to $\mathbb{T}_2$. Indeed in this case we have the explicit "slider" solution

$$b_1(t) := \frac{e^{-it}\omega}{\sqrt{1 + e^{2\sqrt{3}t}}}; \quad b_2(t) := \frac{e^{-it}\omega^2}{\sqrt{1 + e^{-2\sqrt{3}t}}} \tag{37}$$

where $\omega := e^{2\pi i/3}$ is a cube root of unity.

This solution approaches $\mathbb{T}_1$ exponentially fast as $t \to -\infty$, and approaches $\mathbb{T}_2$ exponentially fast as $t \to +\infty$. One can translate this solution in the j parameter, and obtain solutions that "slide" from $\mathbb{T}_j$ to $\mathbb{T}_{j+1}$. Intuitively, the proof of the Sliding Theorem for higher N should then proceed by concatenating these slider solutions....This though cannot work directly because each solution requires an infinite amount of time to connect one hoop to the next. It turned out though that a suitably perturbed or "fuzzy" version of these slider solutions can in fact be glued together.

4.4. **The Approximation lemma and the scaling argument.** There are still two steps we need to complete to prove Theorem 4.3. The first is to show that a solution to the Toy Model (35) is a good approximation for the solution to the original problem (31). This is accomplished with the following approximation lemma.

Lemma 4.5. *[Approximation Lemma] Let $\Lambda \subset \mathbb{Z}^2$ introduced above. Let $B \gg 1$ and $\delta > 0$ small and fixed. Let $t \in [0, T]$ and $T \sim B^2 \log B$. Suppose there exists $b(t) \in l^1(\Lambda)$ solving $RFNLS_\Lambda$ such that*

$$\|b(t)\|_{l^1} \lesssim B^{-1}. \tag{38}$$

Then there exists a solution $a(t) \in l^1(\mathbb{Z}^2)$ of FNLS (31) such that for any $t \in [0, T]$

$$a(0) = b(0), \quad \text{and} \quad \|a(t) - b(t)\|_{l^1(\mathbb{Z}^2)} \lesssim B^{-1-\delta}.$$

The proof for this lemma is pretty standard. The main idea is to check that the "non resonant" part of the nonlinearity remains small enough, see [22] for details.

The last ingredient before we proceed to the proof of our main result is *the scaling argument*. What we proved so far is that we can find a solution of mass one that a time zero is localized in Λ_3 and if we wait long enough will be localized in Λ_{N-2}. But what Theorem 4.3 asks is a solution that is "small" at time zero. This is why we need to introduce scaling. It is easy to check that if $b(t)$ solves $RFNLS_\Lambda$ (33) then the rescaled solution

$$b^\lambda(t) = \lambda^{-1} b\left(\frac{t}{\lambda^2}\right)$$

solves the same system with datum $b_0^\lambda = \lambda^{-1} b_0$.

We then pick the complex vector $b(0)$ that was found in the Sliding Theorem 4.4 above. For simplicity let us assume here that $b_j(0) = 1-\epsilon$ if $j=3$ and $b_j(0)=\epsilon$ if $j \neq 3$ and then we fix

$$a_n(0) = \begin{cases} b_j^\lambda(0) & \text{for any } \ n \in \Lambda_j \\ 0 & \text{otherwise .} \end{cases} \tag{39}$$

We are now ready to finish the proof of Theorem 4.3. For simplicity we recast it with all the notations and reductions introduced so far:

Theorem 4.6. *For any $0<\sigma \ll 1$ and $K \gg 1$ there exists a complex sequence (a_n) such that*

$$\left(\sum_{n\in\mathbb{Z}^2} |a_n|^2 |n|^{2s} \right)^{1/2} \lesssim \sigma$$

and a solution $(a_n(t))$ of FNLS and $T>0$ such that

$$\left(\sum_{n\in\mathbb{Z}^2} |a_n(T)|^2 |n|^{2s} \right)^{1/2} > K.$$

4.5. **Proof of Theorem 4.6.** We start by estimating the size of $(a_n(0))$. By definition

$$\left(\sum_{n\in\Lambda} |a_n(0)|^2 |n|^{2s} \right)^{1/2} = \frac{1}{\lambda} \left(\sum_{j=1}^{M} |b_j(0)|^2 \left(\sum_{n\in\Lambda_j} |n|^{2s} \right) \right)^{1/2} \sim \frac{1}{\lambda} Q_3^{1/2},$$

where the last equality follows from defining

$$\sum_{n\in\Lambda_j} |n|^{2s} = Q_j,$$

and the definition of $a_n(0)$ given in (39). At this point we use the proprieties of the set Λ to estimate $Q_3 \sim C(N)R^{2s}$, where R is the inner radius of Λ. We then conclude that

$$\left(\sum_{n\in\Lambda} |a_n(0)|^2 |n|^{2s} \right)^{1/2} = \lambda^{-1} C(N) R^s \sim \sigma,$$

for a large enough R.

Now we want to estimate the size of $(a_n(T))$. Take $B=\lambda$ and $T=\lambda^2\tau$ in Lemma 4.5 and write

$$\|a(T)\|_{H^s} \geq \|b^\lambda(T)\|_{H^s} - \|a(T) - b^\lambda(T)\|_{H^s} = I_1 - I_2.$$

We want $I_2 \ll 1$ and $I_1 > K$. For I_2 we use the Approximation Lemma 4.5

$$I_2 \lesssim \lambda^{-1-\delta} \left(\sum_{n\in\Lambda} |n|^{2s} \right)^{1/2} \lesssim \lambda^{-1-\delta} C(N) R^s.$$

At this point we need to pick λ and N so that

$$\|a(0)\|_{H^s} = \lambda^{-1} C(N) R^s \sim \sigma \ \text{ and } \ I_2 \lesssim \lambda^{-1-\delta} C(N) R^s \ll 1$$

and thanks to the presence of $\delta > 0$ this can be achieved by taking λ and R large enough.

Finally we estimating I_1. It is important here that at time zero one starts with a fixed non zero datum, namely $\|a(0)\|_{H^s} = \|b^\lambda(0)\|_{H^s} \sim \sigma > 0$. In fact we will show that

$$I_1^2 = \|b^\lambda(T)\|_{H^s}^2 \geq \frac{K^2}{\sigma^2}\|b^\lambda(0)\|_{H^s}^2 \sim K^2.$$

If we define for $T = \lambda^2 t$

$$R = \frac{\sum_{n\in\Lambda} |b_n^\lambda(\lambda^2 t)|^2 |n|^{2s}}{\sum_{n\in\Lambda} |b_n^\lambda(0)|^2 |n|^{2s}},$$

then we are reduce to showing that $R \gtrsim K^2/\sigma^2$. Recall the notation

$$\Lambda = \Lambda_1 \cup ..... \cup \Lambda_N \quad \text{and} \quad \sum_{n\in\Lambda_j} |n|^{2s} = Q_j.$$

Using the fact that by the Sliding Theorem 4.4 one obtains $b_j(T) = 1-\epsilon$ if $j = N-2$ and $b_j(T) = \epsilon$ if $j \neq N-2$, it follows that

$$\begin{aligned} R &= \frac{\sum_{i=1}^N \sum_{n\in\Lambda_i} |b_i^\lambda(\lambda^2 t)|^2 |n|^{2s}}{\sum_{i=1}^N \sum_{n\in\Lambda_i} |b_i^\lambda(0)|^2 |n|^{2s}} \\ &\geq \frac{Q_{N-2}(1-\epsilon)}{(1-\epsilon)Q_3 + \epsilon Q_1 + ... + \epsilon Q_N} \sim \frac{Q_{N-2}(1-\epsilon)}{Q_{N-2}\left[(1-\epsilon)\frac{Q_3}{Q_{N-2}} + ... + \epsilon\right]} \\ &\gtrsim \frac{(1-\epsilon)}{(1-\epsilon)\frac{Q_3}{Q_{N-2}}} = \frac{Q_{N-2}}{Q_3} \end{aligned}$$

and the conclusion follows from the "Wide Spreading" property (34) of Λ_j:

$$Q_{N-2} = \sum_{n\in\Lambda_{N-2}} |n|^{2s} \gtrsim \frac{K^2}{\sigma^2} \sum_{n\in\Lambda_3} |n|^{2s} = \frac{K^2}{\sigma^2} Q_3.$$

5. Periodic Schrödinger equations as infinite dimension Hamiltonian systems

In this section we are going to view some Schrödinger equations as infinite dimension Hamiltonian systems. We will show two results generalizing to the infinite dimensional setting two important concepts such as the invariance of the Gibbs measure and the non-squeezing lemma of Gromov [24]. In the next subsection we recall these two concepts in more details.

5.1. **The finite dimension case.**

Hamilton's equations of motion have the antisymmetric form

$$\dot{q}_i = \frac{\partial H(p,q)}{\partial p_i}, \qquad \dot{p}_i = -\frac{\partial H(p,q)}{\partial q_i} \tag{40}$$

the Hamiltonian $H(p,q)$ being a first integral:

$$\frac{dH}{dt} := \sum_i \frac{\partial H}{\partial q_i}\dot{q}_i + \frac{\partial H}{\partial p_i}\dot{p}_i = \sum_i \frac{\partial H}{\partial q_i}\frac{\partial H}{\partial p_i} + \frac{\partial H}{\partial p_i}(-\frac{\partial H}{\partial q_i}) = 0.$$

By defining $y := (q_1, \dots, q_k, p_1, \dots, p_k)^T \in \mathbb{R}^{2k}$ $(2k = d)$ we can rewrite (40) in the compact form

$$\frac{dy}{dt} = J\nabla H(y), \qquad J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}.$$

We now recall the following theorem giving a sufficient condition under which a flow map preserves the volume:

Theorem 5.1. *[Liouville's Theorem] Let a vector field $f : \mathbb{R}^d \to \mathbb{R}^d$ be divergence free. If the flow map Φ_t satisfies*

$$\frac{d}{dt}\Phi_t(y) \; = \; f(\Phi_t(y)),$$

then Φ_t is a volume preserving map for all t.

In particular if f is associated to a Hamiltonian system then automatically $div\, f = 0$. As a consequence the Lebesgue measure on $\mathbb{R}^{2k}$ is invariant under the Hamiltonian flow of (40).

There are other measures that remain invariant[10] under the Hamiltonian flow: the Gibbs measures. In fact we have

Theorem 5.2. *[Invariance of Gibbs measures] Assume that Φ_t is the flow generated by the Hamiltonian system* (40)*. Then the Gibbs measures*

$$d\mu := e^{-\beta H(p,q)} \prod_{i=1}^{d} dp_i \, dq_i$$

with $\beta > 0$, are invariant under the flow Φ_t.

The proof is trivial since from conservation of the Hamiltonian H the functions $e^{-\beta H(p,q)}$ remain constant, while, thanks to the Liouville's Theorem 5.1 the volume $\prod_{i=1}^{d} dp_i \, dq_i$ remains invariant as well.

