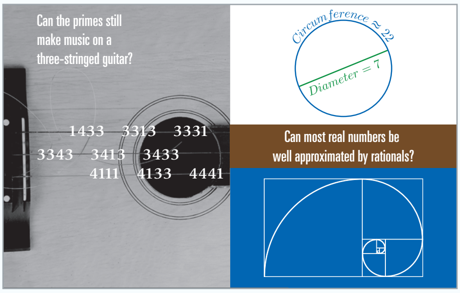
AMERICAN MATHEMATICAL SOCIETY

Current Events Bulletin

Friday, January 6, 2023
2:00–6:00 pm

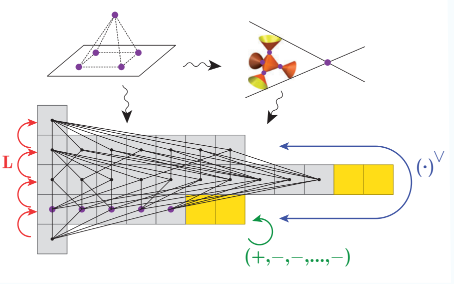
Ballroom AB, Hynes Convention Center | Joint Mathematics Meetings, Boston, MA



2:00 pm | **Andrew Granville**
Université de Montréal

Missing digits, and good approximations

What wonder will be next in the ancient study of the sequence of primes?



3:00 pm | **Christopher Eur**
Harvard University

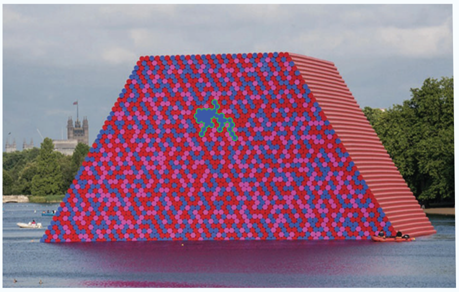
An essence of independence: recent works of June Huh on combinatorics and Hodge theory

Matroids, an abstract setting for linear independence are a backbone of combinatorics. Now they have fused with a central part of algebraic geometry over the complex numbers.

4:00 pm | **Henry Cohn**
Massachusetts Institute of Technology

From sphere packing to Fourier interpolation

What's with dimension 8 that makes it so special?



5:00 pm | **Martin Hairer**
Imperial College London

A stroll around the critical Potts model

Phase transitions are all around us. Perhaps this is a phase transition in the theory of phase transitions!

Organized by **David Eisenbud**, *University of California, Berkeley*

AMS American Mathematical Society

jmm 2023 Joint Mathematics Meetings
Boston • January 4–7

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
Mathematical Sciences Research Institute
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.

Thank you

The AMS is deeply grateful for the support of the Bose, Datta, Mukhopadhyay and Sarkar Fund, which sponsors one lecture at the Current Events Bulletin annually.

The Fund's purpose is to bring appreciation for mathematics to a broader audience. It was established in 2021 and made possible by the generosity of Dr. Salilesh Mukhopadhyay.

MISSING DIGITS, AND GOOD APPROXIMATIONS

ANDREW GRANVILLE

Abstract. James Maynard has taken the analytic number theory world by storm in the last decade, proving several important yet fun theorems, resolving questions that had seemed far out of reach. He is perhaps best known for his work on small and large gaps between primes (which were discussed, hot off the press, in my 2014 CEB lecture). In this talk we discuss two other breakthroughs:

– Mersenne numbers take the form $2^n - 1$ and so appear as $111\dots111$ in base 2, having no digit '0'. It is a famous conjecture that there are infinitely many such primes. More generally it was, until Maynard's work, an open question as to whether there are infinitely many primes that miss any given digit, in any given base. We will discuss Maynard's beautiful ideas that went into mostly resolving this question.

— In 1926, Khinchin gave remarkable conditions for when real numbers can usually be "well approximated" by infinitely many rationals. However Khinchin's theorem regarded 1/2, 2/4, 3/6 as distinct rationals and was unable to cope, say, with a restriction to fractions with prime denominators. In 1941 Duffin and Schaefer proposed an appropriate but significantly more general analogy involving approximation by reduced fractions (which is much more useful). We will discuss its recent resolution by Maynard together with Dimitris Koukoulopoulos.

This year's Current Events Bulletin highlights the work of the 2022 Fields medallists. In James Maynard's case there are a surprising number of quite different breakthroughs that could be discussed.[1] In my 2014 CEB lecture I described the work of Yitang Zhang [32] on bounded gaps between primes and noted that a first-year postdoc, James Maynard, had taken a different, much simpler but related approach, to also get bounded gaps [24] (and a similar proof had been found, independently, by Terry Tao, and given on his blog). Versions of both Zhang's proof and the Maynard-Tao proof appear in my article [12], where it is also announced that Maynard had within months made another spectacular breakthrough, this time on the largest known gaps between consecutive primes [25] (and a rather different proof [8] had been found by Ford, Green, Konyagin and Tao, the two proofs combining to give an even better result [9]). It has been like this ever since with Maynard, many breakthrough results, some more suitable for a broad audience, some of primary importance for the technical improvements. Rather than attempt

Received by the editors 31/10/2022.

2020 *Mathematics Subject Classification.* Primary .

Thanks to Dimitris Koukoulopoulos, Sun-Kai Leung and Cihan Sabuncu for their comments on a draft of this article, and to James Maynard for sharing his graphics. The author is partially supported by NSERC of Canada, both by a Discovery Grant and by a CRC..

[1]In my forthcoming textbook about the distribution of primes, starting from the basics, about one-sixth of the book is dedicated to various Maynard theorems. This, in one of the oldest and most venerable subjects of mathematics.

to summarize these all, I have selected two quite different topics, in both of which Maynard proved spectacular breakthroughs on questions that had long been stuck.

1. Primes missing digits

Does a given arithmetically natural set $\mathcal{A}$ of integers contain infinitely many primes? We believe so unless there is an obvious reason why not (like say, if $\mathcal{A}$ was the set of even integers, or the set of values of a reducible polynomial). Well known examples include,

- $\mathcal{A}$ is the set of all integers;
- $\mathcal{A}$ is the set of all integers in a given arithmetic progression (like $a \pmod q$ with $(a,q)=1$);
- $\mathcal{A}=\{p-2 : p \text{ is prime}\}$, which is a way to ask for twin primes;
- $\mathcal{A}=\{n^2+1 : n\in\mathbb{Z}\}$.

The first two questions are resolved and we even know an asymptotic estimate for how many such primes there are up to a given x, while the second two questions are (wide) open.

1.1. Guessing at the number of primes in $\mathcal{A}$.

The prime number theorem asserts that there are $\sim \frac{x}{\log x}$ primes $\leq x$ (so roughly 1 in $\log x$ of the integers around x are prime).[2] As a first guess we might think that the primes are equidistributed amongst the arithmetic progressions mod q and so the answer to the second question is $\sim \frac{1}{q}\cdot\frac{x}{\log x}$; however (a,q) divides any element of $a \pmod q$ and so if $(a,q)>1$ then this arithmetic progression contains only finitely many primes. Therefore we should restrict our attention to a with $(a,q)=1$. There are $\phi(q)$ such progressions, and so we should adjust our guess so that if $(a,q)=1$ then there are $\sim \frac{1}{\phi(q)}\cdot\frac{x}{\log x}$ primes $\leq x$ that are $\equiv a \pmod q$. This is the *prime number theorem for arithmetic progressions.*[3]

Let's put this heuristic in a broader context, for convenience letting $\mathcal{A}(x)$ denote the set of integers in $\mathcal{A}$ up to x, and $\pi_{\mathcal{A}}(x)$ the number of primes in $\mathcal{A}(x)$. We could have guessed at our answer by first estimating the size of $\mathcal{A}(x)$, (in this case x/q), and then as a first guess supposing that roughly 1 in $\log x$ of these integers are prime (so $\pi_{\mathcal{A}}(x)$ would be roughly $\sim \frac{|\mathcal{A}(x)|}{\log x}$). However this fails to take into account "local densities". That is, for each small prime p, we need to adjust by the density of integers in $\mathcal{A}$ that are coprime to p, divided by the density of all integers that are coprime to p. One sees that $1-\frac{1}{p}$ of all integers are coprime to p, and also in our $\mathcal{A}$ if $p \nmid q$. If p divides q and $(a,q)=1$ then all of the integers in $\mathcal{A}$ are

---

[2]To prove such a result it helps to include a weight $\log p$ at each prime p and prove instead that $\sum_{p \text{ prime}, p\leq x}\log p \sim x$, since x is a more natural function to work with than $\frac{x}{\log x}$. The prime number theorem can be deduced by the technique of "partial summation" which allows one to add or remove smooth weights from an estimate like this.

[3]First claimed by de la Vallée Poussin in 1899 based on ideas from his proof of the prime number theorem, and Dirichlet's proof of the infinitude of primes in arithmetic progressions. Thanks to Siegel and Walfisz this can be given, when x is large enough compared to q, as follows: Fix reals $A,B>0$. If $q\leq(\log x)^A$ then the number of primes $\equiv a \pmod q$ up to x,

$$\pi(x;q,a)=\frac{\pi(x)}{\phi(q)}\bigg(1+O\bigg(\frac{1}{(\log x)^B}\bigg)\bigg) \text{ whenever } (a,q)=1.$$

coprime to p. Combining these "probabilities" gives us the guess

$$\pi_{\mathcal{A}}(x) \approx \frac{|\mathcal{A}(x)|}{\log x} \cdot \prod_{p \nmid q} \frac{1-1/p}{1-1/p} \cdot \prod_{p|q} \frac{1}{1-1/p} = \frac{q}{\phi(q)} \cdot \frac{|\mathcal{A}(x)|}{\log x} = \frac{1}{\phi(q)} \cdot \frac{x}{\log x},$$

so we recover the correct prediction. This outlines a general strategy for guessing at $\pi_{\mathcal{A}}(x)$.

1.2. **Sparse sets of primes.** The first three questions above involve sets $\mathcal{A}$ that are quite dense amongst the integers. These well-worn methods usually have limited traction with sets $\mathcal{A}$ that are *sparse* like

- $\mathcal{A} = \{n \in (x, x + x^{.99}]\}$;
- $\mathcal{A} = \{n \equiv a \pmod q : n \leq x := q^{100}\}$ for given integer q and $(a, q) = 1$;
- $\mathcal{A} = \{n \leq x : \alpha n \pmod 1 \in [0, x^{-.01}]\}$ for a given real, irrational α.

In each of these examples, $|\mathcal{A}| \sim x^{.99}$, a rather sparse set, yet each was shown to have more-or-less the expected number of primes over 50 years ago (Theorems of Hoheisel, Linnik and Vinogradov, respectively), albeit all known proofs are rather difficult. However if we change ".99" to an exponent $< \frac{1}{2}$ then these questions are far beyond our current state of knowledge.[4]

A family of sparse arithmetic sequences are given by the sets of values of polynomials (perhaps in several variables). Examples for which infinitely many primes have been found include

$\mathcal{A} = \{c^2 + d^4 : c, d \geq 1\}$ which has $|\mathcal{A}(x)| \asymp x^{3/4}$ (see [10]); and

$\mathcal{A} = \{a^3 + 2b^3 : a, b \geq 1\}$ which has $|\mathcal{A}(x)| \asymp x^{2/3}$ (see [17]).

This last set is an example of the set of values of a *norm-form* as a^3+2b^3 is the norm an element, $a+2^{1/3}b$ of the ring of integers of $\mathbb{Q}(2^{1/3})$. For a number field $K/\mathbb{Q}$, with ring of integers $\mathbb{Z}[\omega_1, \ldots, \omega_d]$ we have $\text{Norm}_{K/\mathbb{Q}}(x_1\omega_1 + \ldots + x_d\omega_d) \in \mathbb{Z}[x_1, \ldots, x_d]$ is a degree d polynomial in the d variables $x_1, \ldots, x_d$. The prime ideal theorem implies that it takes on infinitely many prime values with the x_i all integers, provided that this polynomial has no fixed prime factor, though these sequences are not so sparse. For example, the norm form $a + 2^{1/3}b + 2^{2/3}c$ yields the expected number of primes; and the last displayed example states that this is true even when $c = 0$ (in which case one has a sparse set). There are infinitely many primes which equal the norm, $m^2 + n^2$, of $m + in$ for some integers m, n, but if we fix $n = 1$ we get the open question of primes of the form m^2+1. In 2002 Heath-Brown and Moroz [18] proved that one can take any cubic norm form with one of the variables equal to 0 (as long the new form is irreducible). Moreover in 2018, Maynard [27] proved such a result for norms of

$$\sum_{i=1}^{r} x_i \omega^i \in \mathbb{Z}(\omega) \text{ where } [\mathbb{Q}(\omega) : \mathbb{Q}] \leq \frac{4}{3} r.$$

1.3. **Primes with missing digits.** Other than short intervals, short arithmetic progressions, and polynomial values perhaps the best known question is to find primes without some explicitly named digit or digits in their decimal expansion. For example, as on the cover of this booklet, we might ask for primes which only

[4] The sparsest sets known in these questions to contain primes are $(x, x + x^{.525}], x = q^5$ and $\alpha n \pmod 1 \in [0, x^{-\frac{1}{3}+\epsilon}]$ due to [1, 31, 23] respectively.

have the digits 1,3 and 4 in their decimal expansions.[5] Our guess is that there are infinitely many such primes; to guess how many up to x, we can follow the above recipe: $\mathcal{A}(10^k) = 3^k$ (as there are three possibilities for each digit in the decimal expansion) and so $|\mathcal{A}(x)| \asymp x^\alpha$ where $\alpha = \frac{\log 3}{\log 10}$. These numbers are equidistributed mod p, except perhaps if p divides 10. Since the last digit is $1, 3$ or 4, the probability that 2 divides an element of $\mathcal{A}$ is $\frac{1}{3}$, and that 5 divides an element of $\mathcal{A}$ is 0, and so we get extra factors $\frac{1-1/3}{1-1/2} = \frac{4}{3}$ and $\frac{1-0}{1-1/5} = \frac{5}{4}$ respectively. Combining this information we guess that $\pi_{\mathcal{A}}(x) \sim \frac{4}{3} \cdot \frac{5}{4} \cdot \frac{|\mathcal{A}(x)|}{\log x} = \frac{5}{3} \cdot \frac{|\mathcal{A}(x)|}{\log x}$.

The above heuristic is a little misleading. For example if the integers of $\mathcal{A}$ only have the digits 4 and 5 in their decimal expansions then 5 is the only prime in $\mathcal{A}$ (since every element of $\mathcal{A}$ is divisible by 2 or 5). Therefore, in general, if $\mathcal{A}$ is the set of integers n which have only digits from $\mathcal{D}$ in their base q expansion let $\mathcal{D}_q = \{d \in \mathcal{D} : (d, q) = 1\}$ and then we predict that

$$\pi_{\mathcal{A}}(x) \sim \frac{|\mathcal{D}_q|/|\mathcal{D}|}{\phi(q)/q} \cdot \frac{|\mathcal{A}(x)|}{\log x}$$

Maynard obtained the first such theorems [26, 28] for certain general families of sparse sets $\mathcal{A}$. His most spectacular result [26] yields the above with $q = 10$ and $|\mathcal{D}| = 9$; that is, Maynard proved that there are roughly the expected number of primes that are missing *one* given digit in decimal. His methods give a lot more (as we will describe). His methods can't quite handle sets as sparse as $\mathcal{D} = \{1, 3, 4\}$ with $q = 10$ from our cover art, that is for another day. We will sketch the easier argument from [28] which gives many results of this type though only for much larger bases than 10.

1.4. **Who cares?** Is this a silly question? It is certainly diverting to wonder whether there are infinitely primes with given missing digits, but how does that impact any other serious questions in mathematics? This is a case of "the proof of the pudding is in the eating", that is its real value can be judged only from the beautiful mathematics that unfolds. The story is two-fold. The relevant Fourier coefficients have an extraordinary structure that allows Maynard to import ideas from Markov processes, which allows us to prove such theorems in bases > 100. To get the base down to 10, Maynard develops his ideas with a virtuosity of all sorts of deep techniques that spin an extraordinary (though technical) tale.

1.5. **Fourier analysis.** If $|n| < N$ then, summing the geometric series we have

$$\frac{1}{N} \sum_{j=0}^{N} e\left(\frac{jn}{N}\right) = \begin{cases} 1 & \text{if } n = 0; \\ 0 & \text{otherwise,} \end{cases}$$

where $e(t) := e^{2i\pi t}$ for any real t. To identify whether prime p equals some $a \in \mathcal{A}$ (in our case, $\mathcal{A}$ is the set of integers missing some given digit in base-q) we can take

[5] Riemann's explicit formula for the primes in terms of the zeros of the Riemann zeta-function, can be expressed in a form that reminds one of harmonics from Fourier analysis. Thus the Riemann Hypothesis has been paraphrased as "Do the primes have music in them?". The first part of our cover drawing represents this variation on that theme.

the above identity for $n = p - a$ and sum over all $a \in \mathcal{A}(N)$, to obtain

$$\pi_{\mathcal{A}}(N) = \sum_{p\le N}\sum_{a\in\mathcal{A}(N)} \frac{1}{N}\sum_{j=0}^{N-1} e\left(\frac{j(p-a)}{N}\right) = \frac{1}{N}\sum_{j=0}^{N-1} S_{\mathcal{P}}\left(\frac{j}{N}\right)S_{\mathcal{A}}\left(\frac{-j}{N}\right) \tag{1.1}$$

where $\mathcal{P}$ denotes the set of primes, and for a given set of integers T, we define the *exponential sum* (or the *Fourier transform* of $T(N)$) by

$$S_T(\theta) := \sum_{n\in T(N)} e(n\theta) \text{ for any real } \theta.$$

To establish a good estimate for $\pi_{\mathcal{A}}(N)$ using (1.1) one needs to identify those j for which the summand on the right-hand side is large; for example, $S_T(0) = |T|$ and so the $j = 0$ term in (1.1) yields $\frac{1}{N}|\mathcal{A}(N)| \cdot \pi(N) \sim \frac{|\mathcal{A}(N)|}{\log N}$ which is the expected order of magnitude of our main term. Other terms where $\frac{j}{N}$ is close to a rational with small denominator often also contribute to the main term, whereas we hope that the combined contribution of all of the other terms is significantly smaller. At first sight this seems unlikely since we only have the trivial bound $|S_T(\theta)| \le |T|$, but the trick is to use the Cauchy-Schwarz inequality followed by Parseval's identity so that

$$\frac{1}{N}\sum_{j=0}^{N-1} |S_T(\tfrac{j}{N})| \le \left(\frac{1}{N}\sum_{j=0}^{N-1} |S_T(\tfrac{j}{N})|^2\right)^{1/2} = |T|^{1/2}.$$

This implies for example that a typical term in the sum on the right-hand side of (1.1) has size $\sqrt{|\mathcal{A}(N)|}\cdot\sqrt{\pi(N)}$ which is significantly bigger than the main term but not as egregiously as when we used the trivial bound.

We have just described the thinking behind the *circle method* used when one sums or integrates over the values of an exponential sum as the variable rotates around the unit circle (that is, $e(\frac{j}{N})$ for $0 \le j \le N-1$, or $e(\theta)$ for $0 \le \theta < 1$). When trying to estimate the sum on the right-hand side of (1.1), we are most interested in those $\theta = \frac{j}{N}$ for which $S_{\mathcal{P}}(\theta)S_{\mathcal{A}}(-\theta)$ is "large". Experience shows that with arithmetic problems, the exponential sums can typically only be large when θ is close to a rational with small denominators, and so we cut the circle up into these *major arcs*, usually those θ near to a rational with small denominator, and *minor arcs*, the remaining θ, bounding the contribution from the minor arcs, and being as precise as possible with the major arcs to obtain the main terms.

Fourier analysis/the circle method is most successful when one has the product of at least three exponential sums to play with. For example the ternary Goldbach problem was more-or-less resolved by Vinogradov 85 years ago, whereas the binary Goldbach problem remains open.[6] For the ternary Goldbach problem, the number of representations of odd N as the sum of three primes is given by

$$\int_0^1 e(-N\theta)S_{\mathcal{P}(N)}(\theta)^3 d\theta,$$

[6] It is known that *almost all* integers n can be written as the sum of two primes in the expected number of ways, since by counting over all integers, one is, in effect, adding another exponential sum.

and the arc of width $\asymp \frac{1}{N}$ around 0 yields a main term of size $\asymp \frac{N^2}{(\log N)^3}$. We have the trivial bound $|S_{\mathcal{P}(N)}(\theta)| \leq \pi(N)$ and we will define here the minor arcs to be

$$\mathfrak{m} := \{\theta \in [0,1] : \ |S_{\mathcal{P}(N)}(\theta)| \leq \pi(N)/(\log N)^2\}.$$

(Since the typical size of $|S_{\mathcal{P}(N)}(\theta)|$ is $\sqrt{\pi(N)} < N^{1/2}$ we expect that all but a tiny set of the θ belong to these minor arcs.) Then

$$\begin{aligned}\left| \int_{\theta \in \mathfrak{m}} e(-N\theta) S_{\mathcal{P}(N)}(\theta)^3 d\theta \right| &\leq \int_{\theta \in \mathfrak{m}} |S_{\mathcal{P}(N)}(\theta)|^3 d\theta \\ &\leq \frac{\pi(N)}{(\log N)^2} \cdot \int_{\theta \in [0,1)} |S_{\mathcal{P}(N)}(\theta)|^2 d\theta \\ &= \frac{\pi(N)^2}{(\log N)^2} \sim \frac{N^2}{(\log N)^4}\end{aligned}$$

which is significantly smaller than the main term. Thus if we can identify which θ belong to $\mathfrak{m}$, then we can focus on evaluating $S_{\mathcal{P}(N)}(\theta)$ on the major arcs $\mathfrak{M} := [0,1) \setminus \mathfrak{m}$. There are strong bounds known for $S_{\mathcal{P}(N)}(\theta)$, as we will see later so this can all be done in practice.

1.6. **The missing digit problem.** To initiate the analogous plan to determine $\pi_{\mathcal{A}}(N)$ we would need to define the minor arcs to be those $\theta = \frac{j}{N}$ for which $|S_{\mathcal{A}}(\theta)| \leq \frac{|\mathcal{A}(N)|}{\sqrt{N}\log N}$. If, say, $|\mathcal{D}| = 9$ (primes with one missing digit) then $|\mathcal{A}(N)| \sim N^{1-2\delta}$ where $2\delta := \frac{\log 10/9}{\log 10} \approx 0.04576$. Therefore the bound desired here is $N^{\frac{1}{2}-2\delta+o(1)}$, which is significantly smaller than the bound $N^{\frac{1}{2}-\delta+o(1)}$ obtained from Parseval (i.e. a power of N smaller, rather than a power of $\log N$). This is what makes the product of only two exponential sums in (1.1) seem impossible. However restricted digit problems in base q are more tractable because the structure of $\mathcal{A}$ leads to an unusual distribution of its exponential sums. While it is true that $S_{\mathcal{P}}(\theta)$ is only large when θ is near to a rational with small denominator (as proved by Vinogradov), $S_{\mathcal{A}}(\theta)$ behaves differently. It is only large when there are many 0's and $q-1$'s in the decimal expansion of θ. Now if $S_{\mathcal{P}}(\theta)S_{\mathcal{A}}(-\theta)$ is large then $S_{\mathcal{P}}(\theta)$ and $S_{\mathcal{A}}(-\theta)$ must both individually be large, and so θ is both near to a rational with small denominator and has many 0's and $q-1$'s in its decimal expansion, something which is very rare. In fact this implies that θ is close to a rational whose denominator is a small power of q. Indeed for $N = q^k$ the major arcs which give the expected main term are those $\frac{j}{q^k} = \frac{r}{q}$ for some integer r, so their contribution to the above sum is

$$\begin{aligned}q^{-k} \sum_{r=0}^{q-1} S_{\mathcal{P}}\left(\frac{r}{q}\right) S_{\mathcal{A}}\left(\frac{-r}{q}\right) &= q^{-k} \sum_{a \in \mathcal{A}, a \leq q^k} \ \sum_{p \text{ prime}, \leq q^k} \ \sum_{r=0}^{q-1} e\left(\frac{r}{q}(p-a)\right) \\ &= q^{1-k} \sum_{a \in \mathcal{A}, a \leq q^k} \ \sum_{\substack{p \text{ prime}, \leq q^k \\ p \equiv a \pmod q}} 1.\end{aligned}$$

Now if a prime p has last digit d then $(d,q) = 1$, and if $d \equiv p \equiv a \pmod q$ then $d \in \mathcal{D}$ so that $d \in \mathcal{D}_q$. There are $|\mathcal{D}|^{k-1}$ integers $a \in \mathcal{A}, a \leq q^k$ with $a \equiv d \pmod q$, and so this sum becomes, using the prime number theorem for arithmetic progressions,

$$q^{1-k}\sum_{d\in\mathcal{D}_q}|\mathcal{D}|^{k-1}\sum_{\substack{p\text{ prime},\le q^k\\ p\equiv d\pmod q}}1\sim q^{1-k}\cdot\frac{|\mathcal{D}_q|}{|\mathcal{D}|}\cdot|\mathcal{A}(q^k)|\cdot\frac{1}{\phi(q)}\frac{q^k}{\log q^k}$$

$$=\frac{|\mathcal{D}_q|/|\mathcal{D}|}{\phi(q)/q}\cdot\frac{|\mathcal{A}(N)|}{\log N},\tag{1.2}$$

which is precisely the prediction we had for $\pi_{\mathcal{A}}(N)$ above.

These major arcs were not difficult to identify and evaluate. The result will follow if we can show that

$$\frac{1}{N}\sum_{\substack{0\le j\le N\\ j\ne\frac{N}{q}i,i\in\mathbb{Z}}}\left|S_{\mathcal{P}}\left(\frac{j}{N}\right)\right|\cdot\left|S_{\mathcal{A}}\left(\frac{-j}{N}\right)\right|\ll\frac{|\mathcal{A}(N)|}{(\log N)^A}\tag{1.3}$$

for some $A>1$. The challenge is to then suitably bound the summand on all of the remaining arcs, the remaining major arcs as well as the minor arcs.

