American Mathematical Society

CURRENT EVENTS BULLETIN

Monday, January 12, 2015, 1:00 PM to 5:00 PM

Room 205 Henry B. Gonzalez Convention Center
Joint Mathematics Meetings, San Antonio, TX

1:00 PM

Jared S. Weinstein, *Boston University*

Exploring the Galois group of the rational numbers: Recent breakthroughs.

There's a deep analogy between number fields and curves over finite fields. Learn about Peter Scholze's work establishing a new and more direct connection.

2:00 PM

Andrea R. Nahmod, *University of Massachusetts, Amherst*

The nonlinear Schrödinger equation on tori: Integrating harmonic analysis, geometry, and probability.

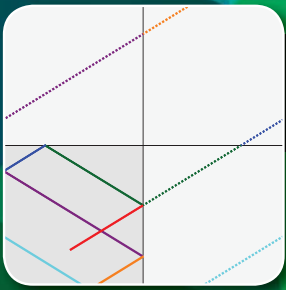
One of the most important and classical partial differential equations, in a broad modern context.

3:00 PM

Mina Aganagic, *University of California, Berkeley*

String theory and math: Why this marriage may last.

Is string theory physics? Is it math? Both? What's going on now? Hear a report from the front!



4:00 PM

Alex Wright, *Stanford University*

From rational billiards to dynamics on moduli spaces.

Dynamical systems in some old and some very modern settings. Come hear about one of the things for which Maryam Mirzakhani won the Fields Medal in 2014!

Organized by David Eisenbud, Mathematical Sciences Research Institute

AMS
American Mathematical Society

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 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. I love the idea of having an expert explain such things to me in a brief, accessible way. 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 varying, 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.

Exploring the Galois group of the rational numbers: recent breakthroughs

Jared Weinstein

1 Motivation: the splitting problem

Suppose $f(x)$ is a monic irreducible polynomial with integer coefficients. If p is a prime number, then reducing the coefficients of $f(x)$ modulo p gives a new polynomial $f_p(x)$, which may be reducible. We say that $f(x)$ is *split modulo* p if $f_p(x)$ is the product of distinct linear factors.

This article is concerned with the following simple question.

Question A. Given an irreducible polynomial $f(x)$ with integer coefficients, is there a rule which, given a prime p, determines whether $f(x)$ is split modulo p?

This motivating question is lifted almost verbatim from B. Wyman's 1972 article, [Wym72], of which the present article is merely an updated version. It may surprise the reader to learn that a large swath of modern number theory known as the *Langlands program* is dedicated to variations on the theme of Question A.

We ought to clarify what is meant by a "rule" in Question A. We are not looking for an algorithm to factor a polynomial modulo a prime. Rather we are seeking a systematic connection to some other part of mathematics. Such a rule will be called a *reciprocity law.* Our search for reciprocity laws can be rephrased as the study of a single group, the absolute Galois group of the field of rational numbers, written $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. The representation theory of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ has been particularly fruitful in answering instances of Question A. In this article we will review reciprocity laws in three successive epochs:

1. The solution of Question A in the case of $f(x) = x^2 + 1$ is due to Fermat. The solution for a general quadratic polynomial was conjectured by Euler and first proved by Gauss; this is the famous quadratic

reciprocity law. Thereafter, many other reciprocity laws followed, due to Eisenstein, Kummer, Hilbert, Artin, and others, leading up to the formulation of *class field theory* in the early 20th century. These reciprocity laws are *abelian*. They only apply to those instances of Question A where the polynomial $f(x)$ is solvable.

2. In the second half of the 20th century, a remarkable link was found between *modular forms* and 2-dimensional representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, due to Eichler, Shimura, and especially Deligne. This made it possible to find reciprocity laws for certain quintic $f(x)$ which are not solvable.

3. The 21st century has seen an explosion of results which link representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ to the geometry of *arithmetic manifolds*. We highlight Scholze's recent work [Sch13], which employs techniques invented within the past three years.

2 Fermat, Gauss, and solvable reciprocity laws

Which positive integers n are the sum of two squares? Fermat settled this question in 1640. Using his method of "descent", he showed that if a prime number p divides a sum of two squares, neither of which is divisible by p, then p is itself a sum of two squares. Also one sees from the identity $(a^2+b^2)(c^2+d^2) = (ad-bc)^2 + (ac+bd)^2$ that the property of being a sum of two squares is preserved under multiplication. From there is simple to check that n is a sum of two squares if and only if $n = p_1 \cdots p_k m^2$, where each of the primes $p_1, \dots, p_k$ are sums of two squares.

Thus we are reduced to the case that $n = p$ is prime. We wish to determine when the congruence $a^2 + b^2 \equiv 0 \pmod{p}$ has a solution for $a, b \not\equiv 0 \pmod{p}$. Recall that the ring $\mathbf{Z}/p\mathbf{Z}$ of integers modulo p is a field. After dividing by b^2 and relabeling, this becomes $x^2 + 1 \equiv 0 \pmod{p}$. Solving it is equivalent to Question A for $f(x) = x^2 + 1$.

Theorem 2.1. *Let p be an odd prime. Then $x^2 + 1 \equiv 0 \pmod{p}$ has a solution if and only if $p \equiv 1 \pmod{4}$.*

Proof. Suppose $x^2 + 1 \equiv 0 \pmod{p}$. Then $x^{p-1} = (x^2)^{(p-1)/2} \equiv (-1)^{(p-1)/2} \pmod{p}$. But by Fermat's Little Theorem, $x^{p-1} \equiv 1 \pmod{p}$, implying that $(-1)^{(p-1)/2} = 1$ and therefore $p \equiv 1 \pmod{4}$.

Conversely, suppose $p \equiv 1 \pmod{4}$. Let $x = ((p-1)/2)!$. We have $x^2 \equiv (-1)^{(p-1)/2}(p-1)! \pmod{p}$ (by pairing up n with $-n$ in the product), which by Wilson's theorem is $\equiv -1 \pmod{p}$. □

Another way of phrasing Thm. 2.1 is that x^2+1 splits modulo a prime p if and only if $p \equiv 1 \pmod 4$. (Note that modulo 2, $(x^2+1) \equiv (x+1)^2$ contains a repeated root, and so is not split as we have defined it. Given $f(x)$, the primes p for which $f_p(x)$ has a repeated factor all divide the discriminant of $f(x)$, and hence are finite in number.)

Thm. 2.1 demonstrates the simplest possible sort of reciprocity law, which namely one where the factorization of $f(x)$ modulo p is determined by a *congruence condition on* p. This is also the case for x^2+x+1, which splits modulo p if and only if $p \equiv 1 \pmod 3$. (Sketch of proof: if $p \equiv 1 \pmod 3$, and g is a generator of the cyclic group $(\mathbf{Z}/p\mathbf{Z})^\times$, then $x = g^{(p-1)/3}$ is a root.) In fact one has a congruence condition whenever $f(x)$ is a quadratic polynomial:

Theorem 2.2 (Quadratic Reciprocity)**.** *Let $f(x) = x^2 + bx + c$ be an irreducible polynomial, so that $d = b^2 - 4c$ is not a square. Then for p not dividing d, the splitting behavior of $f(x)$ modulo p is determined by the congruence class of p modulo d.*

Note that $f(x)$ factors modulo p if and only if d is congruent to a square modulo p. One introduces the *Legendre symbol* $\left(\frac{d}{p}\right)$ for any prime p and any integer d prime to p, defined to be 1 if d is a square modulo p and -1 otherwise. Thus for instance Thm. 2.1 is the statement that $\left(\frac{-1}{p}\right) = (-1)^{(p-1)/2}$. In elementary number theory texts one learns a more precise version of Thm. 2.2: if $q \neq p$ is an odd prime then

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{\frac{p-1}{2}\frac{q-1}{2}},$$

which implies that the splitting behavior of $x^2 - q$ modulo p depends on the congruence class of p modulo $4q$. The symmetry between p and q explains the term "reciprocity" for such laws.

Let us return for a moment to Fermat's theorem on sums of squares. Could it apply to the representation of integers by other quadratic forms, such as a^2+5b^2? Thm. 2.2 shows that a prime $p \neq 2, 5$ divides an integer of the form $x^2 + 5$ if and only if p satisfies a congruence condition modulo 20, which happens to be the condition that $p \equiv 1, 3, 7, 9 \pmod{20}$. But such primes (for instance 7) are not necessarily of the form $a^2 + 5b^2$. It turns out that Fermat's method of descent fails in this context; phrased in modern terms, the culprit is the failure of $\mathbf{Z}[\sqrt{-5}]$ to be a principal ideal domain. In fact $p = a^2 + 5b^2$ if and only if $p \equiv 1, 9 \pmod{20}$. For a fascinating account of the problem of classifying primes of the form $x^2 + ny^2$, see Cox's book of the same title, [Cox89].

What about polynomials $f(x)$ of higher degree? A little experimentation will reveal that the factorization behavior of a "random" cubic will be influenced, but not completely determined by, a congruence condition modulo p. There are special cases: for instance the polynomial $x^3 + x^2 - 2x + 1$ splits modulo p if and only if $p \equiv \pm 1 \pmod 7$. So when is the splitting behavior of a polynomial determined by congruence conditions?

For a clue, let $m \geq 1$ and consider the polynomial $x^m - 1$. It splits modulo p if and only if the multiplicative group $(\mathbf{Z}/p\mathbf{Z})^\times$ contains m distinct elements of order dividing m. Since $(\mathbf{Z}/p\mathbf{Z})^\times$ is a cyclic group of order $p-1$, this happens exactly when $p \equiv 1 \pmod m$. This logic extends to show that the splitting behavior of $f(x)$ is determined by congruence conditions whenever $f(x)$ is a *cyclotomic polynomial*, that is, an irreducible divisor of some $x^m - 1$. Using some algebraic number theory, one even gets congruence conditions for those $f(x)$ whose roots are contained in a *cyclotomic field* $\mathbf{Q}(\zeta_m)$, where $\zeta_m = \exp(2\pi i/m)$. (The roots of x^3+x^2-2x+1, for instance, are $2\cos(2\pi k/7)$, where $k = 1, 2, 3$.)

What would a reciprocity law look like if it isn't a congruence condition? As with the quadratic reciprocity law, the following theorem was conjectured by Euler and proved by Gauss.

Theorem 2.3. *The polynomial $x^4 - 2$ splits modulo p if and only if $p = a^2 + 64b^2$ for integers a and b.*

Unlike the case of $a^2 + b^2$, the representation of p by the quadratic form a^2+64b^2 is not determined by a congruence condition on p. But in fact there is a disguised congruence condition in Thm. 2.3, which was well known to Gauss. If x^4-2 splits modulo p, then the quotient of two of its roots in $\mathbf{Z}/p\mathbf{Z}$ must be a square root of -1, so that by Thm. 2.1 we have $p \equiv 1 \pmod 4$. By Fermat's theorem $p = a^2 + b^2$. Without loss of generality, assume that a is odd and b is even. We now pass to the ring $\mathbf{Z}[i]$ of *Gaussian integers*, the subring of $\mathbf{C}$ consisting of those $a + bi$ with $a, b \in \mathbf{Z}$. In $\mathbf{Z}[i]$, p is no longer prime; we have $p = \omega\overline{\omega}$, where $\omega = a + bi$. Thm. 2.3 says that $x^4 - 2$ splits modulo p if and only if ω is congruent to a rational integer modulo 8. Indeed, this condition translates into the statement that $b = 8b_0$ for an integer b_0, in which case $p = a^2 + 64b_0^2$. Thus the splitting behavior of $x^4 - 2$ modulo a prime $p \equiv 1 \pmod 4$ is determined by a congruence condition on a prime of $\mathbf{Z}[i]$ which divides p.

At this point it is appropriate to introduce some basic notions from algebraic number theory. If $f(x)$ is an irreducible polynomial with rational coefficients, then $K = \mathbf{Q}[x]/f(x)$ is an *algebraic number field.* Let $\mathcal{O}_K$ be the integral closure of $\mathbf{Z}$ in K. It is a basic fact of algebraic number theory that $\mathcal{O}_K$ is a *Dedekind domain.* This means that even though $\mathcal{O}_K$ may not have the property of unique factorization into prime elements, it does have the corresponding property for *ideals.* As an example, in the ring $\mathbf{Z}[\sqrt{-5}]$, the element 6 admits the two factorizations $2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5})$ into irreducible elements, none of which divide any other. However, the ideal $6\mathcal{O}_K$ admits a factorization into prime ideals in one way only: $6\mathcal{O}_K = (2, 1 + \sqrt{-5})^2(3, 1 + \sqrt{-5})(3, 1 - \sqrt{-5})$.

If p is a prime number, then $p\mathcal{O}_K = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_k^{e_k}$ is a product of powers of distinct prime ideals. In fact we have $e_i = 1$ for all i unless p belongs to a finite list of *ramified primes.* For each i, the quotient $\mathcal{O}_K/\mathfrak{p}_i$ is a finite field extension of $\mathbf{F}_p$; the *degree* of $\mathfrak{p}_i$ is defined as the degree of this extension. We say that p is *split* in K if $p\mathcal{O}_K$ is the product of distinct prime ideals of degree 1. These concepts all have relative notions with respect to an extension of number fields L/K.

The "correct" generalization of Question A is then:

Question B. Let L/K be an extension of number fields. Is there a rule for determining when a prime ideal of K is split in L?

Question B is inextricably linked with Galois theory. Recall that if K is a field, $f(x)$ an irreducible polynomial with coefficients in K, and $L = K[x]/f(x)$, then L/K is *Galois* if it is normal (meaning that L contains all roots of $f(x)$) and separable (meaning that $f(x)$ contains no repeated roots; this is automatically satisfied in characteristic 0). If L/K is Galois, one defines the Galois group $\mathrm{Gal}(L/K)$ as the group of field automorphisms of L which act as the identity on K. Its cardinality is the same as the degree of L/K.

As an example, the splitting field of the polynomial $x^4 - 2$ over $\mathbf{Q}$ is $L = \mathbf{Q}(i, 2^{1/4})$. The Galois group $\mathrm{Gal}(L/\mathbf{Q})$ is the dihedral group of order 8, generated by two elements σ and τ, defined by the table

$$\begin{array}{ll} \sigma(2^{1/4}) = i2^{1/4}, & \tau(2^{1/4}) = 2^{1/4}, \\ \sigma(i) = i, & \tau(i) = i. \end{array}$$

These generators satisfy the relations $\sigma^4 = 1$, $\tau^2 = 1$, and $\tau\sigma\tau = \sigma^{-1}$.

If L/K is Galois and $\mathfrak{p}$ is a prime ideal of K which is unramified in L, let $\mathfrak{P}$ be a prime ideal of L dividing $\mathfrak{p}$. The number of elements of $\mathcal{O}_K/\mathfrak{p}$ is denoted $N\mathfrak{p}$. The Galois group $\mathrm{Gal}((\mathcal{O}_L/\mathfrak{P})/(\mathcal{O}_K/\mathfrak{p}))$ is cyclic of order equal to the degree of $\mathfrak{P}|\mathfrak{p}$, with a distinguished generator $x \mapsto x^{N\mathfrak{p}}$. It turns out that there exists a unique element $\mathrm{Frob}_{\mathfrak{P}|\mathfrak{p}} \in \mathrm{Gal}(L/K)$, the *Frobenius element*, which lifts this generator. That is:

$$\mathrm{Frob}_{\mathfrak{P}|\mathfrak{p}}(x) \equiv x^{N\mathfrak{p}} \pmod{\mathfrak{P}}$$

for all $x \in \mathcal{O}_L$. If a different prime $\mathfrak{P}'$ dividing $\mathfrak{p}$ is chosen, the resulting Frobenii $\mathrm{Frob}_{\mathfrak{P}}$ and $\mathrm{Frob}_{\mathfrak{P}'}$ are conjugate in $\mathrm{Gal}(L/K)$. Thus one can talk about $\mathrm{Frob}_{\mathfrak{p}}$ as a well-defined *conjugacy class* in $\mathrm{Gal}(L/K)$. An important observation is that

$$\mathrm{Frob}_{\mathfrak{p}} = 1 \text{ if and only if } \mathfrak{p} \text{ is split in } L.$$

Class field theory refers to the complete solution of Question B in the case that $\mathrm{Gal}(L/K)$ is *abelian*. Roughly speaking, it predicts that for a prime $\mathfrak{p}$ of K which is unramified in L, the element $\mathrm{Frob}_{\mathfrak{p}} \in \mathrm{Gal}(L/K)$ is determined by "congruence conditions" on $\mathfrak{p}$, where the modulus is an ideal $\mathfrak{f}$ of K divisible only by ramified primes. Rather than making this precise, we spell out the example relevant to Gauss' $a^2 + 64b^2$ theorem.

Example 2.4. Let $K = \mathbf{Q}(i)$ and $L = K(2^{1/4})$, so that $\mathrm{Gal}(L/K)$ is a cyclic group of order 4 generated by σ. Here $\mathfrak{f} = (8)$. Class field theory shows that if $\mathfrak{p} = (\omega)$ is a prime of K which is relatively prime to 2, then $\mathrm{Frob}_{\mathfrak{p}}$ is determined by the image of ω in $(\mathbf{Z}[i]/8\mathbf{Z}[i])^{\times}$. There exists a unique surjective homomorphism

$$r\colon (\mathbf{Z}[i]/8\mathbf{Z}[i])^{\times} \to \mathrm{Gal}(L/K)$$

which is trivial on $(\mathbf{Z}/8\mathbf{Z})^{\times}$ and i, and which sends $1 + 2i$ to σ. Then $\mathrm{Frob}_{\mathfrak{p}} = r(\omega)$. As a result, $\mathrm{Frob}_{\mathfrak{p}} = 1$ if and only if $\mathfrak{p} = (p)$ for $p \equiv 3 \pmod 4$ or else if $\mathfrak{p} = (a + 8bi)$ with $p = a^2 + 64b^2$ prime.

Class field theory allows us to answer Question B in the case that the polynomial $f(x)$ is *solvable*, meaning that its roots lie in a tower of number fields $\mathbf{Q} = K_0 \subset K_1 \subset \cdots \subset K_n = K$, with each K_{i+1}/K_i abelian. A prime p splits in K if and only if p splits in K_1, a prime above p in K_1 splits in K_2, and so on, with each splitting being governed by congruences. In example 2.4, the relevant tower was $\mathbf{Q} \subset \mathbf{Q}(i) \subset \mathbf{Q}(i, 2^{1/4})$.

It is immensely useful to talk about all of the extensions of $\mathbf{Q}$ at once, as living in an algebraic closure $\overline{\mathbf{Q}}$. One considers the *absolute Galois group* $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$; this is just the automorphism group of the field $\overline{\mathbf{Q}}$. More to the point, we have

$$\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) = \varprojlim_K \mathrm{Gal}(K/\mathbf{Q}),$$

where K ranges over finite Galois extensions of $\mathbf{Q}$. Written this way, $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ becomes a topological group, whose open subgroups are exactly the subgroups $\mathrm{Gal}(\overline{\mathbf{Q}}/K)$ consisting of automorphisms which act trivially on a finite extension $K/\mathbf{Q}$. Focus can then shift from particular number fields $K/\mathbf{Q}$ to *representations* of the group $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$.

Example 2.5. The group D_8 has a two-dimensional representation which sends σ to $\begin{pmatrix} & 1 \\ -1 & \end{pmatrix}$ and τ to $\begin{pmatrix} & 1 \\ 1 & \end{pmatrix}$. The character of this representation is 2 on the identity of D_8, -2 on σ^2, and 0 everywhere else. Thus we can construct a 2-dimensional Galois representation

$$\rho\colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(\mathbf{C})$$

which factors through $\mathrm{Gal}(\mathbf{Q}(i, 2^{1/4})/\mathbf{Q})$. This will have the property that for all odd primes p, ρ is unramified at p, meaning that the fixed field of the kernel of ρ is unramified at p. Consequently $\rho(\mathrm{Frob}_p)$ is well defined. We have

$$\mathrm{tr}\,\rho(\mathrm{Frob}_p) = \begin{cases} 2, & p = a^2 + 64b^2, \\ -2, & p = a^2 + 16b^2,\ b \text{ odd}, \\ 0, & \text{otherwise.} \end{cases} \tag{2.1}$$

3 Elliptic modular forms

The theory of modular forms developed in a context completely unrelated to the arithmetic questions posed in this article. They arose in relation to the elliptic functions investigated by Abel and Jacobi in the early 19th century, which in turn arose in association with finding the arc length of an ellipse. For an introduction to the subject, we recommend the book [Ser73].

In brief, a modular form is a certain kind of holomorphic function on the upper half-plane $\mathcal{H} = \{\tau | \mathrm{Im}\ \tau > 0\}$, which we view simultaneously as a complex manifold and as a Riemannian manifold equipped with a hyperbolic metric. The automorphism group of $\mathcal{H}$ is the group of *Möbius transformations* $z \mapsto (az+b)/(cz+d)$, where $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbf{R})$. In brief, a modular form is a holomorphic function $f(\tau)$ on $\mathcal{H}$ which transforms in a certain way under a subgroup of $\mathrm{SL}_2(\mathbf{R})$.

For a nonzero integer N, let $\Gamma_0(N)$ denote the subgroup of $\mathrm{SL}_2(\mathbf{Z})$ consisting of matrices which are upper-triangular modulo N.

Definition 3.1. Let $N, k \geq 1$ be integers, and let $\chi\colon (\mathbf{Z}/N\mathbf{Z})^\times \to \mathbf{C}^\times$ be a homomorphism. A *modular form of weight k, level N and character χ* is a holomorphic function g on $\mathcal{H}$ which satisfies

$$g\left(\frac{a\tau+b}{c\tau+d}\right) = \chi(d)(c\tau+d)^k g(\tau)$$

for all $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N)$, and which is "holomorphic at the cusps".

In particular $g(\tau+1) = g(\tau)$, so that g is a function of the parameter $q = e^{2\pi i\tau}$. "Holomorphic at the cusps" means that the Fourier expansion of $g(\tau)$, *a priori* a series of the form $\sum_{n\in\mathbf{Z}} a_n(g)q^n$, has $a_n(g) = 0$ for $n < 0$; a similar condition is imposed for all functions $g((a\tau+b)/(c\tau+d))$ with $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbf{Z})$. We say g is a *cusp form* if it is zero at the cusps, meaning that $a_0(g) = 0$ as well.

Some modular forms known as *theta functions* arise from sums involving rings of integers in quadratic fields, such as $\mathbf{Z}[i]$. Suppose $\mathfrak{f} \subset \mathbf{Z}[i]$ is an ideal, and $\chi\colon (\mathbf{Z}[i]/\mathfrak{f})^\times \to \mathbf{C}^\times$ is a nontrivial homomorphism. Extend χ to a function on $\mathbf{Z}[i]$ by declaring it 0 on elements which are not prime to $\mathfrak{f}$. Then

$$\theta_\chi(\tau) = \frac{1}{4}\sum_{\alpha\in\mathbf{Z}[i]} \chi(\alpha)q^{N(\alpha)}$$

is a modular form of weight 1 and level $4N(\mathfrak{f})$.

Example 3.2. Let $\chi\colon (\mathbf{Z}[i]/8\mathbf{Z}[i])^\times \to \mathbf{C}^\times$ be the homomorphism which is trivial on i and $(\mathbf{Z}/8\mathbf{Z})^\times$, and which sends $1+2i$ to i. Then θ_χ is a modular form of weight 1; for a prime p, its pth Fourier coefficient is

$$a_p(\theta_\chi) = \begin{cases} \chi(a+bi) + \chi(a-bi), & p \equiv 1 \pmod 4,\ p = a^2+b^2, \\ 0, & p \equiv 3 \pmod 4 \text{ or } p = 2 \end{cases}$$

Now if $p \equiv 1 \pmod 4$, we can write $p = a^2+b^2$ with a odd and b even. A short calculation shows that

$$a_p(\theta_\chi) = \begin{cases} 2, & 8|b \\ -2, & 4|b \text{ but } 8\nmid b, \\ 0, & 4\nmid b. \end{cases}$$

Referring back to Eq. (2.1), we find that

$$a_p(\theta_\chi) = \operatorname{tr} \rho(\mathrm{Frob}_p)$$

for the Galois representation ρ constructed in Example 2.5. This equation hints at an extraordinary relationship between modular forms and representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$.

The space of cusp forms of weight N and level k is finite-dimensional. For each prime p not dividing N one defines a *Hecke operator* T_p on this space, which has the following effect on q-expansions:

$$T_p g(\tau) = \sum_{n\geq 0} c_{pn} q^n + p^{k-1} \sum_{n \geq 0} c_n q^{pn}.$$

(There are similar operators for primes dividing N.) These operators commute with one another, and so it makes sense to attempt to diagonalize them simultaneously. A modular form is an *eigenform* if it is an eigenvector for all Hecke operators. If $g = \sum_{n\geq 1} a_n(g) q^n$ is a cuspidal eigenform with $a_1(g) = 1$, then the eigenvalue of T_p on g is just $a_p(g)$.

Theorem 3.3. *Let $g(\tau) = \sum_{n\geq 1} a_n(f) q^n$ be a cuspidal eigenform of weight k and level N with character χ. Let E a number field containing the $a_n(g)$.*

1. *Suppose $k \geq 2$. Then for all prime ideals λ of $\mathcal{O}_E$ there exists an odd irreducible Galois representation*

   $$\rho_{f,\lambda} \colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(\mathcal{O}_{E,\lambda})$$

   such that for all p prime to $N\lambda$, $\rho_{g,\lambda}$ is unramified at p, and the characteristic polynomial of $\rho_{g,\lambda}(\mathrm{Frob}_p)$ is $x^2 - a_p(g)x + \chi(p)p^{k-1}$.

2. *Suppose $k = 1$. Then there exists an odd irreducible Galois representation*

   $$\rho_f \colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(\mathbf{C})$$

   such that for all p prime to N, ρ_g is unramified at p, and the characteristic polynomial of $\rho_g(\mathrm{Frob}_p)$ is $x^2 - a_p(g)x + \chi(p)$.

These two statements are proved in [Del71] and [DS74], respectively. In the first statement, $\mathcal{O}_{E,\lambda}$ is the completion of $\mathcal{O}_E$ with respect to the ideal λ; the image of $\rho_{g,\lambda}$ is infinite. In the second statement, where $k = 1$, the image of ρ_g is finite. A Galois representation ρ is *odd* if $\det \rho(c) = -1$, where c is complex conjugation.

Example 3.4 (An icosahedral form)**.** The following example is due to Joe Bühler, [Buh78]. Let

$$f(x) = x^5 + 10x^3 - 10x^2 + 35x - 18.$$

The discriminant of $f(x)$ is $2^6 5^8 11^2$, a square number. This means that the Galois group of f is contained in the icosahedral group A_5; in fact it equals A_5. The group A_5 doesn't have any irreducible 2-dimensional representations, but there exists a 4-fold cover $\widetilde{A}_5$ which does. It can be shown that there is an odd irreducible representation $\rho\colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(\mathbf{C})$ whose image is $\widetilde{A}_5$, such that in the diagram

$$\begin{array}{ccc} & & \mathrm{GL}_2(\mathbf{C}) \\ & \overset{\rho}{\nearrow} & \downarrow \\ \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) & & \\ & \underset{P\rho}{\searrow} & \\ & & \mathrm{PGL}_2(\mathbf{C}) \end{array}$$

the fixed field of the kernel of $P\rho$ is the splitting field K of f. Artin's conjecture predicts a weight 1 cusp form $g(\tau)$ of level 800 and character χ associated to ρ, where χ is a Dirichlet character of conductor 100 and order 10. Indeed there is one:

$$g(\tau) = q - iq^3 - ijq^7 - q^9 + jq^{13} + (i - ij)q^{19} - jq^{21} + \dots,$$

where $i = \sqrt{-1}$ and $j = (1+\sqrt{5})/2$. A prime $p \neq 2, 5$ splits in K if and only if $\rho(\mathrm{Frob}_p)$ is a scalar matrix. Since $\rho(\mathrm{Frob}_p)$ has finite order, it is semisimple, and therefore it is scalar if and only if its characteristic polynomial has zero discriminant. But the characteristic polynomial is $x^2 - a_p(g)x + \chi(p)$, with discriminant $a_p(g)^2 - 4\chi(p)$. Therefore we have the following answer to Question B: p splits in K if and only if $a_p(g)^2 = 4\chi(p)$.

Example 3.5 (The Ramanujan Δ-function)**.** The product

$$\Delta(\tau) = q \prod_{n\geq 1} (1 - q^n)^{24} = \sum_{n\geq 1} \tau(n) q^n$$

defines a cuspidal eigenform of weight 12 and level 1, and so Thm. 3.3 associates to it an ℓ-adic representation $\rho_{\Delta,\ell}$ for all primes ℓ. This can be reduced modulo ℓ to obtain a mod ℓ Galois representation $\overline{\rho}_{\Delta,\ell}\colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to$

$\mathrm{GL}_2(\mathbf{Z}/\ell\mathbf{Z})$, whose kernel cuts out a number field which is ramified only at ℓ. It is a difficult computational problem to compute this number field. For some small primes ℓ this has been carried out in [Bos11], at least for the associated projective representation $P\overline{\rho}_{\Delta,\ell}\colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{PGL}_2(\mathbf{Z}/\ell\mathbf{Z})$. For instance if $\ell = 11$, the fixed field of the kernel of $P\overline{\rho}_{\Delta,\ell}$ is the splitting field of

$$\begin{aligned} f(x) &= x^{12} - 4x^{11} + 55x^9 - 165x^8 + 264x^7 - 341x^6 \\ &\quad +330x^5 - 165x^4 - 55x^3 + 99x^2 - 41x - 111. \end{aligned}$$

From this we can conclude the following reciprocity law, valid for for almost all p: if $f(x)$ splits modulo p then $\tau(p)^2 \equiv 4p \pmod{11}$.

4 The cohomology of arithmetic manifolds

Modular forms are holomorphic forms on $\mathcal{H}$ which admit symmetries with respect to a finite-index subgroup $\Gamma_0(N) \subset \mathrm{SL}_2(\mathbf{Z})$. It stands to reason that they correspond to objects defined on the quotient $Y_0(N) = \Gamma_0(N)\backslash\mathcal{H}$, a (non-compact) Riemann surface. For instance, if $f(\tau)$ is a modular form of weight 2, level N, and trivial character, then $f(\tau)d\tau$ is *invariant* under $\Gamma_0(N)$, and so descends to a differential form on $Y_0(N)$. If $f(\tau)$ happens to be a cusp form, then $f(\tau)d\tau$ extends to a differential form on $X_0(N)$, the smooth compactification of $Y_0(N)$. In fact the space of cusp forms of weight 2 is isomorphic to the space $H^0(X_0(N), \Omega^1_{X_0(N)/\mathbf{C}})$ of holomorphic differential forms on $X_0(N)$.

On the other hand, the Hodge decomposition for the compact Riemann surface $X_0(N)$ shows that the singular cohomology $H^1(X_0(N), \mathbf{C})$ is the direct sum of $H^0(X_0(N), \Omega^1_{X_0(N)/\mathbf{C}})$ and its complex conjugate. All of these spaces come equipped with actions by the Hecke operators T_p. The conclusion is that systems of Hecke eigenvalues coming from weight 2 forms are present already in the singular cohomology $H^1(X_0(N), \mathbf{C})$. (There is a similar statement for forms of higher weight; one replaces the $\mathbf{C}$ in $H^1(X_0(N), \mathbf{C})$ with a non-constant coefficient system.) Therefore one could have phrased Thm. 3.3 (at least the part pertaining to forms of weight $k \geq 2$) in terms of *Hecke eigenclasses* in the singular cohomology of $X_0(N)$. (Equivalently, one can phrase it in terms of the group cohomology $H^1(\Gamma_0(N), \mathbf{C})$.)

One might seek to generalize Thm. 3.3 to higher dimension as follows. The upper halfplane $\mathcal{H}$ is the quotient $\mathrm{SL}_2(\mathbf{R})/\mathrm{SO}(2)$, so let us put $\mathcal{H}_n = \mathrm{SL}_n(\mathbf{R})/\mathrm{SO}(n)$; this is a manifold with a left action by $\mathrm{SL}_n(\mathbf{R})$. Let us abuse notation and write $\Gamma_0(N) \subset \mathrm{SL}_n(\mathbf{Z})$ for the subgroup of matrices whose first column is congruent to $(*, 0, \ldots, 0)$ modulo N. Then one can form the quotient $\Gamma_0(N)\backslash\mathcal{H}_n$, an *arithmetic manifold*. The cohomology $H^j(\Gamma_0(N)\backslash\mathcal{H}_n, \mathbf{C})$ admits actions by Hecke operators. For each prime $p \nmid N$, there isn't just one Hecke operator T_p but rather $n-1$ operators $T_{p,1}, \ldots, T_{p,n-1}$. These operators commute with one another, and so one can talk about *eigenclasses* in $H^j(\Gamma_0(N)\backslash\mathcal{H}_n, \mathbf{C})$ for all the $T_{p,i}$. Do these correspond to n-dimensional Galois representations?