Next result, much more difficult to prove, is the non-squeezing theorem:

Theorem 5.3 (Non-squeezing [24])**.** *Assume that Φ_t is the flow generated by the Hamiltonian system* (40)*. Fix $y_0 \in \mathbb{R}^{2k}$ and let $B_r(y_0)$ be the ball in $\mathbb{R}^{2k}$ centered at y_0 and radius r. If $C_R(z_0) := \{y = (q_1, \dots, q_k, p_1, \dots, p_k) \in \mathbb{R}^{2k} / |q_i - z_0| \leq R\}$, a cylinder of radius R, and $\Phi_t(B_r(y_0)) \subset C_R(z_0)$, it must be that $r \leq R$.*

We now would like to see if Theorem 5.2 and Theorem 5.3 can be generalized to an infinite dimensional setting.

5.2. **Periodic Schrödinger equations and Gibbs measures.** Let us go back to (1). One can use $H(u, \bar{u})$ and check that equation (1) is equivalent to

$$\dot{u} = i\frac{\partial H(u, \bar{u})}{\partial \bar{u}}$$

where $H(t)$ is the Hamiltonian defined in (2), and one can think of u as the infinite dimension vector given by its Fourier coefficients $(\hat{u}(k))_{k\in\mathbb{Z}^n} = (a_k, b_k)_{k\in\mathbb{Z}^n}$.

Lebowitz, Rose and Speer [34] considered the Gibbs measure *formally* given by

$$\text{“}d\mu = Z^{-1} \exp\left(-\beta H(u)\right) \prod_{x\in\mathbb{T}} du(x)\text{”} \tag{41}$$

and showed that μ is a well-defined probability measure on $H^s(\mathbb{T})$ for any $s < \frac{1}{2}$, see Remark 5.6.

[10] A measure μ remains invariant under a flow Φ_t if for any A, subset of the support of μ, one has

$$\mu(\Phi_t(A)) = \mu(A).$$

Bourgain [5] proved the invariance of this measure and almost surely global well-posedness of the associated initial value problem[11]. For example, for $p = 4$ in (1) he proved:

Theorem 5.4. *Consider the NLS initial value problem*

$$\begin{cases} (i\partial_t + \Delta)u = \lambda |u|^4 u \\ u(0,x) = u_0(x), \quad \text{where } x \in \mathbb{T}. \end{cases} \tag{42}$$

If $\lambda = 1$ (defocusing case) the measure μ (41) is well defined in H^s, $0 < s < 1/2$ and there exists $\Omega \subset H^s$ such that $\mu(\Omega) = 1$ and (42) is globally well-posed in Ω. Moreover the measure μ is invariant under the flow given by (42). If $\lambda = -1$ (focusing case), then a similar result holds for

$$d\mu = Z^{-1} \chi_{\{\|u\|_L^2 \le B\}} \exp\left(-\beta H(u)\right) \prod_{x \in \mathbb{T}} du(x)$$

with B small enough.

Remark 5.5. *If one considers the IVP (1) in the focusing case, then Theorem 5.4 only holds for $p \le 5$, but if $p < 5$ we can take $B > 0$ arbitrary, see* [5].

After Bourgain's result recalled above, almost surely global well-posedness for a variety of IVP has been studied by introducing invariant measures. See for example Burq and Tzevtkov for subcubic and subquartic radial NLW on 3D balls [12, 13], T. Oh for the periodic KdV-type and Schrödinger-Benjamin-Ono coupled systems [35, 36, 37], Oh-Nahmod-Rey-Bellet-Staffilani [38] and Thomann and Tzevtkov [48] for the periodic derivative NLS equation. This last one will be the subject of Subsection 5.5.

5.3. **Gaussian measures and Gibbs measures.** In this subsection I would like to elaborate a little more on Gaussian and Gibbs measures by using as an example the measure that is naturally attached to the IVP (42) above. Note that the quantity

$$H(u) + \frac{1}{2} \int |u|^2(x)\, dx$$

is conserved. Then the best way to make sense of the Gibbs measure μ formally defined in (41) is by writing it as

$$d\mu = Z^{-1} \chi_{\|u\|_{L^2} \le B} \exp\left(\frac{1}{6} \int |u|^6\, dx\right) \exp\left(-\frac{1}{2} \int (|u_x|^2 + |u|^2)\, dx\right) \prod_{x \in \mathbb{T}} du(x).$$

In this expression

$$d\rho = \exp\left(-\frac{1}{2} \int (|u_x|^2 + |u|^2)\, dx\right) \prod_{x \in \mathbb{T}} du(x)$$

is the Gaussian measure and

$$\frac{d\mu}{d\rho} = \chi_{\|u\|_{L^2} \le B} \exp\left(\frac{1}{6} \int |u|^6\, dx\right),$$

corresponding to the nonlinear term of the Hamiltonian, is understood as the Radon-Nikodym derivative of μ with respect to ρ.

[11] The remarkable fact is that this statement is true both in the focusing and defocusing case, modulo of course the restriction on the L^2 norm in Remark 5.5.

The Gaussian measure ρ is defined as the weak limit of the finite dimensional Gaussian measures

$$d\rho_N = Z_{0,N}^{-1} \exp\Big(- \frac{1}{2} \sum_{|n|\le N} (1+|n|^2)|\widehat{u}_n|^2\Big) \prod_{|n|\le N} da_n db_n \, .$$

For a precise definition of Gaussian measures on Hilbert and Banach spaces in general see [25, 33]. Here we briefly recall how one shows that Sobolev spaces $H^s(\mathbb{T})$ are supports for ρ only if $s<\frac{1}{2}$. Consider the operator $\mathcal{J}_s=(1-\Delta)^{s-1}$. Then

$$\sum_n (1+|n|^2)\,|\widehat{u}_n|^2 \;=\; \langle u,u\rangle_{H^1} \;=\; \langle \mathcal{J}_s^{-1} u, u\rangle_{H^s} \, .$$

The operator $\mathcal{J}_s : H^s \to H^s$ has the set of eigenvalues $\{(1+|n|^2)^{(s-1)}\}_{n\in\mathbb{Z}}$ and the corresponding eigenvectors $\{(1+|n|^2)^{-s/2}e^{inx}\}_{n\in\mathbb{Z}}$ form an orthonormal basis of H^s. For ρ to be *countable additive* we need $\mathcal{J}_s$ to be of *trace class* which is true if and only if $s<\frac{1}{2}$. Then ρ is a countably additive measure on H^s for any $s<\frac{1}{2}$. See again [25, 33].

The following remark is meant to explain the probabilistic aspect of Gibbs measures.

Remark 5.6. *The measure ρ_N above can be regarded as the induced probability measure on $\mathbb{R}^{4N+2}$ under the map*

$$\omega \longmapsto \left\{ \frac{g_n(\omega)}{\sqrt{1+|n|^2}} \right\}_{|n|\le N} \qquad \textit{and} \quad \widehat{u}_n = \frac{g_n}{\sqrt{1+|n|^2}},$$

where $\{g_n(\omega)\}_{|n|\le N}$ are independent standard complex Gaussian random variables on a probability space $(\Omega,\mathcal{F},P)$.

In a similar manner, we can view ρ as the induced probability measure under the map

$$\omega \mapsto \left\{ \frac{g_n(\omega)}{\sqrt{1+|n|^2}} \right\}_{n\in\mathbb{Z}} .$$

5.4. **Bourgain's Method.** Above we stated Bourgain's theorem for the quintic focusing periodic NLS. Here we give an outline of Bourgain's idea in a general framework, and discuss how to prove almost surely global well-posedness and the invariance of a measure starting with a local well-posedness result.

Consider a dispersive nonlinear Hamiltonian PDE with a k-linear nonlinearity, possibly with derivative:

$$\begin{cases} u_t = \mathcal{L}u + \mathcal{N}(u) \\ u|_{t=0} = u_0, \end{cases} \tag{43}$$

where $\mathcal{L}$ is a (spatial) differential operator like $i\partial_{xx}$, ∂_{xxx}, etc. Let $H(u)$ denote the Hamiltonian of (43). Then (43) can also be written as

$$u_t = J\,\frac{dH}{du} \quad \text{if } u \text{ is real-valued,} \quad u_t = J\,\frac{\partial H}{\partial \overline{u}} \quad \text{if } u \text{ is complex-valued,}$$

for an appropriate operator J. Let μ denote a measure on the distributions on $\mathbb{T}$, whose invariance we would like to establish. We assume that μ is (formally) given by

$$\text{“}\, d\mu = Z^{-1} e^{-F(u)} \prod_{x\in\mathbb{T}} du(x)\,\text{”},$$

where $F(u)$ is conserved[12] under the flow of (43) and the leading term of $F(u)$ is quadratic and nonnegative. Now, suppose that there is a good local well-posedness theory, that is there exists a Banach space $\mathcal{B}$ of distributions on $\mathbb{T}$ and a space $X_\delta \subset C([-\delta,\delta];\mathcal{B})$ of space-time distributions in which one proves local well-posedness by a fixed point argument with a time of existence δ depending on $\|u_0\|_B$, say $\delta \sim \|u_0\|_B^{-\alpha}$ for some $\alpha > 0$. In addition, suppose that the Dirichlet projections P_N – the projection onto the spatial frequencies $\leq N$ – act boundedly on these spaces, uniformly in N. Then for $\|u_0\|_B \leq K$ the finite dimensional approximation to (43)

$$\begin{cases} u_t^N = \mathcal{L}u^N + P_N\big(\mathcal{N}(u^N)\big) \\ u^N|_{t=0} = u_0^N := P_N u_0(x) = \sum_{|n|\leq N} \widehat{u}_0(n)e^{inx}. \end{cases} \tag{44}$$

is locally well-posed on $[-\delta,\delta]$ with $\delta \sim K^{-\alpha}$, independent of N. We need two more important assumptions on (44): that (44) is Hamiltonian with $H(u^N)$ i.e.

$$u_t^N = J\,\frac{dH(u^N)}{d\overline{u}^N} \tag{45}$$

and that

$$\frac{d}{dt}F(u^N(t)) = 0, \tag{46}$$

that is $F(u^N)$ is still conserved under the flow of (44).

Note that the first holds for example when J commutes with the projection P_N, (e.g. $J = i$ or ∂_x.). In general however the two assumptions above are not guaranteed and may not necessarily hold. See Subsection 5.5.

At this point we have:

- By Liouville's theorem and (45) the Lebesgue measure $\prod_{|n|\leq N} da_n db_n$, where $\widehat{u^N}(n) = a_n + ib_n$, is invariant under the flow of (44).
- Using (46) - the conservation of $F(u^N)$- the finite dimensional version μ_N of μ:
$$d\mu_N = Z_N^{-1}e^{-F(u^N)}\prod_{|n|\leq N} da_n db_n$$
is also invariant under the flow of (44).

The next ingredient we need is:

Lemma 5.7 (Fernique-type tail estimate)**.** *For K sufficiently large, we have*
$$\mu_N\big(\{\|u_0^N\|_{\mathcal{B}} > K\}\big) < Ce^{-CK^2},$$
where all constants are independent of N.

This lemma and the invariance of μ_N imply the following estimate controlling the growth of the solution u^N to (44) [5].

Proposition 5.8. *Given $T < \infty$, $\varepsilon > 0$, there exists $\Omega_N \subset \mathcal{B}$ such that $\mu_N(\Omega_N^c) < \varepsilon$ and for $u_0^N \in \Omega_N$, (44) is well-posed on $[-T,T]$ with the growth estimate:*
$$\|u^N(t)\|_{\mathcal{B}} \lesssim \Big(\log\frac{T}{\varepsilon}\Big)^{\frac{1}{2}}, \ \textit{for } |t| \leq T.$$

[12] $F(u)$ could be the Hamiltonian, but not necessarily!