1.7. **Arcs.** The usual way to dissect the circle is to pick a parameter $1<M<N$ and recall that, by Dirichlet's Theorem (see the discussion in Part II), for every $\alpha\in[0,1]$ there exists a reduced fraction r/s with $s\le M$ for which

$$\left|\alpha-\frac{r}{s}\right|\le\frac{1}{sM}$$

(and the right-hand side is $\le 1/s^2$). Therefore we may cover $[0,1]$ (and so cover the circle, by mapping $t\to e(t)$) with the intervals (arcs),

$$\bigcup_{s\le M}\bigcup_{\substack{0\le r\le s\\ (r,s)=1}}\left[\frac{r}{s}-\frac{1}{sM},\ \frac{r}{s}+\frac{1}{sM}\right].$$

The arcs with s small are usually the major arcs, those s large are the minor arcs, though in this case it is a little more complicated: The major arcs will be given by

$$\bigcup_{s\le(\log N)^A}\bigcup_{\substack{0\le r\le s\\ (r,s)=1}}\left[\frac{r}{s}-\frac{(\log N)^A}{N},\ \frac{r}{s}+\frac{(\log N)^A}{N}\right].$$

1.8. **Other major arcs, when all prime factors of s divide q.** Throughout this subsection we assume that $p|s\implies p|q$ so that s divides $N=q^k$ for all sufficiently large k, and so r/s may be written as j/N for some integer j. We also assume that $s\le(\log N)^A$. The prime number theorem for arithmetic progressions gives

$$S_{\mathcal{P}}\left(\frac{r}{s}\right)=\sum_{p\le N}e\left(\frac{pr}{s}\right)=\sum_{a:(a,s)=1}e\left(\frac{ar}{s}\right)\pi(N;s,a)$$

$$=\frac{\pi(N)}{\phi(s)}\sum_{a:(a,s)=1}e\left(\frac{ar}{s}\right)+O\left(\frac{\pi(N)}{(\log N)^B}\right)=\pi(N)\left(\frac{\mu(s)}{\phi(s)}+O\left(\frac{1}{(\log N)^B}\right)\right)\tag{1.4}$$

Therefore, by partial summation, if $1 \le |i| \le (\log N)^A$, or if $i = \mu(s) = 0$,

$$S_{\mathcal{P}}\left(\frac{r}{s} + \frac{i}{q^k}\right) = \pi(N)\frac{\mu(s)}{\phi(s)}\int_0^N e\left(\frac{it}{N}\right)dt + O\left(\frac{i\pi(N)}{(\log N)^B}\right) \ll \frac{\pi(N)}{(\log N)^{B-A}}.$$

Therefore, since $|S_{\mathcal{A}}(\frac{-j}{N})| \le |\mathcal{A}(N)|$ trivially, taking $B = 5A - 1$ with $A \ge 2$ we obtain

$$\frac{1}{N} \sum_{\substack{s \le (\log N)^A \\ p|s \implies p|q}} \sum_{\substack{0 \le r < s \\ (r,s)=1}} \sum_{\substack{j: \\ \mu(s)^2 \le |j - \frac{r}{s}N| \le (\log N)^A}} \left| S_{\mathcal{P}}\left(\frac{j}{N}\right) S_{\mathcal{A}}\left(\frac{-j}{N}\right) \right| \ll \frac{|\mathcal{A}(N)|}{(\log N)^A}$$

which is much smaller than the main term in (1.2).

The only remaining such terms are at r/s where s is squarefree and all its prime factors divide q, which imply that s divides q, and these terms were already included in the sum in the previous subsection that led to (1.2).

The calculations in this subsection accounted for the major arcs,

$$\bigcup_{\substack{s \le (\log N)^A \\ p|s \implies p|q}} \bigcup_{\substack{0 \le r \le s \\ (r,s)=1}} \left[\frac{r}{s} - \frac{(\log N)^A}{N}, \ \frac{r}{s} + \frac{(\log N)^A}{N}\right].$$

1.9. **The extraordinary structure of these exponential sums.** If $\mathcal{A}$ is the set of integers missing the digit b in base q, and $N = q^k$, we can write

$$\mathcal{A}(N) = \left\{ n = \sum_{i=0}^{k-1} a_i q^i : \text{ Each } a_i \in \mathcal{D} := \{0, 1, \ldots, q-1\} \setminus \{b\} \right\}.$$

Since $e(n\theta) = \prod_{i=0}^{k-1} e(a_i q^i \theta)$, therefore

$$S_{\mathcal{A}}(\theta) = \sum_{\text{Each } a_i \in \mathcal{D}} \prod_{i=0}^{k-1} e(a_i q^i \theta) = \prod_{i=0}^{k-1} \left(\sum_{a_i \in \mathcal{D}} e(a_i q^i \theta) \right)$$

$$= \prod_{i=0}^{k-1} \left(\frac{e(q^{i+1}\theta) - 1}{e(q^i\theta) - 1} - e(bq^i\theta) \right). \tag{1.5}$$

Write $\theta = \sum_{j \ge 1} t_{j-1}/q^j$ in base q (i.e. the $t_i \in \{0, 1, \ldots, q-1\}$) so that

$$q^i\theta \mod 1 = \frac{t_i}{q} + \frac{t_{i+1}}{q^2} + \cdots = \frac{t_i + (q^{i+1}\theta \mod 1)}{q}.$$

Thus $q^i\theta \mod 1 \in [\frac{t_i}{q}, \frac{t_i+1}{q})$ and so $e(q^i\theta) \approx e(t_i/q)$.

1.10. **Base q, where q is large.** We begin with the bounds

$$\left| \frac{e(q^{i+1}\theta) - 1}{e(q^i\theta) - 1} - e(bq^i\theta) \right| \le \min\left\{ q - 1, 1 + \frac{1}{2\|q^i\theta\|} \right\}$$
$$\le \min\left\{ q - 1, 1 + \frac{q}{2\min\{t_i, q-1-t_i\}} \right\},$$

and so, by (1.5),

$$|S_{\mathcal{A}}(\theta)| \le \prod_{i=0}^{k-1} \min\left\{ q - 1, 1 + \frac{q}{2\min\{t_i, q-1-t_i\}} \right\}. \tag{1.6}$$

Another bound (which is easier to work with) begins by noting that

$$|e(a\phi) + e((a+1)\phi)|^2 = 2 + 2\cos(2\pi\phi) < 4\exp(-2\|\phi\|^2),$$

so that $|e(a\phi)+e((a+1)\phi)| \leq 2\exp(-\|\phi\|^2)$. If $q > 3$ then there are two consecutive integers in $\mathcal{D}$ and so

$$\sum_{a\in\mathcal{D}} e(a\phi) \leq q - 3 + 2\exp(-\|\phi\|^2) \leq (q-1)\exp\Big(-\frac{\|\phi\|^2}{q}\Big),$$

and therefore, by (1.5),

$$|S_{\mathcal{A}}(\theta)| \leq |\mathcal{A}(N)| \exp\Big(-\frac{1}{q}\sum_{i=0}^{k-1}\|q^i\theta\|^2\Big) \tag{1.7}$$

We use this to deal with the remaining possible major arcs.

These arguments are far from sharp and both (1.6) and (1.7) can be sharpened.

1.11. **Major arcs, where s has a prime factor that does not divide q.** Suppose that prime $p|s$ but $p \nmid q$. Then p divides the denominator of the reduced fraction for $q^i \cdot \frac{r}{s}$ so that $\|q^i \cdot \frac{r}{s}\| \geq \frac{1}{p}$. Moreover if $|\theta - \frac{r}{s}| \leq \frac{1}{2pN^{1/2}}$ and $i \leq \frac{k}{2}$ then

$$\|q^i\theta\| \geq \|q^i \cdot \tfrac{r}{s}\| - q^i|\theta - \tfrac{r}{s}| \geq \tfrac{1}{p} - \tfrac{q^{k/2}}{2pN^{1/2}} = \tfrac{1}{2p}.$$

Now if $\|q^i\theta\| < \frac{1}{2q}$ then $\|q^{i+1}\theta\| = q\|q^i\theta\|$. Therefore, for every integer i there exists an integer $j, i \leq j \leq i + \lfloor\frac{\log p}{\log q}\rfloor$ for which $\|q^j\theta\| \geq \frac{1}{2q}$, which implies that

$$\sum_{i=0}^{k/2}\|q^i\theta\|^2 \geq \frac{1}{4q^2}\#\{j \in [0, \tfrac{k}{2}) : \|q^j\theta\| \geq \tfrac{1}{2q}\} \geq \frac{1}{4q^2}\frac{\log q^{k/2}}{\log pq} \geq \frac{k}{8mq^2}$$

for $s \leq q^m$ and $m \in \mathbb{Z}$, since then $\lfloor\frac{\log p}{\log q}\rfloor \leq m-1$. Here we let $m = \lfloor\sqrt{k}/9q^3\rfloor$ and assume that $k \geq 100q^6$.

Thus $|S_{\mathcal{A}}(\theta)| \leq |\mathcal{A}(N)|\exp(-\frac{k}{8mq^3})$ by (1.7), and $|S_{\mathcal{P}}(\theta)| \leq \pi(N)$ trivially, so that as $2q^{2m} \leq N^{1/2}$ then

$$\begin{aligned}\frac{1}{N}\sum_{\substack{s\leq q^m\\ \exists p|s,p\nmid q}}\ \sum_{\substack{0\leq r<s\\ (r,s)=1}}\ \sum_{j:|j-\frac{r}{s}N|\leq q^m}\Big|S_{\mathcal{P}}\Big(\frac{j}{N}\Big)S_{\mathcal{A}}\Big(\frac{-j}{N}\Big)\Big| &\ll \frac{|\mathcal{A}(N)|}{\log N}q^{3m}\exp(-\tfrac{k}{8mq^3})\\ &\ll \frac{|\mathcal{A}(N)|}{\log N}e^{-\sqrt{k}},\end{aligned}$$

which is much smaller than the main term in (1.2).

This subsection accounts for the major arcs,

$$\bigcup_{\substack{s\leq q^m\\ \exists p|s \text{ such that } p\nmid q}}\ \bigcup_{\substack{0\leq r\leq s\\ (r,s)=1}}\Big[\frac{r}{s} - \frac{q^m}{N},\ \frac{r}{s} + \frac{q^m}{N}\Big].$$

and $q^m = c_q^{\sqrt{k}}$ for some $c_q > 1$, which is larger $(\log N)^A$ for k sufficiently large.

1.12. **The minor arcs.** The remaining challenge comes from the minor arcs. Now

$$\sum_{t=0}^{q-1} \min\left\{q-1, 1+\frac{q}{2\min\{t, q-1-t\}}\right\} = 2(q-1)+(q-2)+\sum_{1\le t\le \frac{q-1}{2}} \frac{q}{t} \le 3q+q\log q,$$

and since the set of values $\{\theta+\frac{j}{q^k} \mod 1 : 0\le j\le q^k-1\}$ run once through all of the $(t_0,\dots,t_{k-1})$ possibilities we obtain

$$\sum_{j=0}^{q^k-1}\left|S_{\mathcal{A}}\left(\theta+\frac{j}{q^k}\right)\right| = \sum_{0\le t_0,\dots,t_{k-1}\le q-1}\left|S_{\mathcal{A}}\left(\theta+\sum_{i=1}^{k}\frac{t_{i-1}}{q^i}\right)\right| \le (3q+q\log q)^k. \tag{1.8}$$

Therefore

$$\int_0^1 |S_{\mathcal{A}}(\alpha)|d\alpha = \int_0^{q^{-k}}\sum_{j=0}^{q^k-1}\left|S_{\mathcal{A}}\left(\theta+\frac{j}{q^k}\right)\right|d\theta \le (3+\log q)^k. \tag{1.9}$$

Notice that $\frac{d}{d\theta}e(n\theta)=2i\pi\cdot ne(n\theta)=2i\pi\cdot\sum_{j=0}^{k-1}a_jq^je(a_jq^j)\prod_{i\ne j}e(a_iq^i)$. We can modify the above argument from bounds for a sum of $|S_{\mathcal{A}}(\cdot)|$-values to a sum of $|S'_{\mathcal{A}}(\cdot)|$-values, by bounding the contribution of the jth term in the product by q^j times

$$\left|\sum_{a=0}^{q-1} a\,e(aq^j\theta)-b\,e(bq^j\theta)\right| \le (q-1)\min\left\{q-1, 1+\frac{1}{2\|q^i\theta\|}\right\}.$$

Therefore, as $(q-1)\sum_{j=0}^{k-1}q^j<q^k$ we obtain

$$\int_0^1 |S'_{\mathcal{A}}(\alpha)|d\alpha \le q^k(3+\log q)^k. \tag{1.10}$$

One can bound a differentiable function $f(\cdot)$ at a point by its values in a neighbourhood by the classical inequality

$$|f(\theta)| \le \frac{1}{2\Delta}\int_{\theta-\Delta}^{\theta+\Delta}|f(\phi)|d\phi+\frac{1}{2}\int_{\theta-\Delta}^{\theta+\Delta}|f'(\phi)|d\phi$$

We can sum this over a set of points (on the unit circle), $\theta_1,\dots,\theta_m$ where $|\theta_i-\theta_j|\le 2\Delta$ if $i\ne j$ so the integrals above do not overlap, to obtain

$$\sum_{i=1}^{m}|f(\theta_i)| \le \frac{1}{2\Delta}\int_0^1|f(\phi)|d\phi+\frac{1}{2}\int_0^1|f'(\phi)|d\phi.$$

Our choice of points is a bit complicated: The θ_i will be selected within $\Delta=\frac{1}{4D^2}$ of the fractions $\frac{r}{s}$ with $(r,s)=1$ and $0\le r<s\le D$ with $(r,s)=1$ displaced by a fixed quantity ξ. The fractions are distinct so any two differ by $|\frac{r}{s}-\frac{r'}{s'}|\ge\frac{1}{ss'}>\frac{1}{D^2}$, and therefore the points differ by $\ge\frac{1}{D^2}-2\Delta=2\Delta$ and so

$$\sum_{s\le D}\sum_{\substack{0\le r<s\\(r,s)=1}}\max_{|\eta|\le\Delta}\left|f\left(\frac{r}{s}+\xi+\eta\right)\right| \le 2D^2\int_0^1|f(\phi)|d\phi+\frac{1}{2}\int_0^1|f'(\phi)|d\phi.$$

We now apply this with $f=S_A$ and use (1.9) and (1.10) to obtain

$$\sum_{s\le D}\sum_{\substack{0\le r<s\\(r,s)=1}}\max_{|\eta|\le\frac{1}{4D^2}}\left|S_A\left(\frac{r}{s}+\xi+\eta\right)\right| \le (2D^2+\tfrac{1}{2}q^k)(3+\log q)^k. \tag{1.11}$$

1.13. **Hybrid estimate.** We need notation that reflects that our sum is up to q^k, since we will now vary k. So let

$$\widehat{A_k}(\theta) := S_{\mathcal{A}}(\theta) = \sum_{n\in\mathcal{A}(q^k)} e(n\theta)$$

Our formula (1.5), implies that if $\ell \le k$ then

$$\widehat{A_k}(\theta) = \widehat{A_{k-\ell}}(\theta)\widehat{A_\ell}(q^{k-\ell}\theta).$$

For $m \le k-\ell$ replace k by $k-\ell$ and $k-\ell$ by m so that

$$\widehat{A_{k-\ell}}(\theta) = \widehat{A_m}(\theta)\widehat{A_{k-\ell-m}}(q^m\theta).$$

and therefore

$$\widehat{A_k}(\theta) = \widehat{A_m}(\theta)\widehat{A_{k-\ell-m}}(q^m\theta)\widehat{A_\ell}(q^{k-\ell}\theta).$$

Since $|\widehat{A_{k-\ell-m}}(q^m\theta)| \le (q-1)^{k-\ell-m}$ this yields

$$|\widehat{A_k}(\theta)| = (q-1)^{k-\ell-m}|\widehat{A_m}(\theta)|\cdot|\widehat{A_\ell}(q^{k-\ell}\theta)|.$$

and so

$$\begin{aligned}\left|\widehat{A_k}\left(\frac{j}{q^k}\right)\right| &\le (q-1)^{k-\ell-m}\left|\widehat{A_m}\left(\frac{j}{q^k}\right)\right|\cdot\left|\widehat{A_\ell}\left(\frac{j}{q^\ell}\right)\right| \\ &\le (q-1)^{k-\ell-m}\left|\widehat{A_\ell}\left(\frac{j}{q^\ell}\right)\right|\cdot\max_{i:|i-q^k\cdot\frac{r}{s}|\le B}\left|\widehat{A_m}\left(\frac{i}{q^k}\right)\right|.\end{aligned}$$

Now assume that $B \le q^\ell/2$ and $B \le q^k/4D^2$ so that

$$\begin{aligned}&\sum_{s\le D}\sum_{\substack{0\le r<s\\(r,s)=1}}\sum_{j:|j-q^k\cdot\frac{r}{s}|\le B}\left|S_{\mathcal{A}}\left(\frac{j}{q^k}\right)\right| \\ &\le (q-1)^{k-\ell-m}\sum_{\substack{0\le r<s\le D\\(r,s)=1}}\max_{i:|i-q^k\cdot\frac{r}{s}|\le B}\left|\widehat{A_m}\left(\frac{i}{q^k}\right)\right|\cdot\sum_{j:|j-q^k\cdot\frac{r}{s}|\le B}\left|\widehat{A_\ell}\left(\frac{j}{q^\ell}\right)\right|.\end{aligned}$$

We extend the final sum to a sum over all $j \pmod{q^\ell}$ so that it is $\le (3q+q\log q)^\ell$ by (1.8). For the next sum

$$\max_{i:|i-q^k\cdot\frac{r}{s}|\le B}\left|\widehat{A_m}\left(\frac{i}{q^k}\right)\right| \le \max_{i:|\eta|\le\frac{B}{q^k}}\left|\widehat{A_m}\left(\frac{r}{s}+\eta\right)\right| \le \max_{i:|\eta|\le\frac{1}{4D^2}}\left|\widehat{A_m}\left(\frac{r}{s}+\eta\right)\right|$$

and so the internal sum is $\le (2D^2+\frac{1}{2}q^m)(3+\log q)^m$ by (1.11). Therefore

$$\sum_{s\le D}\sum_{\substack{0\le r<s\\(r,s)=1}}\sum_{j:|j-q^k\cdot\frac{r}{s}|\le B}\left|S_{\mathcal{A}}\left(\frac{j}{q^k}\right)\right| \le |\mathcal{A}(q^k)|(2D^2/q^m+\tfrac{1}{2})\mathcal{L}_q^{\ell+m}$$

where $\mathcal{L}_q := \frac{q(3+\log q)}{q-1}$. Select ℓ to be the smallest integer for which $q^\ell \ge 2B$ and m to be the largest integer with $q^m \le \min\{D^2, q^{k-\ell}\}$, so that $q^\ell = Bq^{O(1)}$ and $q^m = D^2q^{O(1)}$, and therefore $2D^2/q^m + \frac{1}{2} \ll q^{O(1)}$.

If $\mathcal{L}_q < q^\tau$ then $\mathcal{L}_q^{\ell+m} < (BD^2)^\tau q^{O(1)}$ so that

$$\sum_{s\le D}\sum_{\substack{0\le r<s\\(r,s)=1}}\sum_{j:|j-q^k\cdot\frac{r}{s}|\le B}\left|S_{\mathcal{A}}\left(\frac{j}{q^k}\right)\right| \ll |\mathcal{A}(N)|q^{O(1)}(BD^2)^\tau.$$

1.14. **Putting it all together.** A well known estimate on exponential sums over primes gives that if $s \asymp D$ and $|j - N \cdot \frac{r}{s}| \asymp B$ with $D < \sqrt{N}$ then

$$\left|S_{\mathcal{P}}\left(\frac{j}{N}\right)\right| \ll \left(N^{4/5} + \frac{N}{(BD)^{1/2}} + (BDN)^{1/2}\right)(\log N)^{-3/4} \tag{1.12}$$

Our arcs have $|j - q^k \cdot \frac{r}{s}| \leq N/sM$ and so we may assume that $B \leq N/DM$. Since we are on the minor arcs we may assume that $BD > (\log N)^{20A}$. We will select $k > q$ so that $q \leq \log N$ with $M = \sqrt{N}$ and $D \leq M$ so that $BD \leq \sqrt{N}$. Combining the last two estimates we obtain

$$\frac{1}{N}\sum_{s\sim D}\sum_{\substack{0\leq r<s\\(r,s)=1}}\sum_{j:|j-q^k\cdot\frac{r}{s}|\sim B}\left|S_{\mathcal{P}}\left(\frac{j}{N}\right)S_{\mathcal{A}}\left(-\frac{j}{N}\right)\right|$$
$$\ll \frac{|\mathcal{A}(N)|}{\log N}\cdot(BD^2)^{\tau}\left(\frac{1}{N^{1/5}} + \frac{1}{(BD)^{1/2}}\right)(\log N)^{O(1)}$$

as $BD \leq \sqrt{N}$. If $\tau < \frac{1}{5}$ then the first term is $< 1/N^{\frac{1}{5}-\tau}$ as $BD^2 \ll N$. Moreover $(BD^2)^{\tau}/(BD)^{1/2} < 1/(BD)^{1/10} \ll (\log N)^{-2A}$. Therefore the contribution of the minor arcs is much smaller than the main term in (1.2).

1.15. **Explicit bounds on q.** Now $\mathcal{L}_q := \frac{q(3+\log q)}{q-1} < q^{1/5}$ for $q \geq 1520573$.

To improve this, note that if $\phi = \frac{t+\delta}{q} \in [0, \frac{1}{2}]$ with $0 \leq \delta < 1$ then

$$\left|\frac{e(q\phi)-1}{e(\phi)-1} - e(b\phi)\right| \leq 1 + \max_{0\leq\delta\leq\frac{1}{2}}\frac{\sin(\pi\delta)}{\sin(\pi\frac{t+\delta}{q})} \leq 1 + \frac{1}{\sin(\pi\frac{t}{q})},$$

and so our proofs work for $q \geq 62893$ since then

$$\sum_{t=0}^{q-1}\min\left\{q-1, 1 + \frac{1}{\sin(\pi\frac{|t|}{q})}\right\} \leq q^{1/5}(q-1).$$

(This sum is $\sim \frac{2}{\pi}q\log q$.) It seems like we will need more ideas to reduce the base from 62893 all the way down to 10. In particular we need to do more than simply obtain bounds on the ith component of (1.5) as a function of t_i.

1.16. **Better bounds on** (1.5) **via a special Markov process.** We approximated the terms in (1.5) using only the first term of the base-q expansion of $q^i\theta$ mod 1. However if we obtain a more precise approximation using, say, the first two terms, t_i and t_{i+1}, of the base-q expansion of $q^i\theta$ mod 1, then the bounds on the ith and $(i+1)$st terms are no longer independent (it was that independence which allowed us to take a sum of the product equal to the product of various smaller sums). In particular we obtain a more accurate approximation using $e(q^i\theta) \approx e(t_i/q + t_{i+1}/q^2)$ by using two terms of the expansion, etc. Substituting this first approximation into (1.5) yields that

$$|S_{\mathcal{A}}(\theta)| \approx \prod_{i=0}^{k-1}F(t_i,t_{i+1}) \text{ where } F(t,u) := \left|\frac{e(\frac{u}{q})-1}{e(\frac{t}{q}+\frac{u}{q^2})-1} - e(b(\tfrac{t}{q}+\tfrac{u}{q^2}))\right| \text{ if } t \neq 0,$$

and $F(0,u) = q-1$. Now the consecutive terms depend on each other so we cannot separate them as before. Instead we can form the q-by-q matrix M with entries

$M_{a,b} := F(a,b)$ for $0 \le a, b \le q-1$. Then for $t_0, t_k \in \{0,\dots,q-1\}$

$$M^k_{t_0,t_k} = \sum_{t_1,\dots,t_{k-1}\in\{0,\dots,q-1\}} \prod_{i=0}^{k-1} F(t_i,t_{i+1}) \approx \sum_{\substack{t_1,\dots,t_{k-1}\in\{0,\dots,q-1\}\\ \theta=\sum_{i=0}^k t_i/q^{i+1}}} |S_{\mathcal{A}}(\theta)|.$$

Summing this over all $t_0, t_k \in \{0,\dots,q-1\}$ gives the complete sum over the $\theta = j/q^k$; that is,

$$\sum_{j=0}^{q^k-1} |S_{\mathcal{A}}(\tfrac{j}{q^k})| \approx (1,1,\dots,1)M^k(1,1,\dots,1)^T \le c_M|\lambda_M|^k$$

where λ_M is the largest eigenvalue of M and $c_M > 0$ is some computable constant.[7] Our proof of the bounds for the minor arcs can be modified and the result follows provided

$$\lambda_M < q^{1/5}.$$

But this is far from the end of the story, since we can be more precise by replacing the transition from the first two terms of the expansion of $q^i\theta$, $\{t_i, t_{i+1}\}$ to the next two $\{t_{i+1}, t_{i+2}\}$ in our "Markov process", to using the transition from the first ℓ terms to the next ℓ. Although this yields a q^ℓ-by-q^ℓ transition matrix, in which each row and column is supported at only q entries. And the larger ℓ is the more precise our bounds. For $q = 10$, Maynard found that $\lambda_M < 2.24190 < 10^{27/77}$; this is not quite down to exponent $\frac{1}{5}$ but it is far smaller than the "trivial" $\frac{1}{2}$.

Another surprising feature of this set up are the rth moments for $r > 0$: If we replace $|S_{\mathcal{A}}(\frac{j}{q^k})|$ by $|S_{\mathcal{A}}(\frac{j}{q^k})|^r$ in the above argument then we simply replace M by M^r and so

$$\sum_{j=0}^{q^k-1} |S_{\mathcal{A}}(\tfrac{j}{q^k})|^r \ \le c_{M,r}|\lambda_M|^{kr}.$$

By combining such estimates for $r = 1$ and $r = \frac{235}{154}$ Maynard proved that the summands for the "generic" j make a negligible contribution and we can restrict attention to a set $\mathcal{E} \subset \mathfrak{m}$ exceptional integers j with $|\mathcal{E}| \ll X^{.36}$.