The main obstacle to generalizing Thm. 3.3 is this: in the $n = 2$ case, $\mathcal{H} = \mathcal{H}_2$ is a complex manifold and $X_0(N)$ is an algebraic curve which even admits a model over the rational numbers. This fact is critical for the construction of Galois representations, which live in the ℓ-adic étale cohomology of $X_0(N)$. However if $n > 2$, $\mathcal{H}_n$ isn't even a complex manifold, and so no quotient of it is going to be an algebraic variety. (For instance, $\mathcal{H}_3$ has dimension 5, which is odd.) Nonetheless, the following theorem was announced in 2012:

Theorem 4.1 ([MHT],[Sch13])**.** *Let g be a Hecke eigenclass in the singular cohomology $H^j(\Gamma_0(N)\backslash\mathcal{H}_n, \mathbf{C})$, and let $a_{p,i}(g)$ be the eigenvalue of $T_{p,i}$ on g for $p \nmid N$ prime. Let E be a number field containing all the $a_{p,i}(g)$, and let λ be a prime of E. Then there exists a continuous semisimple Galois representation*

$$\rho_{g,\lambda} \colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_n(E_\lambda)$$

which is associated to g in the sense that for all primes p which are prime to $N\lambda$, $\rho_{g,\lambda}$ is unramified at Frob_p*, and the characteristic polynomial of $\rho_{g,\lambda}(\mathrm{Frob}_p)$ is*

$$x^n + \sum_{k=1}^{n-1} (-1)^k p^{k(k-1)/2} a_{p,k}(g) x^{n-k} + (-1)^n p^{n(n-1)/2}.$$

The results of [MHT] and [Sch13] are rather stronger than this: they show that "every cuspidal regular algebraic automorphic representation of GL_n over a totally real or CM field F has an associated Galois representation". It would take us rather far afield to define the terms in the preceding sentence, but suffice it to say that Thm. 4.1 is the special case $F = \mathbf{Q}$.

In fact the results of [Sch13] are stronger still. Thm. 4.1 concerns the singular cohomology $H^j(\Gamma_0(N)\backslash\mathcal{H}_n, \mathbf{C})$ with complex coefficients, but we could also have considered the integral cohomology $H^j(\Gamma_0(N)\backslash\mathcal{H}_n, \mathbf{Z})$, a finitely generated abelian group equipped with the action of Hecke operators $T_{p,i}$. When $n = 2$, $Y_0(N) = \Gamma_0(N)\backslash\mathcal{H}$ is a surface; the integral cohomology groups of a surface are known to be torsion-free. But for $n > 2$, the cohomology $H^j(\Gamma_0(N)\backslash\mathcal{H}_n, \mathbf{Z})$ can contain a large torsion subgroup. This torsion subgroup is also preserved by the Hecke operators, and one may ask whether the *torsion* eigenclasses have corresponding *torsion* Galois representations. In fact they do:

Theorem 4.2 ([Sch13])**.** *Let ℓ be prime, and let g be a Hecke eigenclass in $H^j(\Gamma_0(N)\backslash\mathcal{H}_n, \overline{\mathbf{F}}_\ell)$. Then there exists a continuous semisimple Galois representation*

$$\rho_g\colon \operatorname{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \operatorname{GL}_n(\overline{\mathbf{F}}_\ell)$$

which is associated to g in the sense of Thm. 4.1.

In prior years, Ash and others had developed Serre-type conjectures which predict that every Galois representation $\rho\colon \operatorname{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \operatorname{GL}_3(\overline{\mathbf{F}}_\ell)$ has a corresponding Hecke eigenclass g in $H^j(\Gamma_0(N), V)$ for an appropriate choice of N, j, and an $\overline{\mathbf{F}}_\ell[\operatorname{SL}_3(\mathbf{Z})]$-module V. See for instance [ADP02], which offers a great deal of numerical evidence in the form of pairs (ρ, g), where ρ and g appear to be associated in the sense that the characteristic polynomial of $\rho(\operatorname{Frob}_p)$ is as in Thm. 4.1 for the first few primes p. Thm. 4.2 shows that there really is a ρ' attached to each g, and then examination of sufficiently many small primes is enough to prove that $\rho = \rho'$. In those cases one has a reciprocity law for the fixed field K of $\ker \rho$: the splitting behavor of primes in K is governed by the Hecke eigenvalues of the eigenclass g.

Thm. 4.2 is truly spectacular. It links Galois representations with torsion classes in the cohomology of arithmetic manifolds, which don't necessarily come from automorphic representations (these had been the starting point for most generalizations of Thm. 3.3). The method of proof is striking. The first step, an idea suggested by Clozel, is to show that the arithmetic manifold $\Gamma_0(N)\backslash\mathcal{H}_n$ appears "at the boundary" of a Shimura variety Sh_N, which implies that an eigenclass g as in Thm. 4.2 shows up as an eigenclass in the cohomology $H^i(\operatorname{Sh}_N, \overline{\mathbf{F}}_\ell)$. The next step is to show that there exists a cuspidal eigenform on some higher level Shimura variety $\operatorname{Sh}_{N\ell^m}$ whose mod ℓ eigenvalues match those of g. This required working with a Shimura variety $\operatorname{Sh}_{N\ell^\infty}$ *at infinite level*.

The space $\operatorname{Sh}_{N\ell^\infty}$ isn't an algebraic variety or even a manifold. Rather, it is a fractal-like entity known as a *perfectoid space*. Perfectoid spaces were also devised by Scholze in [Sch12] for completely different ends; in this application, Scholze proves a comparison theorem for perfectoid spaces which links mod ℓ étale cohomology and coherent cohomology. This comparison theorem furnishes the required cusp form, and with it the Galois representation.

Despite these remarkable advances, there are still major unsolved problems in our search for reciprocity laws. We conclude with a list of open questions.

- By Thm. 3.3, modular forms of weight 1 correspond to *odd* 2-dimensional Galois representations $\rho\colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(\mathbf{C})$. The general philosophy of Langlands predicts that *even* 2-dimensional Galois representations ought to correspond to algebraic Maass forms. A *Maass form* is an analytic (not holomorphic) function on $\Gamma_0(N)\backslash\mathcal{H}$ which is an eigenvector for the Laplacian $-y^{-2}(\partial^2/\partial x^2 + \partial^2/\partial y^2)$; it is *algebraic* if the eigenvalue is $1/4$. Nobody has any idea how the correspondence works in either direction, outside of the "solvable" cases.
- If $\rho\colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_n(\overline{\mathbf{Q}}_\ell)$ is an irreducible Galois representation, subject to a suitable condition at ℓ, must it arise from an eigenclass g as in Thm. 4.1? This is a generalization of the Fontaine-Mazur conjecture, [FM95], which for $n = 2$ was proved by Kisin, [Kis09], save some exceptional cases. This is a question of *modularity* of Galois representations, of which there is a large amount of literature. We remark in passing that a special case of modularity was key to Wiles' attack on Fermat's last theorem.
- If $\rho\colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_n(\overline{\mathbf{F}}_\ell)$ is an irreducible Galois representation, must it arise from an eigenclass g as in Thm. 4.2? The case of $n = 2$ and ρ odd is *Serre's conjecture*, proved by Khare and Wintenberger, [KW09].

References

[ADP02] Avner Ash, Darrin Doud, and David Pollack, *Galois representations with conjectural connections to arithmetic cohomology*, Duke Math. J. **112** (2002), no. 3, 521–579. MR 1896473 (2003g:11055)

[Bos11] Johan Bosman, *Polynomials for projective representations of level one forms*, Computational aspects of modular forms and Galois representations, Ann. of Math. Stud., vol. 176, Princeton Univ. Press, Princeton, NJ, 2011, pp. 159–172. MR 2857091

[Buh78] Joe P. Buhler, *Icosahedral Galois representations*, Lecture Notes in Mathematics, Vol. 654, Springer-Verlag, Berlin-New York, 1978. MR 0506171 (58 #22019)

[Cox89] David A. Cox, *Primes of the form $x^2 + ny^2$*, A Wiley-Interscience Publication, John Wiley & Sons, Inc., New York, 1989, Fermat, class field theory and complex multiplication. MR 1028322 (90m:11016)

[Del71] Pierre Deligne, *Formes modulaires et représentations l-adiques*, Séminaire Bourbaki. Vol. 1968/69: Exposés 347–363, Lecture Notes in Math., vol. 175, Springer, Berlin, 1971, pp. Exp. No. 355, 139–172. MR 3077124

[DS74] Pierre Deligne and Jean-Pierre Serre, *Formes modulaires de poids 1*, Ann. Sci. École Norm. Sup. (4) **7** (1974), 507–530 (1975). MR 0379379 (52 #284)

[FM95] Jean-Marc Fontaine and Barry Mazur, *Geometric Galois representations*, Elliptic curves, modular forms, & Fermat's last theorem (Hong Kong, 1993), Ser. Number Theory, I, Int. Press, Cambridge, MA, 1995, pp. 41–78. MR 1363495 (96h:11049)

[Kis09] Mark Kisin, *The Fontaine-Mazur conjecture for* GL_2, J. Amer. Math. Soc. **22** (2009), no. 3, 641–690. MR 2505297 (2010j:11084)

[KW09] Chandrashekhar Khare and Jean-Pierre Wintenberger, *On Serre's conjecture for 2-dimensional mod p representations of* $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, Ann. of Math. (2) **169** (2009), no. 1, 229–253. MR 2480604 (2009m:11077)

[MHT] Richard Taylor Michael Harris, Kai-Wen Lan and Jack Thorne, *On the rigid cohomology of certain Shimura varieties*, Preprint.

[Sch12] Peter Scholze, *Perfectoid spaces*, Publ. Math. Inst. Hautes Études Sci. **116** (2012), 245–313. MR 3090258

[Sch13] ______, *Torsion in the cohomology of locally symmetric spaces*, Preprint, Bonn, 2013.

[Ser73] J.-P. Serre, *A course in arithmetic*, Springer-Verlag, New York-Heidelberg, 1973, Translated from the French, Graduate Texts in Mathematics, No. 7. MR 0344216 (49 #8956)

[Wym72] B. F. Wyman, *What is a reciprocity law?*, Amer. Math. Monthly **79** (1972), 571–586; correction, ibid. 80 (1973), 281. MR 0308084 (46 #7199)

THE NONLINEAR SCHRÖDINGER EQUATION ON TORI: INTEGRATING HARMONIC ANALYSIS, GEOMETRY AND PROBABILITY

ANDREA R. NAHMOD

Abstract. The field of nonlinear dispersive and wave equations has undergone significant progress in the last twenty years thanks to the influx of tools and ideas from nonlinear Fourier and harmonic analysis, geometry, analytic number theory and most recently probability, into the existing functional analytic methods. In these lectures we concentrate on the semilinear Schrödinger equation defined on tori and discuss the most important developments in the analysis of these equations. In particular, we will discuss recent work by Bourgain and Demeter proving the full range of Strichartz estimates on regular and irrational tori and thus settling an important earlier conjecture by Bourgain.

1. Introduction

The nonlinear Schrödinger equation plays an ubiquitous role as a model for dispersive wave-phenomena in nature. Roughly speaking, dispersion means that when no boundary is present, waves of different wavelengths travel at different phase speeds: long wavelength components propagate faster than short ones. This is the reason why over time dispersive waves spread out in space as they evolve in time, while conserving some form of *energy*. This phenomenon is called *broadening* of the wave packet. Dispersive wave-phenomena should be contrasted with transport phenomena where all frequencies move at the same velocity or dissipative phenomena (heat equation) in which frequencies gradually taper to zero, that is they do not propagate.

The nonlinear Schrödinger equation serves as a mathematical model for the large class of so-called dispersive partial differential equations [1, 67]. It naturally arises in connection to a variety of different physical problems on flat space, tori and other manifolds. One of them is nonlinear optics in a so-called *Kerr medium* where one considers electromagnetic waves in a material (eg. glass fiber) whose time evolution are governed by Maxwell's equations on $\mathbb{R}^3$. The nonlinear Maxwell equations however have disparate scales and understanding their dynamics is a difficult problem. As a first attempt one looks for further simplifications: asymptotic methods then become useful. A natural *ansatz* is to write the electric field E as a Taylor series whose leading term is a small amplitude wave packet of the form

$$A(t,x)e^{i(\xi_0\cdot x-\omega_0 t)}+\overline{A(t,x)}e^{-i(\xi_0\cdot x-\omega_0 t)} \tag{1.1}$$

2010 *Mathematics Subject Classification.* Primary 42, 35.

The author would like to thank Larry Guth for making the insightful notes from his course [43] available online. We borrow heavily from them in Section 4 on ℓ^2 decouplings. The author also thanks her PhD student Michael Boratko for stimulating discussions on multilinear Kakeya, the work of Bourgain and Demeter, as well as for his insights and great help in the preparation of Section 4.

with the wave vector $\xi_0 \in \mathbb{R}^3$, the frequency $\omega_0 \in \mathbb{R}$ and A is a small amplitude and slowly varying function. After inserting back into the nonlinear Maxwell equations, formal calculations, transformations and multiple scale analysis yield a cubic nonlinear Schrödinger equation where time corresponds to the coordinate of the direction of propagation of the wave along the material for the (transformed) amplitude [1, 55]. Essentially the same type of approximations can be made in other problems such as e.g. water waves. In this context one seeks solutions in which the interface of the fluid region is to leading order a wave packet of the same form as (1.1), i.e. with small $O(\epsilon)$ amplitude and slow spatial variation that are balanced. Lengthy formal calculations then suggest that the envelopes of these wave packets evolve on $O(\epsilon^{-2})$ time scales according to a version of a cubic nonlinear Schrödinger equation [25]. It often turns out that the nonlinear Schrödinger equation (approximately) describes the evolution of envelopes of wave packets on the appropriate NLS time scales; for a precise description, see [69, 70]. Other examples of cubic NLS arising from other physical situations can be found in [61].

The nonlinear Schrödinger equations also arise as the equations governing Bose-Einstein condensates. Bose Einstein condensation phenomena were predicted by S.N Bose [3] and by A. Einstein [29] (1924-1925); it is a fascinating phenomena predicted by quantum statistical mechanics. Bose Einstein condensation however was experimentally achieved only in 1995 by Cornell and Wieman [24] and by W.S. Ketterle [49] who produced the first gaseous condensate. For this they were awarded the 2001 Nobel Prize in Physics.

A Bose-Einstein condensate (BEC) is the state of matter of a gas of weakly interacting bosonic atoms confined by an external potential and cooled to temperatures very near *absolute zero* (0 Kelvin). In his 2001 Nobel lecture, W. S. Ketterle described *how profoundly the properties of a gas of bosonic atoms changes when you cool down the gas. Then the wave nature of matter tells us that the wave packets which describes an atom, this fuzzy object, becomes larger and larger and when the wave packet expand to the size that the waves of neighboring atoms overlap then all atoms start to oscillate in concert and form what you may regard a giant matter wave. And this is the Bose Einstein condensate.* [51]. In other words, all bosons occupy the same quantum state and can thus be described by a single wave function $u(t,x)$. The pointwise density of this *gas* at time t is represented by $|u(x,t)|^2$. The interactions between the bosons lead to nonlinear contributions to the Schrödinger equation for this quantum system. Considering only binary collisions between the bosons, one sees that u satisfies a cubic nonlinear Schrödinger equation, which in this context is often called the *Gross-Pitaevski equation* after work by Gross [39] and by Pitaevskii [58]. Physically, it makes sense to study the problem both in the periodic and the non-periodic setting. Recently there has been been intense activity and breakthrough results, particular by L. Erdös, B. Schlein and H.T. Yau in the (mathematically) rigorous derivation of the defocusing cubic NLS, both on $\mathbb{R}^3$ as well as $\mathbb{T}^3$ from the dynamics of many-body quantum systems. We refer the interested reader to [30, 31, 32, 33, 53, 20, 19, 22, 21, 38, 59] and references therein.

Bose-Einstein condensation is based on the wave nature of particles, which is at the heart of quantum mechanics. In a simplified picture, bosonic atoms in a gas may be regarded as quantum-mechanical wave-packets with an extension of their thermal *de Broglie wavelength* (the position uncertainty associated with the thermal momentum distribution). The lower the temperature, the longer is the

Graphically, we can visualize this as follows [49, 50]:

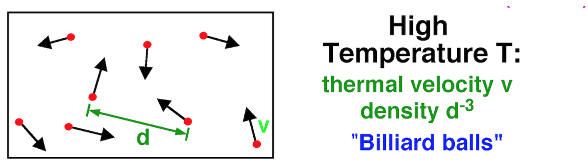


Figure 1. Gas at high temperature, treated as a system of billiard balls, with thermal velocity v and density d^{-3}, where d is the distance between bosonic particles.

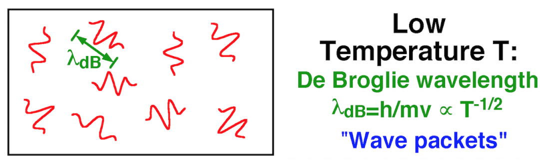


Figure 2. Simplified quantum description of gas at low temperature, in which the particles are regarded as wave packets with a spatial extent of the order of the *de Broglie* wavelength, λ_{dB}.

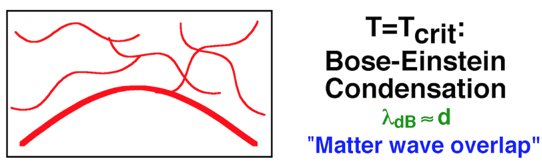


Figure 3. Gas at the transition temperature for Bose-Einstein condensation, when λ_{dB} becomes comparable to d. The wave packets overlap and a Bose-Einstein condensate forms.

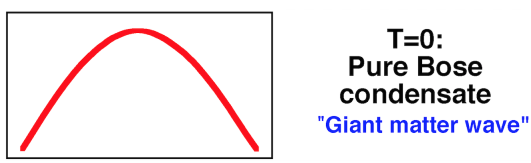


Figure 4. Pure Bose condensate (giant matter wave), which remains as the temperature approaches absolute zero and the thermal cloud disappears.

de Broglie wavelength. When atoms are cooled to the point where the thermal de Broglie wavelength is comparable to the interatomic separation, then the atomic wave-packets overlap and *the indistinguishability of particles* becomes important. Bosons undergo a phase transition and form a Bose-Einstein condensate, a dense and coherent cloud of atoms all occupying the same quantum mechanical state [49].

As mentioned above the nonlinear Schrödinger equation arising from many-body quantum bosonic atoms makes physical sense for bosons in a three-dimensional cube with periodic boundary conditions (or from an experimental perspective a rectangular box). The nonlinear Schrödinger equation however, has very different behavior on $\mathbb{T}^d$ from that on $\mathbb{R}^d$ since *dispersion is weaker* on (periodic) domains. Furthermore, by now we have several examples of nonlinear Schrödinger and wave equations defined $\mathbb{R}^d$ for which it is mathematically proven that dispersion sets in and after a time long enough solutions settle into a purely linear behavior. This phenomenon is often referred to as *scattering* (asymptotic stability). Since linear solutions, energy at any given frequency does not migrate to higher or lower frequencies, that is there is no *forward* or *backward cascade.* Hence, as a consequence of scattering certain nonlinear solutions in $\mathbb{R}^d$ also will avoid these cascades. The situation is believed to be quite different for dispersive equations on compact domains. For example in the periodic case energy cascades and out-of-equilibrium dynamics are expected [23, 40, 45]. It is then not surprising that understanding the time dynamics of solutions to the nonlinear Schrödinger on tori one needs to bring to bear tools and ideas from many different other areas of mathematics, from nonlinear Fourier and harmonic analysis, geometry, probability, analytic number theory, dynamical systems, and others. In these notes we will touch upon a few of these connections by explaining some key results and focus on the spectacular resolution of the ℓ^2 decoupling conjecture by J. Bourgain and C. Demeter [14, 15]. Their results in turn (and in particular) solve a 1993 conjecture by Bourgain and yield the predicted full range of dispersive estimates (known as Strichartz estimates) for solutions to the Cauchy initial value problem for the nonlinear Schrödinger equation (p-NLS) on general rectangular d-dimensional tori:

$$\begin{cases} iu_t + \Delta u = \lambda |u|^{p-1}u, \\ u(0,x) = \phi(x), \quad x \in \Lambda_d(\theta), \end{cases} \tag{1.2}$$

where ϕ is the initial profile, $\lambda = \pm 1$, $p > 1$, $u : \mathbb{R} \times \Lambda_d(\theta) \to \mathbb{C}$ and for $\beta := (\beta_1, \beta_2, \ldots, \beta_d), \beta_j > 0$, $j = 1, \ldots d$ we define the d-dimensional tori by:

$$\Lambda_d(\beta) := (\mathbb{R}/\beta_1\mathbb{Z}) \times (\mathbb{R}/\beta_2\mathbb{Z}) \times \cdots \times (\mathbb{R}/\beta_d\mathbb{Z}).$$

When $\beta_j = 1$ we have $\Lambda_d(\beta) = \mathbb{T}^d := (\mathbb{R}/\mathbb{Z})^d$, the square d-dimensional torus $\mathbb{T}^d$ of 1-periodic functions. If $\beta_j \in \mathbb{Q}$ for all $j = 1, \ldots d$, we call $\Lambda_d(\beta)$ a *rational* torus while if at least one of the ratios $\frac{\beta_j}{\beta_{j'}} \notin \mathbb{Q}$, $j \neq j'$ we call $\Lambda_d(\beta)$ an *irrational* torus[1]. The latter comes up naturally experimentally as well as in KAM theory and Hamiltonian chaos ([10] and references therein).

In the context of nonlinear dispersive equations the general rectangular tori $\Lambda_d(\beta)$ were first studied by Bourgain [4] where he noted that the methods from analytic number theory, previously employed in [6] to obtain for the first time *some* dispersive estimates for the Schrödinger equation on the square $\mathbb{T}^d$ could not be used in the general rectangular case. It is fairly straightforward to see that the dispersive estimates known for $\mathbb{T}^d$ imply those for the rational torus as well. This left open two questions: 1) obtaining the full range of expected dispersive estimates on $\mathbb{T}^d$ and 2) proving dispersive estimates for irrational tori; which was thought to be a much harder question from the harmonic analysis and analytic number theory point of view adopted in [4]. These questions were recently answered by Bourgain and Demeter [14, 15] who rather than analytic number theory rely on decoupling inequalities (in discrete restriction phenomena) and sophisticated arguments from multilinear harmonic analysis (adapted wave packet decompositions,

[1]Without any loss of generality we can assume $\frac{1}{2} < \beta_j < 2, j = 1, \ldots d$.

parabolic rescaling, bilinear square function estimates, multilinear Kakeya, multi-scale bootstrap), and also implicitly on ideas from incidence theory. Their work is at the core of these notes, and will be the focus of Section 4. In Section 5 we discuss the interplay of deterministic and probabilistic approaches in the well posedness theory of nonlinear Schrödinger equations. In Section 2 we start with some preliminaries.

Notation We use $A \lesssim B$ to denote an estimate of the form $A \le CB$ for some $C > 0$ which may depend on the underlying dimension as well as on fixed parameters such as p or s. However we record dependence on variable parameters such as ε using the notation $\lesssim_\varepsilon$. The asymptotic notation $A \gtrsim B$ is defined analogously, and $A \sim B$ will mean $A \lesssim B$ and $B \lesssim A$. By H^s (resp. $\dot{H}^s$) we denote the usual inhomogeneous (resp. homogeneous) Sobolev spaces. Given a function $u = u(t,x)$ depending on time t and the space variable x, we denote by $\|u\|_{L^q_t L^r_x} := \| \|u\|_{L^r_x} \|_{L^q_t}$ the mixed space-time Lebesgue norm. For a fixed time interval I, the spaces $L^\infty(I;\, H^s)$ (resp. $C(I;\, H^s)$) denote the space of functions which are in L^∞ in t (resp. continuous in t) with values in the Banach space H^s.

2. Preliminaries

Whether the underlying space is $\mathbb{R}^d$, a torus or some other manifold, a basic question when studying the Cauchy initial value problem (1.2) is that of *well-posedness*, that is: i) *existence*, ii) *uniqueness* and iii) *stability* of solutions for initial data in a given Banach space, which in these notes we assume to be the Sobolev space H^s. To solve this question, the idea is to use a fixed point theorem on a space of functions whose norm is dictated by strong estimates for $v(t,x) := S(t)\phi(x)$, the solution of the associated linear problem,

$$\begin{cases} iv_t + \Delta v = 0 \\ v(0,x) = \phi(x). \end{cases} \tag{2.1}$$

One should of course recall that under reasonable regularity assumptions, (1.2) is equivalent to (2.2) below, thanks to the Duhamel principle. Formally, well posedness is defined as follows:

Definition 2.1. We say that the Cauchy initial value problem (1.2) is locally well posed in H^s if for any ball $\mathcal{B}$ in H^s there exists a time $T > 0$ and a Banach space of functions $X^s \subset L^\infty([-T,T]; H^s)$ such that for each initial data $\phi \in \mathcal{B}$, there exists a unique solution $u \in X^s \cap C([-T,T];\, H^s)$ of the integral equation

$$u(x,t) = S(t)\phi(x) + c_\lambda \int_0^t S(t-t')\, |u(t',x)|^{p-1} u(t',x)\, dt'. \tag{2.2}$$

Moreover, the map $\phi \longmapsto u$ is continuous from H^s into $C([-T,T];\, H^s)$. If $T > 0$ can be taken arbitrarily large, then we say that the initial value problem is *globally* well posed.

Remark 2.2. Note that the definition above yields uniqueness on $X^s \cap C([-T,T];\, H^s)$ but not necessarily on $C([-T,T];\, H^s)$. Proving uniqueness on $C([-T,T];\, H^s)$ requires additional work and when it holds, the local well posedness is said to be *unconditional.*

2.1. **A tour onto $\mathbb{R}^d$.** In these notes we will primarily focus on the periodic setting as in (1.2) where the problems associated to p-NLS are harder and less understood than when the underlying domain[2] is $\mathbb{R}^d$. Before doing so however let us recall a few important ideas about p-NLS on $\mathbb{R}^d$,

$$\begin{cases} iu_t + \Delta u = \lambda |u|^{p-1} u, \\ u(0,x) = \phi(x), \quad x \in \mathbb{R}^d, \end{cases} \tag{2.3}$$

where $\phi \in H^s$, $\lambda = \pm 1$, $p > 1$, $u : \mathbb{R} \times \mathbb{R}^d \to \mathbb{C}$. The nonlinear Schrödinger equation (2.3) enjoys many symmetries (c.f. [67]) among which we highlight:
Time-reversal symmetry:

$$\phi(x) \longmapsto \overline{\phi(x)}, \qquad u(t,x) \longmapsto \overline{u(-t,x)} \tag{2.4}$$

Scaling symmetry:

$$\phi(x) \longmapsto \mu^{-\frac{2}{(p-1)}} \phi(\frac{x}{\mu}) =: \phi_\mu(x), \qquad u(t,x) \longmapsto \mu^{-\frac{2}{(p-1)}} u(\frac{t}{\mu^2}, \frac{x}{\mu}) \tag{2.5}$$

for any dilation factor $\mu > 0$. From (2.5) we immediately notice that if the initial datum is in $\dot{H}^{s_c}(\mathbb{R}^d), s_c := \frac{d}{2} - \frac{2}{p-1}$ then $\|\phi_\mu\|_{\dot{H}^{s_c}} = \|\phi\|_{\dot{H}^{s_c}}$ and (2.3) is scale invariant. The Sobolev regularity s_c is then called the *critical* scaling regularity. Note that the criticality of H^s depends on both the power p and the dimension d. In fact since we have that

$$\|\phi_\mu\|_{\dot{H}^s} \sim \mu^{s_c - s} \|\phi\|_{\dot{H}^s}$$

we can classify the difficulty of the p-NLS (2.3) above in terms of regularity of its data. When $s > s_c$ note that as $\mu \to \infty$, the norm of $\|\phi_\mu\|_{\dot{H}^s}$ gets smaller; the space H^s is called *subcritical* in this case. On the other hand if $s < s_c$ we have that as $\mu \to \infty$ the norm of $\|\phi_\mu\|_{\dot{H}^s}$ gets larger; the space H^s is then called *supercritical.* When $s = s_c$, the space $\dot{H}^{s_c}$ is *critical* since as we noted above, as $\mu \to \infty$ the norm of $\|\phi_\mu\|_{\dot{H}^s}$ does not change. Accordingly, the local well posedness theory for equation (2.3) is fairly well understood in the subcritical regime- when the equation can be treated as a perturbation of the linear one- while in the supercritical regime only nondeterministic results are available. We return to the latter in Section 5.

Remark 2.3. In proving local well posedness for (2.2) in the *subcritical regime* $s > s_c$ one shows via a fixed point argument that the time of existence T is roughly like $\|\phi\|_{H^s}^{-\alpha}$ for some $\alpha > 0$.

We also note p-NLS conserves both mass and the Hamiltonian; i.e:

$$\mathcal{M}(u(t)) := \int |u(t,x)|^2 \, dx = \int |\phi(x)|^2 \, dx =: \mathcal{M}(\phi(t)) \tag{2.6}$$

and the Hamiltonian; ie.

$$\begin{aligned} H(u(t)) \; &:= \; \frac{1}{2} \int |\nabla u(t,x)|^2 \, dx + \frac{2\lambda}{p+1} \int |u(t,x)|^{p+1} dx \\ &= \; \frac{1}{2} \int |\nabla \phi(x)|^2 \, dx + \frac{2\lambda}{p+1} \int |\phi(x)|^{p+1} dx =: H(\phi) \end{aligned} \tag{2.7}$$

[2]There is by now a substantial body of work on NLS on $\mathbb{R}^d$. The interested reader might want to consult [18, 67].

Remark 2.4. If $\lambda = 1$ (2.7) and (2.6) give a global in time bound for the H^1 norm of $u(t,x)$; p-NLS is called *defocusing* in this case. On the other hand, if $\lambda = -1$ the energy could be negative and blow up may occur; p-NLS is then called *focusing*.

In particular, we have that for those p and d for which the H^1 space is subcritical (e.g. cubic $p = 3$ in $d = 2$), Remark 2.3 and (2.7) allow one to obtain global well posedness by iterating the local theory.

To prove local well posedness for p-NLS one needs to find a suitable Banach space X^s as in Definition 2.1 on which to prove that the map

$$\Phi : \mathfrak{u} \longmapsto S(t)\phi + c_\lambda \int_0^t S(t-t')\,|\mathfrak{u}(t',x)|^{p-1}\mathfrak{u}(t',x)\,dt'$$

is a contraction, whence the solution $u(t,x)$ is as in (2.2). Determining a good choice of X^s is part of the problem. It is dictated by being able to have sufficiently good estimates for the linear evolution $S(t)\phi$ on such space so that then, at least in the subcritical regime and on short time intervals, one can show that $\Phi(\mathfrak{u})$ - hence the solution u- satisfy similar estimates. The most basic and at the same time important space-time estimates that $S(t)\phi$, the solution to the linear problem (2.19), satisfy are the so called Strichartz estimates.

2.1.1. *The Strichartz Estimates on* $\mathbb{R}^d$. The Strichartz estimates are intimately related to the $(L^{\mathfrak{p}}, L^2)$ restriction problem for the Fourier transform (to the paraboloid in the case of the Schrödinger equation). To understand the connection it is illustrative to review Strichartz original argument [62, 56]. Let us first briefly recall the restriction question. Consider $1 \le \mathfrak{p} \le 2$, $f \in L^{\mathfrak{p}}(\mathbb{R}^d)$ and $\mathcal{S}$ a hypersurface on $\mathbb{R}^d$. The $(L^{\mathfrak{p}}, L^2)$-restriction for the Fourier transform asks whether the map

$$\mathcal{R} : f \longmapsto \widehat{f}\big|_{\mathcal{S}}$$

extends to a bounded operator from $L^{\mathfrak{p}}(\mathbb{R}^d) \longmapsto L^2(\mathcal{S}, d\sigma)$; ie. whether

$$\|\widehat{f}\big|_{\mathcal{S}}\|_{L^2(d\sigma)} \le C_{d,\mathfrak{p}}\,\|f\|_{L^{\mathfrak{p}}(\mathbb{R}^d)}, \tag{2.8}$$

where $d\sigma$ is a canonical measure[3] associated to $\mathcal{S}$. If $\mathfrak{p} = 1$, this is always true by the Riemann Lebesgue lemma. But if $\mathfrak{p} = 2$, then $\widehat{f}\big|_{\mathcal{S}}$ is meaningless since $\widehat{f} \in L^2(\mathbb{R}^d)$ and the d-dimensional Lebesgue measure of $\mathcal{S}$ is zero. Moreover, if $\mathcal{S}$ is a plane, then no $\mathfrak{p} > 1$ is allowed, while if $\mathcal{S} = \mathbb{S}^{d-1}$, the unit sphere in $\mathbb{R}^d$, then P. Tomas and E.M. Stein gave an affirmative answer for $1 \le \mathfrak{p} \le \frac{2(d+1)}{(d+3)}$ (and can be shown to fail for any $\mathfrak{p}$ larger). In general, the answer to this question depends on the *curvature* of $\mathcal{S}$. Indeed, if $\mathcal{S}$ is a compact hypersurface with non-vanishing Gaussian curvature, one can show using stationary phase methods that for every $z \in \mathbb{R}^d$,

$$|\widehat{\sigma_{\mathcal{S}}}(z)| \lesssim (1+|z|)^{-\frac{d-1}{2}}. \tag{2.9}$$

By a TT^* argument, (2.8) is equivalent to prove that $\|\widehat{\sigma} * f\|_{L^{\mathfrak{p}'}(\mathbb{R}^d)} \le C_{d,\mathfrak{p}}\,\|f\|_{L^{\mathfrak{p}}(\mathbb{R}^d)}$, which follows from (2.9) in conjunction with Littlewood-Paley and (complex) interpolation (or fractional integration in a direction transverse to $\mathcal{S}$ and convex interpolation). See [60, 72, 65, 66, 56] and references therein for this problem and the more general $(L^{\mathfrak{p}}, L^{\mathfrak{q}})$-restriction conjecture problem.