Proof. Let $\Phi_N(t)$ be the flow map of (44), and define

$$\Omega_N = \cap_{j=-[T/\delta]}^{[T/\delta]} \Phi_N(j\delta)(\{\|u_0^N\|_{\mathcal{B}} \leq K\}).$$

By invariance of μ_N,

$$\mu(\Omega_N^c) = \sum_{j=-[T/\delta]}^{[T/\delta]} \mu_N\left(\Phi_N(j\delta)(\{\|u_0^N\|_{\mathcal{B}} > K\}\right) = 2[T/\delta]\mu_N(\{\|u_0^N\|_{\mathcal{B}} > K\})$$

This implies $\mu(\Omega_N^c) \lesssim \frac{T}{\delta}\mu_N(\{\|u_0^N\|_{\mathcal{B}} > K\}) \sim TK^\theta e^{-cK^2}$, and by choosing $K \sim \left(\log \frac{T}{\varepsilon}\right)^{\frac{1}{2}}$, we have $\mu(\Omega_N^c) < \varepsilon$. By its construction, $\|u^N(j\delta)\|_{\mathcal{B}} \leq K$ for $j = 0, \cdots, \pm[T/\delta]$ and by local theory,

$$\|u^N(t)\|_{\mathcal{B}} \leq 2K \sim \left(\log \frac{T}{\varepsilon}\right)^{\frac{1}{2}} \text{ for } |t| \leq T.$$

□

One then needs to prove that μ_N converges weakly to μ. This is standard and one can check the work of Zhidkov [50] for example. Going back to (43), essentially as a corollary of Proposition 5.8 one can then prove:

Corollary 5.9. *Given $\varepsilon > 0$, there exists $\Omega_\varepsilon \subset \mathcal{B}$ with $\mu(\Omega_\varepsilon^c) < \varepsilon$ such that for $u_0 \in \Omega_\varepsilon$, the IVP* (43) *is globally well-posed and*
(a) $\|u - u^N\|_{C([-T,T];\mathcal{B}')} \to 0$ as $N \to \infty$ uniformly for $u_0 \in \Omega_\varepsilon$, where $\mathcal{B}' \supset \mathcal{B}$.
(b) One has the growth estimate

$$\|u(t)\|_{\mathcal{B}} \lesssim \left(\log \frac{1+|t|}{\varepsilon}\right)^{\frac{1}{2}}, \textit{ for all } t \in \mathbb{R}.$$

One can prove (a) and (b) by estimating the difference $u - u^N$ using the local well-posedness theory and a standard approximation lemma, and then applying Proposition 5.8 to u^N. Finally if $\Omega := \bigcup_{\varepsilon>0}\Omega_\varepsilon$, clearly $\mu(\Omega) = 1$ and (43) is almost surely globally well-posed. At the same time one also obtains the invariance of μ.

5.5. **The periodic, one dimensional derivative Schrödinger equation.** It is now time to introduce another infinite dimensional system: the derivative NLS equation (DNLS)

$$\begin{cases} u_t - i\, u_{xx} = \lambda(|u|^2u)_x, \\ u\big|_{t=0} = u_0, \end{cases} \tag{47}$$

where $(x,t) \in \mathbb{T} \times (-T,T)$ and λ is real. Below we will take $\lambda = 1$ for convenience. We note that DNLS is an Hamiltonian PDE. In fact, it is completely integrable [28]. The first three conserved integrals are:

$$\text{Mass:} \quad m(u) = \frac{1}{2\pi}\int_{\mathbb{T}} |u(x,t)|^2\,dx$$

$$\text{'Energy':} \quad E(u) = \int_{\mathbb{T}} |u_x|^2\,dx + \frac{3}{2}\mathrm{Im}\int_{\mathbb{T}} u^2\overline{uu_x}\,dx + \frac{1}{2}\int_{\mathbb{T}}|u|^6\,dx =: \int_{\mathbb{T}}|u_x|^2\,dx + \mathcal{K}(u) \tag{48}$$

$$\text{Hamiltonian:} \quad H(u) = \mathrm{Im}\int_{\mathbb{T}} u\overline{u}_x\,dx + \frac{1}{2}\int_{\mathbb{T}}|u|^4\,dx.$$

We would like now to explore the possibility of extending Bourgain's approach to the periodic DNLS (47). We should immediately say that Thomann and Tzvetkov [48] already proposed a measure for this problem. Unfortunately though the presence of the derivative term in the nonlinearity, in particular $|u|^2u_x$, makes it impossible to prove the needed multilinear estimates of the type presented in Section 3, that are the fundamental tools to show both invariance and almost surely global well-posedness. For this reason one needs to remove the term $|u|^2u_x$ by gauging via the transformation [28, 46]

$$G(f)(x) := \exp(-iJ(f))\, f(x) \tag{49}$$

where

$$J(f)(x) := \frac{1}{2\pi}\int_0^{2\pi}\int_\theta^x \left(|f(y)|^2 - \frac{1}{2\pi}\|f\|^2_{L^2(\mathbb{T})} \right) dy\, d\theta \tag{50}$$

is the unique 2π-periodic mean zero primitive of the map

$$x \longrightarrow |f(x)|^2 - \frac{1}{2\pi}\|f\|^2_{L^2(\mathbb{T})}.$$

Then, for $u \in C([-T,T]; L^2(\mathbb{T}))$ the adapted periodic gauge is defined as

$$\mathcal{G}(u)(t,x) := G(u(t))(x - 2\,t\,m(u)).$$

Note that the difference between $\mathcal{G}$ and G is a space translation by $2\,t\,m(u)$ and this is introduced simply to remove an extra linear term that would have appeared in the gauged equation if one had only used G. We have that

$$\mathcal{G} : C([-T,T]; H^\sigma(\mathbb{T})) \to C([-T,T]; H^\sigma(\mathbb{T}))$$

is a homeomorphism for any $\sigma \geq 0$. Moreover, $\mathcal{G}$ is locally bi-Lipschitz on subsets of functions in $C([-T,T]; H^\sigma(\mathbb{T}))$ with prescribed L^2-norm. The same is true if we replace $H^\sigma(\mathbb{T})$ by $\mathcal{F}L^{s,r}$, the Fourier-Lebesgue spaces defined in (55) below. If u is a solution to (47) then $v := \mathcal{G}(u)$ is a solution to the gauged DNLS initial value problem, here denoted GDNLS:

$$v_t - iv_{xx} = -v^2\overline{v}_x + \frac{i}{2}|v|^4v - i\psi(v)v - im(v)|v|^2v \tag{51}$$

with initial data $v(0) = \mathcal{G}(u(0))$, where

$$\psi(v)(t) := -\frac{1}{\pi}\int_{\mathbb{T}} \operatorname{Im}(v\overline{v}_x)\, dx\; + \frac{1}{4\pi}\int_{\mathbb{T}} |v|^4 dx - m(v)^2$$

and

$$m(u) = m(v) := \frac{1}{2\pi}\int_{\mathbb{T}} |v|^2(x,t)dx = \frac{1}{2\pi}\int_{\mathbb{T}} |v(x,0)|^2(x)dx$$

is the conserved mass. One can also check that if

$$\begin{aligned} \mathcal{E}(v) \;&:=\; \int_{\mathbb{T}} |v_x|^2\, dx - \frac{1}{2}\operatorname{Im}\int_{\mathbb{T}} v^2\overline{v\,v_x}\, dx + \frac{1}{4\pi}\left(\int_{\mathbb{T}} |v(t)|^2\, dx\right)\left(\int_{\mathbb{T}} |v(t)|^4\, dx\right), \\ \mathcal{H}(v) \;&:=\; \operatorname{Im}\int_{\mathbb{T}} v\overline{v}_x - \frac{1}{2}\int_{\mathbb{T}} |v|^4\, dx + 2\pi m(v)^2 \end{aligned}$$

and

$$\tilde{\mathcal{E}}(v) := \mathcal{E}(v) + 2m(v)\mathcal{H}(v) - 2\pi\, m(v)^3 \tag{52}$$

then all are conserved integrals. For convenience let us write

$$\tilde{\mathcal{E}}(v) = \int_{\mathbb{T}} |v_x|^2\, dx + \mathcal{N}(v), \tag{53}$$

where $\mathcal{N}(v)$ represents the part of the energy that comes from the nonlinearity. We now define, at least formally the measure μ as

$$\text{“}d\mu = Z^{-1}\chi_{\{\|v\|_{L^2}<B\}}e^{\mathcal{N}(v)}d\rho\text{”}, \tag{54}$$

where the cut-off function with respect to the L^2 norm is suggested by Remark 5.5 and the fact that equation (51) has a quintic term in it. The plan is then is to show that for B small this measure is well defined and invariant for the GDNLS, that GDNLS is almost surely global well-posedness with respect to it and finally that one can un-gauge to go back to the DNLS (47). Unfortunately there are several obstacles that one needs to overcome to implement this plan. The first is that (51) is ill posed[13] in $H^s, s < \frac{1}{2}$, see [3]. On the other hand Grünrock-Herr [27] proved local well-posedness for initial data $v_0 \in \mathcal{F}L^{s,r}(\mathbb{T})$ and $2 \leq r < 4$, $s \geq \frac{1}{2}$, where

$$\|v_0\|_{\mathcal{F}L^{s,r}(\mathbb{T})} := \| \langle n \rangle^s \hat{v}_0 \|_{\ell^r_n(\mathbb{Z})} \qquad r \geq 2, \tag{55}$$

avoiding in this way L^2 based Sobolev spaces. These spaces scale like the Sobolev spaces $H^\sigma(\mathbb{T})$, where $\sigma = s + 1/r - 1/2$. For example for $s = \frac{2}{3}-$ and $r = 3$ one has that $\sigma < \frac{1}{2}$. As a result one can use Gaussian measures on Banach spaces [14] $\mathcal{F}L^{s,r}(\mathbb{T})$. The next issue is the fact that when one projects (51) via P_N the resulting IVP

$$v^N_t = iv^N_{xx} - P_N((v^N)^2\overline{v^N_x}) + \frac{i}{2}P_N(|v^N|^4v^N) - i\psi(v^N)v^N - im(v^N)P_N(|v^N|^2v^N) \tag{56}$$

with initial data $v^N_0 = P_N\, v_0$ is no longer in an Hamiltonian form that one can recognize and one needs to prove Liouville's theorem by hands. The final, and probably the most serious problem is that the energy $\tilde{\mathcal{E}}(v)$ in (53), that is conserved for (51), is no longer conserved when one projects via P_N. Fortunately though Bourgain's argument can be made more general, in particular it is enough to show that $\tilde{\mathcal{E}}(v^N)$ is *almost* conserved. At the end one can show

Theorem 5.10. *[Almost sure global well-posedness of GDNLS* (51) *and invariance] The measure μ in* (54) *is well defined on $\mathcal{F}L^{\frac{2}{3}-,3}(\mathbb{T})$. Moreover there exists $\Omega \subset \mathcal{F}L^{\frac{2}{3}-,3}(\mathbb{T})$, $\mu(\Omega^c) = 0$ such that GDNLS* (51) *is globally well-posed in Ω and μ is invariant on Ω.*

The last step, a pretty straightforward one, is going back to the un-gauged equation DNLS (47) by pulling back the gauge, that is by defining $\nu := \mu\circ\mathcal{G}$ and in so doing obtaining a theorem like Theorem 5.10 for the initial value problem DNLS (47) and the measure ν, see [38] for details. Now the interesting question is to understand what $\nu = \mu \circ \mathcal{G}$ really represents. Is ν absolutely continuous with respect to the measure that can be naturally constructed for DNLS by using its energy E in (48), as done by Thomann-Tzevtkov [48]?