1.17. **Extra ideas in Base 10.** When q is small the above calculations are not strong enough to complete the proof. Instead Maynard needed to more thoroughly investigate many aspects of this discussion. To extend the range of bases q, Maynard forwent the asymptotic when $q = 10$ but rather obtained upper and lower bounds that are the expected amount times a positive constant. The upper bound is easy sieve theory, but the lower bound is much more subtle. First off, Maynard deploys delicate sieve methods to (in effect) replace needing to understand how often primes are written with the digits from $\mathcal{D}$ in base q, to understanding when integers divisible by a product of large primes in certain given intervals are represented. This allowed him to, in effect, improve the upper bounds for exponential sums over

[7]We need to change the "$\approx$" in $|S_{\mathcal{A}}(\theta)| \approx \prod_{i=0}^{k-1} F(t_i,t_{i+1})$ above to a precise inequality, like

$$|S_{\mathcal{A}}(\theta)| \le \prod_{i=0}^{k-1} F(t_i,t_{i+1}), \text{ where } F(t,u) := \max_{0\le\eta\le 1/q^2} \left| \frac{e(\frac{u+\eta}{q})-1}{e(\frac{t}{q}+\frac{u+\eta}{q^2})-1} - e(b(\tfrac{t}{q}+\tfrac{u+\eta}{q^2}))\right|$$

primes (as in (1.12)), since now he is working with a more malleable set of the integers, and so he requires bounds like $\lambda_M < q^{\frac{27}{77}}$ rather than $\lambda_M < q^{\frac{1}{5}}$ (as above).

For the $j \in \mathcal{E}$ (the very exceptional minor arcs) to have an important effect on our sum, the fractions j/q^k will have to simultaneously have several surprising Diophantine features, which Maynard proves are mostly incompatible (when $q = 10$). The following diagram exhibits the tools used in the whole proof, but especially when dealing with these exceptional arcs.

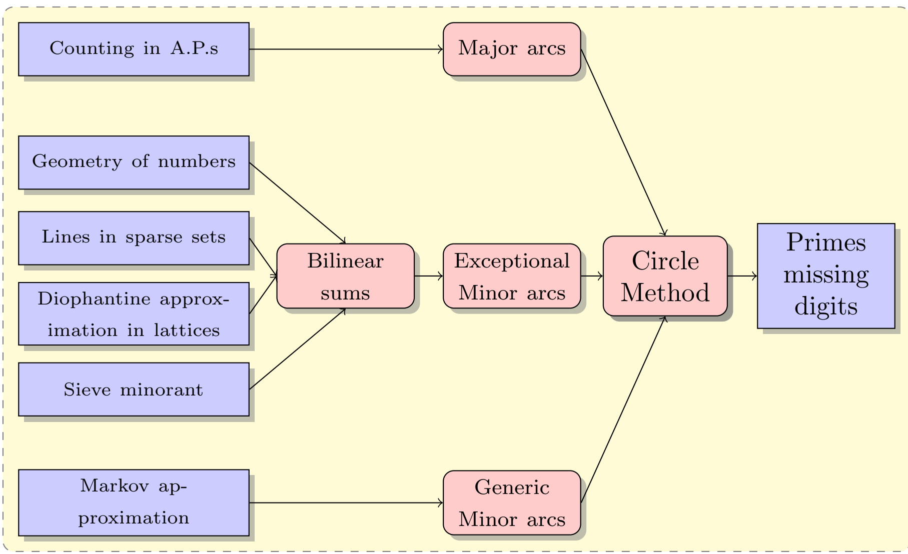


FIGURE 1. Outline of steps to prove primes with missing digits

1.18. **Further reductions of the base.** For the theorems that we have proved here we will try to reduce the base from 62893 to as small as possible, by better using and understanding the transition matrices M (following Karwatowski [20]).

Let $F(s) := |\frac{1}{|\mathcal{D}|}\sum_{n\in\mathcal{D}} e(ns)| \leq 1$ and note that if $\mathcal{D} = \{0, \dots, q-1\} \setminus \{a_0\}$ then

$$F(s) \leq F_1(s) := \frac{1}{q-1}\left(1 + \frac{\sin \pi \|qs\|}{\sin \pi \|s\|}\right). \tag{1.13}$$

Now if $|n| \leq q$ with $|\eta| \leq 1/q$ then $|e(n\eta) - 1| \ll q\eta \leq 1$ and so

$$e(n(\theta + \eta)) = e(n\theta) + (e(n\eta) - 1)e(n\theta) = e(n\theta) + O(q\eta).$$

Therefore $F(\theta + \eta) = F(\theta) + O(q\eta)$.

The general transition matrix M is indexed by J digits in base q and $M_{i,j}$ can only be non-zero if

$$i = (t_1, \dots, t_J), j = (t_2, \dots, t_{J+1}) \text{ for some base-}q\text{ digits } t_1, \dots, t_{J+1}.$$

If $\theta = \sum_{i=1}^{J+1} t_i/q^i$, the entry is $G(t_1, \dots, t_{J+1})^r$ where

$$G(t_1, \dots, t_{J+1}) := \max_{0 \leq \eta \leq 1/q^{J+1}} F(\theta + \eta) = F(\theta) + O(q^{-J}).$$

Now we can take the rth power so that

$$G(t_1,\ldots,t_{J+1})^r = (F(\theta)+O(q^{-J}))^r = F(\theta)^r + \sum_{j=1}^{r}\binom{r}{j}F(\theta)^{r-j}O(q^{-J})^j$$
$$= F(\theta)^r + O\bigg(\sum_{j=1}^{r}\binom{r}{j}\cdot 1\cdot q^{-J}\bigg) = F(\theta)^r + O(2^r q^{-J}).$$

Now we take the sth root, using that $(x+y)^{1/s}\le x^{1/s}+y^{1/s}$ so that

$$G(t_1,\ldots,t_{J+1})^{r/s} \le F(\theta)^{r/s} + O_{r,s}(q^{-J/s}).$$

The largest (real) eigenvalue $\lambda_{r/s,J}$ of the above matrix is bounded by the largest of the column sums; that is,

$$\lambda_{r/s,J} \le \max_{t_2,\ldots,t_{J+1}\in\{0,\ldots,q-1\}} \sum_{t_1=0}^{q-1} G(t_1,\ldots,t_{J+1})^{r/s}$$
$$\le \max_{\substack{t_2,\ldots,t_{J+1}\in\{0,\ldots,,q-1\}\\ \phi=\sum_{i=2}^{J+1}t_i/q^i}} \sum_{t_1=0}^{q-1} F\Big(\frac{t_1}{q}+\phi\Big)^{r/s} + O_{r,s}(q^{1-J/s})$$
$$\le \max_{0\le\phi<1/q}\sum_{n=1}^{q-1}\bigg(\frac{1}{q-1}\bigg(1+\frac{\sin\pi\|q\phi\|}{\sin\pi\|\frac{n}{q}+\phi\|}\bigg)\bigg)^{r/s} + 1 + O_{r,s}(q^{1-J/s}) \tag{1.14}$$

by (1.13), using the bound $|F|\le 1$ on the $n=0$ term. In this range these sine-values are positive, $\sin\pi\|q\phi\|\le 1$, and $\sin\pi\|\frac{n}{q}+\phi\|$ is decreasing with for $n\le\frac{q-1}{2}$ and as ϕ increases, and so by symmetry the above sum is

$$\le 2\sum_{n=1}^{\lfloor\frac{q}{2}\rfloor}\bigg(\frac{1}{q-1}\bigg(1+\frac{1}{\sin\pi\|\frac{n}{q}\|}\bigg)\bigg)^{r/s}$$

These sums can be estimated numerically to deduce that if J is sufficiently large then $\lambda_{1,J} < q^{\frac{27}{77}}$ for $q\ge 102$ and $\lambda_{\frac{235}{154},J} < q^{\frac{59}{433}}$ for $q\ge 174$. If we go back to (1.14) then these inequalities hold for $q\ge 72$ and $q\ge 89$, respectively. One can follow a little more of Maynard's plan [26] than we have done here, and use this to deduce the "primes missing one digit in base-q" result for all bases $q\ge 89$. Karwatowski [20] develops these ideas further and deduces the "primes missing one digit in base-q" result for all bases $q\ge 10$ (whereas Maynard focussed on $q=10$), and he seems to now have extended this to base 9. Now each new smaller base q will almost certainly be susbtantially harder than the last and require significant new ideas.

1.19. **Further results.** The proof that we gave here can be extended (as in [28]) to prove that there are infinitely many such primes if $\mathcal{D}$ contains $\ge q-q^{1/5-o(1)}$ elements, or if $\mathcal{D}$ contains $\ge q^{1-1/5+o(1)}$ consecutive integers. The "$\frac{1}{5}$" was improved to "$\frac{1}{4}$" in [28], and even "$\frac{23}{80}$" if one just wants a lower bound of the correct order of magnitude.

2. Approximating most real numbers

Given any real number θ let m_θ be the integer nearest to θ so that $(\theta) := \theta - m_\theta \in (-\frac{1}{2}, \frac{1}{2}]$. Dirichlet observed that if $\alpha \in [0,1)$ then $0, (\alpha), (2\alpha), \cdots, (N\alpha)$ all belong to an interval of length 1 so two of them $(i\alpha)$ and $(j\alpha)$ must differ by $< \frac{1}{N}$ (by the pigeonhole principle).[8] Now if $n = |j-i|$ then $n \leq N$ and

$$\pm n\alpha = (j-i)\alpha = m_{j\alpha} + (j\alpha) - m_{i\alpha} - (i\alpha) = M + \Delta$$

where $|\Delta| = |(j\alpha) - (i\alpha)| < \frac{1}{N}$ and $M = m_{j\alpha} - m_{i\alpha} \in \mathbb{Z}$ so that $m_{n\alpha} = \pm M$. Writing $m = m_{n\alpha}$ we find that $|n\alpha - m| < \frac{1}{N}$ so that

$$\left|\alpha - \frac{m}{n}\right| < \frac{1}{nN} \leq \frac{1}{n^2}.$$

This is a close approximation to α by rationals, and one wonder whether one can do much better? In general, no, since the continued fraction of the golden ratio $\phi := \frac{1+\sqrt{5}}{2}$ implies that the best approximations to ϕ are given by $F_{n+1}/F_n, n \geq 1$ where F_n is the nth Fibonacci number. One can show that

$$\left|\phi - \frac{F_{n+1}}{F_n}\right| \sim \frac{1}{\sqrt{5}} \cdot \frac{1}{F_n^2},$$

and so all approximations to ϕ by rationals p/q miss by $\geq \{1 + o(1)\}\frac{1}{\sqrt{5}} \cdot \frac{1}{q^2}$.

This led researchers at the end of the 19th century to realize that if the partial quotients in the continued fraction for irrational α are bounded, say by B (note that $\phi = [1,1,1,\dots]$) then there exists a constant $c = c_B > 0$ such that $|\alpha - \frac{m}{n}| \geq \frac{c_B}{n^2}$. However there are very few such α under any reasonable measure. If the partial quotients aren't bounded then how good can approximations be? And how well can one approximate famous irrationals like π? (still a very open question).[9]

An easy argument shows that the set of $\alpha \in [0,1)$ with infinitely many rational approximations $\frac{m}{n}$ for which $|\alpha - \frac{m}{n}| \leq \frac{1}{n^3}$ has measure 0. Indeed if there are infinitely many such rational approximations then there is one with $n > x$ (an integer). Now for each n the measure of $\alpha \in [0,1)$ with $|\alpha - \frac{m}{n}| \leq \frac{1}{n^3}$ is $\frac{1}{n^3}$ for $m = 0$ or n, $\frac{2}{n^3}$ for $1 \leq m \leq n-1$ and 0 otherwise, a total of $\frac{2}{n^2}$, and summing that over all $n > x$ gives $\sum_{n>x} \frac{2}{n^2} < \int_x^\infty \frac{2}{t^2} dt = \frac{2}{x}$. Letting $x \to \infty$ we see that the measure is 0. Obviously the analogous result holds for $|\alpha - \frac{m}{n}| \leq \frac{1}{(n \log n)^2}$, and any other such bounds that lead to convergence of the infinite sum.

More generally we should study, for a given function $\psi : \mathbb{Z}_{\geq 1} \to \mathbb{R}_{\geq 0}$, the set $\mathcal{L}(\psi)$ which contains those $\alpha \in [0,1)$ for which there are infinitely many rationals

---

[8]And by embedding the interval onto the circle by the map $t \to e(t) := e^{2i\pi t}$ we see that they must differ by $< \frac{1}{N+1}$.

[9]If α has continued fraction $[a_0, a_1, \dots]$ and $q^2|\alpha - \frac{b}{q}| < \frac{1}{2}$ then $b/q = b_j/q_j$ a convergent of the continued fraction, and then one can show that $\frac{1}{2} \leq q_j q_{j+1}|\alpha - \frac{b_j}{q_j}| < 1$. Now $q_{j+1} = a_j q_j + q_{j-1}$ and $q_{j-1} < q_j$ so that $a_j q_j \leq q_{j+1} < (a_j+1)q_j$, and therefore the best approximations have $q_j^2|\alpha - \frac{b_j}{q_j}| \asymp 1/a_j$; that is, we get better approximations the large the a_j in the continued fractions (especially in comparison to the q_j). Thus this problem seems resolved except that we do not understand well the continued fractions of most real numbers α, so we have simply transferred the difficulty of the problem into a different domain. See appendix 11B of [13] for more on continued fractions.

m/n for which

$$\left|\alpha - \frac{m}{n}\right| \leq \frac{\psi(n)}{n^2}.$$

We have seen that $\mathcal{L}(1) = [0,1)$ whereas if $c < 1/\sqrt{5}$ then $\phi - 1 \notin \mathcal{L}(c)$ so $\mathcal{L}(c) \neq [0,1)$. Moreover if $\sum_n \psi(n)/n$ is convergent then $\mu(\mathcal{L}(\psi)) = 0$ where $\mu(\cdot)$ is the Lebesgue measure. In each case that we worked out, $\mu(\mathcal{L}(\psi)) = 0$ or 1, and Cassels [3] showed that this is always true (using the Birkhoff Ergodic Theorem)! So we need only decide between these two cases.

The first great theorem in *metric Diophantine approximation* was due to Khinchin who showed that if $\psi(n)$ is a decreasing function then

$$\mu(\mathcal{L}(\psi)) = \begin{cases} 0 \\ 1 \end{cases} \quad \text{if and only if } \sum_{n\geq 1} \frac{\psi(n)}{n} \text{ is } \begin{cases} \text{convergent} \\ \text{divergent} \end{cases}.$$

Thus measure 1 of reals α have approximations $\frac{m}{n}$ with $|\alpha - \frac{m}{n}| \leq \frac{1}{n^2 \log n}$, and measure 0 with $|\alpha - \frac{m}{n}| \leq \frac{1}{n^2 (\log n)^{1+\epsilon}}$

The hypothesis "$\psi(n)$ is decreasing" is too restrictive since, for example, one can't determine anything from this about rational approximations where the denominator is prime. So can we do without it? Our proof above that if $\sum_{n\geq 1} \frac{\psi(n)}{n}$ is convergent then $\mu(\mathcal{L}(\psi)) = 0$, works for general ψ. Indeed we follow the usual proof of the first Borel-Cantelli lemma: Let E_n be the event that $\alpha \in [\frac{m}{n} - \frac{\psi(n)}{n^2}, \frac{m}{n} + \frac{\psi(n)}{n^2}] \cap [0,1]$ for some $m \in \{0, 1, \dots, n\}$, where we have selected α randomly from $[0,1]$, and we established that $\sum_n \mathbb{P}(E_n) = \sum_n \frac{\psi(n)}{n} < \infty$. Then, almost surely, only finitely many of the E_j occur, and so $\mu(\mathcal{L}(\psi)) = 0$.

The second Borel-Cantelli lemma states that if the E_n are independent and $\sum_n \mathbb{P}(E_n)$ diverges then almost surely infinitely many of the E_j occur. Our E_n are far from independent (indeed compare E_n with E_{2n}) but this nonetheless suggests that perhaps with the right notion of independence it is feasible that Khinchin's theorem holds without the decreasing condition.

2.1. Duffin and Schaefer's example.

However Duffin and Schaefer constructed a (complicated) example of ψ for which $\sum_{n\geq 1} \frac{\psi(n)}{n}$ diverges but $\mu(\mathcal{L}(\psi)) = 0$;. Their example uses many representations like $\frac{1}{3} = \frac{2}{6}$, that is, non-reduced fractions:

We begin with ψ_0 where $\psi_0(q) = 0$ unless $q = q_\ell := \prod_{p\leq \ell} p$ is the product of the primes up to some prime ℓ, in which case $\psi_0(q_\ell) = \frac{q_\ell}{\ell \log \ell}$. Therefore

$$\sum_q \frac{\psi_0(q)}{q} = \sum_\ell \frac{1}{\ell \log \ell}$$

which converges by the prime number theorem, and so $\mu(\mathcal{L}(\psi_0)) = 0$ as we just proved in the last subsection.

Now we construct a new ψ for which if q is squarefree integer with largest prime factor ℓ (so that q divides q_ℓ), then $\psi(q) = q^2/(q_\ell \ell \log \ell)$, and $\psi(q) = 0$ otherwise. Now if $|x - \frac{a}{q}| \leq \frac{\psi(q)}{q^2}$ then for $A = a(q_\ell/q)$ we have

$$\left|x - \frac{A}{q_\ell}\right| = \left|x - \frac{a}{q}\right| \leq \frac{\psi(q)}{q^2} = \frac{\psi(q_\ell)}{q_\ell^2} = \frac{\psi_0(q_\ell)}{q_\ell^2}$$

so that $\mathcal{L}(\psi) = \mathcal{L}(\psi_0)$ which has measure 0. On the other hand

$$\sum_q \frac{\psi(q)}{q} = \sum_\ell \frac{1}{\ell \log \ell} \sum_{\ell | q | q_\ell} \frac{q}{q_\ell} = \sum_\ell \frac{1}{\ell \log \ell} \prod_{p<\ell} \left(1 + \frac{1}{p}\right) \gg \sum_\ell \frac{1}{\ell}$$

by Mertens' Theorem, which diverges.

2.2. **A revised conjecture.** Duffin and Schaefer's example uses many representations like $\frac{1}{3} = \frac{2}{6}$, which suggests that we should restrict attention to *reduced fractions* $\frac{m}{n}$ with $(m, n) = 1$. We let E_n^* be the event that $\alpha \in [\frac{m}{n} - \frac{\psi(n)}{n^2}, \frac{m}{n} + \frac{\psi(n)}{n^2}] \cap [0, 1]$ for some $m \in \{0, 1, \dots, n\}$ with $(m, n) = 1$.

Therefore Duffin and Schaefer defined $\mathcal{L}^*(\psi)$ to be those $\alpha \in [0, 1)$ with infinitely many reduced fractions m/n for which

$$\left| \alpha - \frac{m}{n} \right| \leq \frac{\psi(n)}{n^2},$$

and conjectured

$$\mu(\mathcal{L}^*(\psi)) = \begin{cases} 0 \\ 1 \end{cases} \quad \text{if and only if } \sum_{n \geq 1} \frac{\phi(n)}{n} \cdot \frac{\psi(n)}{n} \text{ is } \begin{cases} \text{convergent} \\ \text{divergent} \end{cases}.$$

Here $\phi(n) = \#\{\frac{m}{n} \in [0, 1) : (m, n) = 1\}$. Now if $\sum_n \mathbb{P}(E_n^*) = \sum_n \frac{\phi(n)}{n} \cdot \frac{\psi(n)}{n} < \infty$, then almost surely, only finitely many of the E_j^* occur, and so $\mu(\mathcal{L}^*(\psi)) = 0$. We therefore can assume that $\sum_{n \geq 1} \frac{\phi(n)}{n} \cdot \frac{\psi(n)}{n}$ is divergent.

Gallagher [11] (in a slight variant of Cassell's result [3]) showed that $\mu(\mathcal{L}^*(\psi))$ always equals either 0 or 1. Therefore we only need to show that $\mu(\mathcal{L}^*(\psi)) > 0$ to deduce that $\mu(\mathcal{L}^*(\psi)) = 1$.

Duffin and Schaefer themselves proved the conjecture in the case that there are arbitrarily large Q for which

$$\sum_{q \leq Q} \frac{\phi(q)}{q} \cdot \frac{\psi(q)}{q} \gg \sum_{q \leq Q} \frac{\psi(q)}{q};$$

which more-or-less implies that the main weight of $\psi(q)$ should not be focussed on integers q with many small prime factors (which are extremely rare), since that is what forces

$$\frac{\phi(q)}{q} = \prod_{p|q} \left(1 - \frac{1}{p}\right) \text{ to be small.}$$

Thus for example, the conjecture follows if we only allow prime q (that is, if $\psi(q) = 0$ whenever q is composite), or even only allow integers q which have no prime factors $< \log q$.

In 2021, Koukoulopoulos and Maynard [22] showed that this Duffin-Schaefer conjecture is true, the end of a long saga. The proof is a blend of number theory, probability theory, combinatorics, ergodic theory, and graph theory combined with considerable ingenuity.

2.3. **Probability.** Assuming that $\sum_{n\geq 1}\frac{\phi(n)}{n}\cdot\frac{\psi(n)}{n}$ is divergent, we want to show that almost surely, infinitely many of the E_j^* occur, where E_q^* is the event that α belongs to

$$[0,1)\cap\bigcup_{(a,q)=1}\left[\frac{a}{q}-\frac{\psi(q)}{q^2},\frac{a}{q}+\frac{\psi(q)}{q^2}\right].$$

The E_q^* are not "independent", but were they independent enough, say if

$$\mu(E_q^*\cap E_r^*)=(1+o_{q,r\to\infty}(1))\,\mu(E_q^*)\,\mu(E_r^*),$$

then we could prove our result; however one can easily find counterexamples to this, for example when $r=2q$. However since we only need to show that $\mu(\mathcal{L}^*(\psi))>0$, we will only need to establish a very weak quasi-independence, on average, like

$$\sum_{Q\leq q\neq r<R}\mu(E_q^*\cap E_r^*)\leq 10^6\bigg(\sum_{Q\leq q<R}\mu(E_q^*)\bigg)^2 \tag{2.1}$$

for arbitrarily large Q and certain R: Since $\sum_{q\geq Q}\mu(E_q^*)=2\sum_{q\geq Q}\frac{\phi(q)}{q}\cdot\frac{\psi(q)}{q}$ diverges, we may select $R\geq Q$ for which $1\leq\sum_{Q\leq q<R}\mu(E_q^*)\leq 2$. Now let $N=\sum_{Q\leq q<R}1_{E_q^*}$ so that $\mathbb{E}[N]=\sum_{Q\leq q<R}\mu(E_q^*)$ and so

$$\begin{aligned}1\leq\bigg(\sum_{Q\leq q<R}\mu(E_q^*)\bigg)^2&=\mathbb{E}[N]^2=\mathbb{E}[1_{N>0}\cdot N]^2\leq\mu\bigg(\bigcup_{Q\leq q<R}E_q^*\bigg)\cdot\mathbb{E}[N^2]\\&=\mu\bigg(\bigcup_{Q\leq q<R}E_q^*\bigg)\sum_{Q\leq q,r<R}\mu(E_q^*\cap E_r^*)\end{aligned}$$

by the Cauchy-Schwarz inequality. Therefore

$$\mu\bigg(\bigcup_{q\geq Q}E_q^*\bigg)\geq\mu\bigg(\bigcup_{Q\leq q<R}E_q^*\bigg)\geq 10^{-6}$$

by (2.1). But this is true for arbitrarily large Q and so $\mu(\mathcal{L}^*(\psi))\geq 10^{-6}$, which implies that $\mu(\mathcal{L}^*(\psi))=1$.

Following Pollington and Vaughan [29] we study $\mu(E_q^*\cap E_r^*)$, assuming $(q,r)=1$ for convenience: If $\alpha\in[\frac{a}{q}-\frac{\psi(q)}{q^2},\frac{a}{q}+\frac{\psi(q)}{q^2}]\cap[\frac{b}{r}-\frac{\psi(r)}{r^2},\frac{b}{r}+\frac{\psi(r)}{r^2}]$ with $(a,q)=(b,r)=1$ then $|\frac{a}{q}-\frac{b}{r}|\leq\frac{\psi(q)}{q^2}+\frac{\psi(r)}{r^2}\leq 2\Delta$ where $\Delta:=\max\{\frac{\psi(q)}{q^2},\frac{\psi(r)}{r^2}\}$ and the overlap will have size $\leq 2\delta$ where $\delta:=\min\{\frac{\psi(q)}{q^2},\frac{\psi(r)}{r^2}\}$. Now the $\frac{a}{q}-\frac{b}{r}$ are in 1-to-1 correspondence with the $\frac{n}{qr}$ as n runs through the reduced residue classes mod qr. Therefore, by the small sieve,

$$\begin{aligned}\mu(E_q^*\cap E_r^*)&\leq 2\delta\#\{n:|n|\leq 2\Delta qr\text{ and }(n,qr)=1\}\ll\delta\Delta qr\prod_{\substack{p|qr\\p\leq\Delta qr}}\left(1-\frac{1}{p}\right)\\&\leq\frac{\phi(q)\psi(q)}{q^2}\cdot\frac{\phi(r)\psi(r)}{r^2}\cdot\exp\bigg(\sum_{\substack{p|qr\\p>\Delta qr}}\frac{1}{p}\bigg)\ll\mu(E_q^*)\mu(E_r^*)\exp\bigg(\sum_{\substack{p|qr\\p>\Delta qr}}\frac{1}{p}\bigg).\end{aligned}$$

(If $(q,r)>1$ then we need only alter this by taking $p|qr/(q,r)^2$ instead of $p|qr$ in the sum over p on the far right of the previous displayed equation.)

Using this one can easily deduce the Duffin-Schaefer conjecture provided $\psi(\cdot)$ does not behave too wildly. For example Erdős and Vaaler [7, 30] proved the

Duffin-Schaefer conjecture provided the $\psi(n)$ are bounded. Key to this is to note that there are $\ll e^{-y}x$ integers $n \le x$ for which

$$\sum_{\substack{p|n \\ p>y}} \frac{1}{p} \ge 1.$$

Therefore we obtain good enough bounds on $\mu(E_q^* \cap E_r^*)$ in the previous displayed equation unless (q,r) is large, and unless q and r are each divisible by a lot of different small prime factors. This reduces the problem to one in the *anatomy* of integers (a concept that is brought to life in the graphic novel [14]).