[3] e.g. if $\mathcal{S} = \mathbb{S}^{d-1}$, $d\sigma$ is surface measure. Of interest for PDE is the case when $d\sigma$ is a measure supported on $\mathcal{S}$.

Consider now the linear Schrödinger equation (2.19) and let us assume that the initial datum ϕ is a smooth function such that $\operatorname{supp}\widehat{\phi} \subset \{|\xi| < 1\}$. By taking the Fourier transform in space of (2.19) and solving the corresponding ODE we have that the solution to the linear problem is defined by $\widehat{v}(t,\xi) := e^{-i|\xi|^2 t}\widehat{\phi}(\xi)$ whence we have that

$$v(t,x) = \int_{\mathbb{R}^d} e^{i(x\cdot\xi + h(\xi)t)}\widehat{\phi}(\xi)\, d\xi = \mathcal{F}_x^{-1}(\widehat{\phi}\sigma)(t,x), \tag{2.10}$$

where we have denoted by $h(\xi) = -|\xi|^2$ and σ is the measure in $\mathbb{R}^{d+1}$ carried by the paraboloid hypersurface

$$\Sigma := \{(\xi,\tau) \in \mathbb{R}^d \times \mathbb{R} \,:\, \tau = h(\xi)\}$$

defined by

$$\int_{\mathbb{R}^{d+1}} \psi(\xi,\tau)\, d\sigma(\xi,\tau) = \int_{\mathbb{R}^d} \psi(\xi, h(\xi))\, d\xi$$

for any ψ continuous on $\mathbb{R}^{d+1}$. In other words, $v(t,x) = \mathcal{R}^*(\phi)(x,t)$, where $\mathcal{R}^*$ is the adjoint of $\mathcal{R}(f) = \widehat{f}\big|_{\Sigma}$, the operator that restricts the Fourier transform on $\mathbb{R}^{d+1}$ to the the paraboloid Σ. Then by the Tomas-Stein endpoint estimate and duality we obtain that

$$\|v(t,x)\|_{L^{\frac{2(d+2)}{d}}(\mathbb{R}^{d+1})} = \|\mathcal{F}_x^{-1}(\widehat{\phi}\sigma)\|_{L^{\frac{2(d+2)}{d}}(\mathbb{R}^{d+1})} \lesssim \|\widehat{\phi}\|_{L^2(\mathbb{R}^d)} = \|\phi\|_{L^2(\mathbb{R}^d)}.$$

Remark 2.5. Note that since Tomas-Stein endpoint estimate is scale invariant by rescaling $\phi(x)$ and (parabolically) $v(t,x)$ one may remove the assumption made above that $\widehat{\phi}$ is supported in the unit frequency ball.

The full range of Strichartz estimates can be derived via a shorter argument -essentially due to Ginibre and Velo and to Yajima- thanks to the explicit form of the linear semigroup (see [18, 67] and references therein). Indeed, from (2.10) we have that

$$v(t,x) = S(t)\phi(x) := e^{it\Delta}\phi(x) = K_t * \phi(x) = \frac{c}{t^{d/2}} \int_{\mathbb{R}^d} e^{i\frac{|x-y|^2}{2t}}\, \phi(y)\, dy, \tag{2.11}$$

whence we immediately obtain that

$$\|S(t)\phi\|_{L^\infty_x(\mathbb{R}^d)} \lesssim \frac{1}{t^{d/2}} \|\phi\|_{L^1_x(\mathbb{R}^d)} \tag{2.12}$$

called the *dispersive estimate*. On the other hand, given the linear semigroup is unitary and commutes with other Fourier multiplies, we clearly have that for any s,

$$\|S(t)\phi\|_{H^s_x(\mathbb{R}^d)} = \|\phi\|_{H^s_x(\mathbb{R}^d)}. \tag{2.13}$$

Interpolating (2.12) and (2.13) with $s = 0$ we have for any $1 \le \mathfrak{p} \le 2$ the *fixed time* estimates:

$$\|S(t)\phi\|_{L^{\mathfrak{p}'}_x(\mathbb{R}^d)} \lesssim \frac{1}{t^{\frac{d}{\mathfrak{p}} - \frac{d}{2}}} \|\phi\|_{L^{\mathfrak{p}}_x(\mathbb{R}^d)} \tag{2.14}$$

where $\frac{1}{\mathfrak{p}} + \frac{1}{\mathfrak{p}'} = 1$. These estimates are not enough since the initial data is usually only assumed to be in an L^2-based Sobolev space; however by combining (2.14) with duality, TT^* arguments, fractional integration in time and interpolation we obtain the full range of Strichartz estimates (c.f. [62, 60, 18, 67, 56] and references therein) which we are now ready to state. For $d \geq 1$ we define $\mathcal{A}$ the set of *admissible* exponents to be those pairs (q, r) such that $2 \leq q, r \leq \infty$ and

$$\frac{2}{q} = \frac{d}{2} - \frac{d}{r}, \qquad (q, r, d) \neq (2, \infty, 2) \tag{2.15}$$

Theorem 2.6 (Strichartz estimates on $\mathbb{R}^d$)**.** *For $(q, r) \in \mathcal{A}$ we have the homogeneous estimate,*

$$\|S(t)\phi\|_{L_t^q L_x^r(\mathbb{R}\times\mathbb{R}^d)} \lesssim \|\phi\|_{L_x^2(\mathbb{R}^d)},$$

and for any other admissible pair $(\tilde{q}, \tilde{r})$ we also have the inhomogeneous estimate:

$$\left\| \int_0^t S(t-t')\, N(u)(t')\, dt' \right\|_{L_t^q L_x^r(\mathbb{R}\times\mathbb{R}^d)} \lesssim \|N(u)\|_{L_t^{\tilde{q}'} L_x^{\tilde{r}'}(\mathbb{R}\times\mathbb{R}^d)},$$

where $\frac{1}{\tilde{q}} + \frac{1}{\tilde{q}'} = 1 = \frac{1}{\tilde{r}} + \frac{1}{\tilde{r}'}$ and $N(\cdot)$ is any Lipschitz continuous function.

Remark 2.7. It can be proved via a standard Knapp example and scaling arguments that the admissibility of (q, r) is a necessary condition. We also note that these estimates hold on $[-T, T] \times \mathbb{R}^d$ in lieu of $\mathbb{R} \times \mathbb{R}^d$.

2.2. **Back to the periodic setting.** We start by rescaling the tori $\Lambda_d(\beta)$ so that we can use coordinates based on the regular square torus $\mathbb{T}^d$ and work with Fourier series based on the standard integer lattice $\mathbb{Z}^d$. In this way we incorporate the geometry of $\Lambda_d(\beta)$ into the Laplace operator, which after such rescaling is defined by

$$\Delta_\theta := \theta_1 \frac{\partial^2}{\partial x_1^2} + \theta_2 \frac{\partial^2}{\partial x_2^2} + \dots \theta_d \frac{\partial^2}{\partial x_d^2}, \qquad \theta_j = \frac{1}{\beta_j^2}, \quad j = 1, \dots, d. \tag{2.16}$$

In other words, for $\mathbf{k} = (k_1, \dots, k_d) \in \mathbb{Z}^d$,

$$\widehat{\Delta f}(\mathbf{k}) := -(\theta_1\, k_1^2 + \cdots + \theta_d\, k_d^2)\, \widehat{f}(n) \tag{2.17}$$

where as usual, we have denoted the Fourier transform and Fourier series as:

$$\widehat{f}(\mathbf{k}) = \int_{\mathbb{T}^d} e^{-2\pi i \mathbf{k}\cdot x} f(x)\, dx,\ (\mathbf{k} \in \mathbb{Z}^d) \quad \text{and} \quad f(x) = \sum_{\mathbf{k}\in\mathbb{Z}^d} e^{2\pi i \mathbf{k}\cdot x} \widehat{f}(\mathbf{k}),\ (x \in \mathbb{T}^d).$$

Our problem (1.2) can then be rewritten on $\mathbb{T}^d$ as follows:

$$\begin{cases} iu_t + \Delta_\theta u = \lambda |u|^{p-1} u, \\ u(0, x) = \phi(x), \quad x \in \mathbb{T}^d, \end{cases} \tag{2.18}$$

We note that the solution $u(t, x)$ to the linear Schrödinger equation,

$$\begin{cases} iu_t + \Delta_\theta u = 0 \\ u(0, x) = \phi(x), \quad x \in \mathbb{T}^d, \end{cases} \tag{2.19}$$

is given by

$$u(t, x) = e^{it\Delta_\theta} u_0 = \sum_{\mathbf{k}\in\mathbb{Z}^d} e^{2\pi i \left(n\cdot x - t \sum_{j=1}^d \theta_j\, k_j^2\right)}\, \widehat{u_0}(\mathbf{k}). \tag{2.20}$$

Of course, equation (2.18) also conserves mass (2.6) and the Hamiltonian (2.7).

Remark 2.8. The periodic cubic NLS equation,

$$\begin{cases} iu_t + \Delta u = \lambda |u|^2 u, \\ u(x, 0) = u_0(x), \quad x \in \mathbb{T}^3 \end{cases}$$

is the one governing Bose Einstein condensation alluded to in the introduction.

3. Strichartz Estimates on Tori

In the periodic setting, the tools from harmonic analysis we used to establish the Strichartz estimates are no longer available. Even in the case of the square torus, $\mathbb{T}^d$, obtaining some Strichartz estimates for $S(t)u_0(x)$ is highly nontrivial and required new ideas. They were introduced by Bourgain in [6] for the square (rational) torus as a conjecture:

Conjecture 3.1. Assume that $\mathbb{T}^d$ is a *rational* torus and for $N \geq 1$, let $\phi \in L^2(\mathbb{T}^d)$ be a smooth function such that the $\operatorname{supp}\hat{\phi} \subset [-N,N]^d \subset \mathbb{Z}^d$. Then for any $\epsilon > 0$ the following estimates should hold:

$$\|S(t)\phi\|_{L^q_tL^q_x(\mathbb{T}^{d+1})} \lesssim C_q \|\phi\|_{L^2_x(\mathbb{T}^d)} \quad \text{if} \quad q < \frac{2(d+2)}{d}$$
$$\|S(t)\phi\|_{L^q_tL^q_x(\mathbb{T}^{d+1})} \ll N^{\epsilon} \|\phi\|_{L^2_x(\mathbb{T}^d)} \quad \text{if} \quad q = \frac{2(d+2)}{d}$$
$$\|S(t)\phi\|_{L^q_tL^q_x(\mathbb{T}^{d+1})} \lesssim C_q N^{\frac{d}{2}-\frac{d+2}{q}} \|\phi\|_{L^2_x(\mathbb{T}^d)} \quad \text{if} \quad q > \frac{2(d+2)}{d}$$

In [6] Bourgain in fact partially proved these bounds in the following cases: i) $d = 1, 2$ and $q > \frac{2(d+2)}{d}$, ii) $d = 3$ and $q > 4$ and iii) $d \geq 4$ and $q > \frac{2(d+4)}{d}$. His proof relies on Weyl's sum estimates, the Hardy-Littlewood circle method and the Tomas-Stein restriction theorem. Partial improvements were obtained in [26, 47].

Remark 3.2. Bourgain also proved in [6] that dispersion is indeed weaker in the periodic setting by proving that when $d = 1$ the endpoint L^6 estimate which holds on $\mathbb{R}$ with constant independent of N is false in the periodic setting. More precisely, he showed

$$\|\sum_{|k|\leq N} a_k e^{i(kx+k^2t)}\|_{L^6(\mathbb{T}\times\mathbb{T})} \geq c(\log N)^{\frac{1}{6}} \left(\sum_{k\in\mathbb{Z}} |a_k|^2\right)^{1/2}.$$

The failure of the endpoint estimate $\|e^{it\Delta}\phi_N\|_{L^4(\mathbb{T}^2\times\mathbb{T})} \lesssim \|\phi_N\|_{L^2(\mathbb{T}^2)}$ when $d = 2$ was established by Takaoka and Tzvetkov [64].

More recently, Bourgain [11] improved his results from [6] by establishing the above Conjecture 3.1 for $d \geq 4$ and $q > \frac{2(d+3)}{d}$ by relying on the multilinear harmonic analysis techniques for the restriction and Kakeya problems developed in [2] and in [16]. These techniques will once again mark the way for the resolution of the full conjecture by Bourgain and Demeter [14] as we will see below.

3.1. Strichartz Estimates and the Discrete Fourier Restriction.

For any given $N \in \mathbb{N}$, let $S_{d,N}$ be the set

$$\left\{(k_1, \cdots, k_d) \in \mathbb{Z}^d : |k_j| \leq N,\ 1 \leq j \leq d\right\}.$$

For $q > 1$, let $A_{q,d,N}$ represent the best constant satisfying

$$\sum_{\mathbf{k}\in S_{d,N}} \left|\widehat{f}(\mathbf{k}, |\mathbf{k}|^2)\right|^2 \leq A_{q,d,N} \|f\|^2_{q'}, \tag{3.1}$$

where $\mathbf{k} = (k_1, \cdots, k_d) \in S_{d,N}$, $|\mathbf{k}| = \sqrt{k_1^2 + \cdots + k_d^2}$ and f is any $L^{q'}$-function on $\mathbb{T}^{d+1}$ and $q' = q/(q-1)$.

As we mentioned above Bourgain [6] obtained (in particular) the following estimate for the square torus $\mathbb{T}^d$:

$$A_{q,d,N} \leq CN^{d-\frac{2(d+2)}{q}+\epsilon} \text{ for } q > \frac{2(d+4)}{d}. \tag{3.2}$$

By duality, it is straightforward to see that the Strichartz estimates,

$$\left\| \sum_{\mathbf{k}\in S_{d,N}} a_{\mathbf{k}} e^{i(\mathbf{k}\cdot x+|\mathbf{k}|^2 t)} \right\|_{L^q(\mathbb{T}^{d+1})} \leq \sqrt{A_{q,d,N}} \left(\sum_{\mathbf{k}\in S_{d,N}} |a_{\mathbf{k}}|^2 \right)^{1/2} \tag{3.3}$$

are in fact equivalent to the discrete Fourier restriction estimates:

$$\left(\sum_{\mathbf{k}\in S_{d,N}} \left| \widehat{f}(\mathbf{k},|\mathbf{k}|^2) \right|^2 \right)^{1/2} \leq \sqrt{A_{q,d,N}} \|f\|_{q'} \tag{3.4}$$

To understand how the rational character of the torus enters in a basic fashion, let us review Bourgain's result for the square torus in the case where $q = 4$ and $d = 2$. We would like to show that $A_{4,2,N} < N^\varepsilon$, $\varepsilon > 0$. Bourgain [6] reduced the problem to estimating the number of representations of an integer as a sum of squares. Let

$$f(\mathbf{x},t) = \sum_{|\mathbf{k}|<N} a_{\mathbf{k}} e^{i(\mathbf{k}\cdot\mathbf{x}+|\mathbf{k}|^2 t)} \text{ with } (\mathbf{x},t) \in \mathbb{T}^2 \times \mathbb{T},$$

and for a given integer j and $\mathbf{p} \in \mathbb{Z}^2$ define,

$$\mathcal{C}_{\mathbf{p},j} := \{\mathbf{k} \in \mathbb{Z}^2 : |\mathbf{k}| \leq N \text{ and } |\mathbf{k}|^2 + |\mathbf{p}-\mathbf{k}|^2 = j\},$$

and let $r_{\mathbf{p},j} = \#\mathcal{C}_{\mathbf{p},j}$. If we square our function f, we obtain the estimate that

$$f(\mathbf{x},t)^2 = \sum_{\mathbf{p}} e^{i(\mathbf{p}\cdot\mathbf{x})} \left[\sum_{\mathbf{k}} a_{\mathbf{k}} a_{\mathbf{p}-\mathbf{k}} e^{i(|\mathbf{k}|^2+|\mathbf{p}-\mathbf{k}|^2 t)} \right] = \sum_{\mathbf{p},j} \left(\sum_{\mathbf{k}\in\mathcal{C}_{\mathbf{p},j}} a_{\mathbf{k}} a_{\mathbf{p}-\mathbf{k}} \right) e^{i(\mathbf{p}\cdot\mathbf{x}+jt)}$$

so that if we take the L^2 norm, we find that

$$\|f^2\|^2_{L^2(\mathbb{T}^2\times\mathbb{T})} \leq \left\{ \max_{\substack{|\mathbf{p}|\leq 2N \\ |j|\leq 2N^2}} r_{\mathbf{k},j} \right\} \left(\sum |a_{\mathbf{k}}|^2 \right)^2. \tag{3.5}$$

We can rewrite $|\mathbf{k}|^2 + |\mathbf{p}-\mathbf{k}|^2 = j$ as

$$(2k_1 - p_1)^2 + (2k_2 - p_2)^2 = 2j - |\mathbf{p}|^2$$

so we have that $r_{\mathbf{k},j}$ is bounded by the number of solutions of

$$X_1^2 + X_2^2 = R^2$$

where $R^2 = 2j - |\mathbf{p}|^2 \lesssim N^2$. Hence the right hand side of (3.5) is bounded by the number of integer lattice points (X_1, X_2) that *that lie on the circle* of radius R. Since there are at most $\exp(C\frac{\log R}{\log\log R}) \sim R^\varepsilon$ such points, we get the desired estimate.

The case is much more difficult when generalizing to any given p and d. Hu and Li in [47] presented a variant of the proof of Bourgain's result (3.4), which just as Bourgain's makes use of the Hardy-Littlewood circle method and estimates on level sets. Their proof of Bourgain's level set estimates is however somewhat simpler. We briefly sketch their proof next. When q is large, the desired estimate (3.4) follows immediately from the following result (cf. [6, 47]):

Theorem 3.3. *For any $\sigma > 0$, any $d \geq 1$, and any $q > \frac{4(d+2)}{d}$, there exists a constant C, independent of N, such that*

$$\sum_{\mathbf{k}\in\mathbb{Z}^d} e^{-\frac{\sigma|\mathbf{k}|^2}{N^2}} \left|\hat{f}(\mathbf{k},|\mathbf{k}|^2)\right|^2 \leq CN^{d-\frac{2(d+2)}{q}} \|f\|_{q'}^2,$$

for all $f \in L^{q'}(\mathbb{T}^{d+1})$.

The proof of this theorem in turn follows from *Hardy-Littlewood circle method*, a tool to count the number of representations of a given integer as an arbitrary sum of powers. Let us recall some simple aspects about it.

Definition 3.4. Let $f(z)$ be an analytic function that converges in the open unit disc $|z| = 1$, where

$$f(z)^s = \sum_{a_1\in A}\cdots\sum_{a_s\in A} z^{a_1}\cdots z^{a_s} = \sum_{N=0}^{\infty} r_{A,s}(N) z^N.$$

Here, $r_{A,s}(N)$ is the number of representations of N as the sum of s elements of $A \subset \mathbb{Z}$. In other words, the number of solutions of the equation

$$N = a_1 + a_2 + \cdots a_s$$

with

$$a_1, a_2, \cdots a_s \in A.$$

We can now apply Cauchy's theorem to the summation above by integration:

$$r_{A,s}(N) = \frac{1}{2\pi i}\int_{|z|=\rho} \frac{f(z)^s}{z^{N+1}} dz$$

for any $\rho \in (0,1)$. This is the original form of the circle method. Note that the integral above counts the number of ways the number N can be written as a sum of arbitrary powers of s. The evaluation of such an integral is not a trivial task, and requires breaking up our circle into *major arcs* and *minor arcs*.

Sketch of the proof of Theorem 3.3. For $r \in \mathbb{N}$ let

$$\mathcal{P}_r := \{y \in \mathbb{N} \,:\, 1 \leq y \leq r,\, (y,r) = 1\}.$$

For $a \in \mathcal{P}_r$, define the interval $J_{a/r}$ as

$$J_{a/r} = \left(\frac{a}{r} - \frac{1}{Nr}, \frac{a}{r} + \frac{1}{Nr}\right).$$

If $r \geq N/10$, then we have a minor arc, and we have a major arc if $r < N/10$. By Dirichlet principle[4], we can then partition the interval $(0,1]$ into a union of major and minor arcs as

$$(0,1] = \bigcup_{1\leq r\leq N, a\in\mathcal{P}_r} J_{a/r} \;=\; \mathcal{M}_1 \cup \mathcal{M}_2$$

[4]Recall Dirichlet Principle states that given any $N \in \mathbb{N}$ and any $x \in (0,1]$ there exist $a, r \in \mathbb{N}$ such that $\left|x - \frac{a}{r}\right| \leq \frac{1}{Nr}$, $1 \leq r \leq N, a \in \mathcal{P}_r$.

where $\mathcal{M}_1$ is the union of all major arcs and $\mathcal{M}_2$ the union of all minor ones. If χ_J is the characteristic function on the set $J_{a/q}$ then set

$$K_{a/r}(\mathbf{x},t) := K_\sigma(\mathbf{x},t)\chi_{J_{a/r}}(t),$$

where

$$K_\sigma(\mathbf{x},t) := \sum_{\mathbf{k}\in\mathbb{Z}^d} e^{-\frac{\sigma|\mathbf{k}|^2}{N^2}} e^{i|\mathbf{k}|^2 t} e^{i\mathbf{k}\cdot\mathbf{x}}. \tag{3.6}$$

The point is that now one can show that for $1 \le r \le N$, $a \in \mathcal{P}_r$, and $q > \frac{2(d+1)}{d}$

$$\|K_{a/r}\|_{L^q} \le C_{r,d,q} N^{d-\frac{d+2}{q}},$$

which then leads to the estimate for $q > \frac{2(d+2)}{d}$, that

$$\|K_\sigma\|_{L^q} \le C_{q,\sigma} N^{d-\frac{d+2}{q}}.$$

Since

$$\sum_{\mathbf{n}\in\mathbb{Z}^\mathbf{d}} e^{-\frac{\sigma|\mathbf{k}|^\mathbf{2}}{N^2}} \left|\hat{f}(\mathbf{k},|\mathbf{k}|^\mathbf{2})\right| = \langle K_\sigma * f, f\rangle,$$

if we apply Hölder's inequality and Hausdorff-Young's inequality, we have that

$$\langle K_\sigma * f, f\rangle \le \|K_\sigma\|_{q/2}\|f\|_{q'}^2;$$

whence since $q > \frac{4(d+2)}{d}$ we have the desired conclusion. □

The estimate for smaller cases of q, follow from level set estimates [6, 47] of the form:

Theorem 3.5 ([47]). *Let F_N be a periodic function on $\mathbb{T}^{d+1}$ such that*

$$F_N(\mathbf{x},t) = \sum_{\mathbf{k}\in S_{d,N}} a_\mathbf{k} e^{i\mathbf{k}\cdot\mathbf{x}} e^{i|\mathbf{k}|^2 t}$$

where $\{a_\mathbf{k}\}$ is a sequence with $\sum_\mathbf{k} |a_\mathbf{k}|^2 = 1$. For any $\lambda > 0$, let

$$E_\lambda = \{(\mathbf{x},t) \in \mathbb{T}^{d+1} : |F_N(\mathbf{x},t)| > \lambda\}.$$

Then for any $Q > 0$ such that $Q \ge N$ we have that

$$\lambda^2 |E_\lambda|^2 \le C_1 Q^{d/2} |E_\lambda|^2 + \frac{C_2 N^\epsilon}{Q}|E_\lambda| \tag{3.7}$$

holds for all λ and $\epsilon > 0$. The constants C_1 and C_2 are independent of N and Q.

Without loss of generality assume Q is a positive integer and consider $N \le Q \le N^2$. The idea is to suitably decompose the kernel K_σ in (3.6) into the sum of two kernels $K_{1,Q} + K_{2,Q}$ such that

$$\|K_{1,Q}\|_{L^\infty} \le C_1 Q^{d/2}$$

and

$$\|\widehat{K_{2,Q}}\|_{L^\infty} \le \frac{C_2 N^\epsilon}{Q},$$

whence the estimates (3.7) follow. To find such decomposition, the key is to choose an appropriate function supported on $[0,1]$ so that if we denote by $\Phi(t)$ its periodic extension we define

$$K_{1,Q}(\mathbf{x},t) = \frac{1}{\widehat{\Phi(0)}} K_\sigma(\mathbf{x},t)\Phi(t) \quad \text{and} \quad K_{2,Q} = K_\sigma - K_{1,Q}.$$

The Φ that works is the periodic extension of the function

$$\sum_{Q\le r<2Q}\sum_{a\in\mathcal{P}_r}\varphi\left(\frac{t-a/r}{1/r^2}\right),$$

where φ is a bump function supported on a small interval -say $[2^{-8},2^{-7}]$. Full details can be found in [47].

The point is that one can then show as corollaries the following estimates in [6]:

(1) If $\lambda \ge CN^{d/4}$ for some suitable $C>0$, then the level set E_λ defined Theorem 3.5 above satisfies $|E_\lambda|\lesssim N^\varepsilon\lambda^{-\frac{2(d+2)}{d}}$

(2) For each positive $\varepsilon>0$ we have that $\sqrt{A_{q,d,N}}\le C_\varepsilon N^{\frac{d}{2}-\frac{d+2}{q}+\varepsilon}$ provided $q>\frac{2(d+4)}{d}$; which in turn immediately yield (3.3) in this case.

3.2. **The Strichartz estimates on general tori.** As mentioned above, the study of the NLS on general rectangular tori was first started in the work of Bourgain [10] where it was shown that certain Strichartz estimates with a loss of derivative hold. Some other partial results for the NLS on irrational tori were obtained in [17, 28, 41, 63]. The combined range of estimates proved for irrational tori in these works are weaker than those proved by Bourgain in [6] due to number-theoretical difficulties. A completely different approach to the problem was recently taken in the work of Bourgain and Demeter [14, 15] - see also prior work by C. Demeter [27]-. Such approach has led to the full range of Strichartz estimates conjectured in [6] (c.f.[10]) up to ϵ-loss for irrational tori. This ϵ loss was removed in subsequent work by Killip and Vişan [52].

In [14] Bourgain and Demeter actually prove a stronger result than the Strichartz estimates. Namely they establish the **ℓ^2-decoupling conjecture** (Theorem 4.1 below) whence- in addition to proving the Strichartz estimates on general (rational or irrational) tori- they also derive perhaps somewhat surprisingly new results in number theory and in incidence geometry theory.

Our interest in these notes is in understating how Bourgain and Demeter establish the Strichartz estimates for general (rational or irrational) tori.

Relabeling the notation: From now through the end of Section 4 we follow the notation in [14] and relabel the dimension d as $n-1$. Hence $\mathbb{T}^{d+1}$ will become $\mathbb{T}^n$. Furthermore, the L^q in Conjecture 3.1 and subsequent presentation above will become L^p (that is we will use p in lieu of q). This p should not be confused with the power nonlinearity of NLS.

Theorem 3.6. *[Strichartz estimates for general tori]. Let $\phi\in L^2(\mathbb{T}^{n-1})$ with $\operatorname{supp}\widehat{\phi}\subset[-N,N]^{n-1}$. Then for each $\varepsilon>0$, $p\ge\frac{2(n+1)}{n-1}$ and each interval $I\subset\mathbb{R}$ with $|I|\gtrsim 1$ we have,*

$$\|e^{it\Delta_\theta}\phi\|_{L^p(\mathbb{R}^{n-1}\times I)}\lesssim_\varepsilon N^{\frac{n-1}{2}-\frac{n+1}{p}+\varepsilon}|I|^{1/p}\|\phi\|_2, \tag{3.8}$$

The implicit constant does not depend on I,N or $\theta:=(\theta_1,\dots,\theta_{n-1})$ as in (2.16).

Remark 3.7. We note that Theorem 3.6 in particular fully establishes Bourgain's Conjecture 3.1. Bourgain showed in [6] Proposition 3.113 how to remove[5] the ε-loss for $p>\frac{2(n+1)}{n-1}$ in the case of square (rational) tori. Recall that the ε-loss is

[5]i.e. obtain scale invariant estimates

necessary when $q = \frac{2(n+1)}{n-1}$ as discussed above in Remark 3.2. Recent work by Killip and Vişan [52] show how to remove the ε-loss for $p > \frac{2(n+1)}{n-1}$ in the case of irrational tori; in fact their argument works for either rational or irrational tori.

Remark 3.8. As a consequence of Theorem 3.6 one can prove in particular that the p-NLS equation (2.18) on general tori is locally well-posed in $H^s(\mathbb{T}^d)$ for any $s > \frac{d}{2} - \frac{2}{p-1}$.

The proof of Theorem 3.6 follows rather quickly once Theorem 4.1 below is proven. The idea is to use the discrete version of the ℓ^2 decoupling theorem, as was done in [11], and apply a change of variables to (3.8) which puts us in the perfect position to apply the discrete estimate. We therefore focus on proving Theorem 4.1 below.

4. ℓ^2 DECOUPLINGS

We provide a brief overview of the proof of the ℓ^2 decoupling conjecture by J. Bourgain and C. Demeter in [14]. We borrow heavily from L. Guth's notes on the topic [43]. In what follows and in order to remain faithful to the literature we relabel the spatial dimension d as $n-1$ so that the space-time dimension will now be n. Hence $n = 2$ means 1 spatial dimension and so forth.

Throughout this section we take S to be a compact C^2 hypersurface in $\mathbb{R}^n$ with positive definite second fundamental form. The typical example we will always refer to is the *truncated elliptic paraboloid*,

$$P^{n-1} := \{(\xi_1, \ldots, \xi_{n-1}, \xi_1^2 + \cdots + \xi_{n-1}^2) \in \mathbb{R}^n : |\xi_i| \leq 1/2\}.$$

We assume $n \geq 2$, and to fix ideas we will frequently give examples where $n = 2$.

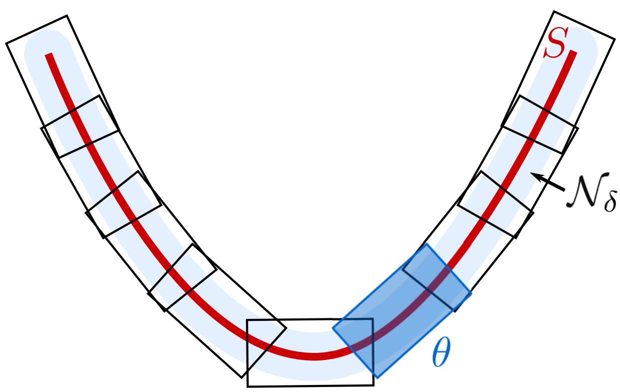


FIGURE 5. The setup for the truncated parabola P^1. Each rectangular region represents a θ "slab", and T is the collection of all such θ.

The main result is the following theorem:

Theorem 4.1 (ℓ^2 Decoupling (Theorem 1.1 in [14])). *Let S be a compact C^2 hypersurface in $\mathbb{R}^n$ with positive definite second fundamental form. Let $\mathcal{N}_\delta S$ be* *the δ-neighborhood of S and let T be a covering of $\mathcal{N}_\delta S$ by blocks θ of dimension $\delta^{1/2} \times \cdots \times \delta^{1/2} \times \delta$. If* $\operatorname{supp}(\widehat{f}) \subseteq \mathcal{N}_\delta S$ *then for* $p \geq \frac{2(n+1)}{n-1}$ *and* $\epsilon > 0$,

$$\|f\|_p \leq C_{p,n,\epsilon}\, \delta^{-\frac{n-1}{4}+\frac{n+1}{2p}-\epsilon} \left(\sum_{\theta \in T} \|f_\theta\|_p^2 \right)^{1/2}. \tag{4.1}$$

Note that we will often switch between using δ and R, where $\delta = R^{-1}$.

4.1. **Main steps.** Let us first note that the subcritical estimate

$$\|f\|_p \lesssim_\epsilon \delta^{-\epsilon} \left(\sum_{\theta \in T} \|f_\theta\|_p^2 \right)^{1/2} \quad \text{where} \quad 2 \leq p \leq \frac{2(n+1)}{n-1},$$

will become possible by a localization argument and interpolation between the trivial $p = 2$ case and the endpoint $p = 2(n+1)/(n-1)$ from Theorem 4.1, which we henceforth refer to as s to distinguish it from general p. The endpoint s here is hinted at in prior discussions of this topic. G. Garrigós and A. Seeger proved in [37] that, up to the ε term, the exponent $-\frac{n-1}{4} + \frac{n+1}{2p} - \varepsilon$ of δ in Theorem 4.1 is optimal. Thus the clear breaking point for when this exponent is a constraint is precisely s. Furthermore, in our argument we will encounter the norm $\|f\|_{L^{\frac{2(n+1)}{n}}}$, and we the algebraic properties of s allow the convenient bound

$$\|f\|_{L^{\frac{2(n+1)}{2}}} \leq \|f\|_{L^2}^{1/2} \|f\|_{L^s}^{1/2}.$$

via the Hölder inequality.

For brevity, we now focus on the endpoint case exclusively, following closely the notes [43] by L. Guth. As mentioned above, proving the endpoint case quickly implies the subcritical estimate as well. The supercritical estimate $p > s$ requires a different argument, but many of the tools introduced here are used in that proof also.