When we un-gauge the measure μ, at least formally we are un-gauging two pieces, the Radon-Nikodym derivative and the Gaussian measure. Treating the Radon-Nikodym derivative is easy. The problem is un-gauging the Gaussian measure ρ. We can ask the following question: What is $\tilde{\rho} := \rho\circ\mathcal{G}$? Is (its restriction to a sufficiently small ball in L^2) absolutely continuous with respect to ρ? If so, what is its Radon-Nikodym derivative?

---

[13]The ill-posedness result has actually been proved only in $\mathbb{R}$ so far and it says that a fixed point argument cannot be used in Sobolev spaces based L^2. It is believed that this negative result is also true in the periodic case.

[14]For this reason at the end of the day one will be talking about weighted Wiener measures instead of Gibbs measures, see [38] for more details.

5.6. **Gaussian measures and gauge transformations.** In order to finish this step one should stop thinking about the solution v as a infinite dimension vector of Fourier modes and instead start thinking about v as a (periodic with period 1) complex Brownian path in $\mathbb{T}$ (Brownian bridge) solving a certain stochastic process. The argument that follows can be found in full details in [40].

We notice from (49) that to un-gauge we need to use

$$\mathcal{G}^{-1}(v)(x) = \exp(iJ(v))\, v(x)$$

where $J(v)(x)$ was defined in (50). It will be important later that $J(v)(x) = J(|v|)(x)$. Then, if v satisfies

$$dv(x) = \underbrace{dB(x)}_{\text{Brownian motion}} + \underbrace{b(x)dx}_{\text{drift terms}}$$

by Ito's calculus and since $\exp(iJ(v))$ is differentiable we have:

$$d\mathcal{G}^{-1}v(x) = \exp(iJ(v))\, dv + iv \exp(iJ(v)) \left(|v(x)|^2 - \frac{1}{2\pi}\|v\|_{L^2}^2 \right) dx + \dots.$$

Substituting above one has

$$d\mathcal{G}^{-1}v(x) = \exp(iJ(v))\, [dB(x) + a(v, x, \omega))\, dx] + \dots$$

where

$$a(v, x, \omega) = iv \left(|v(x)|^2 - \frac{1}{2\pi}\|v\|_{L^2}^2 \right). \tag{57}$$

What could help? Certainly the fact that $\exp(iJ(v))$ is a unitary operator and that one can prove Novikov's condition:

$$E\left[\exp\left(\frac{1}{2}\int a^2(v, x, \omega)dx\right)\right] < \infty. \tag{58}$$

In fact this last condition looks exactly like what we need for the following theorem:

Theorem 5.11 (Girsanov [39])**.** *If we change the drift coefficient of a given Ito process in an appropriate way, see* (57)*, then the law of the process will not change dramatically. In fact the new process law will be absolutely continuous with respect to the law of the original process and we can compute explicitly the Radon-Nikodym derivative.*

Unfortunately though Girsanov's theorem doesn't save the day.... at least not immediately. If one reads the theorem carefully one realizes that an important condition is that $a(v, x, \omega)$ is *non anticipative*, in the sense that it only depends on the BM v up to "time" x and not further. This unfortunately is not true in our case. The new drift term $a(v, x, \omega)$ involves the L^2 norm of $v(x)$, see (57), and hence it is *anticipative*. A different strategy is needed and conformal invariance of complex BM comes to the rescue

We use the well known fact that if $W(t) = W_1(t) + iW_2(t)$ is a complex Brownian motion, and if ϕ is an analytic function then $Z = \phi(W)$ is, after a suitable time change, again a complex Brownian motion[15], [39]. For $Z(t) = \exp(W(s))$ the time change is given by

$$t = t(s) = \int_0^s |e^{W(r)}|^2 dr, \qquad s(t) = \int_0^t \frac{dr}{|Z(r)|^2}.$$

[15] In what follows one should think of $Z(t)$ to play the role of our complex BM $v(x)$.

We are interested on $Z(t)$ for the interval $0 \leq t \leq 1$ and thus we introduce the stopping time

$$\mathcal{S} = \inf\left\{ s\,;\ \int_0^s |e^{W(r)}|^2 dr = 1 \right\}$$

and remark the important fact that the stopping time $\mathcal{S}$ depends only on the real part $W_1(s)$ of $W(s)$ (or equivalently only $|Z|$). If we write $Z(t)$ in polar coordinate $Z(t) = |Z(t)|e^{i\Theta(t)}$, we have

$$W(s) = W_1(s) + iW_2(s) = \log|Z(t(s)| + i\Theta(t(s))$$

and W_1 and W_2 are real independent Brownian motions. If we define

$$\tilde{W}(s) := W_1(s) + i\left[W_2(s) + \int_0^{t(s)} h(|Z|)(r)dr\right] = W_1(s) + i\left[W_2(s) + \int_0^{t(s)} h(e^{W_1})(r)dr,\right]$$

then have

$$e^{\tilde{W}(s)} = \tilde{Z}(t(s)) = \mathcal{G}^{-1}(Z)(t(s)).$$

In terms of W, the gauge transformation is now easy to understand: it gives a complex process in which the real part is left unchanged and the imaginary part is translated by the function $J(Z)(t(s))$ in (50) which depends only on the real part (i.e. on $|Z|$, which has been fixed) and in that sense is deterministic. It is now possible to use the Cameron-Martin-Girsanov's Theorem [14, 39] only for the law of the imaginary part and conclude the proof. Then if η denotes the probability distribution of W and $\tilde{\eta}$ the distribution of $\tilde{W}$ we have the absolute continuity of $\tilde{\eta}$ and η whence the absolute continuity between $\tilde{\rho}$ and ρ follows with the *same Radon-Nikodym derivative* (re-expressed back in terms of t). All in all then we prove that our un-gauged measure ν is in fact essentially (up to normalizing constants) of the form

$$d\nu(u) = \chi_{\|u\|_{L^2} \leq B} e^{-\mathcal{K}(u)} d\rho,$$

where $\mathcal{K}(u)$ was introduced in (48), that is the weighted Wiener measure associated to DNLS (constructed by Thomann-Tzvetkov [48]). In particular we prove its invariance.

Remark 5.12. *The sketch of the argument above needs to be done carefully for complex Brownian bridges (periodic BM) by conditioning properly. See* [40].

5.7. **Periodic dispersive equations and the non-squeezing theorem.** In Theorem 5.3 we recalled like a finite dimensional Hamiltonian flow Φ_t cannot squeeze a ball into a cylinder with a smaller radius. Generalizing this kind of result in infinite dimensions has been a long project of Kuksin [32] who proved, roughly speaking, that compact perturbations of certain linear dispersive equations do indeed satisfy the non-squeezing theorem. It is easy to show that the L^2 space equipped with the form

$$\omega(f, g) = \langle if, g\rangle_{L^2} \tag{59}$$

is a symplectic space for the cubic, defocusing NLS equation on $\mathbb{T}$ and its global flow $\Phi(t)$ is a symplectomorphism. One can show that this setting does not satisfy the conditions in [32]. Nevertheless Bourgain proved the following theorem:

Theorem 5.13. *[Non-squeezing* [5]*] Assume that Φ_t is the flow generated by the cubic, periodic, defocusing NLS equation in L^2. If we identify L^2 with l^2 via Fourier transform and we let $B_r(y_0)$ be the ball in l^2 centered at $y_0 \in l^2$ and radius r, $C_R(z_0) := \{(a_n) \in l^2 / |a_i - z_0| \leq R\}$ a cylinder of radius R and $\Phi_t(B_r(y_0)) \subset C_R(z_0)$, at some time t, then it must be that $r \leq R$.*

The proof of this theorem is based on projecting the IVP onto finitely many frequencies via the projection operator P_N as was done in (44). In this case the new projected problem is a finite dimensional Hamiltonian system and Gromov's Theorem 5.3 can be applied. The difficult part then is to show that the flow $\Phi_N(t)$ of the projected IVP approximates well the flow $\Phi(t)$ of the original problem. In this case this can be done thanks to strong multilinear estimates based on the Strichartz estimates recalled in Theorem 18; see [5] for the complete proof. We should mention here that unfortunately Bourgain's argument may not work for other kinds of dispersive equations. For example in [21], where the KdV problem was studied, the lemma in Bourgain's work that gives the good approximation of the flow $\Phi(t)$ by $\Phi_N(t)$ does not hold. This has to do with the number of interacting waves in the nonlinearity. There it was proved that still the non-squeezing theorem holds, but the proof was indirect and it had to go through the Miura transformation.

References

[1] D. Benney and A. Newell. *Random wave closures*, Stud. Appl. Math., Vol. 48 (1969), 29-53.

[2] D. Benney and P. Saffman. *Nonlinear interactions of random waves in a dispersive medium*, Proc. Roy. Soc. A **289** (1966), 301-320.

[3] H.A. Biagioni and F. Linares. *Ill-posedness for the derivative Schrödinger and generalized Benjamin-Ono equations*, Trans. AMS **353** no. 9 (2001), 3649-3659.

[4] J. Bourgain. *Fourier transform restriction phenomena for certain lattice subsets and applications to nonlinear evolution equations, I:Schrödinger equations*, Geom. Funct. Anal. **3** (1993), 107-156.

[5] J. Bourgain. *Periodic nonlinear Schrödinger equation and invariant measures*, Commun. Math. Phys. **166** (1994), 1-26.

[6] J. Bourgain. *Aspects of longtime behavior of solutions of nonlinear Hamiltonian evolution equations*, GAFA **5** (1995), 105–140.

[7] J. Bourgain. *On the growth in time of higher Sobolev norms of smooth solutions of Hamiltonian PDE*, Int. Math. Research Notices **6** (1996), 277-304.

[8] J. Bourgain. *Global Solutions of Nonlinear Schrödinger Equations*, AMS Colloquium Publications **46**, AMS, Providence, RI, 1999.

[9] J. Bourgain. *On Strichartz inequalities and the nonlinear Schrödinger equation on irrational tori*, Mathematical aspects of nonlinear dispersive equations, 120, Ann. of Math. Stud., 163, Princeton Univ. Press, Princeton, NJ, 2007.

[10] J. Bourgain. *Moment inequalities for trigonometric polynomials with spectrum in curved hypersurfaces* . Preprint 2011.

[11] N. Burq, P. Gérard and N. Tzvetkov. *Bilinear eigenfunction estimates and the nonlinear Schrödinger equation on surfaces*, Invent. Math. **159** (2005), 187-223.

[12] N. Burq and N. Tzvetkov. *Random data Cauchy theory for supercritical wave equations. I. A local theory,* Invent. Math. **173**, no. 3 (2008) 449 -475.

[13] N. Burq and N. Tzvetkov. *Random data Cauchy theory for supercritical wave equations. II. A global theory*, Invent. Math. **173**, no. 3 (2008) 477-496.