2.4. **The anatomy of integers.** By partitioning $[Q,R]$ into dyadic intervals and studying the contribution of the integers in such intervals to the total we find ourselves drawn towards the following

Model Problem *Fix* $\eta \in (0,1]$. *Suppose that* S *is a set of* $\gg \eta Q/B$ *integers in* $[Q,2Q]$ *for which there are at least* $\eta|S|^2$ *pairs* $q,r \in S$ *such that* $(q,r) \ge B$. *Must there be an integer* $g \ge B$ *which divides* $\gg_\eta Q/B$ *elements of* S*?*

The model problem is false but a technical variant, which takes account of the $\phi(q)/q$-weights, is true.[10] Using this one can reduce the problem to the Erdős-Vaaler argument, by anatomy of integers arguments, and prove the theorem.

To attack the (variant of the) Model Problem, Koukoulopoulos and Maynard view it as a question in graph theory:

2.5. **Graph Theory.** Consider the graph G, with vertex set S and edges between vertices representing pairs of integers with gcd$> B$.

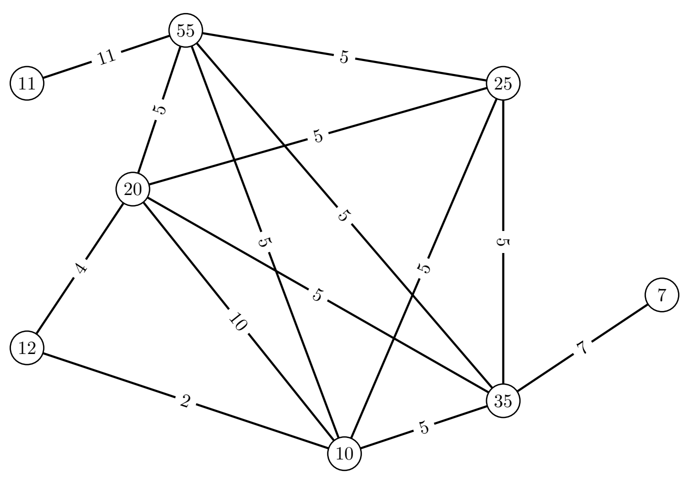


FIGURE 2. Vertices = The integers in our set.
. Edges = Pairs of integers with a large GCD.

[10]Let $Q = \prod_{p \le 2y} p$ and $S := \{Q/p : y < p \le 2y\}$. If $q = Q/p, r = Q/\ell \in S$ then $(q,r) = Q/p\ell \ge B := Q/4y^2$, but any integer $\ge B$ divides no more than two elements of S. (This is adapted from an idea of Sam Chow.)

Beginning with such a graph for which the edge density is η, we wish to prove that there is a "dense subgraph" H whose vertices are each divisible by a fixed integer $\geq B$. To locate this structured subgraph H, Koukoulopoulos and Maynard use an iterative "compression" argument, inspired by the papers of Erdös-Ko-Rado [6] and Dyson [5]: with each iteration, they pass to a smaller graph but with more information about which primes divide the vertices. This is all complicated by the weights $\phi(q)/q$. The details are complicated (see a vague sketch in the next subsection); and the reader is referred to [21], where the original proof of [22] is better understood from more recent explorations of Green and Walker [15], who gave an elegant proof of the following important variant:

If $R \subset [X, 2X]$ and $S \subset [Y, 2Y]$ are sets of integers for which $(r, s) \geq B$ for at least $\delta|R||S|$ pairs $(r, s) \in R \times S$ then $|R||S| \ll_\epsilon \delta^{-2-\epsilon} XY/B^2$.

Although this has a slightly different focus from the model problem, it focuses on the key question of how large such sets can get and takes account of the example of footnote 8 (unlike the model problem).

2.6. **Iteration and graph weights.** The key to any such iteration argument is to develop a measure of how close one is getting to the goal, which can require substantial ingenuity. In their paper Koukoulopoulos and Maynard [22] begin with two copies of S and construct a bipartite graph $V_0 \times W_0$ with edges in-between $q \in V_0 = S$ and $r \in W_0 = S$ if $(q, r) \geq B$. The idea is to select distinct primes $p_1, p_2, \dots$ and then $V_j = \{v \in V_{j-1} : p_j \text{ divides } v\}$ or $V_j = \{v \in V_{j-1} : p_j \text{ does not divide } v\}$, and similarly W_j, so that p_j divides all $(v_j, w_j), v_j \in V_j, w_j \in W_j$ or none. If we terminate at step J then there are integers a_J, b_J, constructed out of the p_j, such that a_J divides every element of V_J and b_J divides every element of W_J. The goal is to proceed so that $(v_J, w_J) \geq B$ for some J, for all $v_J \in V_J, w_J \in W_J$ such that all of the prime divisors of any (v_J, w_J) appears amongst the p_j. Hence, if say all the integers in S are squarefree, then $(a_J, b_J) = (v_J, w_J) \geq B$. So how do we measure progress in this algorithm?

One key measure is δ_j, the proportion of pairs $v_j \in V_j, w_j \in W_j$ with $(v_j, w_j) \geq B$, another the size of the sets V_j and W_j. Finally we want to measure how much of the $a_j b_j$ are given by prime divisors not dividing (a_j, b_j), so we use $a_j b_j/(a_j, b_j)^2$. Koukoulopoulos and Maynard [22] found, after some trial and error, that the measure

$$\delta_j^{10} \cdot |V_j| \cdot |W_j| \cdot \frac{a_j b_j}{(a_j, b_j)^2}$$

fits their needs, allowing them eventually to restrict their attention to $v, w \in S$ for which a_j divides v, b_J divides w and

$$\sum_{\substack{p | vw/(v,w)^2 \\ p > y}} \frac{1}{p} \approx 1.$$

Koukoulopoulos and Maynard then finish the proof by applying a relative version of the Erdös-Vaaler argument to the pairs $(v/a_J, w/b_J)$.

2.7. **Hausdorff dimension.** If $\sum_{n \geq 1} \phi(n) \cdot (\psi(n)/n^2)$ is convergent then $\mu(\mathcal{L}^*(\psi)) = 0$ so we would like to get some idea of the true size of $\mathcal{L}^*(\psi)$. Using a result of Beresnevich and Velani [2], one can deduce that the Hausdorff dimension of $\mathcal{L}^*(\psi)$

is given by the infimum of the real $\beta > 0$ for which

$$\sum_{n \geq 1} \phi(n) \cdot \left(\min \left\{ \frac{\psi(n)}{n^2}, \frac{1}{2} \right\} \right)^{\beta} \text{ is convergent.}$$

References

1. Roger Baker, Glynn Harman and Janos Pintz, *The difference between consecutive primes, II*, Proc. Lond. Math. Soc. 83 (2001), 532–562
2. V. Beresnevich and S. Velani, *A mass transference principle and the Duffin-Schaeffer conjecture for Hausdorff measures*, Ann. of Math. (2) 164 no. 3 (2006), 971–992.
3. J.W.S. Cassels, *Some metrical theorems in Diophantine approximation* I. Proc. Cambridge Philos. Soc. 46 (1950), 209–218.
4. R. J. Duffin and A.C. Schaeffer, *Khintchine's problem in metric Diophantine approximation*, Duke Math. J. 8 (1941), 243–255.
5. F. J. Dyson, *A theorem on the densities of sets of integers*. J. London Math. Soc. 20 (1945), 8–14. [
6. P. Erdös, C. Ko, R. Rado, *Intersection theorems for systems of finite sets*, Quart. J. Math. Oxford Ser. (2) 12 (1961), 313–320.
7. P. Erdös, *On the distribution of the convergents of almost all real numbers*, J. Number Theory 2 (70), 425–441
8. Kevin Ford, Ben Green, Sergei Konyagin, and Terence Tao, *Large gaps between consecutive prime numbers*, Ann. of Math. 183 (2016), 935–974.
9. Kevin Ford, Ben Green, Sergei Konyagin, James Maynard and Terence Tao, *Long gaps between primes*, J. Amer. Math. Soc. 31 (2018), no. 1, 65–105.
10. J. Friedlander and H. Iwaniec, *The polynomial X2+Y4 captures its primes*, Ann. of Math. (2) 148(3) (1998), 945–1040.
11. P.X. Gallagher, *Approximation by reduced fractions*, J. Math. Soc. Japan 13 (1961), 342?345.
12. Andrew Granville, *Primes in intervals of bounded length*, Bulletin of the American Mathematical Society, 52 (2015), 171–222.
13. Andrew Granville, *Number Theory Revealed: A Masterclass*, American Mathematical Society, Providence, RI, 2019.
14. Andrew Granville, Jennifer Granville and Robert J. Lewis, *Prime Suspects: The anatomy of integers and permutations*, Princeton University Press, Princeton, NJ, 2019.
15. B. Green, A. Walker, *Extremal problems for GCDs*, arXiv:2012.02078.
16. D. R. Heath-Brown, *Primes represented by* $x^3 + 2y^3$, Acta Math. 186(1) (2001), 1–84.
17. D. R. Heath-Brown and X. Li, *Prime values of* $a^2 + p^4$ Invent. Math. 208(2) (2017), 441–499.
18. D. R. Heath-Brown and B. Z. Moroz, *Primes represented by binary cubic forms*, Proc. Lond. Math. Soc. (3) 84(2) (2002), 257–288.
19. D. R. Heath-Brown and B. Z. Moroz *On the representation of primes by cubic polynomials in two variables*, Proc. Lond. Math. Soc. (3) 88(2) (2004), 289–312.
20. Fabian Karwatowski, *Primes with one excluded digit*, Acta Arithmetica **202** (2022), 105–121
21. Dimitris Koukoulopoulos, *Rational approximations of irrational numbers*, to appear in the ICM 2022 proceedings.
22. Dimitris Koukoulopoulos and James Maynard, *On the Duffin-Schaeffer conjecture*, Ann. of Math. **192** (2020), 251–307.
23. Kaisa Matomäki, *The distribution of* αp *modulo one* Math. Proc. Camb. Phil. Soc. 147 (2009), 267–283
24. James Maynard, *Small gaps between primes*, Ann. of Math. **181** (2015), 1–31.
25. James Maynard, *Large gaps between primes*, Ann. of Math. **183** (2016), 915–933.
26. James Maynard, *Primes with restricted digits*, Invent. Math. **217** (2019), 127–218.
27. James Maynard, *Primes represented by incomplete norm forms* Forum Math. Pi 8 (2020), e3, 128 pp.
28. James Maynard, *Primes and polynomials with restricted digits*, Int. Math. Res. Not. IMRN 14 (2022), 1–23
29. A. D. Pollington and R. C. Vaughan, *The k-dimensional Duffin and Schaeffer conjecture*, Mathematika 37 no. 2 (1990), 190–200.

30. J. D. Vaaler, *On the metric theory of Diophantine approximation*, Pacific J. Math. 76 no. 2 (1978), 527–539.
31. Triantafyllos Xylouris, *Über die Nullstellen der Dirichletschen L-Funktionen und die kleinste Primzahl in einer arithmetischen Progression*. PhD thesis, Universität Bonn, 2011.
32. Yitang Zhang, *Bounded gaps between primes*, Ann. of Math.. 179 (2014), 1121–1174.

Départment de Mathématiques et Statistique, Université de Montréal, CP 6128 succ Centre-Ville, Montréal, QC H3C 3J7, Canada.
Email address: andrew.granville@umontreal.ca

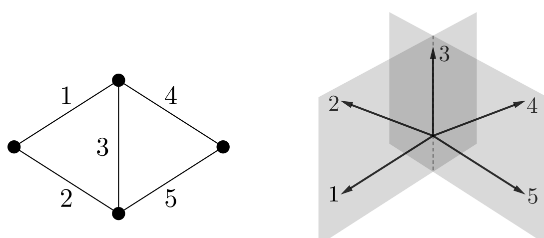
AN ESSENCE OF INDEPENDENCE: RECENT WORKS OF JUNE HUH ON COMBINATORICS AND HODGE THEORY

CHRISTOPHER EUR

ABSTRACT. Matroids are combinatorial abstractions of independence, a ubiquitous notion that pervades many branches of mathematics. June Huh and his collaborators recently made spectacular breakthroughs by developing a Hodge theory of matroids that resolved several long-standing conjectures in matroid theory. We survey the main results in this development and ideas behind them.

1. INTRODUCTION

The notion of "independence" resides everywhere, for example in graphs, vector configurations, field extensions, hyperplane arrangements, matchings, and discrete optimizations. Matroid theory captures the combinatorial essence of "independence" shared in these structures. For example, let us consider the following graph G with edges labelled $\{1, \ldots, 5\}$ and the set of vectors $\{v_1, \ldots, v_5\}$.



$$\begin{array}{ccccc} v_1 & v_2 & v_3 & v_4 & v_5 \end{array}$$
$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

FIGURE 1.

We observe a common combinatorial structure: A subset of edges in G is acyclic if and only if the corresponding subset of vectors is linearly independent. This combinatorial structure is encoded as a matroid, introduced by Whitney [Whi32].

Definition 1.1. A **matroid** $\mathrm{M} = (E, \mathcal{I})$ consists of a finite set $E = \{1, \ldots, n\}$, called its **ground set**, and a nonempty collection $\mathcal{I}$ of subsets of E, called the **independent sets** of M, such that

if $I \in \mathcal{I}$ and $J \subseteq I$, then $J \in \mathcal{I}$, and
if $I, J \in \mathcal{I}$ and $|I| < |J|$, then there exists an element $j \in J \setminus I$ such that $I \cup \{j\} \in \mathcal{I}$.

The definition implies that every maximal independent set of M has the same cardinality r, which we call the **rank** of M.

2020 *Mathematics Subject Classification.* 05B35, 05E14, 14F43, 14C17.

Graphs and vector spaces give prototypical examples of matroids:

- When E is identified with the set of edges of a finite graph G, setting
$$\mathcal{I} = \{I \subseteq E : \text{the subset } I \text{ of edges in } G \text{ is acyclic}\}$$
defines a matroid $\mathrm{M} = (E, \mathcal{I})$. Matroids arising in this way are called **graphical matroids**.
- When E is identified with a finite set of vectors spanning a vector space V, setting
$$\mathcal{I} = \{I \subseteq E : \text{the subset } I \text{ of vectors in } V \text{ is linearly independent}\}$$
defines a matroid $\mathrm{M} = (E, \mathcal{I})$. Matroids arising in this way are called **realizable matroids**.

We see that the graph G and the set of vectors in Figure 1 define the same matroid.

1.1. **Combinatorial sequences from a matroid.** Several long-standing conjectures in matroid theory, recently resolved by June Huh and his collaborators, concern the behavior of sequences of invariants of a matroid. For a sequence $(a_0, a_1, \ldots, a_m)$ of nonnegative real numbers, we say that it

- is *unimodal* if there exists $0 \leq k \leq m$ such that
$$a_0 \leq a_1 \leq \cdots \leq a_k \geq a_{k+1} \geq \cdots \geq a_m,$$
- is *log-concave* if $a_i^2 \geq a_{i-1}a_{i+1}$ for all $1 \leq i \leq m-1$,
- has *no internal zeros* if $a_i a_j \neq 0$ implies $a_k \neq 0$ for all $0 \leq i \leq k \leq j \leq 0$, and
- is *top-heavy* if $a_i \leq a_{d-i}$ for all $0 \leq i \leq \frac{d}{2}$ where d is the largest index such that $a_d \neq 0$.

Note that a log-concave sequence is unimodal if and only if it has no internal zeroes. For a survey of unimodality and log-concavity in combinatorics, see [Sta89, Bre94].

We consider the following sequences of invariants of a rank r matroid M. For some of them, we describe them only for a graphical or a realizable matroid, postponing their descriptions for arbitrary matroids to Section 2.

(a) For $0 \leq i \leq r$, let I_i be the number of independent sets of M of cardinality i. In other words, the sequence $(I_0, \ldots, I_r)$ is the f-vector of the simplicial complex whose faces are the independent sets of M.

(b) We may also consider the h-vector. That is, for $0 \leq i \leq r$, let I_i' be defined by the identity $\sum_{i=0}^r I_i' q^{r-i} = \sum_{i=0}^r I_i (q-1)^{r-i}$.

(c) Suppose M is the graphical matroid of a finite connected nontrivial graph G. The **chromatic polynomial** $\chi_G(q)$ of G is defined as
$$\chi_G(q) = \text{the number of proper colorings of } G \text{ with at most } q \text{ colors},$$
where a coloring of the vertices is proper if no two vertices of an edge share the same color. It is polynomial in q of degree $r+1$, and is divisible by $q(q-1)$. Let $(w_0, \ldots, w_r)$ be the absolute values of the coefficients of $\frac{1}{q}\chi_G(q)$, starting from the highest degree term.

(d) Continuing the assumption that M is the graphical matroid of G, we define $(w'_0, \ldots, w'_{r-1})$ as the absolute values of the coefficients of $\frac{1}{q(q+1)}\chi_G(q+1)$.

(e) Suppose M is the realizable matroid of a set of vectors $\{v_1, \ldots, v_n\}$ spanning a vector space V. Let $(W_0, \ldots, W_r)$ be a sequence defined by setting for each $0 \leq i \leq r$,

$$W_i = \text{the number of } i\text{-dimensional linear subspaces } V' \text{ in } V \text{ such that } V' \text{ is the span of a subset of the vectors } \{v_1, \ldots, v_n\}.$$

We leave it as an exercise to check that for the matroid associated to the graph or the vector configuration in Figure 1, we have:

$$(I_0, I_1, I_2, I_3) = (1, 5, 10, 8), \qquad (w_0, w_1, w_2, w_3) = (1, 5, 8, 4),$$
$$(I'_0, I'_1, I'_2, I'_3) = (1, 2, 3, 2), \qquad (w'_0, w'_1, w'_2) = (1, 2, 1),$$
$$(W_0, W_1, W_2, W_3) = (1, 5, 6, 1).$$

Notice in this example that every sequence is unimodal, log-concave, and top-heavy. Several conjectures from the 70's posited that these sequences are unimodal, log-concave, or top-heavy for an arbitrary matroid. We describe these conjectures and their history more fully in Section 2.2.

1.2. **An approach from algebraic geometry.** After decades of little progress, a breakthrough happened when many of these conjectures were resolved for realizable matroids using algebraic geometry: Huh and Katz [Huh12, HK12] showed that the sequence (c) is log-concave (with no internal zeros), Huh [Huh15] showed that (d) is log-concave (with no internal zeros), and Huh and Wang [HW17] showed that (e) is top-heavy. These developments were particularly significant in light of the following phenomena in matroid theory:

The geometry of realizable matroids often inspires purely combinatorial constructions for all matroids. Certain geometric properties, a priori applicable only to realizable matroids, persist to all matroids through these purely combinatorial constructions. This is surprising because almost all matroids are not realizable [Nel18], but such a creative tension between geometry and combinatorics is a recurring theme in matroid theory.

A recent spectacular example of this phenomenon is the development of the Hodge theory of matroids by June Huh and his collaborators [AHK18, ADH22, BHM+22, BHM+], which successfully resolved conjectures about log-concavity or top-heaviness of the sequences (a),...,(e) for arbitrary (not necessarily realizable) matroids. They established that matroids satisfy combinatorial analogues of certain Hodge-theoretic properties in algebraic geometry, known sometimes as the "Kähler package":

Definition 1.2. Let $A^\bullet = \bigoplus_{i=0}^{d} A^i$ be a finite-dimensional graded real vector space with a symmetric bilinear form $P : A^\bullet \times A^{d-\bullet} \to \mathbb{R}$, and let $\mathcal{K}$ be a convex subset of graded linear operators $L : A^\bullet \to A^{\bullet+1}$ satisfying $P(Lx, y) = P(x, Ly)$ for all $x, y \in A^\bullet$. The triple $(A^\bullet, P, \mathcal{K})$ is said to satisfy the **Kähler package** if the following three properties hold for all nonnegative integers $i \leq \frac{d}{2}$:

(PD) The pairing $P : A^i \times A^{d-i} \to \mathbb{R}$ is non-degenerate (Poincaré duality).

(HL) For any $L_1, \dots, L_{d-2i} \in \mathcal{K}$, the linear map

$$A^i \to A^{d-i} \quad \text{given by} \quad x \mapsto L_1 \cdots L_{d-2i} x$$

is an isomorphism (hard Lefschetz property in degree i).

(HR) For any $L_0, L_1, \dots, L_{d-2i} \in \mathcal{K}$, the symmetric bilinear pairing

$$A^i \times A^i \to \mathbb{R} \quad \text{given by} \quad (x, y) \mapsto (-1)^i P(x, L_1 \cdots L_{d-2i} y)$$

is positive definite when restricted to the kernel of the map $A^i \to A^{d-i+1}$ given by $x \mapsto L_0 L_1 \cdots L_{d-2i} x$ (Hodge-Riemann relations in degree i).

Classical Hodge theory tells us that these properties are satisfied when $A^\bullet$ is the cohomology ring of real (p,p)-forms on a complex projective manifold, P is the Poincaré duality pairing, and $\mathcal{K}$ consists of multiplication by ample divisor classes (see [Huy05] or [Voi02]).

The geometry behind realizable matroids led to purely combinatorial constructions for various "cohomologies" of a matroid. These constructions include the Chow ring of a matroid [FY04, AHK18], the conormal Chow ring of a matroid [ADH22], and the intersection cohomology of a matroid [BHM$^+$]. For a matroid realizable over $\mathbb{C}$, all three satisfy the Kähler package due to classical algebraic geometry. The incredible result of June Huh and his collaborators—Karim Adiprasito, Federico Ardila, Tom Braden, Graham Denham, Eric Katz, Jacob Matherne, Nick Proudfoot, Botong Wang—is that the Kähler package continues to hold for these "cohomologies" of a matroid even when the matroid is not realizable.

We survey this remarkable development in matroid theory and its connection to algebraic geometry in four parts. In Section 2, we give a brief introduction to matroids, and describe the long-standing conjectures resolved by the Hodge theory of matroids. In Section 3, we explain how the conjectures in the case of realizable matroids were resolved using algebraic geometry. In Section 4, we discuss the Kähler package for Chow rings of fans and matroids, and how the validity of (HR) implies the conjectures about log-concavity. In Section 5, we discuss the intersection cohomology of a matroid, and explain its implication to top-heaviness.

Several interesting topics had to be omitted, even though they are closely related to the topics discussed here. A partial list includes the following:

- The Kazhdan-Lusztig theory of matroids [EPW16], which was an inspiration behind the construction of intersection cohomology of a matroid. We point to [Pro18] for a survey of Kazhdan-Lusztig-Stanley polynomials in a more general context.
- The study of matroids through the polyhedral geometry of their base polytopes, their subdivisions, and the geometry of the Grassmannian [GGMS87, Laf03]. We point to the survey [Ard22, Section 4] and references therein.
- Other approaches to the "cohomology" of a matroid in the broader context of tropical geometry, for instance [IKMZ19, AP].

We hope that this survey will spark the reader's general interest in this active field of the study of matroids from an algebro-geometric perspective.

Notation. Throughout, let $E = \{1, \dots, n\}$ be a finite set of cardinality n. For a subset $S \subseteq E$, we denote by $\mathbf{e}_S = \sum_{i\in S} \mathbf{e}_i$ the sum of standard basis vectors in $\Bbbk^E$, where the field $\Bbbk$ will be clear in context. An algebraic variety is reduced and irreducible (over an algebraically closed field).

Acknowledgements. The author thanks June Huh for helpful conversations, and Matt Baker, Andrew Berget, Tom Braden, Eric Katz, and Matt Larson for helpful suggestions and comments on a preliminary draft of this article.

2. Background in matroid theory

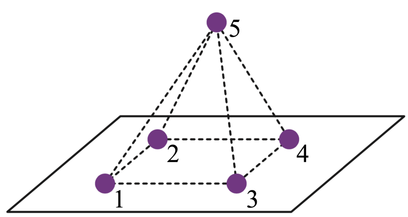
Here we give a minimal introduction to matroids. In addition to standard references on matroids such as [Wel76, Oxl11], we point to [Ard22, Bak18, Huh18, Kat16] for surveys tailored towards studying matroids from an algebro-geometric viewpoint.

2.1. **Constructions.** Since subsets of independent sets are independent, we may specify a matroid by its maximal independent sets, called the **bases** of the matroid.

Example 2.1. For an integer $0 \le r \le n$, the **uniform matroid** of rank r on E is the matroid $\mathrm{U}_{r,n}$ whose bases are all subsets of cardinality r. When $n = r$, we say that $\mathrm{U}_{n,n}$ is the **Boolean matroid** on E. The Boolean matroid $\mathrm{U}_{0,0}$, i.e. when $E = \varnothing$ so $n = r = 0$, is called the **trivial matroid**. Any uniform matroid $\mathrm{U}_{r,n}$ is realizable over any infinite field $\Bbbk$ as a general collection of n vectors in $V = \Bbbk^r$.

Example 2.2. We may visualize a collection of vectors in a 4-dimensional vector space as a collection of points in the projective 3-space $\mathbb{P}^3$. For example, the 5 column vectors of the matrix

$$\begin{bmatrix} 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$



can be visualized as the purple points, four of which lie in a common projective plane. The bases of this matroid are $\{1235, 1245, 1345, 2345\}$.

We define the **dual matroid** $\mathrm{M}^\perp$ of a matroid M on ground set E by declaring

$$\text{the set of bases of } \mathrm{M}^\perp = \{E \setminus B : B \text{ a basis of } \mathrm{M}\}.$$

For example, we have $\mathrm{U}_{r,n}^\perp = \mathrm{U}_{n-r,n}$. For the matroid M in Example 2.2, its dual $\mathrm{M}^\perp$ has the set of bases $\{1, 2, 3, 4\}$. Many notions in matroid theory come in pairs via matroid duality. For instance, an element $e \in E$ is a **loop** of M if it is in no bases, and is a **coloop** if it is in every basis of M. The matroid in Example 2.2 has no loops and has a coloop 5, or equivalently, its dual matroid has no coloops and has a loop 5.