4.1.1. *Decoupling Norms.* We begin by inspecting the right side of the ℓ^2 decoupling inequality (4.1) further. For any f such that $\operatorname{supp} \widehat{f} \subseteq \mathcal{N}_\delta S$ and $\Omega \subseteq \mathbb{R}^n$ any domain we fix a covering T of $\mathcal{N}_\delta S$ and define

$$\|f\|_{L^{p,\delta}(\Omega)} := \left(\sum_{\theta \in T} \|f_\theta\|_{L^p(\Omega)}^2 \right)^{1/2} = \big\| \|f_\theta\|_{L^p(\Omega)} \big\|_{\ell^2(T)}.$$

This turns out to be a norm with some similar properties to the L^p norms, in particular it satisfies the Hölder-type inequality

$$\|f\|_{L^{q,\delta}(\Omega)} \leq \|f\|_{L^{q_1,\delta}(\Omega)}^{1-\alpha} \|f\|_{L^{q_2,\delta}(\Omega)}^{\alpha} \tag{4.2}$$

for $1 \leq q, q_1, q_2 \leq \infty$, $0 < \alpha < 1$, and $\frac{1}{q} = (1-\alpha)\frac{1}{q_1} + \alpha \frac{1}{q_2}$.

It is useful to establish the following superadditive property, which is proven using the Minkowski inequality for the $\ell^{p/2}$ norm.

Lemma 4.2. *If Ω is a disjoint union of Ω_j and $p \geq 2$, then for any δ and any f with* $\operatorname{supp} \widehat{f} \subseteq \mathcal{N}_\delta S$, *we have*

$$\sum_j \|f\|_{L^{p,\delta}(\Omega_j)}^p \leq \|f\|_{L^{p,\delta}(\Omega)}^p$$

The benefit of this lemma is that it allows us to break Ω into disjoint pieces, and any decoupling norm which holds on each piece then holds on their union. This is what L. Guth calls *parallel decoupling* (see[43], page 2).

Lemma 4.3 (Parallel decoupling)**.** *Suppose that Ω is a disjoint union of Ω_j,* $\operatorname{supp}\widehat{f} \subseteq \mathcal{N}_\delta S$*, and $p \geq 2$. Suppose that for each j we have the inequality*

$$\|f\|_{L^p(\Omega_j)} \leq M\|f\|_{L^{p,\delta}(\Omega_j)}.$$

Then we also have the inequality

$$\|f\|_{L^p(\Omega)} \leq M\|f\|_{L^{p,\delta}(\Omega)}.$$

4.1.2. *Decoupling Constant.* We define the decoupling constant $D_p(R)$ as

$$D_p(R) := \inf\|f\|_{L^p(B_R)}/\|f\|_{L^{p,1/R}(B_R)},$$

where the infimum is taken over all f with $\operatorname{supp}\widehat{f} \subset \mathcal{N}_{1/R}S$. We note that $D_p(R)$ also depends on S, but we will ignore this point for now. The claim is that, at the endpoint s,

$$D_p(R) \lesssim R^\epsilon.$$

4.1.3. *Multiple Scales.* We consider the problem at multiple scales in Fourier space. Instead of breaking $\mathcal{N}_{1/R}S$ into pieces at the scale of θ one asks what happens if one starts with a function supported in $\tau \subseteq \mathcal{N}_{1/R}S$ and then breaks τ into θ caps. The result is the following proposition.

Proposition 4.4. *If $\tau \subseteq \mathcal{N}_{1/R}S$ is a $r^{-1/2}$ cap for some $r \leq R$,* $\operatorname{supp}\widehat{f} \subseteq \tau$*, and $\theta \subseteq \mathcal{N}_{1/R}S$ are $R^{-1/2}$ caps as before, then*

$$\|f\|_{L^p(\Omega)} \lesssim D_p(R/r)\left(\sum_{\theta\subseteq\tau}\|f_\theta\|^2_{L^p(B_R)}\right)^{1/2}.$$

The proof of this proposition is based on parabolic rescaling, in which we apply a linear transformation so that the region τ has diameter 1 and then use our parallel decoupling Lemma 4.3 from earlier. As a corollary of Proposition 4.4 we get the following estimate:

Corollary 4.5. *For any radii $R_1, R_2 \geq 1$, we have*

$$D_p(R_1R_2) \lesssim D_p(R_1)D_p(R_2).$$

As a result, we see that there is a unique $\gamma = \gamma(n,p)$ such that for all R, ϵ we have

$$R^{\gamma-\epsilon} \lesssim D_p(R) \lesssim R^{\gamma+\epsilon}.$$

We want to prove that $\gamma = 0$ at the endpoint $p = \frac{2(n+1)}{n-1}$.

4.1.4. *Multilinear versus Linear decoupling.* Perhaps inspired by the tractability of multilinear Kakeya and restriction over their linear counterparts, which we will discuss in Section 4.2, we now consider a multilinear version of the problem. We first define the notion of *transversality.*

Definition 4.6. A collection of $S_j \subset \mathbb{R}^n$ hypersurfaces are *transverse* if for any point $\omega \in S_j$, the normal vector $N_{S_j}(\omega)$ obeys

$$\text{Angle}(N_{S_j}(\omega), j^{\text{th}} \text{ coordinate axis}) \leq (10n)^{-1}.$$

Definition 4.7. We say that functions $f_1, \ldots, f_n$ on $\mathbb{R}^n$ obey the *multilinear decoupling setup* (MDS)[6] if

- For $i = 1, \ldots, n$, $\operatorname{supp} \widehat{f_i} \subseteq \mathcal{N}_{1/R} S_i$
- $S_i \subseteq \mathbb{R}^n$ are compact positively curved C^2 hypersurfaces.
- The surfaces S_j are transverse.

We define $\widetilde{D}_{n,p}(R)$ to be the smallest constant so that whenever f_i obey (MDS),

$$\left\| \prod_{i=1}^{n} |f_i|^{1/n} \right\|_{L^p(B_R)} \leq \widetilde{D}_{n,p}(R) \prod_{i=1}^{n} \|f_i\|_{L^{p,1/R}(B_R)}^{1/n}.$$

Bourgain and Demeter go on to prove the following relationship between linear decoupling and multilinear decoupling:

Theorem 4.8. *Suppose that in dimension $n-1$, the decoupling constant $D_{n-1,p}(R) \lesssim R^\epsilon$ for any $\epsilon > 0$. Then for any $\epsilon > 0$,*

$$D_{n,p}(R) \lesssim R^\epsilon \widetilde{D}_{n,p}(R).$$

The idea of the proof is choose small enough K^{-1} caps τ such that $|f_\tau|$ is morally constant on cubes of side length K in B_R. Then we cover B_R with cubes Q_K and classify them as broad or narrow depending on which τ make a significant contribution to $f|_{Q_K}$. The broad cubes can be controlled simply by the multillinear decoupling inequality, and the narrow ones are controlled by parallel decoupling and parabolic rescaling.

Note that we always have $\widetilde{D}_{n,p}(R) \leq D_{n,p}(R)$ for any n, p, R. Then, using induction on the dimension n, if the decoupling theorem holds in dimension $n-1$ for the endpoint s, then we have shown that

$$\widetilde{D}_{n,p}(R) \sim D_{n,p}(R) \sim R^\gamma.$$

In other words, the linear decoupling problem is equivalent to the multilinear decoupling problem. This is quite surprising, as other problems such as linear Kakeya currently seem harder to prove than multilinear Kakeya. Thus, for decoupling, we can attack the problem using multilinear methods, which we will now leverage to our advantage.

4.2. **Multilinear Kakeya.** This section is a description of the argument in [42], in which as Guth succintly states, 'the multilinear Kakeya inequality is a geometric estimate about the overlap pattern of cylindrical tubes in $\mathbb{R}^n$ pointing in different directions'. We will use it to prove the multilinear restriction estimate which will then allow us to prove the ℓ^2 decoupling conjecture.

Theorem 4.9 (Multilinear Kakeya)**.** *Suppose that $\{\ell_{j,a}\}$ is a finite collection of lines in $\mathbb{R}^n$, where $j \in \{1, \ldots, n\}$ and $a \in \{1, \ldots, N_j\}$ such that each line $\ell_{j,a}$ makes an angle of at most $(10n)^{-1}$ with the x_j-axis. Let $T_{j,a}$ be the characteristic function of the $1-$neighborhood of $\ell_{j,a}$, and let Q_S denote any cube of side length S. Then for any $\varepsilon > 0$ and any $S \geq 1$, the following integral inequality holds:*

$$\int_{Q_S} \prod_{j=1}^{n} \left(\sum_{a=1}^{N_j} T_{j,a} \right)^{\frac{1}{n-1}} \lesssim_\varepsilon S^\varepsilon \prod_{j=1}^{n} N_j^{\frac{1}{n-1}} \tag{4.3}$$

[6]Here we are directly quoting Guth's notes [43], page 5.

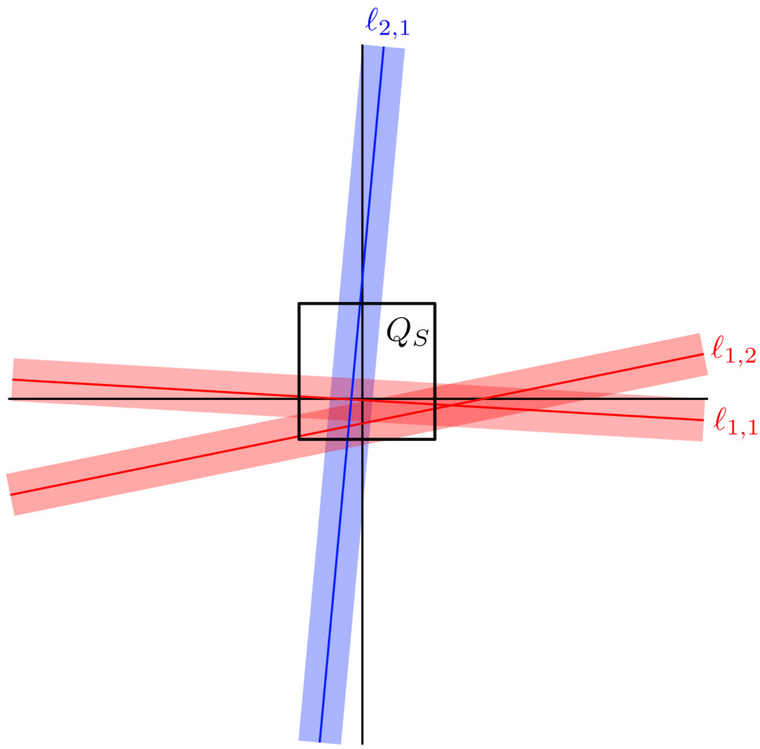


FIGURE 6. An example of the setup for Multilinear Kakeya.

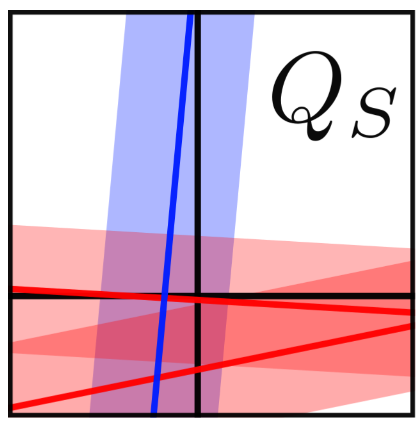


FIGURE 7. Zooming in on the square Q_S.

Figure 9 is an example of a setup for Multilinear Kakeya. The area being considered is simply that within the square Q_S. In addition, considering the values in the inequality (4.3) we note that $\sum_{a=1}^{N_j} T_{j,a}$ represents the color density of our overlayed transparencies (see Figure 7).

In the darker red portion, $\sum_{a=1}^{2} T_{1,a} = 2$, whereas for the lighter red portion this value will be 1, and for areas with no red the value is 0. Since there is only one line close to the x_2-axis in our example, $\sum_{a=1}^{1} T_{2,a} = T_{2,1} = 1$ for the blue areas in Figure 7 and 0 otherwise. Since there is a product on the left side of (4.3), the only portion which is being counted at all is the purple region where the 1-neighborhood of $\ell_{2,1}$ intersects the 1-neighborhood of $\ell_{1,1}$ or $\ell_{1,2}$.

4.2.1. *Nearly axis parallel.* The method of proving the Multilinear Kakeya inequality (Theorem 4.9) which was established by Bennett, Carbery, and Tao in [2] and is also followed by Guth in [42], is to first reduce to nearly axis parallel tubes:

Theorem 4.10. *For every $\varepsilon > 0$ there is some $\delta > 0$ so that the following holds.*

Suppose that $\ell_{j,a}$ are lines in $\mathbb{R}^n$, and that each line $\ell_{j,a}$ makes an angle of at most δ with the x_j-axis. Then for any $S \geq 1$ and any cube Q_S of side length S, we have:

$$\int_{Q_S} \prod_{j=1}^{n} \left(\sum_{a=1}^{N_j} T_{j,a} \right)^{\frac{1}{n-1}} \lesssim_\varepsilon S^\varepsilon \prod_{j=1}^{n} N_j^{\frac{1}{n-1}}$$

The claim is that Theorem 4.10 implies Theorem 4.9. Suppose Theorem 4.10 is true, then if $\delta \geq (10n)^{-1}$ (for a given ε) then we are easily done. If $\delta < (10n)^{-1}$ however, we would like to stretch along an axis whose lines are not within δ, bringing the lines closer to the axis. The only problem with this idea is that doing so also inevitably pulls other lines away from their axes. Clearly if one axis has lines which make too much of an angle, and the other axes are well within δ, we may be able to stretch the space so that all the lines are within delta.

One problem with this idea, however, is that the amount we stretch relies on knowing information about the lines which we don't have. Obviously the other issue is that this does not help us if more than one set of lines makes an angle of more than δ. The technique to handle both problems will be to split up over all possible contributions from various possible directions of lines and scale them each independently.

Assume that for $\varepsilon > 0$ the corresponding $\delta > 0$ from Theorem 4.10 is less than $(10n)^{-1}$. Then we split the spherical cap S_j of radius $(10n)^{-1}$ into caps $S_{j,\beta}$ of radius $\delta/10$, and then apply a linear change of coordinates to each cap centering it on the standard unit vector e_j.

In this case, the specific angle each $\ell_{j,a}$ makes is not important, as we know it is bounded by $(10n)^{-1}$, and so as we center each $S_{j,\beta}$ this linear change of coordinates has a controlled effect on lengths and areas and we can bound the overall integral by a sum of all combinations of contributions from these transformed systems, each of which is controlled by Theorem 4.10.

4.2.2. *Axis parallel (Loomis-Whitney).* The idea of the rest of the argument will be to further simplify matters by zooming in sufficiently close so that nearly axis parallel tubes look almost like axis parallel tubes. In this case we can get the bound we want using the Loomis-Whitney inequality, proven in [54], which states

Theorem 4.11 (Loomis-Whitney)**.** *Suppose that $f_j : \mathbb{R}^{n-1} \to \mathbb{R}$ are measurable functions, and let $\pi_j : \mathbb{R}^n \to \mathbb{R}^{n-1}$ be the linear map that forgets the j^{th} coordinate:*

$$\pi_j(x_1, \ldots, x_n) = (x_1, \ldots, x_{j-1}, x_{j+1}, \ldots x_n).$$

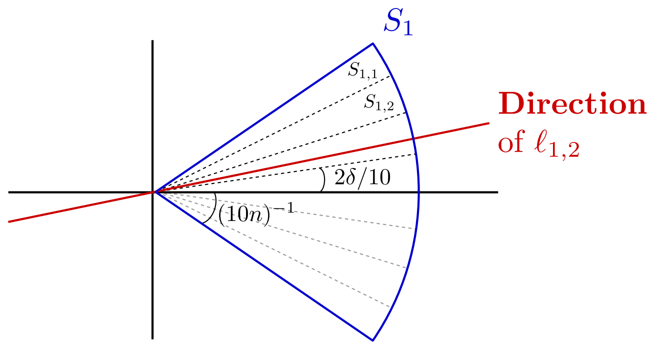


FIGURE 8. How to reduce Multilinear Kakeya to the Nearly Axis Parallel case: We split S_j into pieces, each of which has smaller radius than δ, and then sum over all contributions where the direction of $\ell_{j,a}$ is in $S_{j,\beta}$ (one from each S_j). Our angles here are not to scale.

Then the following inequality holds:

$$\int_{\mathbb{R}^n} \prod_{j=1}^{n} f_j(\pi_j(x))^{\frac{1}{n-1}} \leq \prod_{j=1}^{n} \|f_j\|_{L^1(\mathbb{R}^{n-1})}^{\frac{1}{n-1}}.$$

The connection between this theorem and the axis-parallel case is that a line parallel to the x_j-axis can be written as $\pi_j(x) = y_a$ for some $y_a \in \mathbb{R}^{n-1}$. Then, as noted in [42] by Guth, $\sum_a T_{j,a}(x) = \sum_\alpha \chi_{B(y_a,1)}(\pi_j(x))$, and applying Loomis-Whitney with $f_j = \sum_a \chi_{B(y_a,1)}$ we have

$$\int_{\mathbb{R}^n} \prod_{j=1} \left(\sum_{a=1}^{N_j} T_{j,a} \right)^{\frac{1}{n-1}} = \int_{\mathbb{R}^n} \prod_{j=1} (f_j(\pi_j(x)))^{\frac{1}{n-1}} \leq \prod_{j=1}^{n} \|f_j\|_{L^1(\mathbb{R}^{n-1})}^{\frac{1}{n-1}} \leq \omega_{n-1} N_j$$

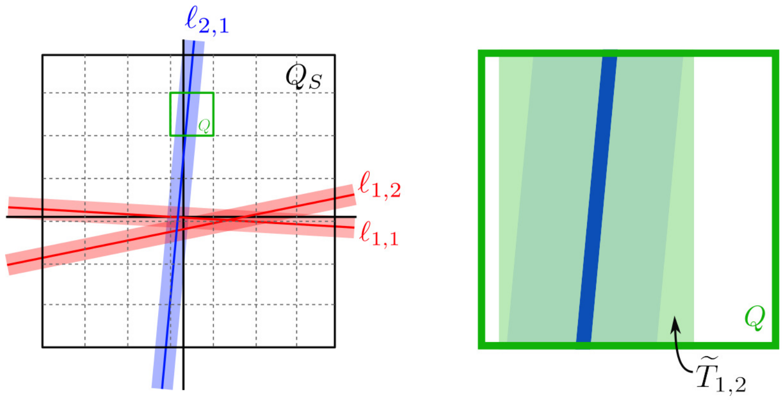
where ω_{n-1} is the volume of the $n-1$ dimensional unit ball. Therefore the axis parallel case does follow quickly from Loomis-Whitney, so we proceed to describe loosely the "zooming in" part of the argument.

Given a cube Q_S, we begin by splitting it up into small enough Q such that each tube $T_{j,a}$ which intersects a small Q can be covered by $\widetilde{T}_{j,a,R}$, an axis-parallel tube with slightly larger radius R.[7] Note that, since the $\widetilde{T}_{j,a,R}$ actually cover the $T_{j,a}$ within Q, we have

$$\int_Q \prod_{j=1}^{n} \left(\sum_a T_{j,a} \right)^{\frac{1}{n-1}} \leq \int_Q \prod_{j=1}^{n} \left(\sum_a \widetilde{T}_{j,a,R} \right)^{\frac{1}{n-1}} \lesssim R^n \prod_{j=1}^{n} N_j(Q)^{\frac{1}{n-1}}$$

where the last inequality follows from using Loomis-Whitney, and $N_j(Q)$ indicates the number of tubes $T_{j,a}$ intersecting Q. In fact, choosing Q small enough, we can make it so that if the tube $T_{j,a}$ intersects Q, the tube $T_{j,a,\delta^{-1}}$ of radius δ^{-1} around

[7] Exact details for what constitutes sufficiently small and slightly larger are contained in [42].



(b) Breaking a large Q_S into sufficiently small Q cubes.

(b) Zooming in on a particular Q, and covering $T_{2,1}$ with a slightly larger axis-parallel tube $\widetilde{T}_{2,1}$

FIGURE 9. Zooming in

$\ell_{j,a}$ is identically 1 on Q. Therefore

$$R^n \prod_{j=1}^{n} N_j(Q)^{\frac{1}{n-1}} \lesssim \frac{R^n}{|Q|} \int_Q \prod_{j=1}^{n} \left(\sum_a T_{j,a,\delta^{-1}} \right)^{\frac{1}{n-1}}.$$

As Guth shows in [42], with the appropriate choice of $|Q|$ and R, we can make $R^n/|Q| \lesssim \delta^n$. Since we can then sum over all Q, this proves the following lemma:

Lemma 4.12. *Suppose that $\ell_{j,a}$ are lines with angle at most δ from the x_j axis. Then if $S \geq \delta^{-1}$ and if Q_S is any cube of sidelength S, then*

$$\int_{Q_S} \prod_{j=1}^{n} \left(\sum_a T_{j,a} \right)^{\frac{1}{n-1}} \lesssim \delta^n \int_{Q_S} \prod_{j=1}^{n} \left(\sum_a T_{j,a,\delta^{-1}} \right)^{\frac{1}{n-1}}.$$

We have essentially traded off making the tubes larger for the δ^n factor. This can be seen as an exploit of the fact that a naive bound for the integrand is to assume that all tubes are identically 1 on Q_S, which yields $\prod_{j=1}^{n} N_j^{\frac{1}{n-1}}$, and so we lose nothing in the trade.

Without loss of generality, assume Q_S is centered at the origin. Now if $S \geq \delta^{-M}$ we begin induction on the scales $\delta^{-1}, \delta^{-2}, \ldots, \delta^{-M}$:

$$\int_{Q_S} \prod_{j=1}^{n} \left(\sum_a T_{j,a}(x) \right)^{\frac{1}{n-1}} dx \leq C_n \delta^n \int_{Q_S} \prod_{j=1}^{n} \left(\sum_a T_{j,a,\delta^{-1}}(x) \right)^{\frac{1}{n-1}} dx \tag{4.4}$$

$$= C_n \int_{\delta Q_S} \prod_{j=1}^{n} \left(\sum_a T_{j,a,\delta^{-1}}(\delta^{-1} x) \right)^{\frac{1}{n-1}} dx \tag{4.5}$$

where (4.5) follows by a change of variables. In the new coordinates, $T_{j,a,\delta^{-1}}(\delta^{-1}x)$ are just unit tubes again, and δQ_S is a cube with side lengths $\geq \delta^{-(M-1)}$, so we repeat the argument. After M repetitions we arrive at

$$\int_{Q_S} \prod_{j=1}^{n} \left(\sum_a T_{j,a}(x) \right)^{\frac{1}{n-1}} dx \leq C_n^M \int_{\delta^M Q_S} \prod_{j=1}^{n} \left(\sum_a T_{j,a,\delta^{-M}}(\delta^{-M} x) \right)^{\frac{1}{n-1}} dx,$$

and we can now use our naive bound to find

$$\int_{Q_S} \prod_{j=1}^{n} \left(\sum_a T_{j,a}(x) \right)^{\frac{1}{n-1}} dx \leq C_n^M (\delta^M S)^n \prod_{j=1}^{n} N_j^{\frac{1}{n-1}}.$$

If we had been working on a cube Q_S such that $S = \delta^{-M}$, at this point we would only need that $C_n^M \leq S^\varepsilon$ to be done. To accomplish this, we solve $S = \delta^{-M}$ for $M = -\log S / \log \delta$, thus we have

$$C_n^M = S^{-\frac{\log C_n}{\log \delta}}.$$

Therefore given $\varepsilon > 0$, we choose $\delta > 0$ such that $-\frac{\log C_n}{\log \delta} < \varepsilon$. Now we have proven the following lemma.

Lemma 4.13. *Given $\varepsilon > 0$, there exists $\delta > 0$ such that if $\ell_{j,a}$ are lines in $\mathbb{R}^n$ which make an angle of at most δ with the x_j-axis then, for every cube Q_S such that $S = \delta^{-M}$ for some integer M,*

$$\int_{Q_S} \prod_{j=1}^{n} \left(\sum_a T_{j,a} \right)^{\frac{1}{n-1}} \leq S^\varepsilon \prod_{j=1}^{n} N_j^{\frac{1}{n-1}}.$$

This is enough to prove Theorem 4.10, since given $\varepsilon > 0$ we take δ as in the above lemma. Then for any S we take M to be the largest integer such that $S \geq \delta^{-M}$, and then we cover S by at most C cubes of sidelength δ^{-M} where, apriori, C depends on both S and δ. By proving Lemma 4.13 for all integers M, however, we have been able to remove the dependence on S, since we can simply cover the cube Q_S with one of side length $\delta^{-(M+1)}$, and then figure out how many cubes of side length δ^{-M} are needed to cover this cube. Consequently the dependence of C is only on δ, which itself depends only on ε. Summing over these cubes yields

$$\int_{Q_S} \prod_{j=1}^{n} \left(\sum_a T_{j,a} \right)^{\frac{1}{n-1}} \lesssim_\varepsilon S^\varepsilon \prod_{j=1}^{n} N_j^{\frac{1}{n-1}}.$$

4.3. **Multilinear Restriction.** At this point we use the multilinear Kakeya inequality to prove multilinear restriction. One way to do this is to use the *method of induction on scales*, similar to the argument in [16], and similar to the argument described above for multilinear Kakeya, which itself was an argument by induction on scales. We first mention the following definition.

Definition 4.14. The $L^2_{avg}(\Omega)$ norm of a function f is defined as

$$\|f\|_{L^2_{avg}(\Omega)} := \left(\frac{1}{m(\Omega)} \int |f|^2 \, dm \right)^{1/2}$$

A key principle to apply induction on scales is to have some way for bounding a desired quantity of one scale by another. In multilinear Kakeya, this was Lemma 4.12. For multilinear restriction, it will be the following.

Lemma 4.15. *If* $\operatorname{supp} f_i \subseteq \mathcal{N}_{1/R} S_i$ *and* S_i *are smooth compact transverse hypersurfaces, and if* $2 \leq p \leq \frac{2n}{n-1}$, *then*

$$\operatorname{Avg}_{B_{R^{1/2}} \subset B_R} \prod_{i=1}^{n} \|f_i\|_{L^2_{avg}(B_{R^{1/2}})}^{p/n} \lesssim R^{\varepsilon} \prod_{i=1}^{n} \|f_i\|_{L^2_{avg}(B_R)}^{p/n},$$

where the average on the left is taken over a finite covering of B_R *by balls of radius* $R^{1/2}$.

By applying Bernstein's inequality at a sufficiently small scale r and then inductively working our way up using Lemma 4.15 we move through scales r^{2^m} until we reach R. This yields a decoupling estimate for $2 \leq p \leq \frac{2n}{n-1}$, as appeared in [11].

4.4. **Using curvature.** In the above argument we used transversality, but we did not use curvature at all. The introduction of curvature allows a more subtle estimate.

Lemma 4.16. *If* $\operatorname{supp} f_i \subseteq \mathcal{N}_{1/R} S_i$, *and* S_i *are compact positively curved transverse hypersurfaces, and* $s = \frac{2(n+1)}{n-1}$ *and* $\delta = R^{-1}$, *then*

$$\operatorname{Avg}_{B_{R^{1/2}} \subset B_R} \prod_{i=1}^{n} \|f_i\|_{L^2_{avg}(B_{R^{1/2}})}^{\frac{s}{n}} \lesssim R^{\varepsilon} \prod_{i=1}^{n} \|f_i\|_{L^2_{avg}(B_R)}^{\frac{1}{2}\cdot\frac{s}{n}} \prod_{i=1}^{n} \|f_i\|_{L^{s,\delta}_{avg}(B_R)}^{\frac{1}{2}\cdot\frac{s}{n}},$$

where the average on the left is taken over a finite covering of B_R *by balls of radius* $R^{1/2}$.

Again, note that this estimate is perfectly suited to an induction on scales type argument. In order to prove this lemma, Bourgain and Demeter show that we can reverse the previous Hölder inequality (4.2) if the function can be broken into a small number of "balanced" pieces, where each piece obeys a reverse Hölder inequality, i.e. if $1 \leq q, q_1, q_2 \leq \infty$ and $\frac{1}{q} = (1-\alpha)\frac{1}{q_1} + \alpha\frac{1}{q_2}$, then

$$\|f\|_{L^{q_1,\delta}(\Omega)}^{1-\alpha} \|f\|_{L^{q_2,\delta}(\Omega)}^{\alpha} \lesssim \|f\|_{L^{q,\delta}(\Omega)}. \tag{4.6}$$

The proof of this fact relies on a wave packet decomposition of f, and essentially interpolation with the $L^{p,\delta}$ norms, which in turn facilitates the proof of Lemma 4.16. Finally, using an induction on scales argument based on Lemma 4.16 as well as parabolic rescaling (discussed in Section 4.1.3) Bourgain and Demeter are able to prove ℓ^2 decoupling for the endpoint s. Some slight adjustments are needed to prove the $p > s$ range for Theorem 4.1 to be complete. In addition, relaxing certain simplifying assumptions we have made, requires weights to be brought into the equations, and a certain amount of care is needed to deal with them also.

5. The Nonlinear Schrödinger Equations: probabilistic methods

As we have seen in previous sections the local well posedness on tori in the subcritical regime[8] is in place once the Strichartz estimates (fully stated in Theorem 3.6) are available. However, in certain critical and more generally in all supercritical regimes *there is no known deterministic local well posedness theory.* However, what

[8]This was defined on $\mathbb{R}^d$ from the scaling symmetry of the equation. Of course on tori such scaling doesn't make sense. It is however still indicative of what to expect in terms of well posedness and so we transfer the the same terminology into the periodic setting.

is within reach is to study the local well posedness of p-NLS (2.18) on the square (rational) torus $\mathbb{T}^d$ from a probabilistic point of view; that is almost surely in the sense of probability. Such approach was first used by Bourgain [7] in the mid 90's to prove that the (Wick ordered) cubic nonlinear Schrödinger equation on $\mathbb{T}^2$ was almost sure locally well-posed in $H^{-\epsilon}(\mathbb{T}^2)$. The latter was the *first result in a supercritical regime*, since $L^2(\mathbb{T}^2)$ is critical for this equation (see also [8, 9]). Recent work by A. Nahmod and G. Staffilani [57] established almost sure locally well-posed in $H^{1-\alpha}(\mathbb{T}^3)$ (some $\alpha > 0$, small) for the quintic nonlinear Schrödinger equation on $\mathbb{T}^3$. This result is also in the supercritical regime since $H^1(\mathbb{T}^3)$ is critical for this equation.

It is worth noting that while for the quintic NLS on $\mathbb{T}^3$, (deterministic) large data well-posedness at the critical $H^1(\mathbb{T}^3)$ regularity is known ([46] for local and [48] for global), to date, there is no known (deterministic) large data well posedness results available for the cubic NLS equation on $\mathbb{T}^2$ at critical $L^2(\mathbb{T}^2)$ regularity.

We explain some of the ideas behind this probabilistic approach below.

5.1. **Random data: a nondeterministic approach.** We start by giving an informal definition of almost sure well posedness. Given μ a probability measure on the space of initial data X (eg. $X = H^s$) we say that the Cauchy initial value problem (IVP) is almost sure locally well-posed if there exists $Y \subset X$, with $\mu(Y) = 1$ and such that for any $\phi \in Y$ there exist $T > 0$ and a unique solution u to the IVP with data ϕ which is in $C([0,T], X)$ with data ϕ that is also stable in the appropriate topology.

The general idea is to consider the Cauchy initial data problem for rough but randomized initial data. To understand why randomization (of the initial data) helps let us recall the following classical result going back to Rademacher, Kolmogorov, Paley and Zygmund proving that random series on the torus enjoy better L^p bounds than deterministic ones[9]. For example, consider *Rademacher Series* :

$$f(\tau) := \sum_{m=0}^{\infty} b_m \, r_m(\tau), \; \tau \in [0,1), \; b_m \in \mathbb{C}$$

where

$$r_m(\tau) := \operatorname{sign} \sin(2^{m+1} \pi \, \tau)$$

Note that if $b_m \in \ell^2$ the sum $f(\tau)$ converges a.e. The following is a classical result which can be found in Zygmund's book.

Theorem 5.1. *If $b_m \in \ell^2$ then the sum $f(\tau)$ belongs to $L^p([0,1))$ for all $p \geq 2$. More precisely,*

$$\Big(\int_0^1 |f|^p \, d\tau \Big)^{1/p} \;\sim\; \| \, b_m \, \|_{\ell^2}$$

The key point is that although randomized initial data live in the same (rough) space as the original (un-randomized) data, their linear flow enjoy *almost surely* improved L^p bounds. These bounds in turn yield improved nonlinear estimates *almost surely* in the analysis of the solution to the difference equation (obtained after subtracting from u the linear evolution of randomized data). More precisely,

[9]Akin to the Kintchine inequalities used to prove the Littlewood-Paley inequalities.

the general scheme is as follows. Consider the Cauchy initial value problem,

$$\begin{cases} iu_t + \Delta u = N(u) & x \in \mathbb{T}^d,\ t > 0 \\ u(0,x) = \phi(x), \end{cases} \tag{5.1}$$

and assume that $\phi \in X^s$. Then if we denote by $a_k := \widehat{\phi}(k)$, to solve (5.1) we proceed as follows:

(1) Randomize ϕ: that is consider $\phi^\omega := \sum_{k\in\mathbb{Z}^d} a_k\, g_k(\omega)\, e^{ix\cdot k}$ where $\{g_k(\omega)\}_k$ are i.i.d. standard (complex/real) centered (Gaussian) random variables on a probability space $(\Omega, \mathcal{F}, P)$.
(2) Let v^ω be the linear evolution with initial datum ϕ^ω.
(3) Prove that v^ω satisfies 'better estimates' than ϕ almost surely.
(4) Show that $w := u - v^\omega$ solves a difference equation and obtain for w a *deterministic* local well-posedness theory in $C([0,T]; X^{s'})$, $s' > s$. That is, prove that almost surely in ω the nonlinear part w is *smoother* than the linear part v^ω.

Remark 5.2. The difference equation that w solves is not an equation at the subcritical/smoother level but rather it is a *hybrid* equation with a nonlinearity containing a mixture of supercritical but random terms plus deterministic (smoother) ones.