[14] R. H. Cameron and W. T. Martin. *Transformation of Wiener integrals under translations*, Annals of Math. **45** no. 2, (1944), 386-396.

[15] T. Cazenave. *Semilinear Schrödinger equations*, Courant Lecture Notes in Mathematics, **10**. New York University, Courant Institute of Mathematical Sciences, AMS, (2003).

[16] F. Catoire and W.-M. Wang. *Bounds on Sobolev norms for the nonlinear Schrödinger equation on general tori*, Commun. Pure Appl. Anal. 9 (2010), no. 2, 483491.

[17] J.Colliander, S. Kwon and T. Oh. *A Remark on Normal Forms and the "Upside-down" I-method for periodic NLS: Growth of Higher Sobolev Norms.* `arXiv:1010.2501`.

[18] J. Colliander, M. Keel, G. Staffilani, H. Takaoka and T. Tao. *Global well-posedness for Schrödinger equations with derivative*, SIAM J.Math.Anal. **33**, no.3 (2001), 649-669.

[19] J. Colliander, M. Keel, G. Staffilani, H. Takaoka and T. Tao. *Polynomial upper bounds for the orbital instability of the 1D cubic NLS below the energy norm*, Commun. Pure Appl. Anal. **2** (2003), no. 1, 33-50.

[20] J. Colliander, M. Keel, G. Staffilani, H. Takaoka and T. Tao. *Sharp global well-posedness for KdV and modified KdV on* $\mathbb{R}$ *and* $\mathbb{T}$, J. Amer. Math. Soc. **16** (2003), no. 3, 705-749. (electronic)

[21] J. Colliander, M. Keel, G. Staffilani, H. Takaoka and T. Tao. *Notes on symplectic non-squeezing of the KdV flow*, Journees "Equations aux Derivee Partielles", Exp. No. XIV, 15 pp., Ecole Polytech., Palaiseau, (2005).

[22] J. Colliander, M. Keel, G. Staffilani, H. Takaoka and T. Tao. *Transfer of energy to high frequencies in the cubic nonlinear Schrödinger equation*, Invent. Math. **181**, No.1 (2010), 39-113.

[23] B. Dodson. *Global well-posedness and scattering for the defocusing,* L^2*- critical, nonlinear Schrödinger equation when* $d = 2$, `http://arXiv:math/1006.1365`.

[24] M. Gromov. *Pseudo-holomorphic curves in symplectic manifolds*, Invent. math., **82** (1985), 307–347.

[25] L. Gross. *Abstract Wiener spaces,* Proc. 5th Berkeley Sym. Math. Stat. Prob. **2**, (1965), 31–42.

[26] E. Grosswald. *Representations of integers as sums of squares* Springer-Verlag (1985).

[27] A. Grünrock and S. Herr. *Low Regularity local well-posedness of the derivative nonlinear Schrödinger equation with periodic data*, SIAM J. Math. Anal. **39**, no. 6 (2008), 1890-1920.

[28] N. Hayashi and T. Ozawa. *On the derivative nonlinear Schrödinger equation*, Physica D **55** no. 1-2, (1992), 14-36.

[29] N.J. Hitchin, G.B. Segal and R.S. Ward. *Integrable Systems*, Oxford Graduate Texts in Mathematics, **4** (1999).

[30] M. Keel and T. Tao. *Endpoint Strichartz estimates*, Amer. J. Math. **120** (1998), 955–980.

[31] C. Kenig, G. Ponce and L. Vega. *Quadratic Forms for the 1-D Semilinear Schrödinger Equation*, Transactions of the AMS, **348**, no.8 (1996), 3323-3353.

[32] S. Kuksin. *Infinite-dimensional symplectic capacities and a squeezing theorem for Hamiltonian PDE's*, CMP **167** (1995), 521–552.

[33] H. Kuo. *Gaussian Measures in Banach Spaces,* Lec. Notes in Math. 463, Springer-Verlag, New York, 1975.

[34] J. Lebowitz, H. Rose and E. Speer. *Statistical mechanics of the nonlinear Schrödinger equation* J. Stat. Phys. **50** no. 3-4, (1988), 657-687.

[35] T. Oh. *Invariant Gibbs measures and a.s. global well-posedness for coupled KdV systems*, Diff. Integ. Eq. **22** no. 7–8, (2009), 637-668.

[36] T. Oh. *Invariance of the white noise for KdV*, Comm. Math. Phys. **292** no. 1 (2009), 217-236 [2009]. *Erratum: "Invariance of the white noise for KdV"*, in preparation.

[37] T. Oh. *Invariance of the Gibbs Measure for the Schrödinger-Benjamin-Ono System,* to appear in SIAM J. Math. Anal.

[38] T. Oh, A. Nahmod, L. Ray-Bellet and G. Staffilani. *Invariant weighted Wiener measures and almost sure global well-posedness for the periodic derivative NLS*, to appear in JEMS.

[39] B. Oksendal. *Stochastic Differential Equations*, Springer-Verlag, Fifth Edition, (1998).

[40] A. Nahmod, L. Ray-Bellet, S. Sheffield and G. Staffilani. *Absolute Continuity of Gaussian measures under certain gauge transformations*, To appear in MRL.

[41] R. Palais. *The symmetries of solitons*, Bull. of AMS, **34** (1997), 339–403.

[42] V. Sohinger. *Bounds on the growth of high Sobolev norms of solutions to 2D Hartree equations*, preprint (2010), `http://arXiv:math/1003.5709`.

[43] G. Staffilani. *On the growth of high Sobolev norms of solutions for KdV and Schrödinger equations*, Duke Math. J. **86** no. 1 (1997), 109-142.

[44] G. Staffilani. *Quadratic Forms for a 2-D Semilinear Schrödinger Equation*, Duke Math. J., **86**, no.1 (1997), 79-107.

[45] E. Stein. *Harmonic Analysis*, Princeton University Press, (1993).

[46] Takaoka, H., *Well-posedness for the one dimensional nonlinear Schrödinger equation with the derivative nonlinearity*, Adv. Differential Equations, **4** , no.4 pp. 561-580 [1999].

[47] T. Tao. *Nonlinear Dispersive Equations: local and global analysis*, CBMS, **106** (2006).

[48] L. Thomann and N. Tzvetkov. *Gibbs measure for the periodic derivative nonlinear Schrödinger equation,* `http://arXiv:1001.4269v1`.

[49] V. E. Zakharov. *Stability of periodic waves of finite amplitude on a surface of deep fluid*, J. Appl. Mech. Tech. Phys., **9** (1968), 190-194.

[50] P. Zhidkov. *KdV and Nonlinear Schrödinger Equations: Qualitative theory*, Lecture Notes in Mathematics **1756** Berlin, Springer, 2001.

[51] S.-J. Zhong. *The growth in time of higher Sobolev norms of solutions to Schrödinger equations on compact Riemannian manifolds*, J.Differential Equations **245** (2008) 359-376.

Department of Mathematics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139

E-mail address: gigliola@math.mit.edu

How does quantum mechanics scale?

Umesh Vazirani*

Abstract

The quantum description of a system of n spins requires 2^n complex numbers. For n = 500 this dwarves estimates for the number of particles in the Universe. This simple observation lies at the heart of the extravagant computing power of quantum computers. It also presents a fundamental obstacle to simulating or "solving" general quantum many body system.

Is it even possible in principle to experimentally verify this exponential scaling? And do quantum states that occur in Nature exhibit this exponential complexity? We discuss recent progress on complexity theoretic formulations of these questions.

1 Introduction

The fundamental lesson that quantum computation teaches is that quantum systems are exponentially complex. The classical description of a general state of an n particle quantum system scales exponentially in n, a phenomenon responsible for the violation by quantum computers of the extended Church-Turing thesis [8, 14], and the remarkable power of quantum algorithms [13]. Indeed, this is among the most counter-intuitive predictions of quantum mechanics, and was missed for a good fraction of a century after the birth of quantum mechanics in the early 1900s. But has this exponential scaling been experimentally tested? Whereas quantum mechanics has been extensively scrutinized and experimentally tested to an exquisite degree of accuracy (some predictions in QED have been verified to better than one part in 10^{12}), it can be argued that these are low complexity quantum mechanics. The complexity measure here is the effective dimension of the Hilbert space in which the state of the system lives. One of the goals of physics research is to test the limits of validity of a theory — in the limit of high energy or at the Planck scale or the speed of light. All this suggests that there is a new regime in which to test quantum mechanics, the limit of high complexity [15].

*Department of Computer Science, UC Berkeley

At a more foundational level, is it even possible in principle to experimentally verify this exponential scaling? The difficulty in doing so is that the exponential scaling presents a fundamental obstacle to solving or simulating a general quantum many body system, and therefore to predicting the outcome to an experiment as a prelude to experimental verification. In other words, there appears to be a tension between the exponential complexity of quantum systems and the standard scientific paradigm of "predict and verify". Nevertheless, Shor's quantum factoring algorithm can be used as the basis of a physical experiment that tests quantum mechanics in this regime. Even though the outputs of the experiment, the prime factors of the input N, are hard to compute assuming the intractability of factoring on a classical computer, the experiment can be validated by checking that the product of the prime factors is indeed N. i.e. by checking a certain consistency condition on the output.

This consistency checking approach can be greatly generalized by appealing to the complexity theoretic framework of interactive proof systems [4]. In an interactive proof system, a computationally weak (polynomial time) verifier, Arthur, can interact with a much more powerful but untrusted prover, Merlin, and determine the correctness of an a claimed answer to a computational problem. For this to be possible, Merlin has to be willing to answer a number of cleverly chosen random questions related to the original claim. Arthur adaptively generates this sequence of questions based on Merlin's answers, and checks Merlin's answers for consistency. The remarkable property of such protocols is that Arthur can only be convinced of the original claim (with non-negligible probability over the choice of questions) if it is in fact a valid claim. Arthur's confidence in the claim does not depend in any way on his trust in Merlin, but rather in the consistency checks that he is able to perform on Merlin's answers. The complexity class of assertions that a polynomial time bounded Arthur can verify in this way is denoted by IP. The class IP generalizes the complexity class NP by allowing both randomness and interaction with the prover. A remarkable result in complexity theory asserts that $IP = PSPACE$ [7]. i.e. a prover who can solve any problem in $PSPACE$ can convince a polynomial time verifier about the correctness of any instance of a problem in $PSPACE$.

Denote by $QPIP$ the class of assertions that a polynomial time bounded Arthur can verify when the prover is restricted to quantum polynomial time (BQP). Clearly $QPIP \subseteq BQP$. The question of interest in the context of experimentally verifying quantum mechanics is whether $BQP \subseteq QPIP$. A positive answer to this question could be interpreted as follows: a (classical) experimentalist wishes to check that the outcome of a particular experiment is consistent with the predictions of quantum mechanics. Unfortunately he cannot theoretically calculate the outcome predicted by quantum mechanics. Instead, he carries out a sequence of interactive experiments with the quantum system. If the outcomes of all these experiments jointly satisfy the consistency checks (specified by the interactive proof system), the experimentalist concludes that the outcome of the original experiment was indeed consistent with the predictions of quantum mechanics. Moreover, his confidence in this conclusion is based only on the outcome of the consistency tests, which he can carry out efficiently. Of course, if the outcomes of the experiments did not pass the consistency tests, then the experimentalist could only conclude that at least one of the experiments failed to meet the predictions of quantum mechanics.