Another useful way of describing a matroid is by its rank function. For a matroid $\mathrm{M} = (E, \mathcal{I})$, its **rank function** $\mathrm{rk}_{\mathrm{M}} : 2^E \to \mathbb{Z}$ is defined by

$$\mathrm{rk}_{\mathrm{M}}(S) = \max\{|I| : I \in \mathcal{I} \text{ and } I \subseteq S\} \quad \text{for every subset } S \subseteq E.$$

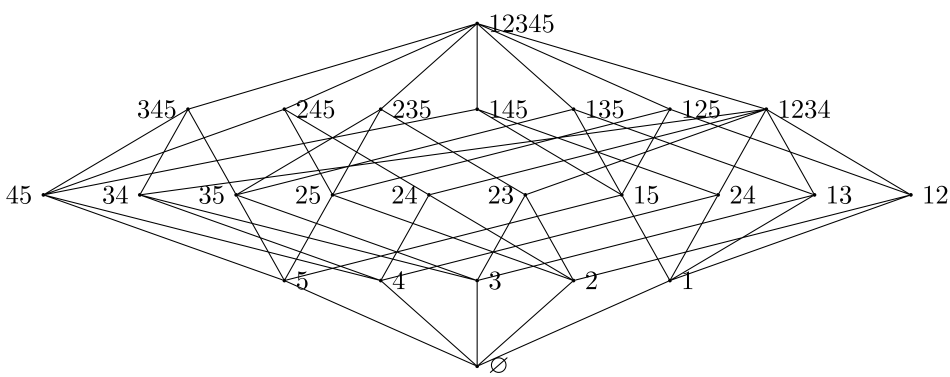
In particular, an independent set of a matroid M is a subset $I \subseteq E$ whose rank $\mathrm{rk}_{\mathrm{M}}(I)$ equals its cardinality $|I|$. That is, independent sets are the minimal subsets of E with respect to a given rank. Considering the maximal subsets leads to the notion of flats of a matroid.

Definition 2.3. A **flat** of a matroid M on E is a subset of E that is maximal for its rank. That is, a subset $F \subseteq E$ is a flat of M if $\mathrm{rk}_{\mathrm{M}}(F \cup \{e\}) > \mathrm{rk}_{\mathrm{M}}(F)$ for all $e \in E \setminus F$.

The set of flats of a matroid M forms a poset under inclusion. This poset is a *lattice* with meet and join defined by

$$F \wedge F' = F \cap F' \quad \text{and} \quad F \vee F' = \text{the smallest flat containing } F \cup F'.$$

For example, the lattice of flats of the matroid in Example 2.2 is



Exercise 2.4.

(1) Show that every subset is contained in a unique flat of the same rank. In particular, the join of flats is well-defined.
(2) How can one recover the bases of a matroid from the lattice of its flats?
(3) Compute the lattice of flats of the matroid of the graph in Figure 1.

We record some linear algebraic interpretations of the notions introduced for a matroid M realized by a set of vectors $\{v_1, \ldots, v_n\}$ spanning a vector space V:

- The bases of M are subsets $B \subseteq E$ such that $\{v_i : i \in B\}$ is a basis of V.
- An element $e \in E$ is a loop if and only if $v_e = 0$.
- We have $\mathrm{rk}_{\mathrm{M}}(S) = \dim \mathrm{span}\{v_i : i \in S\}$ for any subset $S \subseteq E$.
- The flats of M are subsets $F \subseteq E$ such that $F = V' \cap \{v_1, \ldots, v_n\}$ for some linear subspace $V' \subseteq V$. That is, the flats correspond to the different spans of subsets of the vectors.

Per the last bullet point, we now define the sequence (e) for arbitrary matroids.

Definition 2.5. For a matroid M of rank r, the **Whitney numbers of the second kind** $(W_0, W_1, \ldots, W_r)$ are defined by

$$W_i = \text{the number of flats of } \mathrm{M} \text{ with rank } i.$$

Exercise 2.6. Show that if M is the graphical matroid of the complete graph on N vertices, then its flats of rank i correspond to partitions of N into $N-i$ (nonempty) parts. In particular, the numbers $(W_0, W_1, \ldots, W_{N-1})$ in this case are known as the Stirling numbers of the second kind.

Matroid duality also admits a linear algebraic interpretation, given in the next exercise.

Exercise 2.7. Let $\Bbbk^E \twoheadrightarrow V$ be the map given by $\mathbf{e}_i \mapsto v_i$, and let K be its kernel. The short exact sequence $0 \to K \to \Bbbk^E \to V \to 0$ dualizes to $0 \to V^\vee \to \Bbbk^E \to K^\vee \to 0$. Show that the surjection $\Bbbk^E \twoheadrightarrow K^\vee$ realizes the dual matroid $\mathrm{M}^\perp$.

An important pair of operations for forming new matroids from a given matroid is the restriction and contraction.

Definition 2.8. For a matroid M on ground set E, and a subset $A \subseteq E$, we define two matroids $\mathrm{M}|A$ and M/A on ground sets A and $E \setminus A$, respectively, by specifying their rank functions:

$$\begin{aligned} \mathrm{rk}_{\mathrm{M}|A}(S) &= \mathrm{rk}_{\mathrm{M}}(S) && \text{for all } S \subseteq A. \\ \mathrm{rk}_{\mathrm{M}/A}(S) &= \mathrm{rk}_{\mathrm{M}}(S \cup A) - \mathrm{rk}_{\mathrm{M}}(A) && \text{for all } S \subseteq E \setminus A. \end{aligned}$$

The matroid $\mathrm{M}|A$ is called the **restriction** of M to A, and the matroid M/A is called the **contraction** of M by A. The **deletion** $\mathrm{M} \setminus A$ of M by A is the restriction $\mathrm{M}|(E \setminus A)$.

These operations behave particularly well for a flat F of a matroid M:

- The set of flats of $\mathrm{M}|F$ is $\{F' : F' \text{ a flat of } \mathrm{M} \text{ contained in } F\}$.
- The set of flats of M/F is $\{F' \setminus F : F' \text{ a flat of } \mathrm{M} \text{ containing } F\}$.

These operations have a graphical and linear algebraic interpretations as well: For the graphical matroid of a graph G, the deletion corresponds to deleting the corresponding edges, and the contraction corresponds to contracting the corresponding edges. When a matroid M is realized by a set of vectors $\{v_i : i \in E\}$, the restriction $\mathrm{M}|A$ is realized by the subset of vectors $\{v_i : i \in A\}$. The contraction M/A is realized by the images of the vectors $\{v_i : i \in E \setminus A\}$ under the quotient by the span of $\{v_i : i \in A\}$.

Exercise 2.9. Show that deletion and contraction are dual notions, that is, we have $(\mathrm{M} \setminus A)^\perp = \mathrm{M}^\perp / A$.

2.2. **Invariants.** Introduced for graphs by Tutte [Tut67] and extended to matroids by Crapo [Cra69], the Tutte polynomial is among the most famous invariants of a matroid. For the proofs of the statements here, as well as a fuller treatment of Tutte polynomials, see [BO92].

Definition 2.10. Tutte polynomial of a matroid M of rank r on ground set E is a bivariate polynomial defined as

$$\mathrm{T}_{\mathrm{M}}(x,y) = \sum_{S\subseteq E} (x-1)^{r-\mathrm{rk}_{\mathrm{M}}(S)}(y-1)^{|S|-\mathrm{rk}_{\mathrm{M}}(S)}.$$

The Tutte polynomial is the universal deletion-contraction invariant:

Theorem 2.11. The Tutte polynomial can be defined recursively by

$$\mathrm{T}_{\mathrm{M}}(x,y) = \begin{cases} x\mathrm{T}_{\mathrm{M}/e}(x,y) & \text{if } e\in E \text{ a coloop in } \mathrm{M} \\ y\mathrm{T}_{\mathrm{M}\setminus e}(x,y) & \text{if } e\in E \text{ a loop in } \mathrm{M} \\ \mathrm{T}_{\mathrm{M}\setminus e}(x,y) + \mathrm{T}_{\mathrm{M}/e}(x,y) & \text{if } e\in E \text{ neither loop nor coloop} \end{cases}$$

with $\mathrm{T}_{\mathrm{U}_{0,0}}(x,y) = 1$. If f is an invariant of matroids with values in a (commutative unital) ring R such that $f(\mathrm{U}_{0,0}) = 1$ and there exists $x_0, y_0, a, b \in R$ satisfying

$$f(\mathrm{M}) = \begin{cases} x_0 f(\mathrm{M}/e) & \text{if } e\in E \text{ a coloop in } \mathrm{M} \\ y_0 f(\mathrm{M}\setminus e) & \text{if } e\in E \text{ a loop in } \mathrm{M} \\ af(\mathrm{M}\setminus e) + bf\mathrm{M}/e) & \text{if } e\in E \text{ neither loop nor coloop} \end{cases}$$

for all matroids M and an element e. Then, we have

$$f(\mathrm{M}) = a^{|E|-\mathrm{rk}_{\mathrm{M}}(E)} b^{\mathrm{rk}_{\mathrm{M}}(E)} \mathrm{T}_{\mathrm{M}}\left(\frac{x_0}{b}, \frac{y_0}{a}\right).$$

The theorem implies the following basic properties of the Tutte polynomial:

- The Tutte polynomial $T_{\mathrm{M}}(x,y)$ of a matroid M has nonnegative coefficients.
- The constant term of $T_{\mathrm{M}}(x,y)$ is zero unless M is the trivial matroid.
- For the dual matroid $\mathrm{M}^{\perp}$, we have $T_{\mathrm{M}^{\perp}}(x,y) = T_{\mathrm{M}}(y,x)$.

Univariate specializations of the Tutte polynomial leads to many interesting combinatorial sequences. For example, we deduce from the definition that for a matroid M of rank r,

$$T_{\mathrm{M}}(q+1,1) = \sum_{i=0}^{r} I_i q^{r-i},$$

where we recall from (a) that I_i is the number of independent sets of M of cardinality i. Consequently, we have that the sequence $(I'_0,\dots,I'_r)$ of (b) is given by

$$T_{\mathrm{M}}(q,1) = \sum_{i=0}^{r} I'_i q^{r-i}.$$

Another important specialization is the **characteristic polynomial** χ_{M} of a matroid M of rank r, defined as

$$\chi_{\mathrm{M}}(q) = (-1)^r T_{\mathrm{M}}(1-q,0).$$

The basic properties of the Tutte polynomial listed above imply that the coefficients of $\chi_{\mathrm{M}}(q)$ have alternating signs, and that $\chi_{\mathrm{M}}(q)$ is divisible by $(q-1)$ unless M is a trivial matroid. Thus, one often divides out $(q-1)$ to define the **reduced characteristic polynomial** $\overline{\chi}_{\mathrm{M}}(q) = \chi_{\mathrm{M}}(q)/(q-1)$. It follows from Theorem 2.11 that $\chi_{\mathrm{M}} = 0$ if M has a loop.

The characteristic polynomial of a graphical matroid essentially equals the chromatic polynomial of the graph in the following way.

Exercise 2.12. Show that if M is the graphical matroid of a finite graph G, and G has c many connected components, then $q^c\chi_{\mathrm{M}}(q) = \chi_G(q)$. (Hint: Appeal to Theorem 2.11 by showing that both satisfies an identical deletion-contraction relation).

We may now state the sequences (c) and (d), which we only stated for graphical matroids, for arbitrary nontrivial matroids: They are the coefficients of $T_{\mathrm{M}}(1+q, 0)$ and $\frac{1}{q}T_{\mathrm{M}}(q, 0)$, respectively.

Summarizing, the have the following sequences for a matroid M of rank r.

(a) The coefficients $(I_0, \dots, I_r)$ of $T_{\mathrm{M}}(q+1, 1) = \sum_{i=0}^{r} I_i q^{r-i}$.
(b) The coefficients $(I'_0, \dots, I'_r)$ of $T_{\mathrm{M}}(q, 1) = \sum_{i=0}^{r} I'_i q^{r-i}$.
(c) The coefficients $(w_0, \dots, w_r)$ of $T_{\mathrm{M}}(q+1, 0) = \sum_{i=0}^{r} w_i q^{r-i}$.
(d) The coefficients $(w'_0, \dots, w'_{r-1})$ of $\frac{1}{q}T_{\mathrm{M}}(q, 0) = \sum_{i=0}^{r} I_i q^{r-1-i}$.
(e) The Whitney numbers of the second kind $(W_0, \dots, W_r)$ of M.

Theorem 2.13. [AHK18, ADH22, BHM+] Let M be a matroid of rank r.

(a) The sequence $(I_0, \dots, I_r)$ is unimodal, log-concave, and top-heavy.
(b) The sequence $(I'_0, \dots, I'_r)$ is unimodal, log-concave, and top-heavy.
(c) The sequence $(w_0, \dots, w_r)$ is unimodal, log-concave, and top-heavy.
(d) The sequence $(w'_0, \dots, w'_{r-1})$ is unimodal, log-concave, and top-heavy.
(e) The sequence $(W_0, \dots, W_r)$ satisfies $W_i \le W_j$ for all $0 \le i \le j \le r - i$. In particular, it is top-heavy.

The statements of the theorem were long-standing conjectures in matroid theory. The unimodality and log-concavity conjectures are due to: Welsh [Wel71] and Mason [Mas72] for (a), Dawson [Daw84] for (b), Read [Rea68] and Hoggar [Hog74] for (c) of graphical matroids, Heron [Her72], Rota [Rot71], and Welsh [Wel76] for (c), and Brylawski [Bry82] for (d). Hibi [Hib92] and Swartz [Swa03] posed the top-heaviness of (b) and (d) (respectively).[1] Dowling and Wilson [DW74, DW75] conjectured the top-heaviness of (e), generalizing a theorem of de Bruijn and Erdös [dBE48] on point-line incidences in projective planes. There are two notable conjectures on (e) that remain open: Rota [Rot71] conjectured its unimodality, and Mason [Mas72] its log-concavity.

Remark 2.14. One may ask if there is a log-concavity statement for the whole Tutte polynomial of a matroid that explains the log-concavity of the four specializations in Theorem 2.13. This was achieved by Berget, Spink, Tseng, and the author [BEST] who showed that the 4-variable transformation

$$(x+y)^{-1}(y+z)^r(x+w)^{|E|-r}\mathrm{T}_{\mathrm{M}}\left(\frac{x+y}{y+z}, \frac{x+y}{x+w}\right)$$

[1] A slightly different terminology of *flawless-ness* appears in [Hib92] and related works [Hib89, JKL18]. A nonnegative sequence $(a_0, \dots, a_m)$ is flawless if it is top-heavy and additionally satisfies $a_0 \le \cdots \le a_{\lfloor d/2 \rfloor}$, where d is the largest index such that $a_d \ne 0$. Note that unimodal and top-heavy sequences are flawless.

of the Tutte polynomial of a matroid M satisfies a multivariate version of log-concavity. We note that without such a transformation, the coefficients of $\mathrm{T}_{\mathrm{M}}(x, y)$ can fail to be unimodal [Sch93].

In this survey, we will explain how the sequences (c) and (d) are shown to be log-concave with no internal zeros,[2] and how the sequence (e) is shown to be top-heavy. Since $\chi_{\mathrm{M}} = 0$ if M has a loop, and deleting loops of a matroid does not change the lattice of its flats, we assume the following:

Assumption. From now on, a matroid is loopless unless specified otherwise.

3. The realizable case

We explain how the statements in Theorem 2.13 can be deduced using algebraic geometry when the matroid in question is realizable. We assume familiarity with algebraic geometry; those who prefer purely combinatorial treatments may skip this section. For simplicity, we consider matroids realizable over $\mathbb{C}$. For other fields, one can run nearly identical arguments using the Chow cohomology ring [Ful98] or the ℓ-adic (intersection) cohomology in place of singular (intersection) cohomology $(I)H^\bullet(-)$ with rational coefficients.

Throughout this section, let M be a (nontrivial) matroid of rank r realized by a set of vectors $\{v_i : i \in E\}$ spanning a vector space $V \simeq \mathbb{C}^r$. The corresponding surjection $\mathbb{C}^E \twoheadrightarrow V$ dualizes to give an r-dimensional linear subspace $V^\vee \subseteq \mathbb{C}^E$.

Notation. Let us denote $L = V^\vee$ to avoid repeated use of the superscript $\vee$.

The set of independent sets of M can then be described also as the collection

$$\mathcal{I} = \{I \subseteq E : \text{the composition } L \hookrightarrow \Bbbk^E \twoheadrightarrow \Bbbk^I \text{ is surjective}\}.$$

We will often projectivize and work with $\mathbb{P}L \subseteq \mathbb{P}^{n-1}$.

3.1. **Hyperplane arrangements.**

We first discuss some structures of the matroid M in terms of its realization as a subspace $L \subseteq \mathbb{C}^E$. For each $i \in E$ let H_i be the i-th coordinate hyperplane of $\mathbb{C}^E$. Our assumption that M is loopless implies that L is not contained in any H_i. We thus have an **hyperplane arrangement** $\mathcal{A}$ on L consisting of the hyperplanes $\{L \cap H_i : i \in E\}$. Dualizing the correspondence between the flats of M and the spans of subsets of the vectors $\{v_i : i \in E\}$, we obtain a correspondence

$$\{\text{flats of } \mathrm{M}\} \longleftrightarrow \{\text{subspaces of } L \text{ arising as intersections of hyperplanes in } \mathcal{A}\}$$

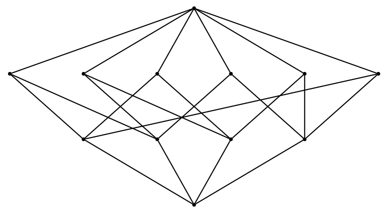
$$F \longleftrightarrow L_F = L \cap \bigcap_{i \in F} H_i.$$

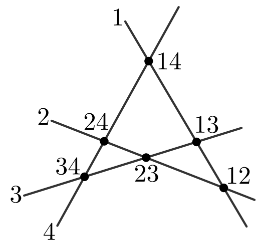
Note that the correspondence is order-reversing, and in particular, a flat of rank $r - i$ maps to the linear subspace L_F of dimension i.

[2] As observed by Lenz [Len13], a result of Brylawski [Bry77, theorem 4.2] implies that the statements for (a) and (b) follow from those for (c) and (d). The top-heaviness of (c) and (d) follows from their unimodality due to [JKL18, Theorem 1.2].

Example 3.1. The columns of the matrix below realizes the matroid $\mathrm{U}_{3,4}$. Equivalently, the embedded subspace $L \subseteq \mathbb{C}^4$, where $L = \{x_1 + x_2 + x_3 + x_4 = 0\}$ is the row-span of the matrix, realizes the matroid $\mathrm{U}_{3,4}$.

$$\begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$





Next to the matrix, we have depicted the lattice of flats of M and the projectivization of the hyperplane arrangement $\mathcal{A}$ in $\mathbb{P}L \simeq \mathbb{P}^2$.

We denote the complement of the hyperplane arrangement by

$$\mathring{L} = L \setminus \bigcup \mathcal{A} = L \cap (\mathbb{C}^*)^E, \quad \text{and likewise,} \quad \mathbb{P}\mathring{L} = \mathbb{P}L \cap \big((\mathbb{C}^*)^E/\mathbb{C}^*\big).$$

Exercise 3.2. Suppose $i \in E$ is not a coloop, and let F be the smallest flat containing i. Show that the subspace L_F is a realization of the contraction M/F, and that the hyperplane arrangement complement $\mathring{L} \setminus L_F$ is a realization of the deletion $\mathrm{M} \setminus F$. What happens when i is a coloop?

The geometric study of (complements of) hyperplane arrangements and its interaction with matroid theory is a rich and on-going research field; some references include [OT92, Dim17]. Here, we only note the following classical fact [OS80], which states that the characteristic polynomial records the dimensions of the cohomologies of the arrangement complement.

Theorem 3.3. Let $L \subseteq \mathbb{C}^E$ realize a matroid M of rank r. Then, we have

$$\mathrm{T}_{\mathrm{M}}(1+q, 0) = \sum_{i=0}^{r} \dim H^i(\mathring{L}) q^{r-i},$$

or equivalently, by applying the Künnuth formula to $\mathbb{P}\mathring{L} \times \mathbb{C}^* \simeq \mathring{L}$,

$$\frac{1}{1+q}\mathrm{T}_{\mathrm{M}}(1+q, 0) = \sum_{i=0}^{r-1} \dim H^i(\mathbb{P}\mathring{L}) q^{r-1-i}.$$

3.2. **Log-concavity via intersection degrees.** We now explain how the log-concavity of the sequences (c) and (d) can be shown using algebraic geometry. We start by describing a general strategy for log-concavity.

Let X be a smooth projective $\mathbb{C}$-variety X of dimension d. When considered as a $2d$-dimensional real compact manifold, Poincaré duality for the cohomology ring $H^\bullet(X)$ provides the isomorphism $\int_X : H^{2d}(X) \to \mathbb{Z}$, called the degree map. Recall that a divisor D on X is ample (resp. semi-ample) if the line bundle $\mathcal{O}_X(mD)$ is very ample (resp. globally generated) for some integer $m >> 0$. The general strategy arises from the following Khovanskii-Teissier inequalities (see [Laz04a, Section 1.6] for a history and a fuller treatment).

Proposition 3.4. Let $\alpha, \beta \in H^2(X)$ be the cohomology classes of two semi-ample (or more generally, nef) divisors on a smooth projective $\mathbb{C}$-variety X of dimension d. Then,

$$\text{the sequence } (a_0, \dots, a_d) \text{ of intersection degrees of } \alpha \text{ and } \beta, \text{ i.e. } a_i = \int_X \alpha^{d-i}\beta^i$$

is log-concave with no internal zeros.

Sketch of the proof. By continuity, one can assume α, β to be ample, and then one reduces to the case of surfaces via the the Bertini theorem. Then, one of the equivalent forms of the Hodge index theorem for surfaces [Har77, Exercise V.1.9] exactly yields the desired log-concavity. We note that the Hodge index theorem for surfaces stated as [Har77, Theorem V.1.9] is exactly the validity of the Hodge-Riemann relations for surfaces. □

Thus, from the realization $L \subseteq \mathbb{C}^E$, one may seek for a smooth projective $\mathbb{C}$-variety equipped with two semi-ample divisor classes α, β such that their intersection degrees yield the appropriate combinatorial sequence. We explain how this is done for the sequences (c) and (d).

3.2.1. *Log-concavity of (c).* We show the slightly stronger statement that the closely related sequence $(\overline{w}_0, \dots, \overline{w}_{r-1})$ defined by

$$\frac{1}{1+q}\mathrm{T}_{\mathrm{M}}(1+q, 0) = \sum_{i=0}^{r-1} \overline{w}_i q^{r-1-i}$$

is log-concave with no internal zeros. Theorem 3.3 states that this sequence is exactly the betti numbers of the arrangement complement $\mathbb{P}\mathring{L}$. The sought-after projective variety is the wonderful compactification of a hyperplane arrangement complement introduced in [DCP95].

Definition 3.5. The **wonderful compactification** W_L is the variety obtained by blowing-up $\mathbb{P}L$ at all the points $\{\mathbb{P}L_F : \mathrm{rk}_{\mathrm{M}}(F) = r-1\}$, then by blowing-up all strict transforms of the lines $\{\mathbb{P}L_F : \mathrm{rk}_{\mathrm{M}}(F) = r-2\}$, and so forth. Let $\pi_L : W_L \to \mathbb{P}L$ be the blow-down map.

By construction π_L is isomorphism on the open loci $\mathbb{P}\mathring{L}$. The boundary $\partial W_L = W_L \setminus \mathbb{P}\mathring{L}$ is a simple-normal-crossing divisor on W_L [DCP95]. For the sought-after divisor classes α, β on W_L, the wonderful compactification for the Boolean matroid plays a special role.

When $\mathrm{M} = \mathrm{U}_{n,n}$, that is, when $L = \mathbb{C}^E$, the wonderful compactification is known as the **permutohedral variety** $X_{A_{n-1}}$. Explicitly, it is obtained from $\mathbb{P}^{n-1}$ by blowing-up all n coordinate points of $\mathbb{P}^{n-1}$, then blowing-up all strict transforms of $\binom{n}{2}$ coordinate lines of $\mathbb{P}^{n-1}$, and so forth. Let $\pi_1 : X_{A_{n-1}} \to \mathbb{P}^{n-1}$ be the blow-down map. Blowing-down the exceptional divisors in a "reverse manner" (see [Huh18] for a detailed description via toric geometry), one obtains a different blow-down map $\pi_2 : X_{A_{n-1}} \to \mathbb{P}^{n-1}$. The resulting birational transformation is the Cremona transformation $\mathrm{crem} : \mathbb{P}^{n-1} \dashrightarrow \mathbb{P}^{n-1}$ given by $[x_1, \ldots, x_n] \mapsto [\frac{1}{x_1}, \ldots, \frac{1}{x_n}]$ in the projective coordinates.

Returning to the case where M is not necessarily Boolean, we note the following: Because the arrangement $\mathcal{A}$ on $\mathbb{P}L$ is the restriction to $\mathbb{P}L$ of the coordinate hyperplane arrangement on $\mathbb{P}^{n-1}$, the universal property of blow-ups implies that W_L is the strict transform of $\mathbb{P}L \subseteq \mathbb{P}^{n-1}$ under the blow-up π_1. Summarizing, we have a commuting diagram:

(†)
$$\begin{array}{ccccc} W_L & \hookrightarrow & X_{A_{n-1}} & & \\ \downarrow{\scriptstyle \pi_L} & & \downarrow{\scriptstyle \pi_1} & \searrow^{\pi_2} & \\ \mathbb{P}L & \hookrightarrow & \mathbb{P}^{n-1} & \overset{\mathrm{crem}}{\dashrightarrow} & \mathbb{P}^{n-1} \end{array}$$

Let h be the hyperplane class of $\mathbb{P}^{n-1}$, and define divisor classes α and β on W_L to be the restrictions of $\pi_1^* h$ and $\pi_2^* h$, respectively. Huh–Katz [HK12] showed that

$$\frac{1}{1+q} T_{\mathrm{M}}(1+q, 0) = \sum_{i=0}^{r-1} \left(\int_{W_L} \alpha^{r-1-i} \beta^i \right) q^{r-1-i}.$$

Since α and β are both hyperplane class pullbacks, they are globally-generated, so Proposition 3.4 implies the desired log-concavity.

How might one think to do this, at least in hindsight? Two key steps are as follows.

(1) The commuting diagram (†) shows that α is also the pullback to W_L of the hyperplane class in $\mathbb{P}L$, so we may loosely interpret multiplication by α as restriction to a general hyperplane H in $\mathbb{P}L$. As a linear subvariety $H \subset \mathbb{P}^{n-1}$, this hyperplane H is again a realization of a matroid, known as the **truncation matroid** $\mathrm{tr(M)}$ of M. With well-known properties of characteristic polynomials [Zas87], it is straightforward to verify that $\frac{1}{1+q}\mathrm{T}_{\mathrm{tr(M)}}(1+q, 0)$ is obtained from $\frac{1}{1+q}\mathrm{T}_{\mathrm{M}}(1+q, 0)$ by erasing the constant term and then dividing by q. That is, the sequence $(\overline{w}_0, \ldots, \overline{w}_{r-2})$ for $\mathrm{tr(M)}$ is obtained from that of M by simply removing the last entry.