Remark 5.3. For $\phi \in H^s, \phi^\omega(x)$ defines almost surely in ω a function in H^s but not in $H^{s'}$ for any $s' > s$. A representative example arises by considering $a_k = \dfrac{1}{|k|^\alpha}$ then $\widehat{\phi^\omega}(k) = \frac{g_k(\omega)}{|k|^\alpha}$ gives rise almost surely in ω to a function in $H^{\alpha - \frac{d}{2} - \epsilon}$ but not in $H^{\alpha - \frac{d}{2}}$.

Randomization does not improve regularity in terms of derivatives. The improvement is with respect to L^p spaces *almost surely.* Another way to rephrase this and the classical Theorem 5.1 above is as follows: Let $\{g_m(\omega)\}$ be a sequence of complex i.i.d. zero mean Gaussian random variables on a probability space $(\Omega, A, \mathbb{P})$ and $(c_m) \in \ell^2$. Define

$$F(\omega) := \sum_m c_m g_m(\omega).$$

Then, there exists $C > 0$ such that for every $\lambda > 0$ we have

$$\mathbb{P}(\{\omega : |F(\omega)| > \lambda\}) \le \exp\left(\frac{-C\,\lambda^2}{\|F\|^2_{L^2(\Omega)}}\right).$$

As a consequence there exists $C > 0$ such that for every $q \ge 2$ and every $(c_m)_m \in \ell^2$,

$$\left\| \sum_m c_m g_m(\omega) \right\|_{L^q(\Omega)} \le C\sqrt{q} \left(\sum_m c_m^2 \right)^{\frac{1}{2}}.$$

More generally one uses the following where k would represent the number of random terms in the multilinear estimate at hand.

Proposition 5.4 (Large Deviation-type)**.** *Let $d \ge 1$ and $c(m_1, \ldots, m_k) \in \mathbb{C}$. Let $\{g_m\}_{1\le m\le d} \in \mathcal{N}_{\mathbb{C}}(0,1)$ be complex centered L^2 normalized independent Gaussians.*

For $k \geq 1$ denote by $A(k,d) := \{(m_1, \dots, m_k) \in \{1, \dots, d\}^k,\, m_1 \leq \dots \leq m_k\}$ and

$$F_k(\omega) = \sum_{A(k,d)} c(m_1, \dots, m_k) g_{m_1}(\omega) \dots g_{m_k}(\omega).$$

Then for $p \geq 2$

$$\|F_k\|_{L^p(\Omega)} \lesssim \sqrt{k+1}(p-1)^{\frac{k}{2}} \|F_k\|_{L^2(\Omega)}.$$

As a consequence from Chebyshev's inequality for every $\lambda > 0$,

$$\mathbb{P}(\{\omega \,:\, |F_k(\omega)| > \lambda\}) \leq \exp\left(\frac{-C\,\lambda^{\frac{2}{k}}}{\|F\|_{L^2(\Omega)}^{\frac{2}{k}}}\right).$$

This result follows from the hyper-contractivity property of the Ornstein-Uhlenbeck semigroup by writing $G_n = H_n + iL_n$ where $\{H_1, \dots, H_d, L_1, \dots L_d\} \in \mathcal{N}_{\mathbb{R}}(0,1)$ are real centered independent Gaussian random variables with the same variance. (c.f. [71, 68])

The key observation is that for given $\delta, r > 0$, the large deviation result above with -say -

$$\lambda = \delta^{-\frac{3}{2}r} \|F_k\|_{L^2(\Omega)}$$

will allow us to replace $|F_k(\omega)|^2$ by $\|F_k\|_{L^2(\Omega)}^2$ on a set $\Omega_\delta \subset \Omega$ with $\mathbb{P}(\Omega_\delta^c) < e^{-\frac{1}{\delta^r}}$. Thus we use the independence and normalization of the random variables to reduce matters to geometric considerations and integer lattice counting.

5.2. **Almost sure local well posedness results for the periodic NLS.** Bourgain's almost sure local well posedness result for the (Wick ordered) cubic NLS on the square/rational torus $\mathbb{T}^2$ reads as follows:

Theorem 5.5 (Bourgain [7])**.**

$$\begin{cases} iu_t + \Delta u = |u|^2 u \;-\; (\int |u|^2 dx)\, u \\ u(0,x) = \phi(x), \quad x \in \mathbb{T}^2, \end{cases}$$

is almost sure locally well-posed below L^2, that is for supercritical data $\phi \in H^{-\varepsilon}(\mathbb{T}^2)$.

Remark 5.6. The typical data considered is $\phi(x) = \sum_{k \in \mathbb{Z}^2} \frac{1}{|k|} e^{ix\cdot k} \in H^{-\epsilon}(\mathbb{T}^2)$ and $\phi^\omega(x) = \sum \frac{g_k(\omega)}{|k|} e^{ix\cdot k} \in H^{-\varepsilon}(\mathbb{T}^2)$ defining almost surely in ω a function in $H^{-\epsilon}(\mathbb{T}^2)$.

In [57] we considered the energy-critical quintic nonlinear Schrödinger equation on the squared/rational torus $\mathbb{T}^3$:

$$\begin{cases} iu_t + \Delta u = \lambda u|u|^4 \quad x \in \mathbb{T}^3 \\ u(0,x) = \phi(x) \quad \in \; H^\gamma(\mathbb{T}^3), \end{cases} \tag{5.2}$$

and established an almost sure local well posedness for random data in $H^\gamma(\mathbb{T}^3), \gamma < 1$; that is in the supercritical regime relative to scaling. The problem we considered is the analogue of Bourgain's Theorem 5.5 mentioned above. In our problem we consider data $\phi \in H^{1-\alpha-\varepsilon}(\mathbb{T}^3)$ for any $\varepsilon > 0$ of the form

$$\phi(x) = \sum_{k \in \mathbb{Z}^3} \frac{1}{\langle k \rangle^{\frac{5}{2}-\alpha}} e^{ik\cdot x}$$

whose randomization is

$$\phi^\omega(x) = \sum_{k\in\mathbb{Z}^3} \frac{g_k(\omega)}{\langle k\rangle^{\frac{5}{2}-\alpha}} e^{ik\cdot x} \tag{5.3}$$

where $(g_k(\omega))_k$ is a sequence of complex i.i.d centered Gaussian random variables on a probability space $(\Omega, A, \mathbb{P})$.

5.2.1. *The difference equation. Heart of the matter.* Assume u solves our IVP, then we define $w := u - S(t)\phi^\omega$, where $S(t)\phi^\omega$ is the linear evolution of the initial profile ϕ^ω. We study the IVP for w which solves a difference equation with nonlinearity,

$$\tilde{N}(w) := |w + S(t)\phi^\omega|^4 (w + S(t)\phi^\omega),$$

and prove that w belongs to H^s for some $s > 1$. The heart of the matter is to prove multilinear estimates for $\tilde{N}(w)$ to then be able to set up a contraction method to obtain well-posedness. The randomness coming from $(g_k(\omega))$ will allow us to say that in a certain space the nonlinearity increases its regularity so that it can hold a bit more than one derivative.

For the quintic NLS equation (5.2) however, multilinear estimates for $\tilde{N}(w)$ can be obtained only after having removed certain resonant terms involved in the nonlinear part of the equation. In Bourgain's case [7] the nonlinearity is *cubic* in 2D and a Wick ordering of the Hamiltonian takes care of bad resonant terms. In our case the nonlinearity is quintic in 3D and Wick ordering is not sufficient to remove the bad *resonant* terms. Instead, a suitable gauge transformation is required [57]. Another difficulty relative to Bourgain's case is that the arithmetic aspects of the problem (integer lattice counting lemmata) in a $3D$ integer lattice are much less favorable than for the $2D$ one.

Let us denote by $\mathcal{X}^s([0,\delta))$ the solution space for the nonlinear part of the solution. Our result reads as follows:

Theorem 5.7 (Nahmod-Staffilani [57])**.** *Let* $0 < \alpha < \frac{1}{12}$, $\;s = s(\alpha) > 1$ *and* ϕ^ω *as in* (5.3) *Then there exists* $0 < \delta_0 \ll 1$ *and* $r = r(s,\alpha) > 0$ *s.t. for any* $\delta < \delta_0$, *there exists* $\Omega_\delta \in A$ *with*

$$\mathbb{P}(\Omega_\delta^c) < e^{-\frac{1}{\delta^r}},$$

and for each $\omega \in \Omega_\delta$ *there exists a unique solution* u *of the quintic NLS* (5.2) *in the space*

$$S(t)\phi^\omega + \mathcal{X}^s([0,\delta)),$$

with initial condition ϕ^ω*.*

The results in [7] and [57] are for the square/rational torus $\mathbb{T}^d$. Despite the fact that we now know the Strichartz estimates on irrational tori by the work of Bourgain and Demeter, these were proved by very different techniques than those in [6]. It is yet not clear how to use Bourgain and Demeter techniques to shed light into, or in lieu of, the necessary integer lattice counting estimates in the irrational setting. For the rational tori these estimates were obtained via analytic number theory results as part of the proof of the (rational) Strichartz estimates established in [6], and these themselves (beyond the Strichartz estimates per se) are a crucial ingredient in proving the necessary estimates in Theorems 5.5 and 5.7. The good news is that as proved by Bourgain and Demeter in [14] the Decoupling Theorem seems amenable to counting solutions of Diophantine inequalities (e.g. Theorems 2.18 and 2.19 in [14]).

References

1. M. Ablowitz, *Nonlinear Dispersive Waves. Asymptotic Analysis and Solitons* Cambridge Texts in Applied Mathematics, Cambridge University Press, 2011.
2. Bennett, J., Carbery, A. and Tao, T. *On the multilinear restriction and Kakeya conjectures* Acta Math. 196 (2006), no. 2, 261-302
3. S. N. Bose, *Plancks Gesetz und Lichtquantenhypothese*, Zeitschrift für Physik, **26** (1924), 178.
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. Bourgain, J. *Eigenfunction bounds for the Laplacian on the n-torus*, Internat. Math. Res. Notices (1993), no. 3, 61-66.
6. J. Bourgain, *Fourier transform restriction phenomena for certain lattice subsets and applications to nonlinear evolution equations. I. Schrödinger equations*, Geom. Funct. Anal. 3 (1993), no. 2, 107-156
7. Bourgain, J., *Invariant measures for the 2D-defocusing nonlinear Schrödinger equation.* Comm. Math. Phys. 176, no. 2, 421- 445, [1996].
8. Bourgain, J., *Nonlinear Schrödinger equations*, IAS/ Park City Math. Series **5** pp. 3-157, [1999].
9. Bourgain, J., *Global solutions of nonlinear Schrödinger equations*, American Mathematical Society, Colloquium Publications **46** [1999].
10. J. Bourgain, *On Strichartz's inequalities and the nonlinear Schrödinger equation on irrational tori*, Mathematical aspects of nonlinear dispersive equations, 1–20, Ann. of Math. Stud., **163**, Princeton Univ. Press, Princeton, NJ, 2007.
11. J. Bourgain, *Moment inequalities for trigonometric polynomials with spectrum in curved hypersurfaces*, Israel J. Math. 193 (2013), no. 1, 441-458.
12. J. Bourgain and C. Demeter, *Improved estimates for the discrete Fourier restriction to the higher dimensional sphere*, Illinois J. Math. 57 (2013), no. 1, 213-227
13. J. Bourgain and C. Demeter, *New bounds for the discrete Fourier restriction to the sphere in four and five dimensions*, to apper in Int. Math. Res. Notices (2014). Preprint at `http://arxiv.org/pdf/1310.5244.pdf`
14. J. Bourgain and C. Demeter, *The proof of the ℓ^2 decoupling conjecture.* To appear in Annals of Math. Preprint at `http://arxiv.org/abs/1403.5335.pdf` (2014).
15. J. Bourgain, C. Demeter, *Decouplings for curves and hypersurfaces with nonzero Gaussian curvature*, Preprint at `http://arxiv.org/abs/1409.1634.pdf` (2014)
16. Bourgain, J. and Guth, L. *Bounds on oscillatory integral operators based on multilinear estimates* Geom. Funct. Anal. 21 (2011), no. 6, 1239-1295
17. F. Catoire, W.-M. Wang, *Bounds on Sobolev norms for the defocusing nonlinear Schrödinger equation on general flat tori*, Commun. Pure Appl. Anal. **9** (2010), no. 2, 483–491.
18. T. Cazenave, *Semilinear Schrödinger equations*, Courant Lecture Notes in Mathematics, 10. New York University, Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 2003. xiv+323 pp.
19. T. Chen, C. Hainzl, N. Pavlović, R. Seiringer, *Unconditional uniqueness for the cubic Gross-Pitaevskii hierarchy via quantum de Finetti*, preprint (2013), `http://arxiv.org/abs/1307.3168.pdf`.
20. T. Chen, N. Pavlović, *Recent results on the Cauchy problem for focusing and defocusing Gross-Pitaevskii hierarchies*, Math. Model. Nat. Phenom. **5** (2010), no. 4, 54–72.
21. T. Chen, N. Pavlović, *The quintic NLS as the mean field limit of a Boson gas with three-body interactions*, J. Funct. Anal. **260** (2011), no. 4, 959–997.
22. T. Chen, N. Pavlović, *Derivation of the cubic NLS and Gross-Pitaevskii hierarchy from many-body dynamics in $d = 2, 3$ based on spacetime norms*, Ann. Henri Poincaré **15** (2014), no. 3, 543–588.
23. J. Colliander, M. Keel, G. Staffilani, H. Takaoka, T. Tao, *Transfer of energy to high frequencies in the cubic nonlinear Schrödinger equation*, Invent. Math. **181** no.1, (2010), 39–113.
24. M.H. Anderson, J.R. Ensher, M.R. Matthews, C.E. Wieman, and E.A. Cornell, *Observations of Bose-Einstein condensation in a dilute atomic vapor*, Science **269** (1995), 198–201.
25. W. Craig, C. Sulem and P.L. Sulem, *Nonlinear modulation of gravity waves: a rigorous approach*, Nonlinearity, **5**, no. 2, (1992), 497–522.

26. D. DaSilva, N. Pavlovic, G. Staffilani and N. Tzirakis, *Global Well-Posedness for a periodic nonlinear Schrdinger equation in 1D and 2D*, Discrete Contin. Dyn. Syst. 19 (2007), no. 1, 3765.
27. C. Demeter, *Incidence theory and discrete restriction estimates*, Preprint at `http://arxiv.org/pdf/1401.1873v1.pdf` (2014).
28. S. Demirbas, *Local Well-posedness for 2-D Schrödinger Equation on Irrational Tori and Bounds on Sobolev Norms*, preprint (2013), `http://arxiv.org/abs/1307.0051`.
29. A. Einstein, *Quantentheorie des einatomigen idealen Gases*, Sitzungsberichte der Preussischen Akademie der Wissenschaften 1: 3. (1925).
30. A. Elgart, L. Erdős, B. Schlein, H.-T. Yau, *Gross-Pitaevskii Equation as the Mean Field Limit of Weakly Coupled Bosons*, Arch. Rational Mech. Anal. **179** (2006), 265–283.
31. L. Erdős, B. Schlein, H.-T. Yau, *Derivation of the Gross-Pitaevskii hierarchy for the dynamics of Bose-Einstein condensate*, Comm. Pure Appl. Math. **59** (2006), no. 12, 1659–1741.
32. L. Erdős, B. Schlein, H.-T. Yau, *Derivation of the cubic non-linear Schrödinger equation from quantum dynamics of many-body systems*, Invent. Math. **167** (2007), no. 3, 515–614.
33. L. Erdős, B. Schlein, H.-T. Yau, *Rigorous derivation of the Gross-Pitaevskii equation*, Phys. Rev. Lett. **98** (2007), no.4, 040404.
34. L. Erdős, B. Schlein, H.-T. Yau, *Rigorous derivation of the Gross-Pitaevskii equation with a large interaction potential*, J. Amer. Math. Soc. **22** (2009), no. 4, 1099–1156
35. L. Erdős, B. Schlein, H.-T. Yau, *Derivation of the Gross-Pitaevskii equation for the dynamics of Bose-Einstein condensate*, Ann. of Math. (2), **172** (2010), no. 1, 291–370.
36. L. Erdős, H.-T. Yau, *Derivation of the nonlinear Schrödinger equation from a many body Coulomb system*, Adv. Theor. Math. Phys. **5** (2001), no.6, 1169–1205.
37. G. Garrigós and A. Seeger, *A mixed norm variant of Wolff's inequality for paraboloids*, Harmonic Analysis and Partial Differential Equations, Contemporary Mathematics, vol. 505, Amer. Math. Soc., Providence, RI, 2010, (pp. 179-197)
38. P. Gressman, V. Sohinger, G. Staffilani, *On the uniqueness of solutions to the 3D periodic Gross-Pitaevskii hierarchy*, J. Funct. Anal. **266** (2014), no. 7, 4705–4764.
39. E. Gross, *Structure of a quantized vortex in boson systems*, Nuovo Cimento **20** (1961), 454–466.
40. M. Guardia M. and V. Kaloshin *Growth of Sobolev norms in the cubic defocusing nonlinear Schrödinger equation*, To appear in JEMS. Preprint at `http://arxiv.org/pdf/1205.5188v1.pdf` (2012).
41. Z. Guo, T. Oh, Y. Wang, *Strichartz estimates for Schrödinger equations on irrational tori*, Proc. London Math. Soc. **109** (2014), no. 4, 975–1013.
42. L. Guth, *A Short Proof of the Multilinear Kakeya Inequality.* Preprint at `arXiv:1409.4683v2` (2014).
43. L. Guth, Mini-course Notes on *Decoupling and multilinear estimates in harmonic analysis*, MIT (2014) `http://math.mit.edu/~lguth/decouplingseminar.html`.
44. Guth, L. *The endpoint case of the Bennett-Carbery-Tao multilinear Kakeya conjecture*, Acta Math. 205 (2010), no. 2, 263-286
45. Z. Hani, B. Pausader, N. Tzvetkov and N. Visciglia, *Modified scattering for the cubic Schrödinger equation on product spaces and applications*, Preprint at `http://arxiv.org/pdf/1311.2275v3.pdf` (2013.
46. S. Herr, D. Tataru and N. Tzvetkov, Global well-posedness of the energy-critical nonlinear Schrödinger equation with small initial data in $H^1(\mathbb{T}^3)$, *Duke Math. J.*, **159** (2011), 329–349.
47. Y. Hu and X. Li, *Discrete Fourier Restriction phenomenon associated with Schrödinger equations*, Rev. Mat. Iberoam. **30** no. 4, (2014),1-22.
48. A. D. Ionescu and B. Pausader, The energy-critical defocusing NLS on $\mathbb{T}^3$. *Duke Math. J.*, **161** (2012), 1581–1612.
49. K.B. Davis, M.O. Mewes, M.R. Andrews, N.J. van Druten, D.S. Durfee, D.M. Kurn, and W. Ketterle, *Bose-Einstein Condensation in a Gas of Sodium Atoms*, Phys. Rev. Let. **75** (1995), no. 22, 3969–3973.
50. W. Ketterle, *When Atoms behave like Waves: Bose-Einstein Condensation and the Atom Laser*, Nobel Lecture (2001). Available at http://`www.nobelprize.org/nobel_prizes/physics/laureates/2001/ketterle-lecture.pdf`

51. W. Ketterle, *When Atoms behave like Waves: Bose-Einstein Condensation and the Atom Laser*, Video of Nobel Lecture (2001). Available at http://www.nobelprize.org/nobel_prizes/physics/laureates/2001/ketterle-lecture.html
52. R. Killip, M. Vişan, *Scale invariant Strichartz estimates on tori and applications*, preprint (2014), http://arxiv.org/abs/1409.3603.
53. S. Klainerman, M. Machedon, *On the uniqueness of solutions to the Gross-Pitaevskii hierarchy*, Comm. Math. Phys. **279**, no.1 (2008), 169–185.
54. L. Loomis and H. Whitney, *An inequality related to the isoperimetric inequality*, Bull. Ameri. Math. Soc 55, (1949). (pp. 961-962)
55. J. Moloney and A. Newell, *Nonlinear optics* Westview Press, 2004.
56. C. Muscalu and W. Schlag, *Classical and Multilinear Harmonic Analysis, Vol. I*, Cambridge Studies in Advanced Mathematics, **137**, Cambridge University Press, 2013.
57. A. R. Nahmod and G. Staffilani, Almost sure well-posedness for the periodic 3D quintic nonlinear Schrödinger equation below the energy space, *J. Eur. Math. Soc.* (JEMS) (to appear), Preprint at http://arxiv.org/pdf/1308.1169.pdf.
58. L. Pitaevskii, *Vortex lines in an imperfect Bose gas*, Sov. Phys. JETP **13** (1961), 451–454.
59. V. Sohinger, *A rigorous derivation of the defocusing cubic nonlinear Schrödinger equation on $\mathbb{T}^3$ from the dynamics of many-body quantum systems*, preprint (2014), http://arxiv.org/abs/1405.3003, to appear in Ann. Inst. H. Poincaré (C), Analyse Non Linéaire.
60. E.M. Stein *Harmonic Analysis: Real-Variable Methods, Orthogonality and Oscillatory Integrals* , Princeton University Press, Princeton NJ, 1993.
61. C. Sulem and P.L. Sulem , *The Nonlinear Schrödinger Equation: Self-Focusing and Wave Collapse*, Series in Mathematical Sciences, **139**, Springer-Verlag, 1999.
62. R.S. Strichartz, *Restrictions of Fourier transforms to quadratic surfaces and decay of solutions of wave equations*, Duke Math. J. 44, (1977), 705–714.
63. N. Strunk, *Strichartz estimates for Schrödinger equations on irrational tori in two and three dimensions*, preprint (2014), http://arxiv.org/abs/1401.6080.
64. H. Takaoka and N. Tzvetkov, *On 2D nonlinear Schrödinger equations with data on $\mathbb{R} \times \mathbb{T}$*, J. Funct. Anal. **182**, (2001), 427–442.
65. T. Tao, *Recent Progress on the Restriction Conjecture*, IAS/Park City Mathematics Series, 2003. Preprint at http://arxiv.org/pdf/math/0311181.pdf.
66. T.Tao, *Some recent progress on the restriction conjecture*, Applied Numerical Harmonic Analysis, Birkhäuser, Boston, MA, (2004), 217–243.
67. T. Tao. *Nonlinear dispersive equations. Local and global analysis*, CBMS Regional Conference Series in Mathematics **106**, 2006.
68. L. Thomann and N. Tzvetkov, *Gibbs measure for the periodic derivative nonlinear Schrödinger equation.* Nonlinearity **23**, no. 11, (2010) 2771–2791.
69. N. Totz and S. Wu, *A Rigorous Justification of the Modulation Approximation to the 2D Full Water Wave Problem*, Comm. Math. Physics **310**, (2012), 817–883.
70. N. Totz, *A Justification of the Modulation Approximation to the 3D Full Water Wave Problem*, Comm. Math. Physics, to appear. Preprint at http://arxiv.org/abs/1309.5995
71. N. Tzvetkov *Construction of a Gibbs measure associated to the periodic Benjamin-Ono equation.* Prob. Theory Related Fields **146** no. 3-4, (2010) 481–514.
72. T. Wolff, *Lectures on Harmonic Analysis*, University Lecture Series, **29**. American Mathematical Society, Providence, RI, 2003

Department of Mathematics and Statistics, University of Massachusetts Amherst, 710 North Pleasant Street, Lederle GRT, Amherst, MA 01003

E-mail address: nahmod@math.umass.edu

String Theory and Math: Why This Marriage May Last

Mathematics and Dualities of Quantum Physics

Mina Aganagic

The relationship between mathematics and physics has a long history. Traditionally, mathematics provides the language physicists use to describe Nature, while physics brings mathematics to life: To discover the laws of mechanics, Newton needed to develop calculus. The very idea of a precise physical law, and mathematics to describe it were born together. To unify gravity and special relativity, Einstein needed the language of Riemannian geometry. He used known mathematics to discover new physics. General relativity has been inspiring developments in differential geometry ever since. Quantum physics impacted many branches of mathematics, from geometry and topology to representation theory and analysis, extending the pattern of beautiful and deep interactions between physics and mathematics throughout centuries.

String theory brings something new to the table: the phenomenon of duality. Duality is the equivalence between two descriptions of the same quantum physics in different classical terms. Ordinarily, we start with a classical system and quantize it, treating quantum fluctuations as small. However, nature is intrinsically quantum. One can obtain the same quantum system from two distinct classical starting points. For every precise question in one description of the theory, there is a corresponding question in the dual description. Duality is similar to a change of charts on a manifold, except it also has the power to map large fluctuations in one description to small fluctuations in the dual, and relate very hard mathematical problems in one are of mathematics to more manageable ones in another. Dualities are pervasive in string theory.

Understanding dualities requires extracting their mathematical predictions and proving the huge set of mathematical conjectures that follow. The best understood duality is *mirror symmetry*. But, mirror symmetry is but one example – many striking dualities have been discovered in quantum field theory (QFT) and many more in string theory over the last 20 years. Duality gives quantum physics, and especially string theory, the power to unify disparate areas of mathematics in surprising ways and provides a basis for a long lasting and profound relationship between the physics and mathematics.

1. Knot theory and Physics

To illustrate these ideas, I will pick one particular area of mathematics, knot theory. The central question of knot theory is: When are two knots (or links) distinct? A knot is an oriented closed loop in $\mathbb{R}^3$. A link consists of several disjoint, possibly tangled knots. Two knots are considered equivalent, if they are homotopic to each other. One approaches the question by constructing knot or link invariants, which depend on the knot up to homotopy.

Knot theory was born out of 19th century physics. Gauss' study of electromagnetism resulted in the first link invariant: the Gauss linking number, which is an invariant of a link with two knot components, K_1 and K_2. One picks a projection of the link onto a plane and defines (twice) the linking number as the number of crossings, counted with signs:

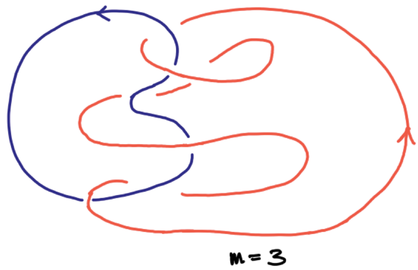


FIGURE 1. To define the sign of a crossing, we approach a crossing along the bottom strand, and assign +1 if the top strand passes from left to right, and −1 otherwise. In the figure, there are 6 crossings, each contributing +1, and so $m = 3$.

$$m(K_1, K_2) = \frac{1}{2} \sum_{\text{crossings}(K_1,K_2)} \text{sign(crossing)}. \tag{1}$$

Gauss discovered the linking number, and gave a beautiful integral formula for it:

$$m(K_1, K_2) = \frac{1}{2\pi} \oint_{K_1} \oint_{K_2} \frac{\vec{x}_1 - \vec{x}_2}{|\vec{x}_1 - \vec{x}_2|^3} \cdot (d\vec{x}_1 \times d\vec{x}_2). \tag{2}$$

Maxwell discovered it independently, some time later, and noted that it is not a very good link invariant – it is easy to find links that are non-trivial, yet the invariant vanishes. Note that, while the first formula (1) for the linking number relies on a choice of a projection, the second one (2) makes it manifest one is studying a link in three dimensional space.

Strikingly, quantum physics enters knot theory. In ′84, Vaughan Jones found a very good polynomial invariant of knots and links, by far the best at the time, depending on one variable q. The Jones polynomial is a Laurent polynomial in $q^{\frac{1}{2}}$; it can be computed in a simple way by describing how it changes as we reconnect the strands and change the knot. One picks a planar projection of the knot, and a neighborhood of a crossing. and defines the value of the Jones polynomial the

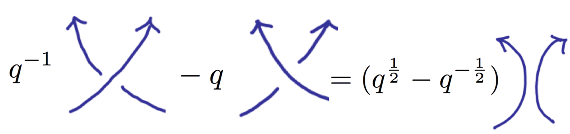


FIGURE 2. Skein relation for the Jones polynomial

skein relation it satisfies:

$$q^{-1} J_{K_+} - q J_{K_-} = (q^{\frac{1}{2}} - q^{-\frac{1}{2}}) J_{K_0}.$$

together with specifying its value for the unknot. While there are examples of distinct knots with the same $J_K(q)$, there is no known examples of non-trivial knots with $J_K(q)$ the same as for the unknot. Despite the ease of construction, the Jones polynomial seems mysterious. Since one has to pick a projection to a plane to define it, it is not obvious at the outset that one obtains an invariant of knots in three dimensional space, rather this is something one must prove. Secondly, what is the meaning of q?

Witten discovered that the Jones polynomial has its origin in quantum field theory: Chern-Simons (CS) gauge theory in three dimensions. Like Yang-Mills theory, Chern-Simons theory on a three-manifold M is written in terms of a connection

$$A = A_i dx^i$$

associated with a gauge group G. The theory is topological from the outset – its classical action is given in terms of Chern-Simons form on M,

$$S_{CS} = \frac{1}{4\pi} \int_M \mathrm{Tr}(A \wedge dA + \frac{2}{3} A \wedge A \wedge A).$$

and hence it is independent of the choice of metric on M. The path integral of the theory

$$Z(M) = \int \mathcal{D}A \exp\Big(ikS_{CS}\Big),$$

where one integrates over spaces of all connections on M and divides by the gauge group is a topological invariant of M. We can introduce a knot K in the theory by inserting a line observable along K,

$$\mathcal{O}_K(R) = \mathrm{Tr}_R\, Pexp\Big(i \oint_K A_i dx^i\Big)$$

in some representation R of the gauge group (P denotes path ordering of the exponential). This preserves topological invariance, so

$$Z(M; K, R) = \int \mathcal{D}A \exp(ikS_{CS})\ O_K(R)$$

is a topological invariant of the knot K in the three manifold M, which depends only on G, R and k. (More precisely, Chern-Simons theory produces an invariant of a framed three-manifold M, with framed knots. Framing is a choice of a homotopy class of trivialization of the tangent bundle of M and K. The need to fix the framing reflects an ambiguity in the phase the partition function [1].) The constant k is required to be an integer, for the integrand to be invariant under "large" gauge transformations, those corresponding to non-trivial elements of $\pi_3(G)$.

Witten made use of the topological invariance of the theory to solve Chern-Simons theory exactly on an arbitrary three manifold M with collection of knots, by cutting the three manifold into pieces, solving the theory on pieces and gluing back together. He showed that, taking M to be an S^3, $G = SU(2)$, and R, the defining two dimensional representation of $SU(2)$

$$M = S^3, G = SU(2), R = \square,$$

a suitably normalized Chern-Simons partition function

$$\langle \mathcal{O}_K \rangle = Z(M; K)/Z(M; \bigcirc)$$

equals the Jones polynomial

$$\langle \mathcal{O}_K \rangle = J_K(q).$$

The normalization we chose corresponds to setting $J_{\bigcirc}(q) = 1$. Chern-Simons theory gave a manifestly three dimensional formulation of the Jones polynomial.

It leads immediately to a vast generalization of Jones' knot invariant, by varying the gauge group G, the representations, and considering knots in an arbitrary three manifold M. Finally, the relation to Chern-Simons theory showed that the Jones polynomial is a quantum invariant: it is a Laurent polynomial in

$$q = exp(i\lambda),$$

where $\lambda = 2\pi/(k+2)$ plays the role of $\hbar$, the Planck constant, in Chern-Simons theory.

Let me pause for a moment to sketch what one means by saying the theory is solvable [1]. It is known that every three manifold M can be related to $S^2 \times S^1$ by a repeated application of surgery. A surgery to produce from M a new three manifold proceeds as follows. One picks an imaginary knot in M, cuts out its solid torus neighborhood, and glues it back in up to an $U \in SL(2,Z)$ transformation of the boundary. If U is not identity one obtains a new manifold M'. Quantum field theory is a functor that associates to a closed three manifold M a complex number $Z(M)$, the value of the path integral on M, and to a manifold with a boundary B a state in the vector space $\mathcal{H}_B$, the Hilbert space of the theory based on B. Vector spaces associated to the same B, with opposite orientation, are canonically dual. Gluing two manifolds over a common boundary B is the inner product of the corresponding states. So surgery on three manifolds translates to a following statement in QFT:

$$Z(M') = \langle 0|M'/K\rangle = \langle 0|U|M/K\rangle$$

Here $\langle 0|$ is the state corresponding to solid torus with no insertions, and $|M'/K\rangle$ the state corresponding to M' with a neighborhood of the knot K cut out. An arbitrary state in $\mathcal{H}_{T^2}$ can be obtained from a solid torus with a line observable colored by a representation R running through it. If we denote the resulting state $\langle R|$, we can write

$$\langle 0|U = \sum_R \langle R|\, U_{0R}.$$

The sum runs over a finite set of representations of G, depending on k. (The Hilbert space $\mathcal{H}_B$ of Chern-Simons theory with gauge group G at level k is the same as the space of conformal blocks of G_k WZW model; the latter is finite dimensional for any B.) This implies

$$Z(M') = \sum_R \langle R|M/K\rangle\, U_{0R} = \sum_R Z(M', K, R)\, U_{0R}$$

where $Z(M', K, R)$ corresponds to the partition function on M' with an actual knot K colored by representation R running through it. In this way, by repeated surgeries, we can reduce any three manifold invariant to that of $S^2 \times S^1$ with a braid running along the S^1. In turn, the later can be computed by

$$Z(S^2 \times S^1, L, R_i) = \mathrm{Tr}_{\mathcal{H}_{S^2,R_i}} B_L,$$

which comes about by first cutting the $S^2 \times S^1$ open into $S^2 \times R$, straightening the braid out, and then recovering the original braid by finding a collection B_L of time-ordered diffeomeorphisms of a sphere S^2 with marked points, which re-braid the braid. Gluing the ends together corresponds to taking the trace of B_L, acting on the Hilbert space $\mathcal{H}_{S^2,R_i}$ of the theory on S^2 with marked points colored by representations R_i determined by the braid.