Whether or not $BQP \subseteq QPIP$ is currently an open question. Indeed, this is one of the most important computational questions about the foundations of quantum mechanics. A closely related result was proved Aharonov, Ben-Or and Eban [4]. They considered the situation where Arthur is not purely classical, but can store and manipulate a constant number (3 to be concrete) of qubits, and can exchange qubits with Merlin, who is restricted to quantum polynomial time (BQP). Denote by $QPIP^*$ the class of assertions that Arthur can verify in this setting. Then $QPIP^* = BQP$. One way to understand this protocol is to imagine that a company QWave claims to have experimentally realized a quantum computer, and wishes to convince a potential buyer that the computer is indeed capable of performing an arbitrary quantum computation on up to n qubits. If the potential buyer has the capability of storing and manipulating 3 qubits, and of exchanging qubits with the quantum computer, then by following the above protocol, he can verify that the computer faithfully carried out any quantum computation of his choice.

One might wonder whether quantum states that occur in Nature are low complexity states. i.e. states with classical descriptions that scale polynomially in n, the number of particles in the quantum system. The exponential complexity of quantum states is directly related to the phenomena of entanglement, which is the fundamental obstacle in the quest to simulate quantum systems on a classical computer. Ground states of quantum many body systems on a lattice, which are ubiquitous in condensed matter physics, provide a natural setting to explore the complexity of quantum states that occur in Nature.

A remarkable conjecture in condensed matter physics dating back about a half century is the Area Law, which strongly bounds the entanglement in ground states of gapped local Hamiltonians (for a survey see [9]). Consider the interaction graph (hypergraph) associated with a local Hamiltonian on a D-dimensional grid. It has a vertex for each particle and an edge for each term of the Hamiltonian. Intuitively and very roughly, an area law says that entanglement is local in this interaction graph in the following sense: for any contiguous region L in the grid, the entangle-

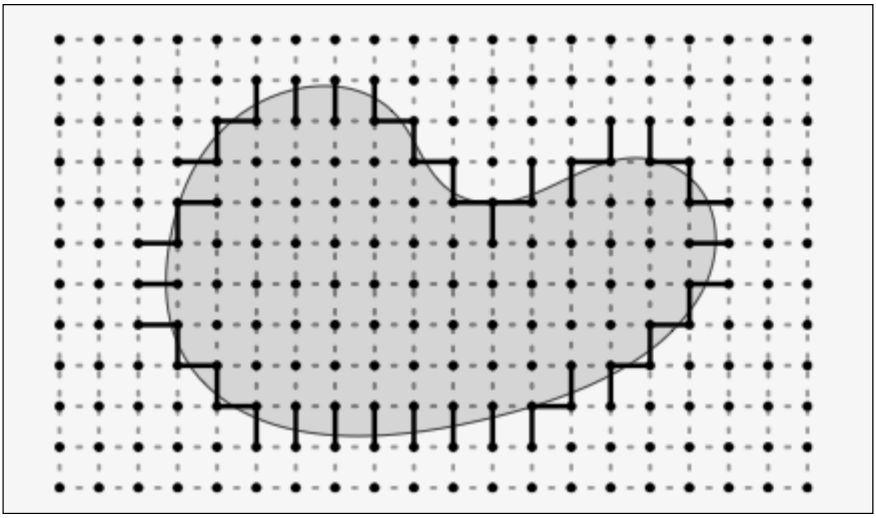


Figure 1: A quantum many-body system on a grid: the grid points represent particles, and the edges represent 2-body interactions. The area law asserts that the entanglement entropy between the particles inside the region L and the particles outside of it is proportional to the surface area of L.

ment between the particles inside L and the particles outside L is mostly due to the local degrees of freedom along the edges that connect these subsets. More precisely, for any region L, let $\overline{L}$ denote its complement. Then a state $|\psi\rangle$ obeys an area law if the *entanglement entropy* between L and $\overline{L}$ is upper-bounded by the order of the number of edges that connect L and $\overline{L}$, which is proportional to the boundary (surface area) of L. This is clearly a much stronger bound on the entropy than the trivial bound (known as a *volume law*), which is proportional to the number of particles (nodes) inside L.

In a seminal paper [11], Hastings proved that ground states of gapped 1D systems obey an area law. For 1D systems the area law asserts that ground state entanglement entropy across any contiguous cut is bounded by a constant, independent of the system size. Specifically, for a nearest neighbor system on a line, $H = \sum_{i=1}^{n-1} H_i$, with particle dimension d, interaction strength $\|H_i\| \leq J$ and a spectral gap $\epsilon > 0$, the theorem states that entanglement entropy across any cut in the chain is upper bounded by $S \leq e^{\mathcal{O}(X)}$ for $X \stackrel{\text{def}}{=} \frac{J \log d}{\epsilon}$.

Hastings' result implies that ground states of gapped 1D systems can be well-approximated by states that admit a classical description of a polynomial size (MPS). This can in turn be used to approximate any local observable efficiently on a classical computer. From a complexity-theoretic point of view it tells us that the local Hamiltonian problem for such 1D systems with a constant spectral gap belongs to the complexity class NP.

Two major issues remained opened following Hastings' paper. The first is an extension of his results to 2 and 3 dimensional systems. The second is the dependence of the bound on X. Hastings shows that S scales exponentially in X, whereas the best lower bounds come from specially crafted 1D Hamiltonians [12, 10], scale as $\epsilon^{-1/4}$. This is important for two reasons: the dependence on ϵ determines how close to the critical point ($\epsilon = 0$) the ground state admits a polynomial size MPS. At the same time, improving the dependence on $\log d$ provides a possible path towards proving an area law in higher dimensions. Indeed, a bound on S that scales as $\mathcal{O}(\log d)$ would imply higher dimensional area laws by the naive reduction from a D-dimensional system to a 1-D system by fusing together particles on surfaces parallel to the boundary.

A completely new and more combinatorial approach to proving the area law was developed in a sequence of papers [2, 3], culminating in a proof by Arad, Landau and Vazirani [6] of a polynomial bound of $\mathcal{O}(X^3 \log^8 X)$ on the entanglement entropy for 1D frustration free systems. Since $\log d$ is a trivial lower bound, and it is known that there are frustration free systems in which the entanglement entropy grows as $\epsilon^{-1/4}$ [10], this bound is tight to within a polynomial factor.

In addition to providing a polynomially optimal upper bound on the 1D entanglement entropy, these results have some relevance in the context of area laws in higher dimensions. A naive application of the improved bound to higher dimensional systems would yield an entropy bound of $S \leq |\partial L|^3 \text{poly}(\log |\partial L|)$, which is still worse than a volume law. However, in 2 or more dimensions, one can further exploit the local properties of the system *along the boundary* ∂L, and improve the bound to $|\partial L|^2 \cdot \text{poly}(\log |\partial L|)$. This bound is at the cusp of being non-trivial; any further improvement that would bound the entropy by $|\partial L|^{2-\delta}$ for any $\delta > 0$, would prove a sub-volume law for 2D systems. Proving the area law for 2D and higher dimensional systems is one of the most important open questions in quantum Hamiltonian complexity.

2 Entanglement Measures:

Entanglement, the feature that a portion of a quantum system cannot be completely described without access to the whole system, is fundamental to quantum mechanics. Given a state $|\phi\rangle$ and a bipartition of the system to two non intersecting sets, R and L, with corresponding Hilbert spaces $\mathcal{H}_L, \mathcal{H}_R$ such that

$\mathcal{H} = \mathcal{H}_L \otimes \mathcal{H}_R$, we can consider the *Schmidt decomposition* of the state along this cut: $|\phi\rangle = \sum_j \lambda_j |L_j\rangle \otimes |R_j\rangle$. Here $\lambda_1 \geq \lambda_2 \geq \ldots$ are the *Schmidt coefficients*. It turns out that if we add the constraint that the sets $\{|L_j\rangle\}$ and $\{|R_j\rangle\}$ are each orthonormal sets and that the $\lambda_j \geq 0$ then such a decomposition can always be done and this decomposition is essentially unique. The collection of coefficients λ_j in the Schmidt decomposition serve to describe the entanglement of $|\phi\rangle$ between R and L.

One straightforward measure of entanglement refers to the minimum number of non-zero terms needed in a decomposition of the type described above. It turns out that the Schmidt decomposition achieves this minimum, and this number, i.e. the number of non-zero λ_j in the Schmidt decomposition, is referred to as the *Schmidt rank* (SR). The usefulness of the SR stems from it being a "worst case" estimate for the entanglement. As such, it is often easy to bound. The following facts are easy to verify:

The following facts are easy to verify:

Fact 2.1

1. *If O is a k-local operator whose support intersects both $\mathcal{H}_L$ and $\mathcal{H}_R$, then it can increase the SR (with respect to the bi-partitioning) of any state by at most a factor of d^k: $\mathrm{SR}(O\phi) \leq d^k \mathrm{SR}(\phi)$. If O intersects only one part of the system, its action cannot increase the SR.*

2. *Consider a 1D system. If r_i and r_j are the SR of $|\phi\rangle$ that correspond to cuts between particles $i, i+1$, and $j, j+1$, then $r_i \leq d^{|i-j|} r_j$.*

An important fact about the SR is the following corollary of the Eckart-Young theorem [16], which states that the truncated Schmidt decomposition provides the best approximation to a vector in the following sense:

Fact 2.2 *Let $|\psi\rangle$ be a vector on a bi-partitioned Hilbert space $\mathcal{H}_L \otimes \mathcal{H}_R$, and let $\lambda_1 \geq \lambda_2 \geq \ldots$ be its corresponding Schmidt coefficients. Then the largest inner product between $|\psi\rangle$ and a normalized vector with Schmidt rank r is $\sqrt{\sum_{j=1}^{r} \lambda_j^2}$.*

Entanglement entropy is a more refined quantitative measure of entanglement between two pieces of a quantum system; it is defined to be the classical entropy of the squares of the Schmidt coefficients $\{\lambda_j^2\}$, i.e. $\sum_j \lambda_j^2 \log(\frac{1}{\lambda_j^2})$. An alternate and perhaps more revealing definition is as follows: form the reduced density matrix $\rho_L = tr_R(|\psi\rangle\langle\psi|)$ of $|\psi\rangle$ on L. We can think of ρ_L as a probabilistic mixture of the eigenstates of ρ_L with the probability of a given eigenstate being its eigenvalue. With this description, a natural quantification of entanglement is the classical entropy of this probability distribution, i.e. the classical entropy of the eigenvalues of

ρ_L. Indeed this quantity is also the entanglement entropy as the the squares of the Schmidt coefficients $\{\lambda_j^2\}$ coincide with the set of eigenvalues ρ_L.

3 1D Area Law:

Let $|\Omega\rangle$ be the ground state of a frustration-free, nearest neighbor Hamiltonian system $H = \sum_{i=1}^{n} H_i$ on a 1D chain of n particles of dimension d. Assume that the system has spectral gap $\epsilon > 0$, and an interaction strength $\|H_i\| \leq J$. For the sake of simplicity, we assume that H_i are projections, and therefore $P_i \stackrel{\text{def}}{=} 1 - H_i$ are projections to the local ground spaces of the different terms.