(2) With the previous step, we now only need compare $\int_{W_L} \beta^{r-1}$ with the constant term $\mathrm{T}_{\mathrm{M}}(1, 0)$. Let $\mathbb{P}L^{-1}$ be the closure of the image of $\mathbb{P}\mathring{L}$ under the Cremona transformation, often known as the **reciprocal linear space**. By the construction of β, the degree of $\mathbb{P}L^{-1}$ as a subvariety of $\mathbb{P}^{n-1}$ equals $\int_{W_L} \beta^{r-1}$. On the other hand, the degree of $\mathbb{P}L^{-1}$ also equals $\mathrm{T}_{\mathrm{M}}(1, 0)$. This last key fact was proven in several contexts [Ter02, PS06, Huh12, HK12]. A topological approach in [Huh12] is as follows: A result of Dimca and Papadima [DP03] from (complex) Morse theory related the Euler characteristic of a hypersurface complement to the degree of the gradient map. The hypersurface $\{x_1 x_2 \cdots x_n = 0\} \subset \mathbb{P}^{n-1}$ is the coordinate hyperplane arrangement that restricts to the hyperplane arrangement $\mathcal{A}$ on $\mathbb{P}L$, and the gradient map of $x_1 x_2 \cdots x_n$ is exactly the Cremona map. Combining these facts with Theorem 3.3, one can deduce $\deg \mathbb{P}L^{-1} = \mathrm{T}_{\mathrm{M}}(1, 0)$.

Exercise 3.6. Let L be as in Example 3.1. Verify that $\mathrm{T_M}(1,0) = 3$, and verify that $\mathbb{P}L^{-1}$ is a cubic surface known as the Cayley nodal cubic. This cubic surface has four singular points, with a line through each pair of points; explain where these come from in terms of the wonderful compactification W_L. (Bonus: This cubic surface has three more lines, for the total of nine; where do they come from?)

3.2.2. *Log-concavity of (d).* The sought-after projective variety is the **variety of critical points** $\mathfrak{X}$, formally introduced in [CDFV11] but implicit in previous works related to Varchenko's problem on critical points of master functions on an affine hyperplane arrangement [Var95]. Here, in order to build upon our previous discussion in Section 3.2.1, we follow [BEST, Section 8] to describe a smooth birational model of $\mathfrak{X}$ in terms of the wonderful compactification W_L, although it differs slightly from the original description in [Huh15].

Consider the embedding $W_L \hookrightarrow X_{A_{n-1}}$ in the diagram (†). Let $\mathcal{N} = \mathcal{N}_{W_L/X_{A_{n-1}}}$ be the normal bundle, and let $\mathfrak{X}_L = \mathbb{P}_{W_L}(\mathcal{N}^\vee)$ be the projectivization[3] of the conormal bundle with the projection map $p : \mathfrak{X}_L \to W_L$. Recall the blow-down map $\pi_L : W_L \to \mathbb{P}L$.

The sought-after divisor classes on $\mathfrak{X}_L$ are as follows. Let γ be the pullback of the hyperplane class in $\mathbb{P}L$ via the composition $\mathfrak{X}_L \xrightarrow{p} W_L \xrightarrow{\pi_L} \mathbb{P}L$. Let $\delta = c_1(\mathcal{O}(1))$ be the first Chern class of the line bundle $\mathcal{O}(1)$ from the construction of $\mathfrak{X}_L$ as a projectivization of a vector bundle, which turns out to be semi-ample. One can then translate [DGS12, Theorem 1.1] to the statement that

$$\frac{1}{q}\mathrm{T_M}(q,0) = \sum_{i=0}^{r-1} \left(\int_{\mathfrak{X}_L} \gamma^{r-1-i}\delta^{n-r-1+i} \right) q^{r-1-i}.$$

Proposition 3.4 now implies the desired log-concavity.

How might one think to do this, at least in hindsight? For the original formulation of $\mathfrak{X}$, maximum likelihood problems in algebraic statistics provided a motivation; see [Huh13, HS14]. For the related construction $\mathfrak{X}_L$ here, we highlight some key steps. In either cases, one uses properties of log-tangent bundles and their characteristic classes; see [Alu05] for an introduction to these tools.

(1) By Theorem 3.3, the constant term equals

$$\left(\frac{1}{q}\mathrm{T_M}(q,0)\right)\bigg|_{q=0} = \left(\frac{1}{1+q}\mathrm{T_M}(1+q,0)\right)\bigg|_{q=-1} = (-1)^{r-1}\chi_{top}(\mathbb{P}\mathring{L}),$$

the signed Euler characteristic of $\mathbb{P}\mathring{L}$. Euler characteristics satisfy the "scissors relation" that $\chi_{top}(X) = \chi_{top}(X \setminus Z) + \chi_{top}(Z)$ for a closed embedding $Z \hookrightarrow X$ of $\mathbb{C}$-varieties. More generally, Chern-Schwartz-MacPherson (CSM) classes [Mac74] of $\mathbb{C}$-varieties are homological objects that respect such scissors relation. One then uses Exercise 3.2 to show that the degrees

[3] Our convention for the projectivization of a vector bundle $\mathcal{E}$ on a variety X is that $\mathbb{P}_X(\mathcal{E}) = \operatorname{Proj}_X \operatorname{Sym}^\bullet(\mathcal{E}^\vee)$, which agrees with [EH16] but is the opposite of [Har77, Laz04b].

of the CSM classes of $\mathbb{P}\mathring{L}$ satisfy the same deletion-contraction relation satisfied by the coefficients of $\frac{1}{-q}\mathrm{T}_M(-q,0)$, so that Theorem 2.11 implies that they are the same.

(2) Having related the CSM classes to $\frac{1}{-q}\mathrm{T}_M(-q,0)$, we now relate powers of δ with the CSM classes of $\mathbb{P}\mathring{L}$. The varieties W_L and $X_{A_{n-1}}$ have simple-normal-crossing boundaries ∂W_L and $\partial X_{A_{n-1}} = X_{A_{n-1}} \setminus ((\mathbb{C}^*)^E/\mathbb{C}^*)$. Under the embedding $W_L \hookrightarrow X_{A_{n-1}}$, one can show that $\partial W_L = W_L \cap \partial X_{A_{n-1}}$ scheme-theoretically. Consequently, one has a short exact sequence (see for instance [EHL, Section 9])

$$0 \to \mathcal{T}_{W_L}(-\log \partial W_L) \to \mathcal{T}_{X_{A_{n-1}}}(-\log \partial X_{A_{n-1}})|_{W_L} \to \mathcal{N} \to 0.$$

Because $X_{A_{n-1}}$ is a toric variety with the dense open torus $(\mathbb{C}^*)^E/\mathbb{C}^*$, the log-tangent bundle $T_{X_{A_{n-1}}}(-\log \partial X_{A_{n-1}})$ is trivial [CLS11, Chapter 8]. Thus, the normal bundle $\mathcal{N}$ is globally-generated so that δ is semi-ample. Moreover, the Segre classes of $\mathcal{N}$, which are given by powers of $-\delta$, equal the Chern classes of $\mathcal{T}_{W_L}(-\log \partial W_L)$, which are the CSM classes of the complement $W_L \setminus \partial W_L = \mathbb{P}\mathring{L}$ [Alu99].

3.3. **The top-heaviness via intersection cohomology.** We start with a general strategy for establishing top-heaviness, which first appeared in [BE09] to establish top-heaviness for Bruhat intervals. Suppose we have a (not necessarily smooth) projective $\mathbb{C}$-variety X with an **affine stratification**: There is a finite collection $\{U_j\}_{j\in J}$ of locally closed subvarieties of X, called the **strata**, such that X is the disjoint union of $\{U_j\}_{j\in J}$, the closure $\overline{U_j}$ of any strata is again a union of strata, and each strata is isomorphic to $\mathbb{C}^m$ for some m.

Theorem 3.7. [BE09, Theorem 3.1] For $0 \le i \le d = \dim X$, let b_i be the number of strata of dimension i. Then, we have $b_i \le b_j$ for all $0 \le i \le j \le d - i$.

Sketch of the proof. Let us assume we have shown that $\dim H^{2i}(X) = b_i$. If X were smooth, the hard Lefschetz theorem implies that $(b_0, \dots, b_d)$ is in fact unimodal and symmetric, so we would be done. Since X may not be smooth, we need consider the intersection cohomology $IH^\bullet(X)$, for which the hard Lefschetz theorem still holds [GM83, BBD82]. There is a natural graded map $H^\bullet(X) \to IH^\bullet(X)$, fitting into a commutative diagram

$$\begin{array}{ccc} H^{2i}(X) & \longrightarrow & IH^{2i}(X) \\ {\scriptstyle \cdot c_1(\mathcal{L})^{j-i}}\downarrow & & \downarrow{\scriptstyle \cdot c_1(\mathcal{L})^{j-i}} \\ H^{2j}(X) & \longrightarrow & IH^{2j}(X) \end{array}$$

for any ample line bundle $\mathcal{L}$ on X, where the injectivity of the right vertical map follows from the validity of the hard Lefschetz property for intersection cohomology. Thus, if $H^\bullet(X) \to IH^\bullet(X)$ is injective, then the left vertical map is necessarily injective, so we can conclude the desired $b_i \le b_j$.

The proof of $\dim H^{2i}(X) = b_i$ and the injection $H^\bullet(X) \hookrightarrow IH^\bullet(X)$ both follow from combining a standard long exact sequence of cohomologies and the result of Weber [Web04] (see also [BE09, Theorem 2.1]) which identified the kernel of $H^k(X) \to IH^k(X)$ in terms of the weight filtration on $H^k(X)$ given by the mixed Hodge structure [Del71]. □

Björner and Ekedahl [BE09] used Theorem 3.7 on Schubert varieties of a generalized flag variety to deduce top-heaviness for Bruhat intervals of a finite crystallographic Coxeter group. In our case, for a realization $L \subseteq \mathbb{C}^E$ of a matroid M, we consider its **matroid Schubert variety** Y_L defined as

$$Y_L = \text{the closure of } L \text{ inside } (\mathbb{P}^1)^n,$$

where $\mathbb{C}^E \subset (\mathbb{P}^1)^n$ via the identification $\mathbb{P}^1 = \mathbb{C} \cup \{\infty\}$. Note that the identification $\mathbb{P}^1 = \mathbb{C} \cup \{\infty\}$ induces an affine stratification of $(\mathbb{P}^1)^n$ with strata $\{\mathbb{C}^S \times \{\infty\}^{E\setminus S} : S \subseteq E\}$. One shows that this stratification restricts to give an affine stratification of Y_L by using the computation of (the Gröbner basis for) the defining ideal of $Y_L \subseteq (\mathbb{P}^1)^n$ given by Ardila and Boocher [AB16].

Theorem 3.8. [HW17, Theorem 14], [PXY18, Lemmas 7.5 & 7.6] The matroid Schubert variety Y_L admits an affine stratification by the strata $\{U^F : F \text{ a flat of } \mathrm{M}\}$ defined by

$$U^F = Y_L \cap \left(\mathbb{C}^F \times \{\infty\}^{E\setminus F}\right).$$

For each flat F, the strata U^F is isomorphic to the image of the composition $L \twoheadrightarrow \mathbb{C}^E \twoheadrightarrow \mathbb{C}^F$, which has dimension $\mathrm{rk}_{\mathrm{M}}(F)$.

The top-heaviness of the sequence (e) now follows from Theorems 3.8 and 3.7.

Exercise 3.9. Verify Theorem 3.8 for a realization of the uniform matroid $\mathrm{U}_{2,3}$.

For Schubert varieties in generalized flag varieties, their intersection cohomology is closely related to Kazhdan-Lusztig theory. The terminology "matroid Schubert variety" was chosen for an analogous relation between the intersection cohomology Y_L and Kazhdan-Lusztig theory of matroid developed in [EPW16, PXY18].

4. Tropical Hodge theory

We will begin by explaining how polyhedral fans and matroids gives rise to "cohomology" rings. We then discuss two fundamental theorems for these "cohomology" rings concerning the validity of the Kähler package (Definition 1.2). We then explain how the log-concavity statements for arbitrary matroids can be deduced from these fundamental theorems. We will assume some familiarity with the basics of polyhedral geometry. For a brief introduction see [Ful93, Chapter 1.2], and see [Zie95] for a fuller treatment.

4.1. **Chow rings of fans and matroids.**

Let N be a lattice, i.e. a finitely generated free abelian group $\mathbb{Z}^m$. Let $N^\vee$ denote its dual lattice. We write $N_\mathbb{R} = N \otimes \mathbb{R}$. Recall that a fan Σ in $N_\mathbb{R}$ is *rational* if each ray ρ in Σ equals $\mathbb{R}_{\geq 0}u$ for some $u \in N$, *simplicial* if every k-dimensional cone in Σ is generated by k rays, and *pure-dimensional* if every maximal cone has the same dimension. For each ray ρ, let $u_\rho \in N$ be the *primitive ray generator*, i.e. the element such that $\rho \cap N = \mathbb{Z}_{\geq 0}u_\rho$. Let $\Sigma(1)$ denote the set of rays of Σ. The *support* of Σ is denoted $|\Sigma|$. A fan Σ in $N_{\mathbb{R}}$ is *complete* if $|\Sigma| = N_{\mathbb{R}}$.

Assumption. All fans we treat will be rational, simplicial, and pure-dimensional, but not necessarily complete.

Definition 4.1. The **Chow (cohomology) ring** (with real coefficients) of a fan Σ in $N_{\mathbb{R}}$ is the graded $\mathbb{R}$-algebra

$$A^\bullet(\Sigma) = \frac{\mathbb{R}[x_\rho : \rho \in \Sigma(1)]}{I_\Sigma + J_\Sigma}$$

where I_Σ and J_Σ are the ideals

$$I_\Sigma = \Big\langle \prod_{\rho \in S} x_\rho : S \subseteq \Sigma(1) \text{ do not form a cone in } \Sigma \Big\rangle \quad \text{and}$$

$$J_\Sigma = \Big\langle \sum_{\rho \in \Sigma(1)} m(u_\rho) x_\rho : m \in N^\vee \Big\rangle.$$

It is an exercise to show that the k-th graded piece $A^k(\Sigma)$ of $A^\bullet(\Sigma)$ is generated by square-free monomials of degree k in the variables. In particular, $A^k(\Sigma) = 0$ for all k greater than the dimension d of Σ, and $A^\bullet(\Sigma) = \bigoplus_{i=0}^d A^i(\Sigma)$ is a finite-dimensional graded real vector space.

Exercise 4.2. For two fans Σ and Σ', show that $A^\bullet(\Sigma \times \Sigma') = A^\bullet(\Sigma) \otimes A^\bullet(\Sigma')$.

Borrowing language from algebraic geometry, let us call a linear combination of the variables x_ρ a **divisor** and its image in $A^1(\Sigma)$ its **divisor class** on Σ. Because Σ is simplicial, a divisor $D = \sum_{\rho \in \Sigma(1)} c_\rho x_\rho$ determines a piecewise-linear function φ_D on $|\Sigma|$ by assigning the value c_ρ to each primitive ray generator u_ρ.

Definition 4.3. A divisor D on a complete fan Σ is **ample** if the piecewise-linear function φ_D is strictly-convex, i.e. $\varphi_D(u) + \varphi_D(v) < \varphi_D(u+v)$ for all $u, v \in N_{\mathbb{R}}$ not in the same cone of Σ. It is **nef** if only the weak inequality $\leq$ is satisfied.

For a not necessarily complete fan Σ, a divisor D is **ample** (resp. **nef**) if φ_D is the restriction of the piecewise-linear function of an ample (resp. nef) divisor on a complete fan containing Σ as a subfan. We denote by $\mathcal{K}(\Sigma) \subset A^1(\Sigma)$ the convex set of the divisor classes of ample divisors on Σ.[4] We often consider $\mathcal{K}(\Sigma)$ as a set of graded linear maps $A^\bullet(\Sigma) \to A^{\bullet+1}(\Sigma)$ given by multiplication.

In terms of toric geometry, the ring $A^\bullet(\Sigma)$ is the Chow cohomology ring of the toric variety X_Σ associated to the fan Σ [Dan78, Bri96]. When Σ is the normal fan of a simple polytope, the toric variety X_Σ is a projective variety with mild (i.e. orbifold) singularities, whose ample cone is $\mathcal{K}(\Sigma)$. Classical results in algebraic geometry then imply the validity of the Kähler package for the ring $A^\bullet(\Sigma)$. Stanley used this to resolve McMullen's g-conjecture on the number of faces of a simple polytope [Sta80]. Afterwards, McMullen [McM93] gave a purely combinatorial

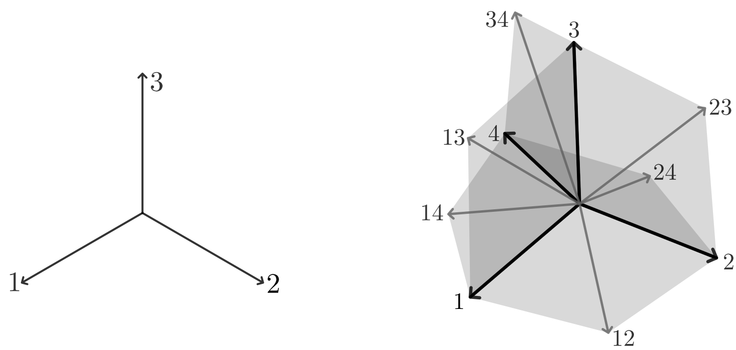
[4] Technically, our definition of ample/nef divisors on non-complete fans differs from that of [ADH22], which makes our $\mathcal{K}(\Sigma)$ a subset of the one in [ADH22], but will suffice for our discussion.

proof of the Kähler package that works even for nonrational fans (with no associated toric variety in the background). In our case, the fans will be generally not complete, so a priori there is no reason to expect any validity of the Kähler package, since the associated toric varieties are generally non-compact. The miraculous results is that certain fans from matroids turn out to enjoy the Kähler package.

For a matroid M on E of rank r, we construct a fan introduced and studied in [Stu02, AK06, Spe08] as tropical geometric analogues of linear spaces. The fan will be in the real vector space over the lattice $\mathbb{Z}^E/\mathbb{Z}\mathbf{e}_E$. For a subset $S \subseteq E$, let us denote by $\overline{\mathbf{e}}_S$ the image of of $\mathbf{e}_S = \sum_{i\in S} \mathbf{e}_i \in \mathbb{R}^E$ under the quotient map $\mathbb{R}^E \to \mathbb{R}^E/\mathbb{R}\mathbf{e}_E$.

Definition 4.4. The **Bergman fan** Σ_M of a rank r matroid M is a pure $(r-1)$-dimensional fan in $\mathbb{R}^E/\mathbb{R}\mathbf{e}_E$ consisting of the maximal cones $\mathbb{R}_{\geq 0}\{\overline{\mathbf{e}}_{F_1}, \dots, \overline{\mathbf{e}}_{F_{r-1}}\}$, one for each maximal chain $\varnothing \subsetneq F_1 \subsetneq F_2 \subsetneq \cdots \subsetneq F_{r-1} \subsetneq E$ of nonempty proper flats of M.

Example 4.5. The Bergman fans of $\mathrm{U}_{2,3}$ and $\mathrm{U}_{3,4}$ are depicted below.



Example 4.6. For the Boolean matroid $\mathrm{U}_{n,n}$, note that the maximal cones of its Bergman fan $\Sigma_{\mathrm{U}_{n,n}}$ correspond to permutations of the ground set E. The fan $\Sigma_{\mathrm{U}_{n,n}}$ is known as the **permutohedral fan** or as the **braid arrangement**, denoted $\Sigma_{A_{n-1}}$. Note that by construction the Bergman fan of any matroid M (on ground set E) is a subfan of $\Sigma_{A_{n-1}}$. See [AA] for a survey of remarkable combinatorial properties of the permutohedral fan.

Definition 4.7. The **Chow ring** $A^\bullet(\mathrm{M})$ **of a matroid** M is the Chow ring $A^\bullet(\Sigma_\mathrm{M})$ of its Bergman fan. Explicitly, it is

$$A^\bullet(\mathrm{M}) = \bigoplus_{i=0}^{r-1} A^i(\mathrm{M}) = \frac{\mathbb{R}[x_F : F \text{ a nonempty proper flat of } \mathrm{M}]}{I_\mathrm{M} + J_\mathrm{M}}$$

where I_M and J_M are the ideals

$$I_\mathrm{M} = \left\langle x_F x'_F : F \not\subseteq F' \text{ and } F \not\supseteq F' \right\rangle \quad \text{and}$$
$$J_\mathrm{M} = \left\langle \sum_{F \ni i} x_F - \sum_{G \ni j} x_G : i, j \in E \right\rangle.$$

Remark 4.8. When M has a realization $L \subseteq \mathbb{C}^E$, the wonderful compactification W_L defined in Section 3.2.1 is the closure of $\mathbb{P}\mathring{L}$ inside the toric variety X_{Σ_M}. The resulting pullback map of cohomologies is an isomorphism between the Chow ring $A^\bullet(\mathrm{M}) = A^\bullet(X_{\Sigma_\mathrm{M}})$ of the toric variety of Σ_M and the cohomology ring of W_L [FY04, DCP95]. This "Chow equivalence" is informed by the theory of *tropical compactifications* [Tev07]; see [MS15, Chapter 6] for an introduction. While the Kähler package for the Chow ring in this realizable case thus follows from classical Hodge theory, [AHK18, Theorem 5.12] states that the existence of such Chow equivalence is equivalent to the realizability of the matroid.

Definition 4.9. The **conormal fan** $\Sigma_{\mathrm{M},\mathrm{M}^\perp}$ of a (loopless and coloopless) matroid M is a fan in $\mathbb{R}^E/\mathbb{R}\mathbf{e}_E \times \mathbb{R}^E/\mathbb{R}\mathbf{e}_E$ whose support equals the support of the product $\Sigma_\mathrm{M} \times \Sigma_{\mathrm{M}^\perp}$.

We omit the precise definition, which involves the intricate and interesting combinatorics of the *bipermutohedron* and *biflags* introduced in [ADH22]. See [ADH22, Section 2.8] for its origin story.

Remark 4.10. Just as the Chow ring $A^\bullet(\mathrm{M})$ of a matroid M is modeled after the wonderful compactification, the conormal Chow ring of a matroid is modeled after the geometry described in Section 3.2.2. For instance, Exercise 2.7 suggests how the product $\Sigma_\mathrm{M} \times \Sigma_{\mathrm{M}^\perp}$ can serve as a polyhedral model of the projectivized conormal bundle $\mathfrak{X}_L = \mathbb{P}_{W_L}(\mathcal{N}^\vee)$.

4.2. **Fundamental theorems of tropical Hodge theory.** We now discuss two fundamental theorems concerning the Kähler package for Chow rings of not necessarily complete fans. To state them, we need a few more terminologies.

For a cone σ of a fan Σ in $N_\mathbb{R}$, the **star** $\mathrm{st}_\sigma(\Sigma)$ is a fan in $N_\mathbb{R}/\operatorname{span}(\sigma)$ whose cones are the images under the projection $N_\mathbb{R} \to N_\mathbb{R}/\operatorname{span}(\sigma)$ of the cones of Σ containing σ. Geometrically, the toric variety $X_{\mathrm{st}_\sigma(\Sigma)}$ of the star is the closure of the torus-orbit of X_Σ corresponding to the cone σ.

Exercise 4.11.

(1) The star of the Bergman fan Σ_M at the ray $\mathbb{R}_{\geq 0}\mathbf{e}_F$ corresponding to a nonempty proper flat F is isomorphic to the product $\Sigma_{\mathrm{M}|F} \times \Sigma_{\mathrm{M}/F}$.
(2) Show that the Chow ring $A^\bullet(\mathrm{st}_\sigma \Sigma)$ of the star is isomorphic to the quotient ring $A^\bullet(\Sigma)/\langle a \in A^\bullet(\Sigma) : a \cdot \prod_{\rho\in\sigma} x_\rho = 0\rangle$.

A **positive Minkowski weight** on a (pure) d-dimensional fan Σ is a linear map $\deg : A^d(\Sigma) \to \mathbb{R}$ such that $\deg(\prod_{\rho\in\sigma} x_\rho) > 0$ for any maximal cone σ of Σ. It defines a symmetric bilinear pairing $A^\bullet \times A^{d-\bullet} \to \mathbb{R}$ by $(x, y) \mapsto \deg(xy)$, which by abuse of notation we also denote $\deg$.

Geometrically, Minkowski weights in general are fundamental objects in tropical intersection theory, serving the role of "Chow homology classes." For their definition and properties, see [FS97, KP08] and [MS15, Chapter 6], as well as [AHK18, Section 5]. Fans arising in the context of tropical compactifications (Remark 4.8) provide many examples of positive Minkowski weights. For instance, the following can be deduced from what is known as the cover-partition property of flats of a matroid.

Proposition 4.12. [AHK18, Proposition 5.2] For a matroid M of rank r, the assignment $x_{F_1} \cdots x_{F_{r-1}} \mapsto 1$ for any maximal chain $\varnothing \subsetneq F_1 \subsetneq \cdots \subsetneq F_{r-1} \subsetneq E$ of nonempty proper flats of M gives a well-defined linear map $\deg_{\mathrm{M}} : A^{r-1}(\mathrm{M}) \to \mathbb{R}$.

Similarly, the conormal fan of a matroid also has a natural positive Minkowski weight $\deg_{\mathrm{M},\mathrm{M}^\perp}$. The key notion for the statement of the two fundamental theorems is the notion of Lefschetz-ness of fans, introduced in [ADH22].

Definition 4.13. A fan Σ of dimension d is said to be **Lefschetz** if the following are satisfied:

(1) $\mathrm{Hom}_{\mathbb{R}}(A^d(\Sigma), \mathbb{R})$ is spanned by a positive Minkowski weight $\deg$.
(2) The triple $(A^\bullet(\Sigma), \deg, \mathcal{K}(\Sigma))$ satisfies the Kähler package (Definition 1.2).
(3) For any positive dimensional cone σ of Σ, the star $\mathrm{st}_\sigma(\Sigma)$ is Lefschetz.

We can now state the two fundamental theorems.

Theorem 4.14. [AHK18, Theorem 1.4] The Bergman fan Σ_{M} of a matroid M is Lefschetz.