To solve the theory one needs only a finite set of data. The $SL(2,\mathbb{Z})$ transformations of the torus are generated by a pair of matrices, S and T satisfying

$$S^4 = 1, \qquad (ST)^3 = S^2,$$

representing the action of $SL(2,\mathbb{Z})$ on $\mathcal{H}_{T^2}$. Similarly, the brading matrix B_L from is obtained from a finite set of data, the braiding matrix B and fusion matrix F on a four punctured sphere [2]. For Chern-Simons theory based on gauge group G, at level k, the S, T, B and F are finite dimensional matrices acting on conformal blocks G_k WZW 2d CFT. Reshetikhin and Turaev formalized this in terms of modular tensor categories [2]. Thus, one can reduce finding knot and three manifold invariants for arbitrary G and k and representations R to matrix multiplication, of a small set of matrices.

2. Gromov-Witten Theory

Quantum physics enters modern mathematics in other places as well. *Gromov-Witten theory* is an example. There, one studies quantum intersection theory of a projective variety X (see [3] for a review, and [4] for a quick overview). Classical intersection corresponds to picking classes

$$\gamma_1, \ldots, \gamma_n \in H_*(X)$$

with degrees $\sum_i \deg(\gamma_i^\vee) = 2d$, where $d = \dim_{\mathbb{C}}(X)$ and computing their intersection numbers, counted with signs:

$$\langle \gamma_1, \ldots, \gamma_n \rangle_{0,0} = \int_X \gamma_1^\vee \wedge \cdots \wedge \gamma_n^\vee, \tag{3}$$

where $\gamma_i^\vee \in H^*(X)$ denotes the Poincare dual of γ_i. Enumerative geometry turns this into a deeper geometric question by counting intersections over algebraic curves, insead over points: one would like to know how many algebraic curves of a give degree $\beta \in H_2(X)$ and genus g meet $\gamma_1, \ldots, \gamma_n$ at points. The corresponding invariant

$$\langle \gamma_1, \ldots, \gamma_n \rangle_{g,\beta}.$$

can be defined by picking a curve Σ of genus g, with n marked points $p_1, \ldots p_n$, and considering intersection theory on the moduli space $\mathcal{M}_{g,n}(X,\beta)$ of holomorphic maps

$$\phi : \Sigma \to X$$

of degree d. More precisely, as explained by Kontsevich, one needs consider *moduli space of stable maps* $\overline{\mathcal{M}}_{g,n}(X)$. This is a compactification of $\mathcal{M}_{g,n}(X,\beta)$ by allowing the domain curve Σ to have "ears", which are additional S^2 that bubble off, and considering *stable maps*, which he defined. Imposing the incidence condition that $\phi(p_i) \subset \gamma_i$ is implemented by pulling back the Poincare dual class $\gamma_i^\vee$ via the evaluation map ev_i. The evaluation map maps a point in the moduli space of maps to $\phi(p_i)$:

$$\langle \gamma_1, \ldots, \gamma_n \rangle_{g,d} = \int_{[\overline{\mathcal{M}}_{g,n}(X,\beta)]} \mathrm{ev}_1^*(\gamma_1^\vee) \cdots \mathrm{ev}_n^*(\gamma_n^\vee), \tag{4}$$

where the brackets [..] denote the (virtual) fundamental class. For genus zero, degree zero curves, the definition agrees with the classical intersection numbers in

(3). At genus zero, it is natural to combine the classical answer (3) and the higher degree data into a generating function of quantum intersection numbers of X,

$$\langle \gamma_1, \ldots, \gamma_n \rangle_{0,Q} = \sum_{\beta \in H_2(X)} \langle \gamma_1, \ldots, \gamma_n \rangle_{0,\beta}\, Q^\beta. \tag{5}$$

For a map of degree β to X, Q^β is the exponent the area of the target curve, $Q^\beta = exp(-\int_\Sigma \phi^* \omega)$, where ω is the Kahler form on X. The leading term in the series is the classical intersection, and the subleading terms are quantum corrections to it.

2.1. **Gromov-Witten Theory and Topological String Theory.** Gromov-Witten theory originates from string theory. It computes the amplitudes of a topological variant of superstring theory, called the *A-model topological string.*

In quantum field theory, to describe a particle propagating on a manifold X one sums over all maps from graphs Γ to X, satisfying certain conditions, where one allows moduli of graph to vary. In string theory, we replace point particles by strings, the maps from graphs Γ by maps from Riemann surfaces Σ to X. In superstring theory, one formulates this in terms of a path integral of a supersymmetric 2d QFT on Σ, describing a string propagating on X. To get topological string theory one modifies the supersymmetry generator Q to square to zero, $Q^2 = 0$ on arbitrary Σ. This turns the 2d QFT into a topological quantum field theory on Σ of cohomological type, with differential Q. The world sheet path integral receives contributions only from configurations that are annihilated by Q. If X is a Calabi-Yau manifold, there are are two inequivalent ways to obtain a TQFT, leading to topological A- and the B-model string theories. They correspond to two distinct generators Q_A and Q_B, each satisfying $Q^2_{A,B} = 0$. Topological A-type string exists for any Kahler manifold X. Restricting to configurations annihilated by Q_A turns out to restrict one to studying holomorphic maps to X only, leading to Gromov-Witten theory. In the B-model, the maps annihilated by Q_B are the constant maps, resulting in a simpler theory, depending on complex structure of X only.

Topological string theory was introduced by Witten in [6, 7], and developed by many (see for e.g. [8, 9] and [3] for a review). The mathematical formulation of

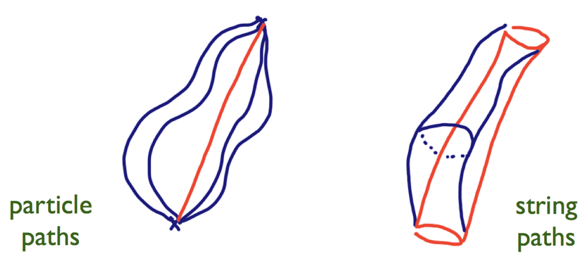


Figure 3. In string theory one sums over all possible paths of a string, leading to sum over surfaces.

Gromov-Witten theory is due to Kontsevich, Manin, Fukaya and many others [3]. The development of Gromov-Witten theory is an example of new mathematics that is inspired by questions in physics.

3. Duality

A quantum system is described by a collection of observables $\mathcal{O}_i$, corresponding to physical quantities in the theory, and expectation values of these observables,

$$\langle \mathcal{O}_1 \dots \mathcal{O}_n \rangle, \tag{6}$$

which physicists call amplitudes, or correlation functions. In Chern-Simons theory, the observables ended up associated to knots in a three manifold M, colored by representations R of the gauge group; in Gromov-Witten theory, the observables were related to homology classes γ_i in X.

To find the correlation functions, one starts with a classical limit of the system, and a quantization procedure. In Gromov-Witten theory of X, one would start with a two dimensional topological theory on a genus g Riemann surface Σ, based on maps to X [3]. The description we are giving assumes quantum fluctuations are small. This opens up a possibility for the same physical system to have different descriptions, with different starting classical points, yet which result in the same set of quantum amplitudes. This expresses the fact that physics is intrinsically quantum – only our descriptions of it rely on classical limits; and, the classical limits need not be unique. The map between the two descriptions of the single physical system, is called a *duality*.

3.1. **Mirror symmetry.** Perhaps the best known example of a duality is mirror symmetry. Mirror symmetry relates topological A-model string on a Calabi-Yau X, to topological B-type string theory on the mirror Calabi-Yau Y (The phenomenon was discovered in [10], for a review see [3]). The underlying Calabi-Yau manifolds are different, even topologically, as mirror symmetry reflects the hodge diamond:

$$h^{p,q}(X) = h^{d-p,q}(Y),$$

yet, the A-model on X and the B-model on Y are the same quantum theory. Here d is the complex dimension of X and Y. The amplitudes of the A-type topological string are computed by Gromov-Witten theory. The B-type topological string is reduces to a quantum field theory on Y which quantizes the variations of complex structures; this is related to the fact that Q_B vanishes on constant maps. In particular, the $g=0$ amplitudes can be read off from classical geometry. The most interesting case is $d=3$, otherwise many amplitudes vanish on dimension grounds. This can be seen from the formula for the (virtual) dimension of the moduli space of stable maps in Gromov-Witten theory: $\dim_{\mathbb{C}}\overline{\mathcal{M}}_{g,n}(X,\beta) = -\beta\cdot c_1(TX)+(1-g)(d-3)+n$. In the Calabi-Yau case, per definition, $c_1(TX)$ vanishes in cohomology; for $d=3$, the moduli space has positive dimension for any g, d and n.

The first prediction of mirror symmetry is that the genus zero amplitudes on X and Y agree. On X, computing this leads to quantum intersection numbers: choosing $\gamma_i, \gamma_j, \gamma_k$ to be three divisors in X, $\gamma^{\vee}_{i,j,k} \in H^2(X)$, one computes

$$\langle \gamma_i, \gamma_j, \gamma_j \rangle_{0,Q} = \sum_{\beta \in H_2(X)} \langle \gamma_i, \gamma_j, \gamma_k \rangle_{0,\beta}\, Q^{\beta}.$$

The $\beta = 0$ term in the sum is the classical intersection number of the three divisor classes, and subsequent terms involve intersection theory on moduli space of stable maps, as we described above. In the mirror B-model on Y, the entire sum is captured [11] by *classical geometry of* Y:

$$\langle\gamma_i,\gamma_j,\gamma_j\rangle_{0,Q} = \int_Y \Omega\wedge \frac{\partial}{\partial t_i}\frac{\partial}{\partial t_j}\frac{\partial}{\partial t_k}\,\Omega.$$

This leads to a striking simplification. Here, $\Omega \in H^{(3,0)}(Y)$ is the unique holomorphic volume form on Y, whose existence is guaranteed by the Calabi-Yau condition. The parameters t_i are suitably chosen moduli of complex structures on Y.

The higher genus amplitudes in the B-model quantize the variations of complex structure on Y. In complex dimension 3, the theory one gets is "Kodaira-Spencer theory of gravity", formulated in [9]. The study of B-model in other dimensions was initiated in [12].

3.2. **Large N duality.** A duality, discovered by Gopakumar and Vafa [14], relates $G = U(N)$ Chern-Simons theory at level k, on

$$M = S^3,$$

with A-model topological string on

$$X_{\mathbb{P}^1} = \mathcal{O}(-1)\oplus\mathcal{O}(-1)\rightarrow\mathbb{P}^1.$$

To complete the statement of the duality conjecture, we need to explain the map of parameters, and the correspondence of observables. In defining Chern-Simons theory on the S^3, we get to chose two parameters, the integers N and k. The Gromov-Witten theory on $X_{\mathbb{P}^1}$ depends on the size of the $\mathbb{P}^1$:

$$t = \int_{\mathbb{P}^1}\omega$$

and λ, the genus counting parameter. The latter enters if, instead of fixing the genus of the Riemann surface g, as we did previously, we want to form a generating function, by summing over g. The duality maps the parameters of Chern-Simons theory to parameters of Gromov-Witten as follows

$$t = \frac{2\pi N}{k+N}, \qquad \lambda = \frac{2\pi}{k+N}.$$

The first prediction of the duality is the equivalence of partition functions before we introduce knots in S^3:

$$Z_{CS} = Z_{GW}. \tag{7}$$

The Chern-Simons partition function on the S^3 can be computed as the matrix element of the S matrix (acting on $SU(N)_k$ WZW model on T^2)

$$Z_{CS}(S^3) = \langle 0|S|0\rangle = S_{00}.$$

This is an example of obtaining a three manifold, in this case $M' = S^3$, from $M = S^2\times S^1$ by surgery. We start by excising a neighborhood of an unknot in $S^2\times S^1$ running around the S^1 and at a point on S^2, which splits M into two solid tori. To recover $S^2\times S^1$ we simply glue the the solid tori back together, with trivial identification $U = 1$; to obtain an S^3 instead, we gluing them with an S transformation of the T^2 boundary.

Gromov-Witten partition function Z_{GW} is defined as the generating function of all maps to $X_{\mathbb{P}^1}$ with no insertions:

$$Z_{GW}(X_{\mathbb{P}^1}) = \exp(\sum_{g=0} \langle 1 \rangle_{g,Q} \lambda^{2g-2}) = \exp(\sum_{g=0,\beta \in H_2(W)} \langle 1 \rangle_{\beta,g} \, Q^{\beta} \lambda^{2g-2})$$

where

$$\langle 1 \rangle_{\beta,g} = \int_{[\overline{\mathcal{M}}_{g,0}(X,\beta)]} 1.$$

$Q^{\beta} = \exp(-\beta t)$. In this case the degree of the curve is captured by a single number, since $X_{\mathbb{P}^1}$ has a single non-trivial 2-cycle class corresponding to the $\mathbb{P}^1$ itself. The Gromov-Witten partition function of $X_{\mathbb{P}}^1$ was computed by Faber and Pandharipande in [15], by computing $\langle 1 \rangle_{g,Q}$ for every g. The Chern-Simons partition function is known, since the S matrix is known explicitly. Gopakumar and Vafa [14] showed that $Z_{CS}(S^3)$ equals $Z_{GW}(X_{\mathbb{P}^1})$, by explicit computation. It is striking that the one sums up infinitely many Gromov-Witten invariants in a single matrix element S_{00} in Chern-Simons theory.

The observables of Chern-Simons theory correspond to line operators associated to knots K colored by irreducible representations R of G. Introducing a knots on S^3 corresponds on $X_{\mathbb{P}^1}$ to allowing maps to have boundaries on a Lagrangian submanifold L_K in $X_{\mathbb{P}^1}$, where L_K gets associated to a knot K in a precise way [16]. If we have several knots on S^3, one will introduce a corresponding Lagrangian for each knot. To explain how these Lagrangians are constructed [17], we must first explain the origin of the duality.

3.2.1. *Chern-Simons Theory as a String Theory.* $SU(N)$ Chern-Simons theory on a three manifold M turns out to compute *open* topological A-model amplitudes on

$$X_M = T^*M,$$

the total space of the cotangent bundle on M. One takes the A-model topological string on X_M, but considers maps with boundaries on M:

$$\phi : \Sigma \to X_M, \qquad \partial\Sigma \to M$$

Allowing boundaries corresponds to considering *open topological A-model*. More precisely, we formally need to take N copies of M in X_M, and keep track of which copy of M a given component of the boundary of Σ falls onto. As in the closed case, only the holomorphic maps end up contributing to amplitudes. In fact, as there are no finite holomorphic curves of any kind in $X_M = T^*M$, only degenerate maps contribute – those where the image curves degenerate to graphs on M. Witten showed that the graph expansion that results is the *Feynman graph expansion* of $SU(N)$ Chern-Simons theory. This means that Chern-Simons theory on M computes open topological string amplitudes in this background, in the same way Gromov-Witten theory on X computes closed A-model topological string amplitudes on X. A mathematical consequence of this is that $G = SU(N)$ Chern-Simons partition function on M must have the following expansion:

$$Z_{CS}(M) = \exp(\sum_{g,h=0} F_{g,h}^{CS} N^h \lambda^{2g-2+h}), \tag{8}$$

where $F_{g,h}^{CS}$ are numbers independent of N, k, which capture contributions of maps from surfaces Σ that have genus g and h boundary components. For every boundary we have a choice of which copy of M it falls on, hence the power N^h, and λ keeps track of the Euler characteristic of such as surface which equals $2-2g-h$. The numbers $F_{g,h}$ (the perturbative Chern-Simons invariants) play a role in knot theory [18, 19] – they are related to Vasilliev invariants and to the Kontsevich integral.

Observables in Chern-Simons theory on M are associated to knots. Introducing a knot K in $U(N)$ Chern-Simons theory on M corresponds to, in topological A model on X_M, to introducing a Lagrangian L_K which is a total space of the conormal bundle to the knot K in T^*M [16]. For every point P on the knot K in M, one takes the tangent vector to the knot, and defines a rank two sub bundle of the cotangent bundle, by taking all cotangent vectors that vanish on it. The conormal condition implies that L_K is Lagrangian; this in turn guarantees that the adding boundaries preserves topological invariance of the A-model. Instead of fixing the representation R coloring the knot, it is better to sum over representations, and consider a formal combination of observables

$$\mathcal{O}_K(U) = \sum_R \mathcal{O}_K(R)\ \mathrm{Tr}_R U \tag{9}$$

where U is an arbitrary unitary matrix of rank m, and the sum runs over arbitrary irreducible representations of $U(N)$. The choice of rank m is the number of copies of L_K we take (similarly to the way we took N copies of M to get $SU(N)$ Chern-Simons theory). This observable probes representations R whose Young diagram has no more than m rows, since otherwise $\mathrm{Tr}_R U$ vanishes. Computing the Chern-Simons partition function in presence of knot K with this observable inserted corresponds to studying A-model on X_{S^3} where one allows boundaries on L_K. The eigenvalues $(u_1, \ldots, u_m)$ of U keep track of which of the m copies of L_K the boundary component lands on: a single boundary on the i-th copy of L_K gets weighted by u_i. The resulting partition function is a symmetric polynomial of the u's (using relation between S_m symmetric polynomials and characters of $U(m)$ in various representations) reflecting the S_m permutation symmetry of m copies of L_K. The open topological A-model is expected to have the same relation to *open Gromov-Witten* theory, as the closed A-model has to closed Gromov-Witten theory – where "closed" refers to absence of boundaries of the domain curves Σ. Unlike the closed Gromov-Witten theory, the foundations of the open Gromov-Witten theory are not entirely in place yet, although progress is being made [20].

3.2.2. *Large N Duality is a Geometric Transition.* Gopakumar and Vafa conjectured that large N duality has a geometric interpretation, as a transition that shrinks the S^3 and grows the $\mathbb{P}^1$:

$$X_{S^3} \quad \to \quad X_* \quad \to \quad X_{\mathbb{P}^1}.$$

In the geometric transition, the S^3 disappears and with it the boundaries of maps. If the conjecture is true, it leads to a extraordinary insight: the transition changes topology of the manifolds classically, becomes a change of description – the theories on X_{S^3} (in presence of boundaries on N copies of the S^3) and on $X_{\mathbb{P}^1}$ are the same. The passage from one description to the other is perfectly smooth, like a change of charts on a single manifold.

When, N becomes large, it is natural to sum over h in (8): while λ still keeps track of the Euler characteristic of the underlying Riemann surface, the explicit N dependence disappears. This reflects the fact that both the boundaries and the S^3 disapear in the large N dual description. The large N duality implies that

$$\sum_h F^{CS}_{g,h} t^h = F^{GW}_g(t), \qquad t = N\lambda \tag{10}$$

which is what Gopakumar and Vafa proved in [14] by showing $Z_{CS} = Z_{GW}$. We defined $F^{GW}_g(t)$ by

$$Z_{GW}(X_{\mathbb{P}^1}) = \exp(\sum_g F^{GW}_g \lambda^{2g-2}).$$

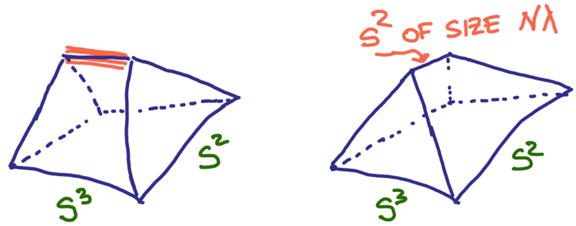


FIGURE 4. Large N duality is a geometric transition that shrinks the S^3 and grows an S^2 of size $t = N\lambda$.

The quantum knot invariants of K are computed by studying open topological A model on X_{S^3} where one allows boundaries on the Lagrangian L_K. The geometric interpretation of the large N duality as a transition between X_{S^3} and $X_{\mathbb{P}^1}$ helps us identify what this corresponds on the dual side. The asymptotic geometry of X_{S^3} and $X_{\mathbb{P}^1}$ are the same – they both approach cones over $S^3 \times S^2$ – they are just filled in differently in the interior. If we first lift L_K off the S^3, it can go through the geometric transition smoothly, to become a Lagrangian on $X_{\mathbb{P}^1}$, which we will denote L_K again. In particular, the topology of the Lagrangian is the same, $\mathbb{R}^2 \times S^1$, both before and after the transition. The construction was made precise in [17]. This leads to a generalization of the basic relation (10) as follows: one considers

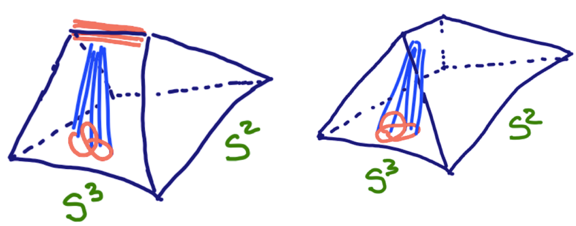


FIGURE 5. The Lagrangian L_K conormal to a knot K in S^3 gets pushed through the transition.

the partition function of Chern-Simons theory, with observable (9) inserted:

$$Z_{CS}(S^3, K, U) = \langle \mathcal{O}_K(U) \rangle \tag{11}$$

This has expansion

(12)
$$Z_{CS}(S^3,K,U)=\exp(\sum_p \sum_{g,h,k_1,\dots k_p=0} F^{CS}_{g,h,k_1,\dots k_p} N^h \lambda^{2g-2+h+p}\,\mathrm{Tr}U^{k_1}\dots\mathrm{Tr}U^{k_p}).$$

The large N duality conjecture states that one can sum over h to get

(13)
$$Z_{GW}(X_{\mathbb{P}^1},L_K,U)=\exp(\sum_p \sum_{g,k_1,\dots k_p=0} F^{GW}_{g,k_1,\dots k_p}(t)\lambda^{2g-2+p}\,\mathrm{Tr}U^{k_1}\dots\mathrm{Tr}U^{k_p}),$$

in other words, that

$$\sum_h F^{CS}_{g,h,k_1,\dots k_p} t^h = F^{GW}_{g,k_1,\dots k_p}(t). \tag{14}$$

For simple knots and links – the unknot and the Hopf-link – one is able to formulate a computation of the open topological A-model amplitude on $X^1_{\mathbb{P}}$ directly in open Gromov-Witten theory [21, 22], and verify the conjecture. For more complicated knots, one needs substantial progress in formulating the open Gromov-Witten side to be able to test the predictions.

The large N duality relating $SU(N)$ Chern-Simons theory to a closed string theory is a part of a family of dualities, whose existence was conjectured by 't Hooft in '70's [23]. He showed that $SU(N)$ gauge theories on general grounds always have Feynman graph expansion of the form (8), with coefficients $F_{g,h}$ that depend on the theory, but not on N or λ. As a consequence, whenever N becomes large, it is natural to re-sum the perturbative series. The result has a form of a closed string Feynman graph expansion; the main question is to identify the dual closed string theory. For Chern-Simons theory on S^3, the closed string theory is Gromov-Witten theory on $X_{\mathbb{P}^1}$. Whenever it exists, the closed string description gets better and better the larger N is, hence the name.

3.3. **Gromov-Witten/Donaldson-Thomas Correspondence.** One is used to studying Gromov-Witten theory by fixing the genus of the Riemann surface Σ. However, Chern-Simons theory and large N duality suggest it is far more economical to consider all genera at once: Chern-Simons amplitudes are the simplest when written in terms of $q=\exp(i\lambda)$, rather than λ. More generally, the fact that one can sum up perturbation series in the gauge theory and solve the theory exactly (at least in principle), is the one of the main reasons why large N duality plays such an important role in physics: the gauge theory allows one to circumvent the usual, genus by genus, formulation of closed string theory.

For Gromov-Witten theory on toric Calabi-Yau three-folds, the theory was indeed solved in this way. Using ideas that originated from large N dualities and Chern-Simons theory, [24] conjectured a solution of Gromov-Witten theory on *any* toric Calabi-Yau threefold by cutting up the Calabi-Yau into $\mathbb{C}^3$ pieces, solving the theory on $\mathbb{C}^3$ exactly, and giving a prescription for how to glue the solution on pieces to a solution of the theory on X. The result is the topological vertex formalism for Gromov-Witten theory of toric Calabi-Yau manifolds, which expresses the partition function, for a fixed degree $\beta\in H_2(X)$ in terms of rational functions of q [24]. The topological vertex conjecture was proven by Maulik, Oblomkov, Okounkov and Pandharipande in [22], who also generalized it away from Calabi-Yau manifolds, to arbitrary toric three-folds.

The resulting invariants of toric three folds turn out to be directly captured by a precise mathematical theory, Donaldson-Thomas theory of X. The theory was introduced in [25], and Okounkov, Maulik, Pandharipande and others provided its foundations [26, 27]. The fact that they are also related to Gromov-Witten invariants of X, is the content of *Gromov-Witten/Donaldson-Thomas Correspondence* [26,27]. Mathematically, Donaldson-Thomas theory also deals with counting Riemann surfaces in X – but it does so in a very different way than Gromov-Witten theory. Instead of describing parameterized curves in X in terms of holomorphic maps $\phi : \Sigma \to X$, as one does in Gromov-Witten theory, in DT theory one describes curves by algebraic equations (see [4] for a review). Let X be a projective variety, $X \subset \mathbb{P}^r$ for some r, and let z_i be the homogenous coordinates of $\mathbb{P}^r$. We can describe the curve C in X as the locus of a set of homogenous polynomials

$$f(z) = 0$$

which vanish on C. The set of all such functions form an ideal $I(C)$ inside $\mathbb{C}[z_0, \ldots z_r]$. We fix the class $\beta \in H_2(X)$, and χ, the holomorphic Euler characteristic of C ($\chi = 1 - g$, were g is the arithmetic genus of C); and denote the moduli space of C by $\mathrm{I}(X, \beta, \chi)$. The moduli space is isomorphic to the Hilbert scheme of curves in X. X being a threefold is special in this case too: the resulting simplifications allow one to construct a (virtual) fundamental cycle in $\mathrm{I}(X, \beta, \chi)$, denoted by $[\mathrm{I}(X, \beta, \chi)]$. The analogue of (4) is

$$\langle \gamma_1, \ldots, \gamma_n \rangle_{\beta,\chi} = \int_{[I(X;\beta,\chi)]} c_2(\gamma_1) \cdots c_2(\gamma_n) \tag{15}$$

To construct $c_2(\gamma)$ takes a special sheaf, the universal ideal sheaf J, $J \in \mathrm{I}(X) \times X$ which has the property that $c_2(J)$ is the locus in $\mathrm{I}(X) \times X$ corresponding to the set (ideal I, point of curve determined by I). $c_2(\gamma)$ is the locus of curves meeting γ – this is the coefficient of $\gamma^\vee$ in the decomposition of $c_2(J) \in H^2(\mathrm{I}(X) \times X)$. Let

$$Z_{DT}(\gamma, q)_\beta = \sum_\chi \langle \gamma_1, \ldots, \gamma_n \rangle^{DT}_{\beta,\chi}\, q^\chi$$

The conjecture equates this, up to normalization, with

$$Z_{GW}(\gamma, \lambda)_\beta = \sum_\chi \langle \gamma_1, \ldots, \gamma_n \rangle^{GW}_{\beta,g}\, \lambda^{2g-2}$$

(In this section we allow disconnected curves domain curves, as this is the natural thing to do if we want to glue the theory on X from pieces. This is also why we do not exponentiate the right hand sides.) More precisely,

$$(-\lambda)^{-\mathrm{vdim}} Z'_{GW}(\gamma; \lambda)_\beta = (-q)^{-\mathrm{vdim}/2} Z'_{DT}(\gamma; q)_\beta$$

where $q = \exp(i\lambda)$, and $'$ denotes dividing by contributions of degree zero curves, which we do on both sides. The Donaldson-Thomas partition function has a beautiful statistical mechanics interpretation in terms of counting boxes stacked up in the toric base of X. One sums over a set of box configurations obeying certain natural conditions and weighs the sum with $q^{\#\text{boxes}}$. Remarkably, the box-counting problem has a saddle point as $q \to 1$, and $\lambda \to 0$. In this limit, the cost of adding a box is small and a limiting shape develops, that dominates the partition function Z_{DT} in the limit. Strikingly, the limiting shape encodes the geometry of the Calabi-Yau Y mirror to X [28, 29].

The duality relating Gromov-Witten theory and Donaldson-Thomas theory has a physical interpretation in M-theory, a quantum theory that underlies and unifies all string theories [30]. Despite the simple appearance – relating counting curves in two different ways – the duality that underlies the Gromov-Witten/Donaldson-Thomas correspondence is far from trivial. In particular, Donaldson-Thomas theory leads to many generalizations that go beyond Gromov-Witten theory. In particular, Donaldson-Thomas theory explains the mysterious integrality of Gromov-Witten invariants which was noticed very early on: while one can express Gromov-Witten invariants in terms of a set of integers, this is not manifest from the definition of the theory – Gromov-Witten theory naturally leads to counts of curves which are rational numbers, not integers, since the underlying moduli spaces are not smooth. One expects that relation of Donaldson-Thomas and Gromov-Witten theories is much like the diagram in Fig.7 – there is a large parameter space of DT theory, the tips of which have Gromov-Witten interpretation.

4. Combining dualities and knot theory

Duality is like a change of charts on a manifold; in particular, we can combine dualities, and get even more mileage from them. For example, one can *combine large N duality and mirror symmetry.* It turns out that this can shed fundamentally new light on knot theory, but to explain this we need to back up to explain the origin of mirror symmetry first.

4.1. **Homological Mirror Symmetry and the SYZ Conjectures.**

We have seen that Gromov-Witten theory computes quantum corrections to the classical geometry of a Calabi-Yau X. Mirror symmetry sums up these corrections, in terms of the geometry of the mirror Calabi-Yau Y. One can make this precise, and give a (conjectural) description for how the classical geometry of Y emerges from quantum geometry of X.

There are two mathematical conjectures that capture aspects of mirror symmetry. *Homological mirror symmetry conjecture* of Kontsevich [31] relates categories of allowed boundary conditions of topological A-model on X and topological B model on Y. The former is captured by the Fukaya category $D^{\mathcal{F}}(X)$ of (X, ω) whose objects are Lagrangian submanifolds $L \subset X$ equipped with a unitary flat connection A:

$$\omega|_L = 0, \qquad F = 0,$$

where $F = dA$ is the curvature of the flat connection A. Holomorphic maps $\phi : \sigma \to X$ are allowed to have boundaries on $L \in X$; the connection on L couples to the boundaries. We have seen examples of this, when $X = T^*M$, where we took $L = M$ or L_K, the Lagrangian associated to the knot. The A connection on M is the Chern-Simons connection. The morphisms in the category are associated to strings with endpoints on pairs of Lagrangians. Kontsevich conjectured that on the mirror Y there is an equivalent category, the bounded derived category of coherent sheaves, $D^b(Y)$. Homological mirror symmetry conjecture was recently proven for a famous example of the quintic Calabi-Yau manifold X and its mirror [32].

Among the objects in $D^b(Y)$, a privileged role is played by the structure sheaf $\mathcal{O}_p$, for p a point in Y. The moduli space of $\mathcal{O}_p$ is Y itself. Mirror symmetry implies that there must be an object in the Fukaya category of X with the same moduli space. Strominger, Yau and Zaslow [33] showed that this fact alone implies that the mirror pair of manifolds (X, Y) must both be T^3 fibrations over a common base B, with fibers that are (special) Lagrangian tori. Let X be a T^3 fibration,

$$T^3 \to X \to B$$

over a base B, and L_p be a T^3 fiber of X above a point in $p \in B$. The moduli space of L_p is the base B itself. The full moduli space is a fibration over this, by moduli of a flat $U(1)$ bundle on T^3. The moduli of a $U(1)$ bundle on T^3 is the dual torus $\hat{T}^3$. More precisely, the resulting moduli can get corrected by "disk instantons" – maps from the disk to X with boundaries on L, and taking this into account results in the mirror manifold:

$$\hat{T}^3 \to Y \to B.$$

This is the *SYZ mirror symmetry conjecture*. This gives a simple geometric picture of mirror symmetry, explicitly constructing the mirror Y from the quantum moduli space of objects on X. The duality that relates string theory on a circle S^1 of radius R to a string theory on a dual circle $\hat{S}^1$ of radius $1/R$ (or a product of circles), is a very basic example of a duality in string theory, called T-duality. Here, we see that mirror symmetry is simply T-duality, applied fiber-wise, over each point in B. For a review of SYZ conjecture, see [34].