Let L be the set of particles 1 through i, and let R be the set of the remaining particles. Then the 1D area law asserts that the entanglement entropy of $|\Omega\rangle$ across every cut $i, i+1$ is bounded by a constant, independent of n. For the remainder of the paper we sketch the 1D area law proved by Arad, Landau and Vazirani [6]:

Theorem 3.1 *Along any cut in the chain, the entanglement entropy of $|\Omega\rangle$ is bounded by* $S_{1D}(\Omega) \leq \mathcal{O}(X^3 \log^8 X)$ *, for* $X \stackrel{\text{def}}{=} \frac{J \log d}{\epsilon}$.

4 Outline of Proof:

A necessary and sufficient condition for a 1D area law is that the largest Schmidt coefficient λ_1 of $|\Omega\rangle$ be a constant. The proof will proceed by establishing a lower-bound on λ_1. We will sketch below why this implies the area law.

Say that an operator K is an (D, Δ)-Approximate Ground State Projector, if it satisfies the following properties:

- **Ground space invariance:** for any ground state $|\Omega\rangle$, $K|\Omega\rangle = |\Omega\rangle$.
- **Shrinking:** for any state $|\Omega^\perp\rangle \in \mathcal{H}^\perp$, also $K|\Omega^\perp\rangle \in \mathcal{H}^\perp$, and $\left\|K|\Omega^\perp\rangle\right\|^2 \leq \Delta$.
- **Entangling:** for any state $|\phi\rangle$, $\mathrm{SR}(K|\phi\rangle) \leq D \cdot \mathrm{SR}(\phi)$.

We refer to D as the SR factor and Δ as the shrinking factor. Clearly, there is a race between these two factors D and Δ. It turns out that when $D \cdot \Delta \leq 1/2$, this can be used to bound λ_1:

Claim 4.1 *If there exists a (D, Δ)-Approximate Ground State Projector K, with $D\Delta \leq 1/2$, then $\lambda_1 \geq 1/\sqrt{2D}$.*

Proof: Consider the state $K|L_1\rangle \otimes |R_1\rangle$. It can be decomposed as $\lambda_1|\Omega\rangle + \delta|\phi\rangle$, where $|\phi\rangle$ is orthogonal to $|\Omega\rangle$, and $\delta^2 \leq \Delta$. Its squared inner product with $|\Omega\rangle$ is therefore at least $\frac{\lambda_1^2}{\lambda_1^2+\delta^2}$. Moreover its Schmidt rank is at most D. By the Eckart-Young theorem, it follows that the squared inner product is at most D times the largest Schmidt coefficient, or $D\lambda_1^2$. Therefore $\frac{\lambda_1^2}{\lambda_1^2+\delta^2} \leq D\lambda_1^2$, or $D(\lambda_1^2 + \delta^2) \geq 1$. But since $D\delta^2 \leq D\Delta \leq 1/2$, it follows that $D\lambda_1^2 \geq 1/2$ and so $\lambda_1 \geq 1/\sqrt{2D}$. ■

Corollary 4.2 *If there exists a* (D, Δ)*-Approximate Ground State Projector* K*, with* $D\Delta \leq 1/2$*, then the ground state entropy is bounded by* $S = O(\log D)$*.*

The corollary follows by applying K repeatedly to the Schmidt vector corresponding to λ_1. This results in a sequence of vectors which geometrically converge to the ground state, while increasing in Schmidt rank by a factor of D with each application. Using the Young-Eckart theorem to bound the higher Schmidt coefficients yields the bound on the entropy.

Detectability Lemma:

As a starting point for an approximate ground state projector, consider first the operator $I - 1/mH$. This operator is bounded between 0 and 1, and leaves $|\Omega\rangle$ invariant. On the other hand, it shrinks any state orthogonal to $|\Omega\rangle$ by a factor of at least $1 - 1/m\epsilon$. Is there a "local" operator that achieves better shrinkage? This is achieved Detectabiility lemma (DL) [1]: partition the projections $\{P_i\}$ into two subsets of even and odd projections, which are called "layers" Inside each layer, the projections commute because they are non-intersecting. Consequently, $\Pi_{odd} \stackrel{\text{def}}{=} P_1 \cdot P_3 \cdot P_5 \cdots$ and $\Pi_{even} \stackrel{\text{def}}{=} P_2 \cdot P_4 \cdot P_6 \cdots$ are the projections into the common eigenspace of the odd and even layers. Then according to the DL, the operator $A \stackrel{\text{def}}{=} \Pi_{even}\Pi_{odd}$ is an *approximation* to the ground state projection. It preserves the ground state, while shrinking its perpendicular subspace by an n-independent factor $\Delta_0(\epsilon) \simeq 1 - c\epsilon$, where c is a geometrical factor. Moreover, each application of A increases the Schmidt rank of our state along any cut in the chain by a constant factor of $D_0 \stackrel{\text{def}}{=} d^2$ (due to the projection that intersects with the cut). Unfortunately, we would expect $D_0\Delta_0 > 1$, so the operator A does not, by itself, suffice as the approximate ground state projector that we wish to construct.

Entanglement Flows:

What about the operator A^ℓ? It achieves a shrinkage factor of Δ_0^ℓ, but now the projection intersecting the cut occurs ℓ times, so the obvious upper bound on the entanglement rank is D_0^ℓ. This does not provide an improvement since $D_0\Delta_0 > 1$ implies $D_0^\ell\Delta_0^\ell > 1$ There is reason to think that this bound on the rate of increase in entanglement is too generous. The repeated application of the projection P_i that intersects with the cut can only increase the SR exponentially if the entanglement created in each iteration is moved away from the boundary to create "more room" for further entanglement. But, intuitively, the remaining projections in A cannot move this entanglement quickly enough to maintain this rate of growth. Although it is not known how to show this for the operator A, it is possible to bound the entanglement growth rate for a new approximate ground state projection operator $\hat{A}$, which is created by suitably modifying A.

To prove this formally, we introduce the notion of entanglement flows. The idea is to analyze the flow of entanglement across many cuts $j, j+1$ in the vicinity the $i, i+1$ cut, as the approximate ground state projection operator is repeatedly applied, and to show that there must be some cut $j, j+1$ across which the SR growth D_j is necessarily small. More precisely we show that the operator $K = \hat{A}^{\ell}$ can be decomposed as the sum of a suitably small number of operators, each of which generates small SR growth D_j at some cut $j, j+1$. The SR growth across the cut of interest $i, i+1$ can then simply be upper bounded by the sum of $D_j d^{|i-j|}$ over all the operators in this decomposition. For suitable choice of ℓ, the shrinkage factor Δ of the operator K and the SR growth D satisfy $D\Delta \leq 1/2$.

Constructing K - coarse graining and Chebyshev polynomials:

To construct the operator K we need several new ideas. First we observe that D_0 and Δ_0 can be replaced by D_0^k and Δ_0^k respectively by coarse graining — fuse k adjacent particles, making them a single particle of dimension d^k. Although this only increases the value of the product, it creates room for the next step, which is to modify the operator A to decrease the factor by which it blows up the Schmidt rank. For concreteness, assume that the even layer contains the projection that intersects with the cut. We will focus on a segment of m projections around the cut, and denote their product by Π_m, so that $\Pi_{even} = \Pi_m \Pi_{rest}$. We will replace the operator Π_{even} with $\hat{\Pi}_m \Pi_{rest}$ that closely approximates Π_{even} while increasing the Schmidt rank by much less than D_0^k (when amortized over several applications).

One of the great benefits of using the DL is that the all projections in a given layer commute, and hence much of the following analysis becomes almost classical. Indeed, the m projections around the cut $\{P_i\}_{i=1}^m$ define a decomposition of the Hilbert space of the system into a direct sum of 2^m eigenspaces, called sectors. Each sector is defined by a string $s = (s_1, \ldots, s_m)$, such that if $|\psi_s\rangle$ is in the s sector, $P_i|\psi_s\rangle = (1 - s_i)|\psi_s\rangle$. A site with $s_i = 1$ is called *a violation*, since it corresponds to a non-zero energy of the corresponding local Hamiltonian term, and $\sum_{i=1}^m s_i$ is the total number of violations in the sector s.

Now, an arbitrary state $|\psi\rangle$ can be decomposed as $|\psi\rangle = |\psi_0\rangle + |\psi_1\rangle$, where

$|\psi_0\rangle$ is its projection on the zero violations sector and $|\psi_1\rangle$ is its projection on the violating sectors. Clearly $\Pi_m|\psi\rangle = |\psi_0\rangle$. To approximate this behavior, we will use the $\{P_i\}$ projections to construct an operator $\hat{\Pi}_m$ that is diagonal in the sectors decomposition, and in addition $\hat{\Pi}_m|\psi\rangle = |\psi_0\rangle + |\psi_1'\rangle$, with $|\psi_1'\rangle$ in the violating sectors and $\||\psi_1'\|^2 \leq \delta\|\psi_1\|^2$. It follows that the operator $\hat{A} \stackrel{\text{def}}{=} \hat{\Pi}_m\Pi_{rest}\Pi_{odd}$ approximates $\Pi_{even}\Pi_{odd}$ in the sense that $\hat{A}|\Omega\rangle = |\Omega\rangle$, $\hat{A}|\Omega^\perp\rangle \in \mathcal{H}^\perp$, and $\|\hat{A}|\Omega^\perp\rangle\|^2 \leq (\Delta_0^k + \delta)\|\Omega^\perp\|^2$. Let $\hat{\Delta} \stackrel{\text{def}}{=} \Delta_0^k + \delta$.

To construct the operator $\hat{\Pi}_m$, first consider the operator $N = \sum_{i=1}^m (1 - P_i)$. The operator N counts the number of violations in a sector: if $|\psi_s\rangle$ belongs to the s sector, $N|\psi_s\rangle = |s| \cdot |\psi_s\rangle$. The operator $\hat{\Pi}_m$ will be a polynomial in N, with the polynomial evaluating to 1 on $|s| = 0$, and less than δ on input with $|s|$ between 1 and m. Three ideas play a critical role in the construction of this polynomial and in bounding the increase in Schmidt rank. The first is the use of a Chebyshev polynomial, which achieves the desired behavior at the $m+1$ points with a degree of only $j = \mathcal{O}(\sqrt{m}\log\delta^{-1})$. The second idea is that it suffices to bound the entanglement across any of the m cuts, and then pay a further penalty of at most $D_I \stackrel{\text{def}}{=} (D_0^k)^m$ to bound the entanglement across the cut of interest. So, if we consider the operator $\hat{A}^\ell$, each term has degree $j\ell$ (i.e. is a product of $j\ell$ of the P_is), and so the typical cut is crossed $j\ell/m$ times, resulting in an Schmidt rank increase by $(D_0^k)^{j\ell/m} \simeq D_0^{k\ell/\sqrt{m}}$. This means that the incremental Schmidt rank per application of a term of $\hat{A}$ is $D_0^{k/\sqrt{m}}$, which can be made arbitrarily small by choosing m large enough. Finally, a recursive grouping argument shows that we do not have to pay a price in Schmidt rank proportional to the number of terms in the polynomial (which would have been catastrophic); instead, we can decompose the operator $\hat{A}^\ell$ as a sum of only $2^{\mathcal{O}(\log^2 j)}$ operators, for each of which there is a (possibly different) cut with entanglement increase of $\simeq D_0^{k\ell/\sqrt{m}}$.