Theorem 4.15. [ADH22, Theorem 1.6] Let Σ and Σ' be fans in $N_{\mathbb{R}}$ with the same support, and suppose $\mathcal{K}(\Sigma)$ and $\mathcal{K}(\Sigma')$ are nonempty. Then Σ is Lefschetz if and only if Σ' is Lefschetz.

The product of two Lefschetz fans is again Lefschetz [AHK18, Section 7.2]. Combining this with the two theorems yields the following.

Corollary 4.16. The conormal fan of the matroid is Lefschetz.

Exercise 4.17. Let M be a matroid of rank 3. Let $\alpha = \sum_{G \ni i} x_G$ for any $i \in E$. Note that $\alpha \in A^\bullet(\mathrm{M})$ is independent of the choice of i due to the linear relations J_{M}.

(1) Show that $\deg_{\mathrm{M}}(x_F^2) = -1$ for any flat F of rank 2.
(2) Show that $\deg_{\mathrm{M}}\left(\alpha^2\right) = 1$ for any element $i \in E$.
(3) Show that $\deg_{\mathrm{M}}(\alpha x_F) = 0$ for any element $i \in E$ and a flat F of rank 2.
(4) Use these steps to establish the Kähler package for $A^\bullet(\mathrm{M})$ with $\mathcal{K} = \mathbb{R}_{\geq 0}\alpha$.

Let us give a broad overview of the proofs of the two theorems. Both employs the following strategy for establishing the Kähler package for a graded $\mathbb{R}$-algebra $A^\bullet$ of "dimension d" in the sense that $A^\bullet = \bigoplus_{i=0}^d A^i$. This general strategy and variations thereof appear in several previous works on the Kähler package across varied mathematical fields, such as the works of McMullen [McM93] on simple polytopes, Elias and Williamson [EW14] on Soergel bimodules, and [dCM09] on the topology of algebraic maps.

(i) It suffices to show the statements of (HL) and (HR) in Definition 1.2 in the special case where $L_0 = L_1 = \cdots = L_{d-2i}$, because this "non-mixed" version of the Kähler package implies the original "mixed" version of the Kähler package [Cat08] (see also [ADH22, Theorem 5.20]).

(ii) Next, one can set up an induction on the "dimension" d as follows. In many situations, the quotient algebra $A^\bullet/\operatorname{ann}(x)$ for certain choices of $x \in A^1$ is again in the family of graded $\mathbb{R}$-algebras that one is seeking to establish the Kähler package for. Here, $\operatorname{ann}(x)$ denotes the annihilator $\{a \in A^\bullet : ax = 0\}$. The key observation then is that

$$\begin{pmatrix}\text{(HR) in degree } i \text{ of } A^\bullet/\operatorname{ann}(x)\\ \text{for sufficiently many } x \in A^1\end{pmatrix} \implies \begin{pmatrix}\text{(HL) in degree } i\\ \text{for the original } A^\bullet\end{pmatrix}$$

(see for instance [BES, Proposition 6.1.6]). Because the quotient $A^\bullet/\operatorname{ann}(x)$ has smaller "dimension", i.e. its d-th graded part is zero, one can now proceed by induction on d.

(iii) By the validity of (HL) from the inductive hypothesis, the validity of (HR) for any single element $L \in \mathcal{K}$ then implies (HR) for all $L \in \mathcal{K}$. Thus, the last step is to finish the induction by establishing (HR) for a well-chosen $L \in \mathcal{K}$. This is often the most intricate step.

Returning to the case of matroids, step (ii) is provided by Exercise 4.11, which showed that $A^\bullet(\mathrm{M})/\operatorname{ann}(x_F) \simeq A^\bullet(\Sigma_{\mathrm{M}|F} \times \Sigma_{\mathrm{M}/F})$ for a nonempty proper flat F of M. Hence, one can induct on the rank of the matroid. In the case of Theorem 4.15, step (ii) is essentially built into the definition that the stars of Lefschetz fans are Lefschetz, so that one can induct on the dimension of the fan.

In the original proof [AHK18] of Theorem 4.14, in order to carry out step (iii), Adiprasito, Huh, and Katz introduced the notion of "flips," inspired by the proof of the Kähler package for simple polytopes by McMullen [McM93]. This process converts the Bergman fan of Σ_{M} through a sequence of fans until it reaches a fan for which (HR) can be verified easily, and the process is set up such that the validity of (HR) for any one fan in the sequence implies (HR) for all fans in the sequence. Geometrically, one may interpret the process of flips as a combinatorial abstraction of the process of constructing the wonderful compactification W_L as a sequence of blow-ups (Definition 3.5).

Afterwards, it was recognized that a key property of *semi-small maps* [dCM02] inspires a strategy that can greatly simplify step (iii). For a map $f : X \to Y$ of smooth projective varieties, the pullbacks to X of ample divisors on Y generally fail (HL) and (HR) in the cohomology ring of X, since the pullbacks are generally not ample on X but only nef (i.e. is a limit of ample divisors). However, a characterizing property of a semi-small map is that the pullbacks still satisfy (HL) and (HR). This inspires the following approach: One can look for a map $\widetilde{A}^\bullet \to A^\bullet$ of graded algebras, behaving like a pullback along a semi-small map. If (HR) is known to hold for $\widetilde{A}^\bullet$, say by induction, step (iii) would follow.

This insight allowed Braden, Huh, Matherne, Proudfoot, and Wang to give a considerably simplified proof of Theorem 4.14 in [BHM+22]. Using that the deletion $\mathrm{M} \setminus e$ of a matroid M by a non-coloop element e behaves like a semi-small map, they carry out step (iii) by an induction that reduces to the case of Boolean matroids. This insight on semi-small maps is also essential in the proof of Theorem 4.15 by Ardila, Denham, and Huh. A key step [ADH22, Theorem 5.9], building upon the works [Wło97, AKMW02], states that any two fans with the same support can be related by a sequence of edge stellar subdivisions, which are operations on fans that play the role of semi-small maps in toric geometry.

4.3. **Applications of Hodge-Riemann relations in degree 1.** The Kähler package gives rise to log-concave sequences in the following way.

Proposition 4.18. Suppose Σ is a Lefschetz fan of dimension d with a positive Minkowski weight $\deg$. Then, for any nef divisor classes α and β,

$$\text{the sequence } (a_0, a_1, \dots, a_d) \text{ defined by } a_i = \deg(\alpha^{d-i}\beta^i)$$

is log-concave with no internal zeros.

Proof. We may assume that α, β are ample, and show that the sequence is strictly positive and log-concave, since a limit of such sequences is necessarily log-concave with no internal zeros. Strict positivity is then implied by Hodge-Riemann relations (HR) in degree 0. For log-concavity, (HR) in degree 1, with $L_1 \cdots L_{d-2} = \alpha^{d-i-1}\beta^{i-1}$, implies that the symmetric bilinear pairing $A^1(\Sigma) \times A^1(\Sigma) \to \mathbb{R}$ given by $(x, y) \mapsto \deg(xy \cdot \alpha^{d-i-1}\beta^{i-1})$ has at most one positive eigenvalue. This implies that the symmetric matrix

$$\begin{bmatrix} \deg(\alpha^2 \cdot \alpha^{d-i-1}\beta^{i-1}) & \deg(\alpha\beta \cdot \alpha^{d-i-1}\beta^{i-1}) \\ \deg(\alpha\beta \cdot \alpha^{d-i-1}\beta^{i-1}) & \deg(\beta^2 \cdot \alpha^{d-i-1}\beta^{i-1}) \end{bmatrix}$$

cannot be positive definite, but it also cannot be negative definite because all of its entries are positive. Hence, the determinant of the matrix is non-positive, or equivalently, $a_i^2 \geq a_{i-1}a_{i+1}$. □

Returning to showing log-concavity for a matroid M of rank r, one now searches for appropriate divisor classes on the Bergman fan Σ_{M} or the conormal fan $\Sigma_{\mathrm{M},\mathrm{M}^\perp}$. This step benefits heavily from the geometry of realizable matroids explained in Section 3.

The divisor classes for the log-concavity of the sequence (c) come from an involutive symmetry of the permutohedral fan $\Sigma_{A_{n-1}}$ (Example 4.6). As $-\overline{\mathbf{e}}_i = \overline{\mathbf{e}}_{E\setminus i}$, the minus map $x \mapsto -x$ on $\mathbb{R}^E/\mathbb{R}\mathbf{e}_E$ gives an involution of $\Sigma_{A_{n-1}}$. This equips $\Sigma_{A_{n-1}}$ with two distinguished coarsenings to a normal fan of a simplex: Let $\Sigma_\triangle$ (resp. be $\Sigma_\triangledown$) be the fan whose rays are $\{\overline{\mathbf{e}}_i : i \in E\}$ (resp. $\{-\overline{\mathbf{e}}_i : i \in E\}$) and whose cones are generated by any subsets of the rays with cardinality $\leq n-1$. The fan $\Sigma_{A_{n-1}}$ is a common refinement of both the fans $\Sigma_\triangle$ and $\Sigma_\triangledown$.

Let us fix an element $i \in E$, and let $\varphi_\triangle$ be the piecewise-linear function on $\Sigma_\triangle$ given by $\overline{\mathbf{e}}_i \mapsto 1$ and $\overline{\mathbf{e}}_j \mapsto 0$ for all $j \neq i$, which is clearly a convex (but not strictly-convex) function. Restricting this piecewise-linear function to Σ_{M}, it defines a divisor on Σ_{M} whose divisor class is denoted $\alpha \in A^1(\mathrm{M})$. Similarly, starting with the fan $\Sigma_\triangledown$, we obtain a divisor class β. Both α and β are nef, and independent of the choice of $i \in E$ we fixed. Algebraically, we may take $\alpha = \sum_{F \ni i} x_F$ and $\beta = \sum_{F \not\ni i} x_F$ for any choice of $i \in E$. We have the following.

Proposition 4.19. [HK12, Proposition 5.2], [AHK18, Proposition 9.5] With the notations as above, we have

$$\frac{1}{1+q}\mathrm{T}_{\mathrm{M}}(1+q,0) = \sum_{i=0}^{r-1} \deg_{\mathrm{M}}(\alpha^{r-1-i}\beta^i)q^{r-1-i}.$$

Combining the proposition with Theorem 4.14 and Proposition 4.18, we obtain that the coefficients of $\frac{1}{1+q}\mathrm{T}_{\mathrm{M}}(1+q,0)$ is log-concave with no internal zeros, which implies the same property for the coefficients of $\mathrm{T}_{\mathrm{M}}(1+q,0)$, i.e. the log-concavity of the sequence (c). One can also consider other kinds of divisor classes on Σ_{M} and their values under $\deg_{\mathrm{M}}$ to study properties of matroids: See for instance [Eur20, BES, BST, DR22].

Remark 4.20. In geometric terms, the toric variety of $\Sigma_{A_{n-1}}$ is the permutohedral variety defined in Section 3.2.1. That the fan $\Sigma_{A_{n-1}}$ coarsens to Σ_Δ and Σ_∇ gives the two blow-down maps $\pi_1, \pi_2 : X_{A_{n-1}} \to \mathbb{P}^{n-1}$, related by the Cremona transformation, which were described in (†). Thus, the divisor classes α and β here and the computations involving them agree with those described in Section 3.2.1.

For the sequence (d), recall that the conormal fan of a matroid M is a fan in $\mathbb{R}^E/\mathbb{R}\mathbf{e}_E \times \mathbb{R}^E/\mathbb{R}\mathbf{e}_E$ with support equal to the support of $\Sigma_{\mathrm{M}} \times \Sigma_{\mathrm{M}^\perp}$. Let

$p : \mathbb{R}^E/\mathbb{R}\mathbf{e}_E \times \mathbb{R}^E/\mathbb{R}\mathbf{e}_E \to \mathbb{R}^E/\mathbb{R}\mathbf{e}_E$ be the projection to the first factor, and

$s : \mathbb{R}^E/\mathbb{R}\mathbf{e}_E \times \mathbb{R}^E/\mathbb{R}\mathbf{e}_E \to \mathbb{R}^E/\mathbb{R}\mathbf{e}_E$ be the addition map $(x,y) \mapsto x+y$.

By pulling back the piecewise-linear function φ_Δ on $\mathbb{R}^E/\mathbb{R}\mathbf{e}_E$ along these two maps, we obtain two divisor classes γ and δ.

Proposition 4.21. [ADH22, Theorem 1.2] With the notations as above, we have

$$\frac{1}{q}\mathrm{T}_{\mathrm{M}}(q,0) = \sum_{i=0}^{r-1} \deg_{\mathrm{M},\mathrm{M}^\perp}(\gamma^{r-1-i}\delta^{n-r-1+i})q^{r-1-i}.$$

Combining the proposition with Corollary 4.16 and Proposition 4.18, we conclude the log-concavity of the sequence (d). The proposition was proved via an intricate combinatorics of biflags in [ADH22] and the Chern-Schwartz-MacPherson classes of matroids [LdMRS20]. By developing a new framework of *tautological classes of matroids*, Berget, Spink, Tseng and the author proved a formula [BEST, Theorem A & Theorem 9.7] that contains both Proposition 4.19 and Proposition 4.21 as special cases.

Note that only a part of the Kähler package, the Hodge-Riemann relations in degrees at most 1, was all that is required for concluding log-concavity. Extracting the essence of the analytic properties behind (HR) in degrees at most 1 leads to the fascinating theory of **Lorentzian polynomials** [BH20] and (equivalently) *completely log-concave polynomials* [ALOGV18]. One powerful feature of this theory is that it often allows one to reduce to "dimension 2" cases, mirroring the feature in classical algebraic geometry that (HR) in degree 1 can be reduced to the case of surfaces. For instance, by reducing to an analysis of rank 2 matroids, Brändén and Huh [BH20] and independently Anari, Liu, Oveis-Gharan, and Vinzant [ALOGV18] proved that the sequence (a) is in fact *ultra*-log-concave, in the sense that

$$\frac{I_i^2}{\binom{n}{i}^2} \geq \frac{I_{i-1}I_{i+1}}{\binom{n}{i-1}\binom{n}{i+1}} \quad \text{for all } i,$$

which was conjectured by Mason [Mas72]. Moreover, by analyzing rank 3 matroids (cf. Exercise 4.17), one can use Lorentzian polynomials to give a simplified proof of the log-concavity of the sequence (c) [BES, BL]. We point to [Huh22, Section 2] for a survey of Lorentzian polynomials and their applications.

5. Intersection cohomology of a matroid

We describe the intersection cohomology of a matroid, and its role in showing the top-heaviness of the sequence (e). We begin by considering the following graded algebra.

Definition 5.1. For a matroid M of rank r, its **Möbius algebra** a graded $\mathbb{R}$-algebra $B^\bullet(\mathrm{M}) = \bigoplus_{i=0}^r B^i(\mathrm{M})$ where $B^i(\mathrm{M})$ has basis $\{y_F : F \text{ a rank } i \text{ flat of } \mathrm{M}\}$ for each $0 \leq i \leq r$, and multiplication is given by

$$y_F \cdot y_{F'} = \begin{cases} y_{F\vee F'} & \text{if } \mathrm{rk}_\mathrm{M}(F) + \mathrm{rk}_\mathrm{M}(F') = \mathrm{rk}_\mathrm{M}(F \vee F') \\ 0 & \text{otherwise.} \end{cases}$$

A strategy for the top-heaviness is to show that there is an injective linear map $B^i(\mathrm{M}) \to B^{r-i}(\mathrm{M})$ for every $i \leq r/2$. The statement of the hard Lefschetz property inspires a candidate for such a map: the multiplication by a power of an element in $B^1(\mathrm{M})$. We then immediately face the difficulty that $B^\bullet(\mathrm{M})$ usually cannot satisfy Poincaré duality (PD) or the hard Lefschetz property (HL), since the sequence $(W_0, \dots, W_r)$ of the dimensions of graded pieces is usually not symmetric.

The intersection cohomology $IH^\bullet(\mathrm{M})$, introduced in [BHM+], is a graded vector space containing $B^\bullet(\mathrm{M})$ that "most efficiently" amends the failure of (PD) and (HL). We give a broad outline of their construction and their properties. The following remark explains some geometric motivation.

Remark 5.2. Recall from Section 3.3 the matroid Schubert variety Y_L of a realization $L \subseteq \mathbb{C}^E$ of a matroid M. One deduces from Theorem 3.8 that the algebra $B^\bullet(\mathrm{M})$ is the cohomology ring (in even degrees) of the *matroid Schubert variety* Y_L (see [HW17, Theorem 14]). The variety Y_L is usually quite singular, which witnesses the failure of (PD) and (HL) for $B^\bullet(\mathrm{M})$. Motivated by the proof of Theorem 3.7, one seeks to understand the intersection cohomology $IH^\bullet(Y_L)$, which contains $B^\bullet(\mathrm{M})$ as a subalgebra.

To do so, let $f : X \to Y_L$ be a resolution of singularities of Y_L. The decomposition theorem of Beilinson, Bernstein, Deligne, and Gabber [BBD82] implies that $H^\bullet(X)$ can be decomposed into a direct sum of $B^\bullet(\mathrm{M})$-modules, and that $IH^\bullet(Y_L)$ is a direct summand. In general, computing these decompositions to get a handle on $IH^\bullet(Y_L)$ can be intractible, but for Y_L there is a resolution $f : \widetilde{Y}_L \to Y_L$ by the *augmented wonderful variety* $\widetilde{Y}_L$ *of* L [BHM+22] (see also [EHL]), whose cohomology ring $H^\bullet(\widetilde{Y}_L)$ and the injection $B^\bullet(\mathrm{M}) \hookrightarrow H^\bullet(\widetilde{Y}_L)$ have explicit combinatorial descriptions in terms of the matroid M.

We first find a bigger graded $\mathbb{R}$-algebra containing $B^\bullet(\mathrm{M})$ that satisfies the Kähler package. The **augmented Bergman fan** [BHM+22, Definition 2.4] of a matroid M of rank r is an r-dimensional fan in $\mathbb{R}^E$ closely related to the Bergman fan Σ_M. Its Chow ring has the following explicit description.

Definition 5.3. The **augmented Chow ring** (with real coefficients) of a matroid M is the graded $\mathbb{R}$-algebra

$$\mathrm{CH}^\bullet(\mathrm{M}) = \frac{\mathbb{R}[y_i, x_F : i \in E,\ F \text{ a (possibly empty) proper flat of } \mathrm{M}]}{\widetilde{I}_\mathrm{M} + \widetilde{J}_\mathrm{M}}$$

where $\widetilde{I}_\mathrm{M}$ and $\widetilde{J}_\mathrm{M}$ are the ideals

$$\widetilde{I}_\mathrm{M} = \Big\langle x_F x_{F'} : F \not\subseteq F' \text{ and } F \not\supseteq F' \Big\rangle + \Big\langle y_i x_F : i \notin F \Big\rangle \quad \text{and}$$

$$\widetilde{J}_\mathrm{M} = \Big\langle y_i - \sum_{F \not\ni i} x_F : i \in E \Big\rangle.$$

The augmented Chow ring has the following useful features:

- The assignment $y_F \mapsto \prod_{i \in F} y_i$ defines an injection $B^\bullet(\mathrm{M}) \hookrightarrow \mathrm{CH}^\bullet(\mathrm{M})$ of graded $\mathbb{R}$-algebras [BHM+22, Proposition 2.28].
- Theorem 4.14 combined with Theorem 4.15 implies that the augmented Bergman fan is Lefschetz, because the support of the augmented Bergman fan of M can be identified with the support of the usual Bergman fan of the *free co-extension matroid* of M (see [EHL, Section 5.3]). Thus, the Chow ring $\mathrm{CH}^\bullet(\mathrm{M})$ satisfies the Kähler package.

Thus, we have found a bigger algebra containing $B^\bullet(\mathrm{M})$ that satisfies the Kähler package. However, we are not done because $\mathrm{CH}^\bullet(\mathrm{M})$ is "too big": To conclude injectivity properties for $B^\bullet(\mathrm{M})$, we need the graded linear operators $\mathcal{K}$ satisfying (HL) on $\mathrm{CH}^\bullet(\mathrm{M})$ to come from $B^1(\mathrm{M})$, but this is almost never the case—a positive linear combination of the y_i's usually does not satisfy (HL) on $\mathrm{CH}^\bullet(\mathrm{M})$. One instead must consider the following $B^\bullet(\mathrm{M})$-submodule of $\mathrm{CH}^\bullet(\mathrm{M})$ that "most efficiently" repairs the the failure of (HL) on $B^\bullet(\mathrm{M})$.

Definition 5.4. Up to isomorphism there is a unique indecomposable $B^\bullet(\mathrm{M})$-module direct summand of $\mathrm{CH}^\bullet(\mathrm{M})$ containing $B^\bullet(\mathrm{M})$. This direct summand is the **intersection cohomology** $IH^\bullet(\mathrm{M})$ of M.

In fact, the authors of [BHM+] establish a canonical decomposition of $\mathrm{CH}^\bullet(\mathrm{M})$ as a $B^\bullet(\mathrm{M})$-module, and identify the direct summand $IH^\bullet(\mathrm{M})$. This decomposition along with the Kähler package for $\mathrm{CH}^\bullet(\mathrm{M})$ is then fed into a highly intricate version of the general strategy for establishing the Kähler package outlined in Section 4.2, resulting in the following main theorem.

Theorem 5.5. [BHM+, Theorem 1.6] The intersection cohomology $IH^\bullet(\mathrm{M})$ of a matroid M satisfies the Kähler package with $\mathcal{K} = \{\sum_{e \in E} c_e y_e : c_e > 0\}$.

As a corollary, for any positive linear combination $\ell = \sum_{e\in E} c_e y_e$ and $0 \leq i \leq j \leq r-i$, we have a commuting diagram

$$\begin{array}{ccc} B^i(\mathrm{M}) & \hookrightarrow & IH^i(\mathrm{M}) \\ {\scriptstyle \cdot\ell^{j-i}}\downarrow & & \downarrow{\scriptstyle \cdot\ell^{j-i}} \\ B^j(\mathrm{M}) & \hookrightarrow & IH^j(\mathrm{M}) \end{array}$$

where the right vertical map is injective by the hard Lefschetz property of $IH^\bullet(\mathrm{M})$. The left vertical map is thus injective, and the desired top-heaviness of the sequence (e) follows.

6. Conclusion

Matroids are combinatorial structures that capture the essence of independence. There were several conjectures about the behavior of sequences of invariants of a matroid, involving log-concavity or top-heaviness. June Huh and his collaborators made fundamental contribution to matroid theory [AHK18, ADH22, BHM$^+$], resolving many of these conjectures. They began by answering the conjectures for realizable matroids using algebraic geometry, a significant step on its own. Then, with considerable effort, they were able to extract the combinatorial heart, and establish Hodge-theoretic properties for arbitrary, not necessarily realizable, matroids. This development of the Hodge theory of matroids forms an integral part of the foundation for studying matroids from an algebro-geometric perspective.