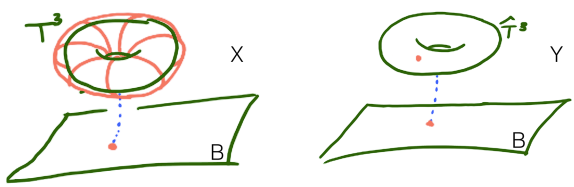


FIGURE 6. SYZ Mirror Symmetry

One can extend the SYZ conjecture away from compact Calabi-Yau manifolds. When X is a toric Calabi-Yau manifold, it is non-compact, and then the T^3 fibration is replaced by an $T^2 \times \mathbb{R}$ fibration over $B = \mathbb{R}^3$. The geometry of the toric Calabi-Yau is, from this perspective, encoded in the geometry of a trivalent graph Γ in B over which the T^2 fiber degenerates to S^1. The role of $L_p = T^3$ in the compact case is replaced by $L_p = S^1 \times \mathbb{R}^2$: the generic $T^2 \times \mathbb{R}$ fiber degenerates, over Γ to a union of two copies of L_p, and we take one of them. The classical moduli space of L_p is the graph Γ; the moduli of the flat connection on L_p is a circle fibered over this, and one still has to take into account disk instanton corrections. The quantum moduli space is a Riemann surface

$$H_\Gamma(x, p) = 0, \qquad x, p \in \mathbb{C}^* \tag{16}$$

The mirror Calabi-Yau is a hypersurface

$$Y: \qquad uv - H_\Gamma(x, p) = 0, \qquad u, v \in \mathbb{C} \tag{17}$$

Mirror to L_p is no longer a structure sheaf on Y, but instead a sheaf supported on a curve, corresponding to choosing a point on the mirror Riemannn surface (16) and picking either $u = 0$ or $v = 0$, depending on which component of the reducible $T^2 \times \mathbb{R}$ fiber we took.

Now, let us describe what this has to do with knot theory.

4.2. **Large N Duality, SYZ Mirror Symmetry and Knot Theory.** Large N duality relates $SU(N)$ Chern-Simons theory on S^3 to Gromov-Witten theory (or A-model topological string) on $X_{\mathbb{P}^1}$; this is a non-compact, toric Calabi-Yau manifold. We can obtain its mirror by application of SYZ mirror symmetry, by finding the quantum moduli space of a Lagrangian in $X_{\mathbb{P}^1}$ of topology of $\mathbb{R}^2 \times S^1$.

In fact, we get one such Lagrangian for *every knot* K in S^3, and with it a distinct mirror Y_K [35]. To construct a Lagrangian L_K in $X_{\mathbb{P}^1}$ of topology of $\mathbb{R}^2 \times S^1$ corresponding to a knot K in S^3, one starts with a Lagrangian in $X_{S^3} = T^*S^3$, which is a total space of the conormal bundle to the knot K in the S^3 base, lifts it off the zero section (so that it does not intersect the singular locus when the S^3 shrinks), and then pushes it through the transition that relates X_{S^3} and $X_{\mathbb{P}^1}$. The mirror depends only on the homotopy type of the knot:

$$Y_K : \qquad uv - H_K(x,p) = 0. \tag{18}$$

The quantum moduli space of L_K in $X_{\mathbb{P}^1}$ is the Riemann surface

$$H_K(x,p) = 0. \tag{19}$$

The pair x and $p = p_K(x)$ that lie on (19) are determined by summing holomorphic disks with boundaries on L_K. Large N duality (14) in turn relates this to a limit of corresponding Chern-Simons amplitude:

$$\log p_K(x) = x\frac{d}{dx}\lim_{\lambda\to 0} \lambda \langle \mathcal{O}_K(x)\rangle$$

where one takes takes $U = x$ to be a rank one matrix, and $Q = \exp(-t)$. For example, taking the knot K to be the unknot, one gets the "conventional" mirror of $X_{\mathbb{P}^1}$, where

$$H_{\bigcirc}(x,p) = 1 - x - p + Qxp.$$

But, taking K to be a trefoil knot instead, as an example, we get a different answer:

$$H_K(x,p) = 1 - Qp + (p^3 - p^4 + 2p^5 - Qp^6 + Q^2p^7)x - (p^9 - p^{10})x^2.$$

Thus, the combination of two string dualities, large N duality and mirror symmetry, gives rise to a new knot invariant, the mirror Calabi-Yau manifold Y_K. Chern-Simons theory produces an infinite list of knot invariants, differing by the representations coloring the knot. To tell knots apart, it is necessary, though maybe not sufficient, to compare the entries of this list. String duality suggests that one can replace the entire list with a single invariant, the mirror Calabi-Yau manifold Y_K, plus presumably a finite set of data needed to define the quantization in this setting. Once the quantization procedure is defined, topological B-model string is a functor, that associates to Y_K quantum invariants. Moreover, unlike knots, Calabi-Yau manifolds are easy to tell apart, simply by comparing the polynomials $H_K(x,p)$. Thus, instead of quantum physics playing the central role in constructing good knot invariants, classical geometry of Y_K becomes the key.

The Riemann surface $H_K(x,p) = 0$ turns out to have an alternative mathematical formulation, as the augmentation variety of the knot [36]. This is one of the knot invariants that arise from knot contact homology. Knot contact homology is the open version, developed by Lenhard Ng [37], of the symplectic field theory approach to counting holomorphic curves, pioneered by Eliashberg and Givental. This provides a relation between two distinct approaches to counting holomorphic curves, one coming from Gromov-Witten theory, and the other from symplectic field theory.

5. M-theory and Homological knot invariants

There is a mysterious aspect of Chern-Simons knot invariants. From the definition of the Jones polynomial $J_K(q)$, one can see that it is always a Laurent polynomial in $q^{1/2}$ with *integer* coefficients. Coefficients of the knot polynomials are always integers, as if they are counting something. What are they counting? Since $q = e^{i\lambda}$, where λ is either the Chern-Simons or the topological string coupling constant, the answer to this question cannot come from Chern-Simons theory or topological string.

Khovanov made this structure manifest in a remarkable way. He constructed a bi-graded homology theory, in such a way that the Jones polynomial arises as the Euler characteristic

$$J_K(q) = \sum_{i,j} (-1)^j \, q^{i/2} \dim \mathcal{H}_{i,j}(K),$$

counting dimensions of knot homology groups,

$$\mathcal{H}_{i,j}(K),$$

with signs. The Poincare polynomial of knot homology

$$P_K(q,t) = \sum_{i,j} t^j q^{i/2} \dim \mathcal{H}_{i,j}(K)$$

has strictly more information about the knot, it is a better knot invariant. One expects that this should have generalizations to all Chern-Simons (Witten-Reshetikhin-Turaev) knot and three manifold invariants, however knot homology theories are extremely complicated. A unified approach to categorification of quantum group invariants was very recently put forward in [39, 40]. As far as we are aware, a fully combinatorial construction of knot homologies is available only for the Jones polynomial itself.

Knot homologies have a physical interpretation within M-theory [41, 42], due to Gukov, Vafa and Schwarz, and later Witten. Knot homologies are Hilbert spaces of states which preserve some supersymmetry in M-theory realization of Chern-Simons theory. To obtain Chern-Simons theory from M-theory, one uses a similar geometry as in topological string. Witten was able to reduce the M-theory construction to computing cohomologies of spaces of solutions to a certain equation, the Kapustin-Witten equation [42, 43], with boundary conditions depending on the knot type, but math and physics are still comparably complex. It was shown in [44] that the approach of [42] leads to the Jones polynomial, once one computes the Euler characteristic. However, this is yet to lead to an explicit construction of knot homologies and $P_K(q,t)$, even in examples.

Physics does provide a powerful insight, if one restricts to three-manifolds and knots respecting a certain circle symmetry. In the presence of the extra symmetry, one can formulate, using M-theory, a three dimensional topological theory, the *refined Chern-Simons theory* [45]. The partition function of the refined Chern-Simons theory (conjecturally) computes a two-variable polynomial $I_K(q,t)$, a close cousin of Poincare polynomial of homology theory categorifying Chern-Simons theory. For three manifolds and knots admitting a (semi-free) circle action, knot homologies corresponding to arbitrary ADE gauge groups and their representations, should admit an additional grade:

$$\mathcal{H}_{i,j}(K) = \oplus_k \mathcal{H}_{i,j,k}(K).$$

This leads to an index, more refined than the Euler characteristic:

$$I_K(q,t) = \sum_{i,j,k} (-1)^k q^{i/2} t^{j+k} \dim \mathcal{H}_{i,j,k}(K),$$

akin to the Hirzebruch χ_y genus. Setting $t = -1$, both $P_K(q,t)$ and $I_K(q,t)$ reduce to Chern-Simons invariants.

The refined Chern-Simons theory, which computes I_K, is solvable explicitly. As in Witten's solution of the "ordinary" Chern-Simons theory – by cutting the three-manifold into pieces, solving the theory on pieces and gluing – one reduces the problem of computing the knot and three manifold invariants to matrix multiplication. In fact, since refined Chern-Simons theory exists for a restricted set of three-manifolds and knots (those admitting a circle symmetry), a smaller set of ingredients enter – all one needs are the S and the T matrix providing a representation of $SL(2,Z)$ on the Hilbert space $\mathcal{H}_{T^2}$. The S and T matrices now depend on both q and t (they are given in Macdonald polynomials of the corresponding ADE group, evaluated at a special point, generalizing the Schur polynomials in Chern-Simons case.) This is immeasurably simpler than constructions of homologies themselves. Even better, for simple representations of $SU(N)$, at large N (corresponding to categorification of the HOMFLY polynomial), the index I_K and the Poincare polynomial P_K of knot homology theory agree. This gives strong evidence that refined Chern-Simons theory indeed computes a new genus on knot homologies, and also evidence that M-theory is indeed behind knot homologies.

It is striking that, even though the refined Chern-Simons theory has been formulated only recently, many connections have already been made. It is known that refined Chern-Simons invariants are related to q-deformation of conformal blocks of W-algebras [46]; they have deep connections to the K-theory of the Hilbert scheme of points on $\mathbb{C}^2$ [47–49]. The knot invariants arising from refined Chern-Simons theory have a direct connection to representation theory of Double Affine Hecke Algebras (DAHA) [49]. There is evidence that the invariants are also related to Donaldson-Thomas invariants of toric three-folds constructed recently in [50].

6. Outlook

Despite the successes of string theory in solving difficult problems in mathematics, this is no doubt just a tip of the iceberg. All string theories are unified in a single theory, M-theory. Genus by genus expansion, on which topological string and superstring theories are based, exists only at the corners of M-theory parameter space. Dualities fill in the rest of the diagram. M-theory has already made an appearance in knot theory context, and in relating Gromov-Witten theory to Donaldson-Thomas theory.

FIGURE 7. M-theory is believed to be the unique quantum theory underlying all string theories. Different descriptions of it, which emerge at the corners of the diagram, are related by dualities.

Mathematical consequences of dualities in M-theory are largely unexplored. Two topological string theories, the A- and the B-model, with their many mathematical uses, capture supersymmetric M-theory partition functions in a very specific background [51,52]. The plethora of mathematical predictions extracted from topological string and its dualities, such as mirror symmetry and large N duality we described provide just a glimpse of the mathematical content of M-theory. Supersymmetric partition functions of M-theory are generalizations of topological string, yet only their exploration has only just begun, see [50].

To be sure, dualities do not require string theory. There are examples of dualities in quantum gauge theories which can be stated without invoking string theory. Even so, string theory often plays the crucial role in discovering the dualities, and in studying them. *Symplectic duality* [56], which plays an important role in knot theory and other areas of mathematics, is a duality of quantum gauge theories in three dimensions. Even though today one can phrase it purely in gauge theory language, the duality was discovered using string theory, in [53, 54], and string theory helps one understand the it better [55]. *Seiberg-Witten (SW) theory*, the celebrated 4d QFT with an important role for 4-manifold invariants [57, 58], turns out to have many dual descriptions [59]. In fact most of the theories in this class turn out not have a conventional description, but need M-theory for their definition. To define them, one considers the 6-dimensional the "theory X" that arises as a part of M-theory, compactified on a Riemann surface C. Only in certain corners of the moduli of C the usual gauge theory description emerges. This observation leads to a precise mathematical prediction: the partition functions of this class of Seiberg-Witten theories are the conformal blocks on C of a class of 2d conformal field theories with W-algebra symmetry [60]. This unifies problems in QFT, geometry and representation theory. Some aspects of this correspondence were recently proven by [61]. The *S-duality* of 4d $\mathcal{N} = 4$ Yang-Mills theory, related to electric-magnetic duality, is believed to be the duality underlying the geometric Langlands program [62–65]. The Langlands program has, for the last 50 years, been one of the key unifying themes in mathematics [66]. Once again, even though one can phrase S duality in terms of gauge theory alone, much of our understanding of it comes from string theory: the 4d $\mathcal{N} = 4$ Yang-Mills theory arises by compactifying the 6d theory X on a torus, and S-duality simply comes from $SL(2, Z)$ symmetry of the torus!

The interacting between the two fields has only really begun in ernest. It is fairly certain that dualities in string theory and quantum field theory hold potential for many new breakthroughs in mathematics, by extracting their mathematical predictions, and proving them. It should also lead to a deeper and sharper understanding of quantum physics. There is a good chance that eventually, our view of mathematics, and quantum physics will have changed profoundly.

7. Acknowledgments

I am grateful to Robbert Dijkgraaf, Tobias Ekholm, Nathan Haouzi, Albrecht Klemm, Markos Marino, Lenhard Ng, Shamil Shakirov, Andrei Okounkov, Cumrun Vafa for years of collaborations that taught me the ideas presented, and the numerous mathematicians and physicists who have worked together to unearth the beautiful structures in our subject.

References

[1] E. Witten, "Quantum Field Theory and the Jones Polynomial," Commun. Math. Phys. **121**, 351 (1989).

[2] N. Reshetikhin and V. G. Turaev, "Invariants of three manifolds via link polynomials and quantum groups," Invent. Math. **103**, 547 (1991).

[3] K. Hori, S. Katz, A. Klemm, R. Pandharipande, R. Thomas, C. Vafa, R. Vakil and E. Zaslow, "Mirror symmetry," (Clay mathematics monographs. 1)

[4] A. Okounkov, "Random surfaces enumerating algebraic curves", arXiv:math-ph/0412008

[5] R. Thomas, "The geometry of mirror symmetry", Encylopedia of mathematical physics, Editor(s): Francoise, Naber, Tsou, Oxford, Elsevier, Pages:-, ISBN:9780125126601;

[6] E. Witten, "Topological Sigma Models," Commun. Math. Phys. **118**, 411 (1988).

[7] E. Witten, "On the Structure of the Topological Phase of Two-dimensional Gravity," Nucl. Phys. B **340**, 281 (1990).

[8] E. Witten, "Mirror manifolds and topological field theory," In *Yau, S.T. (ed.): Mirror symmetry I* 121-160 [hep-th/9112056].

[9] M. Bershadsky, S. Cecotti, H. Ooguri and C. Vafa, "Kodaira-Spencer theory of gravity and exact results for quantum string amplitudes," Commun. Math. Phys. **165**, 311 (1994) [hep-th/9309140].

[10] W. Lerche, C. Vafa and N. P. Warner, "Chiral Rings in N=2 Superconformal Theories," Nucl. Phys. B **324**, 427 (1989).

[11] P. Candelas, X. C. De La Ossa, P. S. Green and L. Parkes, "A Pair of Calabi-Yau manifolds as an exactly soluble superconformal theory," Nucl. Phys. B **359** (1991) 21.

[12] K. J. Costello and S. Li, "Quantum BCOV theory on Calabi-Yau manifolds and the higher genus B-model," arXiv:1201.4501 [math.QA].

[13] E. Witten, "Chern-Simons gauge theory as a string theory," Prog. Math. **133**, 637 (1995) [hep-th/9207094].

[14] R. Gopakumar and C. Vafa, "On the gauge theory / geometry correspondence," Adv. Theor. Math. Phys. **3** (1999) 1415 [hep-th/9811131].

[15] Faber, C., Pandharipande, R. 1998, "Hodge integrals and Gromov-Witten theory", arXiv:math/9810173

[16] H. Ooguri and C. Vafa, "Knot invariants and topological strings," Nucl. Phys. B **577** (2000) 419 [hep-th/9912123].

[17] C. H. Taubes, "Lagrangians for the Gopakumar-Vafa conjecture," Adv. Theor. Math. Phys. **5** (2001) 139 [math/0201219 [math-dg]].

[18] D. Bar-Natan, "On the Vassiliev knot invariants," Topology 34 (1995) 423.

[19] M. Marino, "Chern-Simons theory, matrix integrals, and perturbative three manifold invariants," Commun. Math. Phys. **253**, 25 (2004) [hep-th/0207096].

[20] K. Fukaya, Talk at String-Math 2013, $http://media.scgp.stonybrook.edu/presentations/201306182fukaya.pdf$

[21] S. H. Katz and C. C. M. Liu, “Enumerative geometry of stable maps with Lagrangian boundary conditions and multiple covers of the disc,” Adv. Theor. Math. Phys. **5** (2002) 1 [math/0103074 [math-ag]].

[22] D. Maulik, A. Oblomkov, A. Okounkov and R. Pandharipande, “Gromov-Witten/Donaldson-Thomas correspondence for toric 3-folds,” arXiv:0809.3976 [math.AG].

[23] G. ’t Hooft, ”A planar diagram theory for strong interactions,” Nucl. Phys. B 72, 461 (1974).

[24] M. Aganagic, A. Klemm, M. Marino and C. Vafa, “The Topological vertex,” Commun. Math. Phys. **254** (2005) 425 [hep-th/0305132].

[25] S. Donaldson and R. Thomas, Gauge theory in higher dimensions, in The geometric universe: science, geometry, and the work of Roger Penrose, S. Huggett et. al eds., Oxford Univ. Press, 1998.

[26] D. Maulik, N. Nekrasov, A. Okounkov, and R. Pandharipande, Gromov- Witten theory and Donaldson-Thomas theory. I, Compos. Math. 142 (2006), 1263?1285.

[27] D. Maulik, N. Nekrasov, A. Okounkov, and R. Pandharipande. Gromov- Witten theory and Donaldson-Thomas theory. II, Compos. Math. 142 (2006), 1286?1304.

[28] A. Okounkov, N. Reshetikhin and C. Vafa, “Quantum Calabi-Yau and classical crystals,” Progr. Math. **244**, 597 (2006) [hep-th/0309208].

[29] A. Iqbal, N. Nekrasov, A. Okounkov and C. Vafa, “Quantum foam and topological strings,” JHEP **0804** (2008) 011 [hep-th/0312022];

[30] R. Dijkgraaf, C. Vafa and E. Verlinde, “M-theory and a topological string duality,” hep-th/0602087.

[31] Kontsevich, M, ”Homological Algebra of Mirror Symmetry”, arXiv:alg-geom/9411018

[32] N. Sheridan, arXiv:1111.0632, ”Homological Mirror Symmetry for Calabi-Yau hypersurfaces in projective space” Y. Nohara, & K. Ueda, ”Homological mirror symmetry for the quintic 3-fold”, Geometry & Topology 16 (2012) 1967–2001, arXiv:1103.4956

[33] A. Strominger, S. T. Yau and E. Zaslow, “Mirror symmetry is T duality,” Nucl. Phys. B **479**, 243 (1996) [hep-th/9606040].

[34] M. Gross, ”Mirror Symmetry and the Strominger-Yau-Zaslow conjecture”, arXiv:1212.4220, Current Developments in Mathematics, 2012, 67 pages

[35] M. Aganagic and C. Vafa, “Large N Duality, Mirror Symmetry, and a Q-deformed A-polynomial for Knots,” arXiv:1204.4709;

[36] M. Aganagic, T. Ekholm, L. Ng and C. Vafa, “Topological Strings, D-Model, and Knot Contact Homology,” arXiv:1304.5778. To appear in Adv. Theor. Math. Phys. Vol. 18.4

[37] L. Ng, “Framed knot contact homology,” Duke Math. J. 141 (2008) no. 2, 365–406, arXiv:math/0407071; L. Ng, “Combinatorial knot contact homology and transverse knots” Adv. Math. 227 (2011) no. 6, 2189–2219, arXiv:1010.0451; T. Ekholm, J. Etnyre, L. Ng, and M. Sullivan, “Knot contact homology,” arXiv:1109.1542.

[38] Khovanov, M., ”A categorification of the Jones polynomial,” Duke Math. J. 101 (2000), no. 3, 359–426, arXiv:math/9908171

[39] Webster, B., “Knot invariants and higher representation theory” arXiv:1309.3796

[40] Webster, B., “Tensor product algebras, Grassmannians and Khovanov homology” arXiv:1312.7357

[41] S. Gukov, A. S. Schwarz and C. Vafa, “Khovanov-Rozansky homology and topological strings,” Lett. Math. Phys. **74**, 53 (2005) [hep-th/0412243].

[42] E. Witten, “Fivebranes and Knots,” arXiv:1101.3216 [hep-th].

[43] E. Witten, “Khovanov Homology And Gauge Theory,” arXiv:1108.3103 [math.GT]. E. Witten, “Two Lectures On The Jones Polynomial And Khovanov Homology,” arXiv:1401.6996 [math.GT].

[44] D. Gaiotto and E. Witten, “Knot Invariants from Four-Dimensional Gauge Theory,” Adv. Theor. Math. Phys. **16**, no. 3, 935 (2012) [arXiv:1106.4789 [hep-th]].

[45] M. Aganagic and S. Shakirov, “Knot Homology from Refined Chern-Simons Theory,” arXiv:1105.5117 [hep-th].

[46] M. Aganagic, N. Haouzi and S. Shakirov, “A_n-Triality,” arXiv:1403.3657 [hep-th].

[47] Oblomkov, A., Rasmussen, J., & Shende, V., ”The Hilbert scheme of a plane curve singularity and the HOMFLY homology of its link”, arXiv:1201.2115

[48] Nakajima, H. 2012, ”Refined Chern-Simons theory and Hilbert schemes of points on the plane” arXiv:1211.5821
[49] Gorsky, E., & Negut, A., ”Refined Knot Invariants and Hilbert Schemes”, arXiv:1304.3328
[50] N. Nekrasov and A. Okounkov, “Membranes and Sheaves,” arXiv:1404.2323 [math.AG].
[51] R. Gopakumar and C. Vafa, “M theory and topological strings. 1.,” hep-th/9809187.
[52] R. Gopakumar and C. Vafa, “M theory and topological strings. 2.,” hep-th/9812127.
[53] A. Hanany and E. Witten, “Type IIB superstrings, BPS monopoles, and three-dimensional gauge dynamics,” Nucl. Phys. B **492**, 152 (1997) [hep-th/9611230].
[54] K. A. Intriligator and N. Seiberg, “Mirror symmetry in three-dimensional gauge theories,” Phys. Lett. B **387**, 513 (1996) [hep-th/9607207].
[55] D. Gaiotto and E. Witten, “Supersymmetric Boundary Conditions in N=4 Super Yang-Mills Theory,” J. Statist. Phys. **135** (2009) 789 [arXiv:0804.2902 [hep-th]]. D. Gaiotto and E. Witten, “S-Duality of Boundary Conditions In N=4 Super Yang-Mills Theory,” Adv. Theor. Math. Phys. **13** (2009) 721 [arXiv:0807.3720 [hep-th]].
[56] Tom Braden, Nicholas Proudfoot, Ben Webster: ”Quantizations of conical symplectic resolutions I: local and global structure”, arXiv:1208.3863; ”Quantizations of conical symplectic resolutions II: category $\mathcal{O}$ and symplectic duality”, arXiv: 1407.0964
[57] N. Seiberg and E. Witten, “Electric - magnetic duality, monopole condensation, and confinement in N=2 supersymmetric Yang-Mills theory,” Nucl. Phys. B **426** (1994) 19 [Erratum-ibid. B **430** (1994) 485] [hep-th/9407087]. N. Seiberg and E. Witten, “Monopoles, duality and chiral symmetry breaking in N=2 supersymmetric QCD,” Nucl. Phys. B **431** (1994) 484 [hep-th/9408099].
[58] E. Witten, “Monopoles and four manifolds,” Math. Res. Lett. **1** (1994) 769 [hep-th/9411102].
[59] D. Gaiotto, “N=2 dualities,” JHEP **1208** (2012) 034 [arXiv:0904.2715 [hep-th]].
[60] L. F. Alday, D. Gaiotto and Y. Tachikawa, “Liouville Correlation Functions from Four-dimensional Gauge Theories,” Lett. Math. Phys. **91** (2010) 167 [arXiv:0906.3219 [hep-th]].
[61] A. Braverman, M. Finkelberg, H. Nakajima, ”Instanton moduli spaces and W-algebras”, arxiv:1406.2381[math:QA]
[62] C. Montonen and D. Olive, Phys. Lett. B72 (1977) 117; P. Goddard, J. Nyuts and D. Olive , Nucl. Phys. B125 (1977) 1.
[63] C. Vafa and E. Witten, “A Strong coupling test of S duality,” Nucl. Phys. B **431** (1994) 3 [hep-th/9408074].
[64] A. Kapustin and E. Witten, “Electric-Magnetic Duality And The Geometric Langlands Program,” Commun. Num. Theor. Phys. **1** (2007) 1 [hep-th/0604151].
[65] R. Donagi, T. Pantev: Langlands duality for Hitchin systems, Inv. Math 189 (2012), 653?735,
[66] E. Frenkel, “Lectures on the Langlands program and conformal field theory,” hep-th/0512172.

FROM RATIONAL BILLIARDS TO DYNAMICS ON MODULI SPACES

ALEX WRIGHT

Abstract. Consider a billiard ball bouncing around in a polygon. This simple system demonstrates remarkable complexity. For example, it is an open problem to prove that there is a periodic billiard trajectory in every polygon. However, if the angles are all rational multiples of π, a great deal is known. This is because such a polygon can be "unfolded" to give a surface with extra structure, and there is an $SL(2,\mathbb{R})$ action on the space of all such surfaces. We will explain the relevance of this action, and state a recent result of Eskin, Mirzakhani, and Mohammadi, which gives that the closure of every $SL(2,\mathbb{R})$ orbit is a manifold. Applications and connections to other areas of mathematics will be mentioned.

1. Rational billiards

Consider a point bouncing around in a polygon. Away from the edges, the point moves at unit speed. At the edges, the point bounces according to the usual rule that angle of incidence equals angle of reflection. If the point hits a vertex, it stops moving. The path of the point is called a billiard trajectory.

The study of billiard trajectories is a basic problem in dynamical systems and arises naturally in physics. For example, consider two points of different masses moving on a interval, making elastic collisions with each other and with the endpoints. This system is modeled by billiard trajectories in a right angled triangle **[MT02]**.

A rational polygon is a polygon all of whose angles are rational multiples of π. Many mathematicians are especially interested in billiards in rational polygons for the following three reasons.

First, without the rationality assumption, few tools are available, and not much is known. For example, it is not even known if every triangle has a periodic billiard trajectory. With the rationality assumption, quite a lot can be proven.

Second, even with the rationality assumption a wide range of interesting behavior is possible, depending on the choice of polygon.

Third, the rationality assumption leads to surprising and beautiful connections to algebraic geometry, Teichmüller theory, ergodic theory on homogenous spaces, and other areas of mathematics.

The assumption of rationality first arose from the following simple thought experiment. What if, instead of letting a billiard trajectory bounce off an edge of a polygon, we instead allowed the trajectory to continue straight, into a reflected copy of the polygon?

This leads us to define the "unfolding" of a polygon P as follows: Let G be the subgroup of $O(2)$ (linear isometries of $\mathbb{R}^2$) generated by the derivatives of reflections in the sides of P. The group G is finite if and only the polygon P is rational (in which case G is a dihedral group). For each $g \in G$, consider the polygon gP. These polygons gP can be translated so that they are all disjoint in the plane. We identify

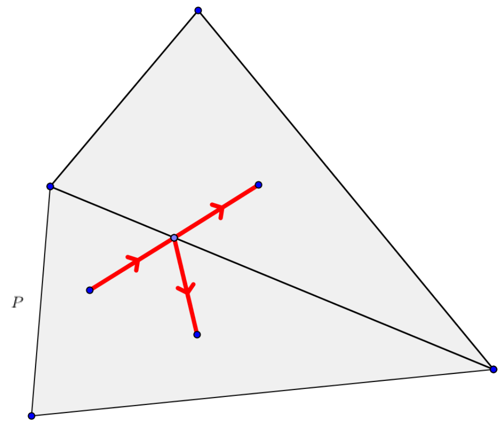


FIGURE 1.1. A billiard trajectory in a polygon P. Instead of allowing the trajectory to bounce off the edge of P, we may allow it to consider straight into a reflected copy of P. A key observation is that the trajectory that continues into the reflected copy of P is in fact the reflection of the trajectory in P that bounces off of the edge.

the edges in pairs in the following way. Suppose r is the derivative of the reflection in one of the edges of hP. Then this edge of hP is identified with the corresponding edge of rhP.

The unfolding construction is most easily understood through examples: see Figures 1.2 and 1.3.

2. Translation surfaces

Unfoldings of rational polygons are special examples of translation surfaces. There are several equivalent definitions of translation surface, the most elementary of which is a finite union of polygons in in the plane with edge identifications, obeying certain rules, up to a certain equivalence relation. The rules are:

(1) The interiors of the polygons must be disjoint, and if two edges overlap then they must coincide and be identified.
(2) Each edge is identified with exactly one other edge, which must be a translation of the first. The identification is via this translation.
(3) When an edge of one polygon is identified with an edge of a different polygon, the polygons must be on "different sides" of the edge. For example, if a pair of vertical edges are identified, one must be on the left of one of the polygons, and the other must be on the right of the other polygon.

Two such families of polygons are considered to be equivalent if they can be related via a string of the following "cut and paste" moves.

(1) A polygon can be translated.
(2) A polygon can be cut in two along a straight line, to give two adjacent polygons.

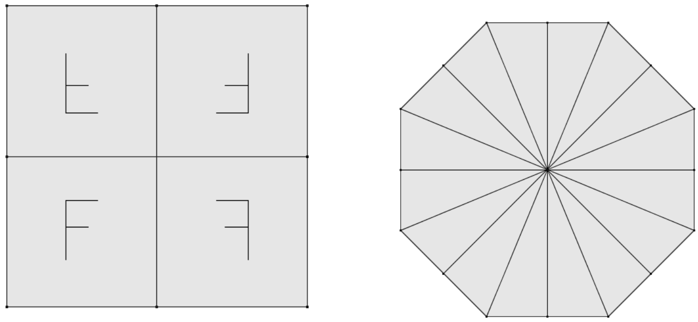


FIGURE 1.2. Left: The unit square unfolds to four squares, with opposite edges identified (a flat torus). (By "opposite edges" in these pictures, we mean pairs of boundary edges that are perpendicularly across from each other, so for example the top left and top right vertical edges on the unfolding of the square are opposite.) When two polygons are drawn with an adjacent edge, by convention this means these two adjacent edges are identified. Here each square has been decorated by the letter F, to illustrate which squares are reflections of other squares. Right: Unfolding the right angled triangle with smallest angle $\pi/8$ gives the regular octagon with opposite sides identified.

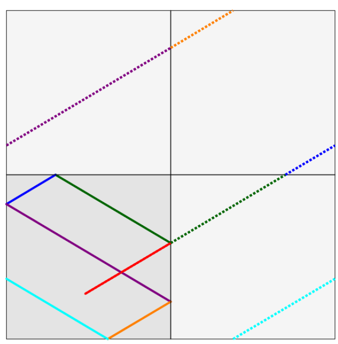


FIGURE 1.3. A billiard trajectory on a rational polygon unfolds to a straight line on the unfolding of the polygon. In this illustration, we have unfolded a billiard trajectory on square (bottom left) to a straight line on a flat torus. The square and its unfolding are superimposed, the billiard trajectory is drawn with a solid line, and the unfolded straight line is drawn with a dotted line.

(3) Two adjacent polygons that share an edge can be glued to form a single polygon.

It is difficult to decide if two collections of polygons as above are equivalent (describe the same translation surface), because each collection of polygons is equivalent to infinitely many others.

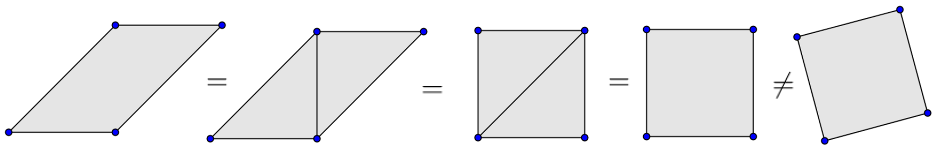


FIGURE 2.1. In all five translation surfaces above, opposite edges are identified. In the leftmost four, each adjacent pair of translation surfaces differs by one of the above three moves, so all four of these pictures give the same translation surface. The rightmost rotated surface is not equal to the other four, since rotation is not one of the three allowed moves.

The requirements above ensure that the union of polygons, modulo edge identifications, gives a closed surface. This surface has flat metric, given by the flat metric on the plane, away from a finite number of singularities. The singularities arise from the corners of the polygons. For example, in the regular 8-gon with opposite sides identified (Figure 1.2, right), the edge identifications imply that the 8 corners of the octagon are in fact all identified, and give a single point on the translation surface. Around this point there is 6π total angle, since at each of the 8 corners of the polygon there is $\frac{3}{4}\pi$ interior angle.

The singularities of the flat metric on a translation surface are always of a very similar conical form, and the total angle around a singularity on a translation surface is always an integral multiple of 2π. Note that, although the flat metric is singular at these points, the underlying topological surface is not singular at any point. (That is, at every single point, including the singularities of the flat metric, the surface is locally homeomorphic to $\mathbb{R}^2$.)

Most translation surfaces do not arise from unfoldings of rational polygons. This is because unfoldings of polygons are exceptionally symmetric, in that they are tiled by isometric copies of the polygon.

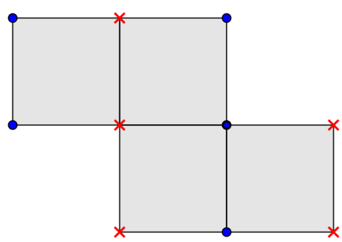


FIGURE 2.2. Consider the translation surface described by the above polygon, with opposite edges identified. This surface has two singularities, each with total angle 4π. One singularity has been labelled with a dot, and the other with an x. An Euler characteristic calculation ($V - E + F = 2 - 2g$) show that it has genus 2. The regular octagon with opposite sides identified also has genus 2, but it has only a single singularity, with total angle 6π.