Putting it all together, we have an operator $K = \hat{A}^\ell$, which increases the Schmidt rank by $D = D_I\hat{D}^\ell$, where $\hat{D} = 2^{\mathcal{O}(\log^2 j)}D_0^{k/\sqrt{m}}$, and achieves a shrinkage factor of $\Delta = \hat{\Delta}^\ell$. We now describe how to choose parameters such that $D\cdot\Delta \leq 1/2$. Clearly it suffices to ensure that $\hat{D}\cdot\hat{\Delta} \leq 1/2$, and to set $\ell = \log D_I + 1$. Roughly, this can be accomplished by setting $2^{\mathcal{O}(\log^2 j)}\hat{\Delta} \ll 1$ and $D_0^{k/\sqrt{m}} \simeq 1$. The first inequality can be satisfied by choosing $\log 1/\delta \simeq \log^2 m$, and $k \simeq \log^2 m/\epsilon$. Ignoring log factors, the second inequality entails choosing m large enough so that $D_0^{1/\epsilon\sqrt{m}}$ is small. This is accomplished by choosing $\sqrt{m} \simeq X = \log D_0/\epsilon$. The increase in Schmidt rank due to an application of the operator K is therefore dominated by the penalty $D_I = (D_0^k)^m$ for shifting back to the cut of interest. This gives a bound of $mk\log D_0 = \mathcal{O}(X^3)$.

References

[1] D. Aharonov, I. Arad, Z. Landau, and U. Vazirani. The detectability lemma and quantum gap amplification *arXiv:0811.3412v1*, 2008.

[2] Aharonov, D., Arad, A., Landau, Z., Vazirani, U., Quantum Hamiltonian complexity and the detectability lemma, arXiv:1011.3445.

[3] Aharonov, D., Arad, A., Landau, Z., Vazirani, U., The 1D Area Law and the Complexity of Quantum States: A combinatorial approach, FOCS '11: Proceedings of the 52nd Annual IEEE Symposium on Foundations of Computer Science (2011).

[4] Aharonov, D., Ben-Or, M., and Eban E., Interactive proofs for quantum computation, Proceedings of Innovations of Computer Science (ICS 2010), China, pp. 453–469 (2010)

[5] Aharonov, D., Vazirani, U., Is Quantum Mechanics Falsifiable? A computational perspective on the foundations of Quantum Mechanics.

[6] Arad, A., Landau, Z., Vazirani, U., An improved 1D area law for frustration-free systems, arXiv:1111.2970v1.

[7] Arora, S. and Barak, B., "Computational Complexity: A Modern Approach", Cambridge University Press, (2009)

[8] Bernstein, E., and Vazirani, U. (1993) Proceedings of Symposium on the Theory of Computing, 1993. Special issue on Quantum Computation of the Siam Journal of Computing, Oct. 1997.

[9] J. Eisert, M. Cramer, and M.B. Plenio. Colloquium: Area laws for the entanglement entropy. *Rev. Mod. Phys.* 82, 277306 (2010)

[10] Gottesman, D., Hastings, M B, Entanglement versus gap for one-dimensional spin systems, New Journal of Physics, vol. 12, 2 (2010), arXiv:0901.1108.

[11] Hastings, MB, An Area Law for One Dimensional Quantum Systems, JSTAT, vol. 8024, (2007) arXiv:0704.3906.

[12] Irani, S., Ground state entanglement in one-dimensional translationally invariant quantum systems, Journal of Mathematical Physics, vol. 51, 2 (2010), arXiv:0901.1107.

[13] Shor, P. W., Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer, SIAM journal on computing, **26**,5, pp. 1484–1509 (1997)

[14] Simon D.R., "On the power of quantum computation", Proceedings of the 35th Annual Symposium on Foundations of Computer Science (FOCS '94), pp 116123, (1994)

[15] Vazirani, U., "Computational constraints on scientific theories: insights from quantum computing" Workshop on the Computational World View and the Sciences, Caltech, (2007)

[16] Eckart, C. and Young, G., The approximation of one matrix by another of lower rank, Psychometrika, vol. 1, pp. 211-218 (1936).

CURRENT EVENTS BULLETIN
Previous speakers and titles

For PDF files of talks, and links to *Bulletin of the AMS* articles, see http://www.ams.org/ams/current-events-bulletin.html.

January 6, 2011 (New Orleans, LA)

Luca Trevisan, Stanford University
Khot's unique games conjecture: its consequences and the evidence for and against it

Thomas Scanlon, University of California, Berkeley
Counting special points: logic, Diophantine geometry and transcendence theory

Ulrike Tillmann, Oxford University
Spaces of graphs and surfaces

David Nadler, Northwestern University
The geometric nature of the Fundamental Lemma

January 15, 2010 (San Francisco, CA)

Ben Green, University of Cambridge
Approximate groups and their applications: work of Bourgain, Gamburd, Helfgott and Sarnak

David Wagner, University of Waterloo
Multivariate stable polynomials: theory and applications

Laura DeMarco, University of Illinois at Chicago
The conformal geometry of billiards

Michael Hopkins, Harvard University
On the Kervaire Invariant Problem

January 7, 2009 (Washington, DC)

Matthew James Emerton, Northwestern University
Topology, representation theory and arithmetic: Three-manifolds and the Langlands program

Olga Holtz, University of California, Berkeley
Compressive sensing: A paradigm shift in signal processing

Michael Hutchings, University of California, Berkeley
From Seiberg-Witten theory to closed orbits of vector fields: Taubes's proof of the Weinstein conjecture

Frank Sottile, Texas A & M University
Frontiers of reality in Schubert calculus

January 8, 2008 (San Diego, California)

Günther Uhlmann, University of Washington
Invisibility

Antonella Grassi, University of Pennsylvania
Birational Geometry: Old and New

Gregory F. Lawler, University of Chicago
Conformal Invariance and 2-d Statistical Physics

Terence C. Tao, University of California, Los Angeles
Why are Solitons Stable?

January 7, 2007 (New Orleans, Louisiana)

Robert Ghrist, University of Illinois, Urbana-Champaign
Barcodes: The persistent topology of data

Akshay Venkatesh, Courant Institute, New York University
Flows on the space of lattices: work of Einsiedler, Katok and Lindenstrauss

Izabella Laba, University of British Columbia
From harmonic analysis to arithmetic combinatorics

Barry Mazur, Harvard University
The structure of error terms in number theory and an introduction to the Sato-Tate Conjecture

January 14, 2006 (San Antonio, Texas)

Lauren Ancel Myers, University of Texas at Austin
Contact network epidemiology: Bond percolation applied to infectious disease prediction and control

Kannan Soundararajan, University of Michigan, Ann Arbor
Small gaps between prime numbers

Madhu Sudan, MIT
Probabilistically checkable proofs

Martin Golubitsky, University of Houston
Symmetry in neuroscience

January 7, 2005 (Atlanta, Georgia)

Bryna Kra, Northwestern University
The Green-Tao Theorem on primes in arithmetic progression: A dynamical point of view

Robert McEliece, California Institute of Technology
Achieving the Shannon Limit: A progress report

Dusa McDuff, SUNY at Stony Brook
Floer theory and low dimensional topology

Jerrold Marsden, Shane Ross, California Institute of Technology
New methods in celestial mechanics and mission design

László Lovász, Microsoft Corporation
Graph minors and the proof of Wagner's Conjecture

January 9, 2004 (Phoenix, Arizona)

Margaret H. Wright, Courant Institute of Mathematical Sciences, New York University
The interior-point revolution in optimization: History, recent developments and lasting consequences

Thomas C. Hales, University of Pittsburgh
What is motivic integration?

Andrew Granville, Université de Montréal
It is easy to determine whether or not a given integer is prime

John W. Morgan, Columbia University
Perelman's recent work on the classification of 3-manifolds

January 17, 2003 (Baltimore, Maryland)

Michael J. Hopkins, MIT
Homotopy theory of schemes

Ingrid Daubechies, Princeton University
Sublinear algorithms for sparse approximations with excellent odds

Edward Frenkel, University of California, Berkeley
Recent advances in the Langlands Program

Daniel Tataru, University of California, Berkeley
The wave maps equation

2012 CURRENT EVENTS BULLETIN
Committee

Hélène Barcelo, *Mathematical Sciences Research Institute*

David Eisenbud, *University of California, Berkeley, Chair*

Susan Friedlander, *University of Southern California*

Robert Ghrist, *University of Pennsylvania*

Richard Karp, *University of California, Berkeley*

Linda Keen, *City University of New York*

Isabella Laba, *University of British Columbia*

László Lovász, *Eötvös Loránd University*

David Nadler, *Northwestern University*

Thomas Scanlon, *University of California, Berkeley*

Richard Taylor, *Harvard University*

Ulrike Tillmann, *Oxford University*

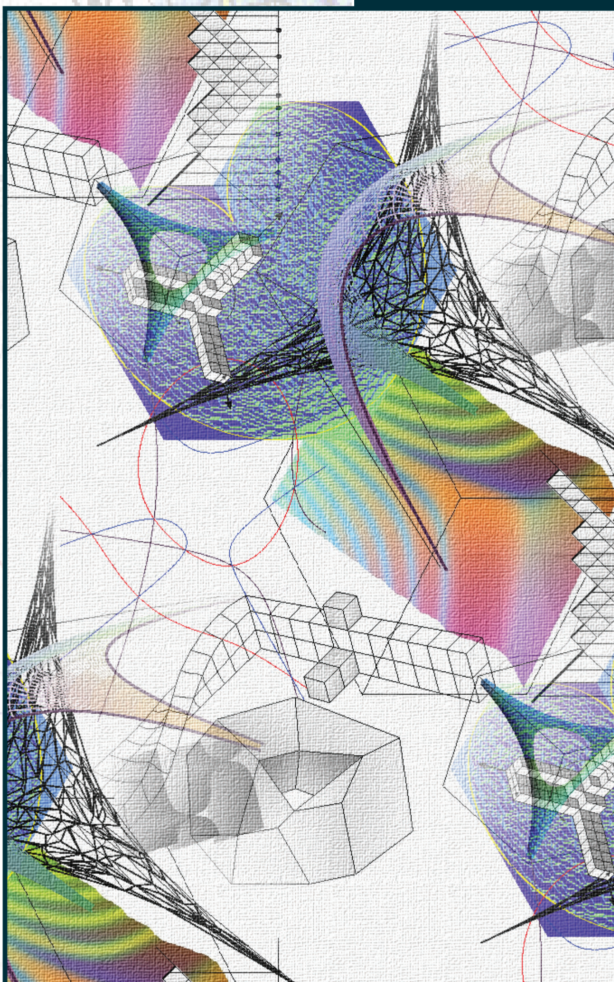
Yuri Tschinkel, *New York University*

Luca Trevisan, *Stanford University*

Akshay Venkatesh, *Stanford University*

Avi Wigderson, *Institute for Advanced Study*

Andrei Zelevinsky, *Northeastern University*



Cover graphic associated with Staffilani's talk is taken from 'Parallel Implementation of Ray Tracing Based Global Illumination Algorithms', Michal Radziszewski, M.Sc. thesis, AGH University of Science and Technology 2006.

Cover graphic associated with Vazirani talk courtesy of IBM Corporation.

The back cover graphic is reprinted courtesy of Andrei Okounkov.