Translation surfaces satisfy a Gauss-Bonnet type theorem. If a translation surface has s singularities with cone angles

$$(1+k_1)2\pi, (1+k_2)2\pi, \ldots, (1+k_s)2\pi,$$

then the genus g is given by the formula $2g-2 = \sum k_i$. (So, in a formal comparison to the usual Gauss-Bonnet formula, one might say that each extra 2π of angle on a translation surface counts for 1 unit of negative curvature.)

Consider now the question of how a given translation surface can be deformed to give other translation surfaces. The polygons, up to translation, can be recorded by their edge vectors in $\mathbb{C}$ (plus some finite amount of combinatorial data, for example the cyclic order of edges around the polygons). Not all edge vectors need be recorded, since some are determined by the rest. Changing the edge vectors (subject to conditions like identified edges should remain parallel and of the same length) gives a deformation of the translation surface.

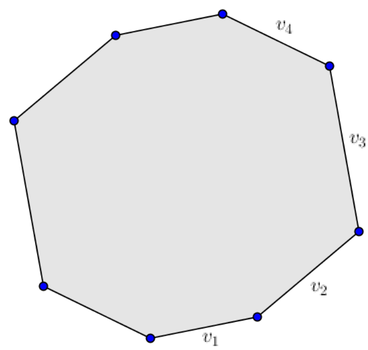


FIGURE 2.3. Any octagon whose opposite edges are parallel can be described by the 4-tuple of its edges vectors $(v_1, v_2, v_3, v_4) \in \mathbb{C}^4$. (Not all choices of v_i give valid octagons.) The coordinates (v_1, v_2, v_3, v_4) are local coordinates for space of deformations of the regular octagon translation surface. These coordinates are not canonical: other equally good coordinates can be obtained by cutting up the octagon and keeping track of different edge vectors.

To formalize this observation, we define moduli spaces of translation surfaces. Given an unordered collection $k_1, \ldots, k_s$ of positive integers whose sum is $2g-2$, the stratum $\mathcal{H}(k_1, \ldots, k_s)$ is defined to be the set of all translation surfaces with s singularities, of cone angles $(1+k_i)2\pi, i = 1, \ldots, s$. The genus of these surfaces must be g by the Gauss-Bonnet formula above. We have

Lemma 2.1. *Each stratum is a complex orbifold of dimension $n = 2g + s - 1$. Each stratum has a finite cover that is a manifold and has an atlas of charts to $\mathbb{C}^n$ with transition functions in $GL(n, \mathbb{Z})$.*

The coordinate charts are called period coordinates. They consist of complex edge vectors of polygons. That strata are orbifolds instead of manifolds is a technical point that should be ignored by non-experts.

Strata are not always connected, but their connected components have been classified by Kontsevich and Zorich [**ZK75**]. There are always at most three connected

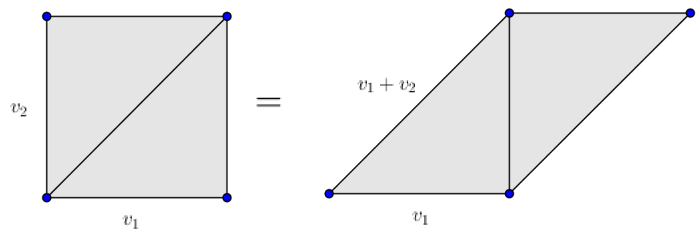


FIGURE 2.4. These two polygons (with opposite sides identified) both describe the same translation surface. Keeping track of the edge vectors in either polygon gives equally good local coordinates for the space of nearby translation surfaces. The two local coordinates thus obtained are related by the linear transformation $(v_1, v_2) \mapsto (v_1, v_1 + v_2)$.

components. The topology and birational geometry of strata is currently not well understood. Kontsevich has conjectured that strata are $K(\pi, 1)$ spaces.

3. $GL(2,\mathbb{R})$ ACTION AND RENORMALIZATION

There is a $GL(2,\mathbb{R})$ action on each stratum, obtained by acting linearly on polygons and keeping the same identification.

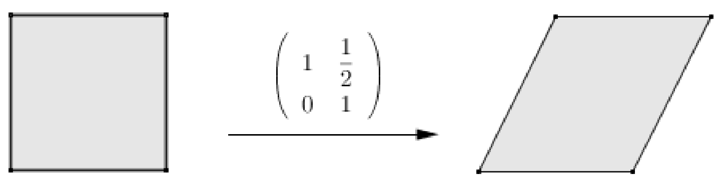


FIGURE 3.1. An example of the $GL(2,\mathbb{R})$ action. In both pictures, opposite edges are identified.

Note that if two edges or polygons differ by translation by a vector v, then their images under the linear map $g \in GL(2,\mathbb{R})$ must differ by translation by gv.

Example 3.1. The stabilizer of the standard flat torus (a unit square with opposite sides identified) is $GL(2,\mathbb{Z})$. For example, Figure 2.1 (near the beginning of the previous section) proves that $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is in the stabilizer. This example illustrates the complexity of the $GL(2,\mathbb{R})$ action: applying a large matrix (say of determinant 1) will yield a collection of very long and thin polygons, but it is hard to know when this collection of polygons is equivalent to a more reasonable one.

Translation surfaces have a well defined area, given by the sum of the areas of the polygons. The action of $SL(2,R)$ of determinant 1 matrices in $GL(2,\mathbb{R})$ preserves the locus of unit area translation surfaces. This locus is not compact, because the polygons can have edges of length going to 0, even while the total area stays constant.

Define

$$g_t = \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix} \in SL(2,\mathbb{R}).$$

Suppose one has a vertical line segment of length L on a translation surface S. For example, if S is the unfolding of a rational polygon, the vertical line might be the unfolding of a billiard trajectory. If one is interested in a line that is not vertical, one can rotate the whole picture (giving a different translation surface) so that it becomes vertical.

This setup is given by a picture: a collection of polygons describing S, with many vertical line segments that match up under the edge identifications and hence give one line segment on the translation surface. Applying g_t to this picture results in a translation surface $g_t(S)$ with a vertical segment of length $e^{-t}L$. We are interested in doing this when L is very large, and $t = \log(L)$ is chosen so the new vertical segment will have length 1. Indeed the point is to do this over and over as L gets longer, giving a family of surfaces $g_t(S)$.

This idea of taking longer and longer trajectories (here a vertical line on the translation surface) and replacing them with a bounded trajectories on new objects is called renormalization, and is a powerful and frequently used tool in the study of dynamical systems. The typical strategy is to transfer some understanding of the sequence of renormalized objects into results on the behavior of the original system. In this case, showing the the geometry of $g_t(S)$ does not degenerate allows good understanding of vertical lines on S.

Theorem 3.2 (Masur's criterion [**Mas92**])**.** *Suppose* $\{g_t(S) : t \geq 0\}$ *does not diverge to infinity in the stratum. Then every infinite vertical line on* S *is equidistributed on* S*.*

"Equidistributed" is a technical term that indicates that the vertical lines becomes dense in S without favoring one part of S over another. Using this, Kerckhoff-Masur-Smillie [**KMS86**] were able to show

Theorem 3.3. *In every translation surface, for almost every slope, every infinite line of this slope is equidistributed.*

There are some surfaces where much more is true. For example, on the unit square with opposite sides identified, any line of rational slope is periodic, and every line of irrational slope is equidistributed. Genus one translation surfaces are quite special, because $GL(2,\mathbb{R})$ acts transitively on the space of genus one translation surfaces. In particular, the $GL(2,\mathbb{R})$ orbit of any genus one translation surface is closed, in a trivial way, since the orbit is the entire moduli space.

Theorem 3.4 (Veech Dichotomy)**.** *If* S *is a translation surface with closed* $GL(2,\mathbb{R})$ *orbit, then for all but countably many slopes, every line with that slope is equidistributed. Moreover every line with slope contained in the countable set is periodic.*

Veech also showed that the regular $2n$-gon with opposite sides identified has closed orbit. However, the property of having a closed orbit is extremely special.

Theorem 3.5 (Masur [**Mas82**], Veech [**Vee82**])**.** *The* $GL(2,\mathbb{R})$ *orbit of almost every translation surface is dense in a connected component of a stratum.*

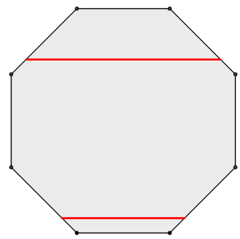


FIGURE 3.2. An example of a periodic line on the regular octagon with opposite sides identified.

For the experts, we remark that in fact Masur and Veech showed the stronger statement that the g_t action on the loci of unit area surfaces in a connected component of a stratum is ergodic, with respect to a Lebesgue class probability measure called the Masur-Veech measure.

This result of Masur and Veech is not satisfactory from the point of view of billiards in rational polygons, since the set of translation surfaces that are unfoldings of polygons is measure 0.

4. Eskin-Mirzakhani-Mohammadi's breakthrough

In light of the idea of renormalization, it is important to understand the $SL(2,\mathbb{R})$ orbit closures of all translation surfaces. It is equivalent to understand the $GL(2,\mathbb{R})$ orbit closures, since the difference is just scaling of the area.

In fact it would also be helpful to know more specific information, such as the g_t orbit closures. However, g_t orbit closures may be fractal objects. While this behavior might at first seem pathological, it is in fact quite typical in dynamical systems. Generally speaking, given a group action it is hugely unrealistic to ask for any understanding of every single orbit, since these may typically be arbitrarily complicated. Thus the following result is quite amazing.

Theorem 4.1 (Eskin-Mirzakhani-Mohammadi **[EM, EMM]**). *The $GL(2,\mathbb{R})$ orbit closure of a translation surface is always a manifold. Moreover, the manifolds that occur are locally defined by linear equations in period coordinates. These linear equations have real coefficients and zero constant term.*

Note that although the local period coordinates are not canonical, if a manifold is cut out by linear equations in one choice of period coordinates, it must also be in any other overlapping choice of period coordinates, because the transition map between these two coordinates is a matrix in $GL(n,\mathbb{Z})$.

Previously orbit closures had been classified in genus 2 by McMullen. (One open problem remains in genus two, which is the classification of $SL(2,\mathbb{Z})$ orbits of square-tiled surfaces in $\mathcal{H}(1,1)$.) The techniques of Eskin-Mirzakhani-Mohammadi's, unlike those of McMullen, are rather abstract, and have surprisingly little to do with translations surfaces. Thus the work of Eskin-Mirzakhani-Mohammadi does not give any information about how many or what sort of submanifolds arise as orbit closures, except for what is given in the theorem statement.

5. Applications of Eskin-Mirzakhani-Mohammadi's Theorem

There are many applications of Theorem 4.1 to translation surfaces, rational billiards, and other related dynamics systems, for example interval exchange transformations. Here we list just a few of the most easily understood applications.

Generalized diagonals in rational polygons. Let P be a rational polygon. A generalized diagonal is a billiard trajectory that begins and ends at a corner of P. If P is a square, an example is the diagonal of P. Let $N_P(L)$ be the number of generalized diagonals in P of length at most L. It is a folklore conjecture that

$$\lim_{L\to\infty} \frac{N_P(L)}{L^2}$$

exists for every P and is non-zero. Previously, Masur had shown that the limsup and liminf are non-zero and finite. Eskin-Mirzakhani-Mohammadi give the best result to date: with some additional Cesaro type averaging, the conjecture is true, and furthermore only countably many real numbers may occur as such a limit.

The illumination problem. Given a polygon P and two points x and y, say that y is illuminated by x if there is a billiard trajectory going from y to x. This terminology is motivated by thinking of P as a polygonal room whose walls are mirrors, and thinking of a candle placed at x. The light rays travel along billiard trajectories. We emphasize that the polygon need not be convex.

Lelièver, Monteil and Weiss have shown that if P is a rational polygon, for every x there are at most finitely many y not illuminated by x [**LMW**].

The Wind Tree Model. This model arose from physics, and is sometimes called the Ehrenfest model. Consider the plane with periodically shaped rectangular barriers ("trees"). Consider a particle (of "wind") which moves at unit speed and collides elastically with the barriers.

Delecroix-Hubert-Lelièvre have determined the divergence rate of the particle for all choices of size of the rectangular barriers [**DHL**]. Without Eskin-Mirzakhani-Mohammadi (and work of Eskin-Chaika [**EC**]), the best that could proven was the existence of a unspecified full measure set of choices of sizes for which such a result holds.

There are many other examples along these lines, where previously results were known to hold for almost all examples without being known to hold in any particular example, and now with Eskin-Mirzakhani-Mohammadi can be upgraded to hold in all cases.

Applications of Eskin-Mirzakhani's proof. The ideas that Eskin-Mirzakhani developed have applications beyond moduli spaces of translation surfaces. They are currently being used by Rodriguez-Hertz and Brown to study random diffeomorphisms on surfaces [**BRH**] and the Zimmer program (lattice actions on manifolds), and are also expected to have applications in ergodic theory on homogeneous spaces.

6. Context from Homogeneous spaces

The primary motivation for Theorem 4.1 is the following theorem.

Theorem 6.1 (Ratner's Theorem)**.** *Let G be a Lie group, and let $\Gamma \subset G$ be a lattice. Let $H \subset G$ be a subgroup generated by unipotent one parameter groups. Then every H orbit closure in G/Γ is a manifold, and moreover is a sub-homogeneous space.*

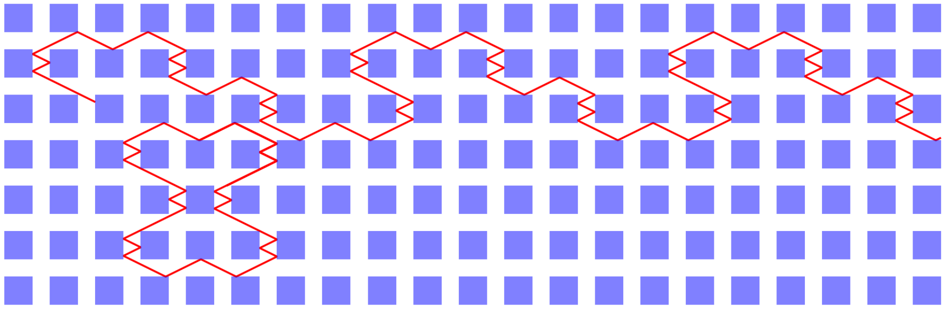


FIGURE 5.1. An example of a trajectory in the Wind Tree Model. Figure courtesy of Vincent Delecroix.

For example, the theorem applies if $G = SL(3,\mathbb{R})$, $\Gamma = SL(3,\mathbb{Z})$, and $H = \{h_t : t \in \mathbb{R}\}$, where

$$h_t = \begin{pmatrix} 1 & t & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \text{or} \qquad h_t = \begin{pmatrix} 1 & t & t^2/2 \\ 0 & 1 & t \\ 0 & 0 & 1 \end{pmatrix}.$$

Ratner's work confirmed conjectures of Raghunathan. Parts of these conjectures had previously been verified in some special cases by Dani and Margulis, for example for the second choice of h_t above [**DM90**].

The basic idea behind such proofs is the strategy of additional invariance. Given a closed H-invariant set, one starts with two points x and y very close together, and applies h_t until the points drift apart. The direction of drift is controlled by another one parameter subgroup, and one tries to show that the closed H set is in fact invariant under the one parameter group that gives the direction of drift. One continues this argument inductively, each time producing another one parameter group the set is invariant under, until one shows that the closed H invariant set is in fact invariant under a larger group L, and is contained in (and hence equal to) a single L orbit. This gives the set in question is homogenous, and in particular a manifold.

Of course, this is in fact very difficult, and complete proofs of Ratner's Theorem are very long and technical. For one thing, as is the case in the work of Eskin-Mirzakhani, it is in fact too difficult to work directly with closed invariant sets, as we have just suggested. Rather, one first must classify invariant measures. Thus the argument takes place in the realm of ergodic theory, which exactly studies group actions on spaces with invariant measures. See [**Mor05, Ein06**] for an introduction.

The fundamental requirement of the proof is that orbits of nearby points drift apart slowly and in a controlled way. This is intimately tied to the fact that unipotent one parameter groups are polynomial, as can be seen in the h_t above. Contrast this to the one parameter group

$$\begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix},$$

whose orbit closures may be fractal sets.

One might hope to study $GL(2,\mathbb{R})$ orbit closures of translation surfaces using the action of

$$u_t = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$$

on strata, in analogy to the proof of Ratner's Theorem. Unfortunately, the dynamics of the u_t action on strata is currently much too poorly understood for this. The best known result on the u_t action on strata is the crude, but still very useful, quantitative recurrence result of Minsky-Weiss [**MW02**]. In particular, this result says that u_t orbits do not diverge.

7. The structure of the proof

The proof of Theorem 4.1 builds on many ideas in homogeneous space dynamics, including the work of Benoist-Quint [**BQ09**], and the high and low entropy methods of Einsiedler-Lindenstrauss-Katok [**EKL06**] (see also the beautiful introduction for a general audience by Venkatesh [**Ven08**]). Lindenstrauss won the Fields medal in 2010 partially for the development of the low entropy method, and Benoist and Quint won the Clay prize for their work.

Entropy measures the unpredictability of a system that evolves over time.

Define P to be the upper triangular subgroup of $SL(2,\mathbb{R})$. The proof of Theorem 4.1 proceeds in two main stages.

In the first, Eskin-Mirzakhani show that any ergodic P invariant measure is in fact a Lebesgue class measure on a manifold cut out by linear equations, and must be $SL(2,\mathbb{R})$ invariant. (An ergodic measure is an invariant measure which is not the average of two other invariant measures in a nontrivial way. Thus the ergodic measures are the building blocks for all other invariant measures.)

In the second stage, Eskin-Mirzakhani-Mohammadi use this to prove Theorem 4.1, by constructing a P-invariant measure on every P-orbit closure. By contrast, it is not possible to directly construct an $SL(2,\mathbb{R})$ invariant measure on each $SL(2,\mathbb{R})$ orbit closure, and this is why the use of P is crucial. The algebraic structure of P makes it possible to average over larger and larger subsets of P and thus produce P invariant measures, whereas the more complicated algebraic structure of $SL(2,\mathbb{R})$ does not allow this. (The relevant property is that P is amenable, while $SL(2,\mathbb{R})$ is not.)

In the paper of Eskin-Mirzakhani, which caries out the first stage, the most difficult part is in fact to show P-invariant measures are $SL(2,\mathbb{R})$ invariant. To do this, extensive entropy arguments are used, partially inspired by the Margulis-Tomanov proof of Ratner's Theorem [**MT94**] and to a lesser extent the high and low entropy methods. This part is the technical heart of the argument, and takes almost 100 pages of delicate arguments. One of the morals is that entropy arguments are surprisingly effective in this context, and can be made to work without the use of an ergodic theorem.

Once Eskin-Mirzakhani show P-invariant measures are $SL(2,\mathbb{R})$ invariant, they build upon ideas of Benoist-Quint to conclude the invariant measure result.

All together, the proof is remarkably abstract. The only facts used about translation surfaces are formulas for Lyapunov exponents due to Forni [**For02**]. (The Lyapunov exponents of a smooth dynamical system, in this case the action of g_t on a stratum, measure the rates of expansion and contraction in different directions.) Forni was awarded the Brin prize partially for these formulas, which are of a remarkably analytic nature and arose from an insight of Kontsevich. Eskin-Mirzakhani also make use of a result of Filip **[Filb]** to handle a certain volume normalization issue at the end of the proof.

Results which classify invariant measures are rare gems. The arguments are abstract, and their purpose is to rule out nonexistent objects, and thus they cannot be guided by examples. To obtain a truly new measure classification result, one must find truly new ideas from among the sea of ideas which don't quite work, and the devil is in the details. This is very much the case in the paper of Eskin-Mirzakhani.

8. Relation to Teichmüller theory and algebraic geometry

Every translation surface has, in particular, the structure of a Riemann surface X. The extra structure is determined by additionally specifying an Abelian differential (a.k.a. holomorphic one form, or global section of the canonical bundle of X). The holomorphic one form is dz on the polygons, where z is the usual coordinate on the plane $\mathbb{C} \simeq \mathbb{R}^2$. It has zeros at the singularities of the flat metric.

Every translation surface can be given as a pair (X, ω). For example, the translation surface given by a square with unit area and opposite edges identified is $(\mathbb{C}/\mathbb{Z}[i], dz)$.

There is a projection map $(X, \omega) \mapsto X$ from a stratum of translation surfaces of genus g to the moduli space $\mathcal{M}_g$ of Riemann surfaces of genus g. Under this map, g_t orbits of translation surfaces project to geodesics for the Teichmüller metric. But it is important to note that there is no $GL(2, \mathbb{R})$ or g_t action on $\mathcal{M}_g$ itself, only on (the strata) of the bundle of Abelian differentials over $\mathcal{M}_g$. This is somewhat analogous to the fact that, given a Riemannian manifold, the geodesic flow is defined on the tangent bundle, and there is no naturally related flow on the manifold itself.

The map $(X, \omega) \mapsto X$ has fibers of real dimension two (given by multiplying ω by any complex number), and thus the projection of a four dimensional $GL(2, \mathbb{R})$ orbit to $\mathcal{M}_g$ is a two real dimensional object. It turns out that this object is an isometrically immersed copy of the upper half plane in $\mathbb{C}$ with its hyperbolic metric: such objects are called complex geodesics or Teichmüller disks. Note that Royden showed that the Teichmüller metric is equal to the Kobayashi metric on $\mathcal{M}_g$.

McMullen showed that every $GL(2, \mathbb{R})$ orbit closure in genus 2 is either a closed orbit, or an eigenform locus, or a stratum **[McM03]**. In particular, every $GL(2, \mathbb{R})$ orbit closure of genus 2 translation surfaces is a quasi-projective variety. The corresponding statement for $\mathcal{M}_2$ is that every complex geodesic is either closed, or dense in a Hilbert modular surface, or dense in $\mathcal{M}_2$.

The appearance of algebraic geometry in the study of orbit closures was unexpected, and arose in very different ways from work of McMullen and Kontsevich. A recent success in this direction is the following, which builds upon Theorem 4.1 and work of Möller **[Möl06b, Möl06a]**.

Theorem 8.1 (Filip **[Fila]**). *In every genus, every $GL(2, \mathbb{R})$ orbit closure is an algebraic variety that parameterizes pairs (X, ω) with special algebro-geometric properties, such as* $\mathrm{Jac}(X)$ *having real multiplication.*

Despite this theorem and Theorem 4.1, it is at present a major problem to classify $GL(2, \mathbb{R})$ orbit closures. Progress is ongoing, see for example **[Wrib, Wria, NW, ANW]**.

9. What to read next

We recommend the two page "What is ... measure rigidity" article by Einseidler **[Ein09]**, as well as the eight page note "The mathematical work of Maryam Mirzakhani" by McMullen **[McM]**. There are a large number of surveys on translation surfaces, for example **[MT02, Zor06]** and the author's recent introduction **[Wric]**. Alex Eskin has a short mini-course on his paper with Mirzakhani, and notes are available on his website **[Esk]**.

Acknowledgements. We are very grateful to Yiwei She, Max Engelstein, Weston Ungemach, Aaron Pollack, and Amir Mohammadi for making helpful comments and suggestions, which lead to significant improvements in this note.

References

[ANW] David Aulicino, Duc-Manh Nguyen, and Alex Wright, *Classification of higher rank orbit closures in* $\mathcal{H}^{\mathrm{odd}}(4)$, preprint, arXiv 1308.5879 (2013), to appear in Geom. Top.

[BQ09] Yves Benoist and Jean-François Quint, *Mesures stationnaires et fermés invariants des espaces homogènes*, C. R. Math. Acad. Sci. Paris **347** (2009), no. 1-2, 9–13.

[BRH] Aaron Brown and Federico Rodriguez Hertz, *Measure rigidity for random dynamics on surfaces and related skew products*, preprint, arXiv 1406.7201 (2014).

[DHL] Vincent Delecroix, Pascal Hubert, and Samuel Lelièvre, *Diffusion for the periodic wind-tree model*, preprint, arXiv 1107.1810 (2011), Ann. ENS to appear.

[DM90] S. G. Dani and G. A. Margulis, *Orbit closures of generic unipotent flows on homogeneous spaces of* $\mathrm{SL}(3, \mathbf{R})$, Math. Ann. **286** (1990), no. 1-3, 101–128.

[EC] Alex Eskin and John Chaika, *Every flat surface is birkhoff and oseledets generic in almost every direction*, preprint, arXiv 1305.1104 (2014).

[Ein06] Manfred Einsiedler, *Ratner's theorem on* $\mathrm{SL}(2, \mathbb{R})$*-invariant measures*, Jahresber. Deutsch. Math.-Verein. **108** (2006), no. 3, 143–164.

[Ein09] ———, *What is ... measure rigidity?*, Notices Amer. Math. Soc. **56** (2009), no. 5, 600–601.

[EKL06] Manfred Einsiedler, Anatole Katok, and Elon Lindenstrauss, *Invariant measures and the set of exceptions to Littlewood's conjecture*, Ann. of Math. (2) **164** (2006), no. 2, 513–560.

[EM] Alex Eskin and Maryam Mirzakhani, *Invariant and stationary measures for the* $SL(2, \mathbb{R})$ *action on moduli space*, preprint, arXiv 1302.3320 (2013).

[EMM] Alex Eskin, Maryam Mirzakhani, and Amir Mohammadi, *Isolation theorems for* $SL(2, \mathbb{R})$*-invariant submanifolds in moduli space*, preprint, arXiv:1305.3015 (2013).

[Esk] Alex Eskin, *Lectures on the* $SL(2, \mathbb{R})$ *action on moduli space*, `math.uchicago.edu/~eskin/luminy2012/lectures.pdf`.

[Fila] Simion Filip, *Semisimplicity and rigidity of the Kontsevich-Zorich cocycle*, preprint, arXiv:1307.7314 (2013).

[Filb] ———, *Splitting mixed Hodge structures over affine invariant manifolds*, preprint, arXiv:1311.2350 (2013).

[For02] Giovanni Forni, *Deviation of ergodic averages for area-preserving flows on surfaces of higher genus*, Ann. of Math. (2) **155** (2002), no. 1, 1–103.

[KMS86] Steven Kerckhoff, Howard Masur, and John Smillie, *Ergodicity of billiard flows and quadratic differentials*, Ann. of Math. (2) **124** (1986), no. 2, 293–311.

[LMW] Samuel Lelièver, Thierry Monteil, and Barak Weiss, *Everything is illuminated*, preprint, arXiv 1407.2975 (2014).

[Mas82] Howard Masur, *Interval exchange transformations and measured foliations*, Ann. of Math. (2) **115** (1982), no. 1, 169–200.

[Mas92] ———, *Hausdorff dimension of the set of nonergodic foliations of a quadratic differential*, Duke Math. J. **66** (1992), no. 3, 387–442.

[McM] Curtis McMullen, *The mathematical work of Maryam Mirzakhani*, `http://www.math.harvard.edu/archive/101_spring_00/www/papers/home/text/papers/icm14/icm14.pdf`.

[McM03] Curtis T. McMullen, *Billiards and Teichmüller curves on Hilbert modular surfaces*, J. Amer. Math. Soc. **16** (2003), no. 4, 857–885 (electronic).

[Möl06a] Martin Möller, *Periodic points on Veech surfaces and the Mordell-Weil group over a Teichmüller curve*, Invent. Math. **165** (2006), no. 3, 633–649.

[Möl06b] ———, *Variations of Hodge structures of a Teichmüller curve*, J. Amer. Math. Soc. **19** (2006), no. 2, 327–344 (electronic).

[Mor05] Dave Witte Morris, *Ratner's theorems on unipotent flows*, Chicago Lectures in Mathematics, University of Chicago Press, Chicago, IL, 2005.

[MT94] G. A. Margulis and G. M. Tomanov, *Invariant measures for actions of unipotent groups over local fields on homogeneous spaces*, Invent. Math. **116** (1994), no. 1-3, 347–392.

[MT02] Howard Masur and Serge Tabachnikov, *Rational billiards and flat structures*, Handbook of dynamical systems, Vol. 1A, North-Holland, Amsterdam, 2002, pp. 1015–1089.

[MW02] Yair Minsky and Barak Weiss, *Nondivergence of horocyclic flows on moduli space*, J. Reine Angew. Math. **552** (2002), 131–177.

[NW] Duc-Manh Nguyen and Alex Wright, *Non-Veech surfaces in $\mathcal{H}^{\mathrm{hyp}}(4)$ are generic*, preprint, arXiv:1306.4922 (2013), to appear in Geom. Funct. Anal.

[Vee82] William A. Veech, *Gauss measures for transformations on the space of interval exchange maps*, Ann. of Math. (2) **115** (1982), no. 1, 201–242.

[Ven08] Akshay Venkatesh, *The work of Einsiedler, Katok and Lindenstrauss on the Littlewood conjecture*, Bull. Amer. Math. Soc. (N.S.) **45** (2008), no. 1, 117–134.

[Wria] Alex Wright, *Cylinder deformations in orbit closures of translation surfaces*, preprint, arXiv 1302.4108 (2013), to appear in Geom. Top.

[Wrib] ———, *The field of definition of affine invariant submanifolds of the moduli space of abelian differentials*, preprint, arXiv 1210.4806 (2012), to appear in Geom. Top.

[Wric] ———, *Translation surfaces and their orbit closures: An introduction for a broad audience*, preprint, arXiv 1411.1827 (2014).

[ZK75] A. N. Zemljakov and A. B. Katok, *Topological transitivity of billiards in polygons*, Mat. Zametki **18** (1975), no. 2, 291–300.

[Zor06] Anton Zorich, *Flat surfaces*, Frontiers in number theory, physics, and geometry. I, Springer, Berlin, 2006, pp. 437–583.

E-mail address: `alexmurraywright@gmail.com`

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 17, 2014 (Baltimore, MD)

Daniel Rothman, Massachusetts Institute of Technology
Earth's Carbon Cycle: A Mathematical Perspective

Karen Vogtmann, Cornell University
The geometry of Outer space

Yakov Eliashberg, Stanford University
Recent advances in symplectic flexibility

Andrew Granville, Université de Montréal
Infinitely many pairs of primes differ by no more than 70 million (and the bound's getting smaller every day)

January 11, 2013 (San Diego, CA)

Wei Ho, Columbia University
How many rational points does a random curve have?

Sam Payne, Yale University
Topology of nonarchimedean analytic spaces

Mladen Bestvina, University of Utah
Geometric group theory and 3-manifolds hand in hand: the fulfillment of Thurston's vision for three-manifolds

Lauren Williams, University of California, Berkeley
Cluster algebras

January 6, 2012 (Boston, MA)

Jeffrey Brock, Brown University
Assembling surfaces from random pants: the surface-subgroup and Ehrenpreis conjectures

Daniel Freed, University of Texas at Austin
The cobordism hypothesis: quantum field theory + homotopy invariance = higher algebra

Gigliola Staffilani, Massachusetts Institute of Technology
Dispersive equations and their role beyond PDE

Umesh Vazirani, University of California, Berkeley
How does quantum mechanics scale?

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)

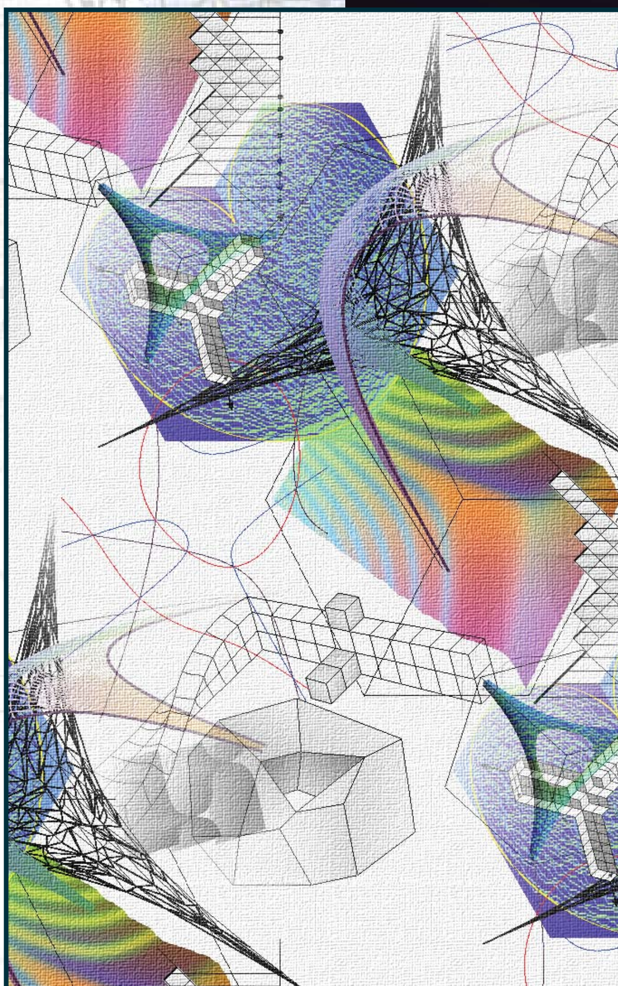
Michael J. Hopkins, MIT
Homotopy theory of schemes

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

2015 CURRENT EVENTS BULLETIN
Committee



The back cover graphic is reprinted courtesy of Andrei Okounkov.

Cover graphic associated with Jared S. Weinstein's talk courtesy of Jeff Weeks.

Cover graphic associated with Andrea R. Nahmod's talk courtesy of Chester Alexander, University of Alabama and Roger McLassus.