References

[AA] Marcelo Aguiar and Federico Ardila. Hopf monoids and generalized permutahedra. *Memoirs of the American Mathematical Society (to appear)*. 18

[AB16] Federico Ardila and Adam Boocher. The closure of a linear space in a product of lines. *J. Algebraic Combin.*, 43(1):199–235, 2016. 16

[ADH22] Federico Ardila, Graham Denham, and June Huh. Lagrangian geometry of matroids. *J. Amer. Math. Soc.*, to appear. DOI: https://doi.org/10.1090/jams/1009, 2022. 3, 4, 9, 17, 19, 20, 22, 23, 26

[AHK18] Karim Adiprasito, June Huh, and Eric Katz. Hodge theory for combinatorial geometries. *Ann. of Math. (2)*, 188(2):381–452, 2018. 3, 4, 9, 19, 20, 21, 23, 26

[AK06] Federico Ardila and Caroline J. Klivans. The Bergman complex of a matroid and phylogenetic trees. *J. Combin. Theory Ser. B*, 96(1):38–49, 2006. 18

[AKMW02] Dan Abramovich, Kalle Karu, Kenji Matsuki, and Jarosław Włodarczyk. Torification and factorization of birational maps. *J. Amer. Math. Soc.*, 15(3):531–572, 2002. 22

[ALOGV18] Nima Anari, Kuikui Liu, Shayan Oveis Gharan, and Cynthia Vinzant. Log-concave polynomials ii: high-dimensional walks and an fpras for counting bases of a matroid. *https://arxiv.org/abs/1811.01816*, 2018. 23, 24

[Alu99] Paolo Aluffi. Differential forms with logarithmic poles and Chern-Schwartz-MacPherson classes of singular varieties. *C. R. Acad. Sci. Paris Sér. I Math.*, 329(7):619–624, 1999. 15

[Alu05] Paolo Aluffi. Characteristic classes of singular varieties. In *Topics in cohomological studies of algebraic varieties*, Trends Math., pages 1–32. Birkhäuser, Basel, 2005. 14

[AP] Omid Amini and Matthieu Piquerez. Hodge theory for tropical varieties. arXiv:2007.07826. 4

[Ard22] Federico Ardila. The geometry of geometries: matroid theory, old and new. arXiv:2111.08726, 2022. 4, 5

[Bak18] Matthew Baker. Hodge theory in combinatorics. *Bull. Amer. Math. Soc. (N.S.)*, 55(1):57–80, 2018. 5

[BBD82] A. A. Beĭlinson, J. Bernstein, and P. Deligne. Faisceaux pervers. In *Analysis and topology on singular spaces, I (Luminy, 1981)*, volume 100 of *Astérisque*, pages 5–171. Soc. Math. France, Paris, 1982. 15, 24

[BE09] Anders Björner and Torsten Ekedahl. On the shape of Bruhat intervals. *Ann. of Math. (2)*, 170(2):799–817, 2009. 15, 16

[BES] Spencer Backman, Christopher Eur, and Connor Simpson. Simplicial generation of Chow rings of matroids. (to appear). arXiv:1905:07114. 21, 23, 24

[BEST] Andrew Berget, Christopher Eur, Hunter Spink, and Dennis Tseng. Tautological classes of matroids. arXiv:2103.08021. 9, 14, 23

[BH20] Petter Brändén and June Huh. Lorentzian polynomials. *Ann. of Math. (2)*, 192(3):821–891, 2020. 23, 24

[BHM$^+$] Tom Braden, June Huh, Jacob Matherne, Nicholas Proudfoot, and Botong Wang. Singular Hodge theory for combinatorial geometries. arXiv:2010.06088. 3, 4, 9, 24, 25, 26

[BHM$^+$22] Tom Braden, June Huh, Jacob P. Matherne, Nicholas Proudfoot, and Botong Wang. A semi-small decomposition of the Chow ring of a matroid. *Adv. Math.*, 409:Paper No. 108646, 49, 2022. 3, 21, 25

[BL] Petter Brändén and Jonathan Leake. Lorentzian polynomials on cones and the heron-rota-welsh conjecture. arXiv:2110.00487. 24

[BO92] Thomas Brylawski and James Oxley. The Tutte polynomial and its applications. In *Matroid applications*, volume 40 of *Encyclopedia Math. Appl.*, pages 123–225. Cambridge Univ. Press, Cambridge, 1992. 7

[Bre94] Francesco Brenti. Log-concave and unimodal sequences in algebra, combinatorics, and geometry: an update. In *Jerusalem combinatorics '93*, volume 178 of *Contemp. Math.*, pages 71–89. Amer. Math. Soc., Providence, RI, 1994. 2

[Bri96] Michel Brion. Piecewise polynomial functions, convex polytopes and enumerative geometry. pages 25–44, 1996. 17

[Bry77] Tom Brylawski. The broken-circuit complex. *Trans. Amer. Math. Soc.*, 234(2):417–433, 1977. 10

[Bry82] Thomas Brylawski. The Tutte polynomial. I. General theory. In *Matroid theory and its applications*, pages 125–275. Liguori, Naples, 1982. 9

[BST] Andrew Berget, Hunter Spink, and Dennis Tseng. Log-concavity of matroid h-vectors and mixed eulerian numbers. *https://arxiv.org/abs/2005.01937*. 23

[Cat08] Eduardo Cattani. Mixed Lefschetz theorems and Hodge-Riemann bilinear relations. *Int. Math. Res. Not. IMRN*, (10):Art. ID rnn025, 20, 2008. 20

[CDFV11] D. Cohen, G. Denham, M. Falk, and A. Varchenko. Critical points and resonance of hyperplane arrangements. *Canad. J. Math.*, 63(5):1038–1057, 2011. 14

[CLS11] David A. Cox, John B. Little, and Henry K. Schenck. *Toric varieties*, volume 124 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2011. 15

[Cra69] Henry H. Crapo. The Tutte polynomial. *Aequationes Math.*, 3:211–229, 1969. 7

[Dan78] V. I. Danilov. The geometry of toric varieties. *Uspekhi Mat. Nauk*, 33(2(200)):85–134, 247, 1978. 17

[Daw84] Jeremy E. Dawson. A collection of sets related to the Tutte polynomial of a matroid. In *Graph theory, Singapore 1983*, volume 1073 of *Lecture Notes in Math.*, pages 193–204. Springer, Berlin, 1984. 9

[dBE48] N. G. de Bruijn and P. Erdös. On a combinatorial problem. *Nederl. Akad. Wetensch., Proc.*, 51:1277–1279 = Indagationes Math. 10, 421–423 (1948), 1948. 9

[dCM02] Mark Andrea A. de Cataldo and Luca Migliorini. The hard Lefschetz theorem and the topology of semismall maps. *Ann. Sci. École Norm. Sup. (4)*, 35(5):759–772, 2002. 21

[dCM09] Mark Andrea A. de Cataldo and Luca Migliorini. The decomposition theorem, perverse sheaves and the topology of algebraic maps. *Bull. Amer. Math. Soc. (N.S.)*, 46(4):535–633, 2009. 20

[DCP95] C. De Concini and C. Procesi. Wonderful models of subspace arrangements. *Selecta Math. (N.S.)*, 1(3):459–494, 1995. 12, 19

[Del71] Pierre Deligne. Théorie de Hodge. II. *Inst. Hautes Études Sci. Publ. Math.*, (40):5–57, 1971. 16

[DGS12] Graham Denham, Mehdi Garrousian, and Mathias Schulze. A geometric deletion-restriction formula. *Adv. Math.*, 230(4-6):1979–1994, 2012. 14

[Dim17] Alexandru Dimca. *Hyperplane arrangements*. Universitext. Springer, Cham, 2017. An introduction. 11

[DP03] Alexandru Dimca and Stefan Papadima. Hypersurface complements, Milnor fibers and higher homotopy groups of arrangments. *Ann. of Math. (2)*, 158(2):473–507, 2003. 13

[DR22] Jeshu Dastidar and Dustin Ross. Matroid psi classes. *Selecta Math. (N.S.)*, 28(3):Paper No. 55, 38, 2022. 23

[DW74] Thomas A. Dowling and Richard M. Wilson. The slimmest geometric lattices. *Trans. Amer. Math. Soc.*, 196:203–215, 1974. 9

[DW75] Thomas A. Dowling and Richard M. Wilson. Whitney number inequalities for geometric lattices. *Proc. Amer. Math. Soc.*, 47:504–512, 1975. 9

[EH16] David Eisenbud and Joe Harris. *3264 and all that—a second course in algebraic geometry*. Cambridge University Press, Cambridge, 2016. 14

[EHL] Christopher Eur, June Huh, and Matt Larson. Stellahedral geometry of matroids. arXiv:2207.10605. 15, 25

[EPW16] Ben Elias, Nicholas Proudfoot, and Max Wakefield. The Kazhdan-Lusztig polynomial of a matroid. *Adv. Math.*, 299:36–70, 2016. 4, 16

[Eur20] Christopher Eur. Divisors on matroids and their volumes. *J. Combin. Theory Ser. A*, 169:105135, 31, 2020. 23

[EW14] Ben Elias and Geordie Williamson. The Hodge theory of Soergel bimodules. *Ann. of Math. (2)*, 180(3):1089–1136, 2014. 20

[FS97] William Fulton and Bernd Sturmfels. Intersection theory on toric varieties. *Topology*, 36(2):335–353, 1997. 19

[Ful93] William Fulton. *Introduction to toric varieties*, volume 131 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 1993. The William H. Roever Lectures in Geometry. 16

[Ful98] William Fulton. *Intersection theory*, volume 2 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]*. Springer-Verlag, Berlin, second edition, 1998. 10

[FY04] Eva Maria Feichtner and Sergey Yuzvinsky. Chow rings of toric varieties defined by atomic lattices. *Invent. Math.*, 155(3):515–536, 2004. 4, 19

[GGMS87] I. M. Gelfand, R. M. Goresky, R. D. MacPherson, and V. V. Serganova. Combinatorial geometries, convex polyhedra, and Schubert cells. *Adv. in Math.*, 63(3):301–316, 1987. 4

[GM83] Mark Goresky and Robert MacPherson. Intersection homology. II. *Invent. Math.*, 72(1):77–129, 1983. 15

[Har77] Robin Hartshorne. *Algebraic geometry*. Springer-Verlag, New York-Heidelberg, 1977. Graduate Texts in Mathematics, No. 52. 12, 14

[Her72] A. P. Heron. Matroid polynomials. pages 164–202, 1972. 9

[Hib89] Takayuki Hibi. Flawless O-sequences and Hilbert functions of Cohen-Macaulay integral domains. *J. Pure Appl. Algebra*, 60(3):245–251, 1989. 9

[Hib92] Takayuki Hibi. Face number inequalities for matroid complexes and Cohen-Macaulay types of Stanley-Reisner rings of distributive lattices. *Pacific J. Math.*, 154(2):253–264, 1992. 9

[HK12] June Huh and Eric Katz. Log-concavity of characteristic polynomials and the Bergman fan of matroids. *Math. Ann.*, 354(3):1103–1116, 2012. 3, 13, 23

[Hog74] S. G. Hoggar. Chromatic polynomials and logarithmic concavity. *J. Combinatorial Theory Ser. B*, 16:248–254, 1974. 9

[HS14] June Huh and Bernd Sturmfels. Likelihood geometry. In *Combinatorial algebraic geometry*, volume 2108 of *Lecture Notes in Math.*, pages 63–117. Springer, Cham, 2014. 14

[Huh12] June Huh. Milnor numbers of projective hypersurfaces and the chromatic polynomial of graphs. *J. Amer. Math. Soc.*, 25(3):907–927, 2012. 3, 13

[Huh13] June Huh. The maximum likelihood degree of a very affine variety. *Compos. Math.*, 149(8):1245–1266, 2013. 14

[Huh15] June Huh. h-vectors of matroids and logarithmic concavity. *Adv. Math.*, 270:49–59, 2015. 3, 14

[Huh18] June Huh. Tropical geometry of matroids. *Current Developments in Mathematics 2016*, pages 1–46, 2018. 5, 12

[Huh22] June Huh. Combinatorics and hodge theory. 2022. 24

[Huy05] Daniel Huybrechts. *Complex geometry*. Universitext. Springer-Verlag, Berlin, 2005. An introduction. 4

[HW17] June Huh and Botong Wang. Enumeration of points, lines, planes, etc. *Acta Math.*, 218(2):297–317, 2017. 3, 16, 24

[IKMZ19] Ilia Itenberg, Ludmil Katzarkov, Grigory Mikhalkin, and Ilia Zharkov. Tropical homology. *Math. Ann.*, 374(1-2):963–1006, 2019. 4

[JKL18] Martina Juhnke-Kubitzke and Dinh Van Le. Flawlessness of h-vectors of broken circuit complexes. *Int. Math. Res. Not. IMRN*, (5):1347–1367, 2018. 9, 10

[Kat16] Eric Katz. Matroid theory for algebraic geometers. pages 435–517, 2016. 5

[KP08] Eric Katz and Sam Payne. Piecewise polynomials, Minkowski weights, and localization on toric varieties. *Algebra Number Theory*, 2(2):135–155, 2008. 19

[Laf03] L. Lafforgue. *Chirurgie des grassmanniennes*, volume 19 of *CRM Monograph Series*. American Mathematical Society, Providence, RI, 2003. 4

[Laz04a] Robert Lazarsfeld. *Positivity in algebraic geometry. I*, volume 48 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]*. Springer-Verlag, Berlin, 2004. Classical setting: line bundles and linear series. 11

[Laz04b] Robert Lazarsfeld. *Positivity in algebraic geometry. II*, volume 49 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]*. Springer-Verlag, Berlin, 2004. Positivity for vector bundles, and multiplier ideals. 14

[LdMRS20] Lucía López de Medrano, Felipe Rincón, and Kristin Shaw. Chern-Schwartz-MacPherson cycles of matroids. *Proc. Lond. Math. Soc. (3)*, 120(1):1–27, 2020. 23

[Len13] Matthias Lenz. The f-vector of a representable-matroid complex is log-concave. *Adv. in Appl. Math.*, 51(5):543–545, 2013. 10

[Mac74] R. D. MacPherson. Chern classes for singular algebraic varieties. *Ann. of Math. (2)*, 100:423–432, 1974. 14

[Mas72] J. H. Mason. Matroids: unimodal conjectures and Motzkin's theorem. pages 207–220, 1972. 9, 24

[McM93] Peter McMullen. On simple polytopes. *Invent. Math.*, 113(2):419–444, 1993. 17, 20, 21

[MS15] Diane Maclagan and Bernd Sturmfels. *Introduction to tropical geometry*, volume 161 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2015. 19

[Nel18] Peter Nelson. Almost all matroids are nonrepresentable. *Bull. Lond. Math. Soc.*, 50(2):245–248, 2018. 3

[OS80] Peter Orlik and Louis Solomon. Combinatorics and topology of complements of hyperplanes. *Invent. Math.*, 56(2):167–189, 1980. 11

[OT92] Peter Orlik and Hiroaki Terao. *Arrangements of hyperplanes*, volume 300 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1992. 11

[Oxl11] James Oxley. *Matroid theory*, volume 21 of *Oxford Graduate Texts in Mathematics*. Oxford University Press, Oxford, 2 edition, 2011. 5

[Pro18] Nicholas Proudfoot. The algebraic geometry of Kazhdan-Lusztig-Stanley polynomials. *EMS Surv. Math. Sci.*, 5(1-2):99–127, 2018. 4

[PS06] Nicholas Proudfoot and David Speyer. A broken circuit ring. *Beiträge Algebra Geom.*, 47(1):161–166, 2006. 13

[PXY18] Nicholas Proudfoot, Yuan Xu, and Ben Young. The Z-polynomial of a matroid. *Electron. J. Combin.*, 25(1):Paper No. 1.26, 21, 2018. 16

[Rea68] Ronald C. Read. An introduction to chromatic polynomials. *J. Combinatorial Theory*, 4:52–71, 1968. 9

[Rot71] Gian-Carlo Rota. Combinatorial theory, old and new. In *Actes du Congrès International des Mathématiciens (Nice, 1970), Tome 3*, pages 229–233. Gauthier-Villars, Paris, 1971. 9

[Sch93] Werner Schwärzler. The coefficients of the Tutte polynomial are not unimodal. *J. Combin. Theory Ser. B*, 58(2):240–242, 1993. 10

[Spe08] David E. Speyer. Tropical linear spaces. *SIAM J. Discrete Math.*, 22(4):1527–1558, 2008. 18

[Sta80] Richard P. Stanley. The number of faces of a simplicial convex polytope. *Adv. in Math.*, 35(3):236–238, 1980. 17

[Sta89] Richard P. Stanley. Log-concave and unimodal sequences in algebra, combinatorics, and geometry. In *Graph theory and its applications: East and West (Jinan, 1986)*, volume 576 of *Ann. New York Acad. Sci.*, pages 500–535. New York Acad. Sci., New York, 1989. 2

[Stu02] Bernd Sturmfels. *Solving systems of polynomial equations*, volume 97 of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, Washington, DC; by the American Mathematical Society, Providence, RI, 2002. 18

[Swa03] Ed Swartz. g-elements of matroid complexes. *J. Combin. Theory Ser. B*, 88(2):369–375, 2003. 9

[Ter02] Hiroaki Terao. Algebras generated by reciprocals of linear forms. *J. Algebra*, 250(2):549–558, 2002. 13

[Tev07] Jenia Tevelev. Compactifications of subvarieties of tori. *Amer. J. Math.*, 129(4):1087–1104, 2007. 19

[Tut67] W. T. Tutte. On dichromatic polynominals. *J. Combinatorial Theory*, 2:301–320, 1967. 7

[Var95] A. Varchenko. Critical points of the product of powers of linear functions and families of bases of singular vectors. *Compositio Math.*, 97(3):385–401, 1995. 14

[Voi02] Claire Voisin. *Hodge theory and complex algebraic geometry. I*, volume 76 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. Translated from the French original by Leila Schneps. 4

[Web04] Andrzej Weber. Pure homology of algebraic varieties. *Topology*, 43(3):635–644, 2004. 16

[Wel71] D. J. A. Welsh. Combinatorial problems in matroid theory. In *Combinatorial Mathematics and its Applications (Proc. Conf., Oxford, 1969)*, pages 291–306. Academic Press, London, 1971. 9

[Wel76] D. J. A. Welsh. *Matroid theory*. Academic Press [Harcourt Brace Jovanovich, Publishers], London-New York, 1976. L. M. S. Monographs, No. 8. 5, 9

[Whi32] Hassler Whitney. The coloring of graphs. *Ann. of Math. (2)*, 33(4):688–718, 1932. 1

[Wło97] Jarosław Włodarczyk. Decomposition of birational toric maps in blow-ups & blow-downs. *Trans. Amer. Math. Soc.*, 349(1):373–411, 1997. 22

[Zas87] Thomas Zaslavsky. The Möbius function and the characteristic polynomial. In *Combinatorial geometries*, volume 29 of *Encyclopedia Math. Appl.*, pages 114–138. Cambridge Univ. Press, Cambridge, 1987. 13

[Zie95] Günter M. Ziegler. *Lectures on polytopes*, volume 152 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1995. 16

HARVARD UNIVERSITY
Email address: `ceur@math.harvard.edu`

FROM SPHERE PACKING TO FOURIER INTERPOLATION

HENRY COHN

Abstract. Viazovska's solution of the sphere packing problem in eight dimensions is based on a remarkable construction of certain special functions using modular forms. Great mathematics has consequences far beyond the problems that originally inspired it, and Viazovska's work is no exception. In this article, we'll examine how it has led to new interpolation theorems in Fourier analysis, specifically a theorem of Radchenko and Viazovska.

1. Sphere packing

The sphere packing problem asks how densely congruent spheres can be packed in Euclidean space. In other words, what fraction of space can be filled with congruent balls, if their interiors are required to be disjoint?[1] Everyone can pack spheres intuitively in low dimensions: the optimal two-dimensional packing is a hexagonal arrangement, and optimal three-dimensional packings are stacks of optimal two-dimensional layers, nestled together as closely as possible into the gaps in the layers (see Figure 1.1).

In fact, these packings are known to be optimal. The two-dimensional problem was solved by Thue [37, 38], with a more modern proof by Fejes Tóth [19], and the three-dimensional problem was solved by Hales [21]. The two-dimensional proof is not so complicated, but the three-dimensional proof is difficult to check, because it relies on both enormous machine calculations and lengthy human arguments in a sequence of papers. To give a definitive demonstration of its correctness, Hales and a team of collaborators have produced a formally verified proof [22], i.e., a proof that has been algorithmically verified using formal logic.

On the one hand, the solution of the three-dimensional sphere packing problem is a triumph of modern mathematics, a demonstration of humanity's ability to overcome even tremendously challenging obstacles. On the other hand, to a general audience it can sound like a parody of pure mathematics, in which mathematicians devote immense efforts to proving an intuitively obvious assertion. It's natural to feel discouraged about the future of a subfield in which it's easy to guess the answer and almost impossible to prove it. For comparison, a rigorous solution of the four-dimensional sphere packing problem remains far out of reach. If the difficulty increases as much from three to four dimensions as it did from two to three, then humanity may never see a proof.

One noteworthy change as we move to higher dimensions is that we lose much of our intuition, and the answer is no longer easy to guess. For example, it is not always true that we can obtain an optimal packing in $\mathbb{R}^n$ by stacking optimal

2020 *Mathematics Subject Classification.* Primary 52C17, 42A15.

[1] To make this question precise, one should take the limit as $r \to \infty$ of the density for packing unit spheres in a sphere of radius r, or a cube of side length r. One obtains the same limit for any reasonable container (see, for example, [6]).

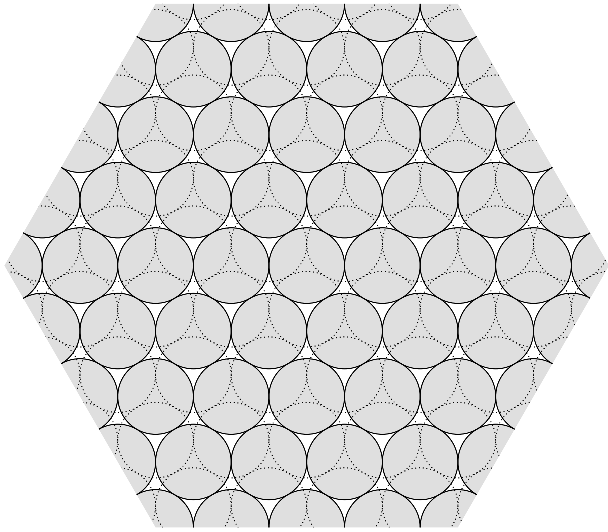


Figure 1.1. A two-dimensional cross section of an optimal three-dimensional sphere packing, with dotted lines indicating spheres in an adjacent layer.

$(n-1)$-dimensional layers (see [14] for details). In sufficiently high dimensions, there are no conjectures for optimal packings, the best upper and lower bounds known for the packing density differ by an exponential factor in the dimension, and we cannot even predict whether the densest packings should be crystalline or disordered. In short, we know shockingly little about how spherical particles behave in high dimensions. Of course this means there are plenty of intriguing phenomena to explore.

Certain dimensions stand out in the midst of this ignorance as having exceptionally dense packings. The most amazing of all are eight and twenty-four dimensions, which feature the E_8 root lattice and the Leech lattice Λ_{24}. (We will not construct these lattices here; see [18, 36, 15] for constructions.) Recall that a lattice in $\mathbb{R}^n$ is just a discrete subgroup of rank n; in other words, for each basis $v_1, \dots, v_n$ of $\mathbb{R}^n$, the set

$$\{a_1 v_1 + \cdots + a_n v_n : a_1, \dots, a_n \in \mathbb{Z}\}$$

is a lattice. Every lattice leads to a sphere packing by centering congruent spheres at the lattice points, with the radius chosen as large as possible without overlap. Lattice packings are common in low dimensions, but there is no reason to expect an optimal packing to have this sort of algebraic structure in general. For example, in $\mathbb{R}^{10}$ the best packing known, the aptly named Best packing [4], has density more than 8% greater than any known lattice packing in $\mathbb{R}^{10}$. By contrast, the E_8 and Leech lattices yield impressively dense packings with extraordinary symmetry groups, and their density and symmetry are so far out of the ordinary that it is difficult to imagine how they could be improved.

In 2016 Maryna Viazovska [39] solved the sphere packing problem in $\mathbb{R}^8$ with an innovative use of modular forms, which was soon extended to $\mathbb{R}^{24}$ as well [12]; both E_8 and the Leech lattice do indeed turn out to be optimal sphere packings. These are the only cases in which the sphere packing problem has been solved above three dimensions. Although the proofs require more machinery than those in two or three dimensions, most notably the theory of modular forms, they are much shorter and simpler than one might fear. Viazovska's proof dispelled the gloomy possibility that higher-dimensional sphere packing could be beyond human understanding, and she was awarded a Fields medal in 2022 for this line of work.

In addition to her breakthrough in sphere packing, Viazovska's modular form techniques have led to unexpected consequences, such as interpolation theorems showing that a radial function f can be reconstructed from the values of f and its Fourier transform $\widehat{f}$ on certain discrete sets of points. Although Fourier interpolation may sound rather far afield from sphere packing, it turns out to be closely connected. In this article, we'll explore how Viazovska's work led to this connection and how to prove a fundamental interpolation theorem of Radchenko and Viazovska [33]. For comparison, [16], [8], [40], [41], and [9] are expositions of her work that focus on other themes.

2. From sphere packing to Fourier analysis

The connection between packing problems and Fourier analysis originated in the work of Delsarte [17] on linear programming bounds for error-correcting codes. For sphere packings in Euclidean space, a continuous analogue of Delsarte's work was developed by Cohn and Elkies [10]. The quality of this bound depends on the choice of an auxiliary function satisfying certain inequalities, and Viazovska's breakthrough amounted to figuring out how to optimize that choice.

We will normalize the Fourier transform of an integrable function $f\colon \mathbb{R}^n \to \mathbb{C}$ by

$$\widehat{f}(y) = \int_{\mathbb{R}^n} f(x) e^{-2\pi i \langle x,y\rangle}\, dx,$$

where $\langle \cdot, \cdot \rangle$ denotes the usual inner product on $\mathbb{R}^n$. We'll generally restrict our attention to Schwartz functions, i.e., infinitely differentiable functions f such that for all real numbers $c > 0$ and nonnegative integers $i_1, \dots, i_n$,

$$\left| \frac{\partial^{i_1+\dots+i_n}}{\partial x_1^{i_1} \cdots \partial x_n^{i_n}} f(x_1, \dots, x_n) \right| = O(|x|^{-c})$$

as $|x| \to \infty$. These smoothness and decay conditions can be somewhat weakened in each application below, but Schwartz functions are the best-behaved case. We'll also frequently study radial functions, i.e., functions f for which $f(x)$ depends only on $|x|$, in which case we will write $f(r)$ for $r \in [0, \infty)$ to denote the value $f(x)$ with $|x| = r$ and f' for the radial derivative of f. Note that the spaces of radial functions and of Schwartz functions are both preserved by the Fourier transform.

The linear programming bound is the following method for producing a density bound from a suitable auxiliary function f. The name "linear programming bound" refers to the fact that optimizing this bound can be recast as an infinite-dimensional linear programming problem (i.e., linear optimization problem).

Theorem 2.1 (Cohn and Elkies [10])**.** *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a radial Schwartz function and r a positive real number such that*

(1) $f(x) \le 0$ *whenever* $|x| \ge r$,
(2) $\widehat{f}(y) \ge 0$ *for all* y, *and*
(3) $f(0) = \widehat{f}(0) = 1$.

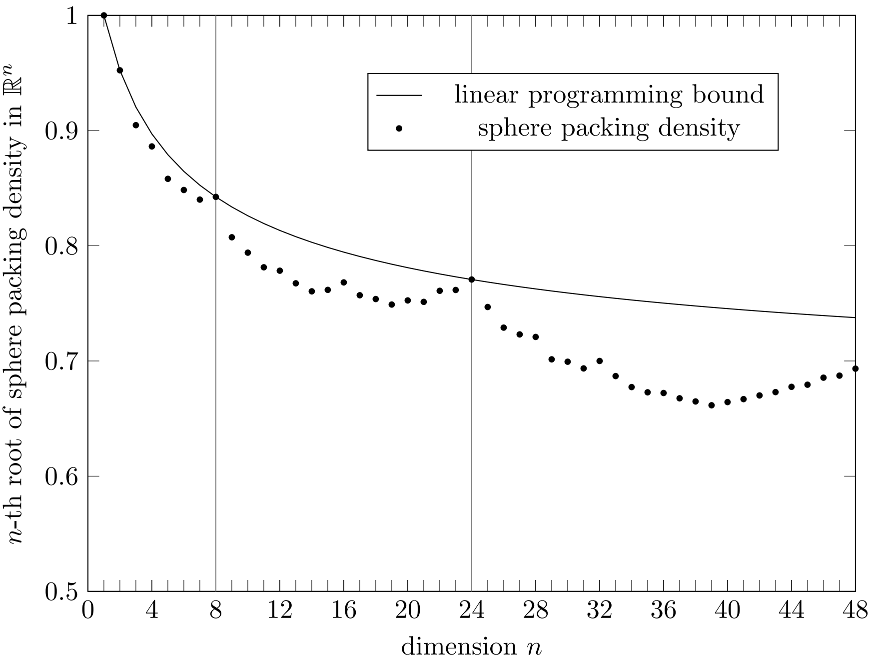


FIGURE 2.1. A plot of the numerically computed linear programming bound [1] and the best sphere packing density currently known [15]. The plot shows the n-th root of the density in dimension n, with $n = 8$ and $n = 24$ marked by vertical lines.

Then the optimal sphere packing density in $\mathbb{R}^n$ *is at most the volume* $\mathrm{vol}(B^n_{r/2})$ *of a ball of radius* $r/2$ *in* $\mathbb{R}^n$.

It is far from obvious how to produce good auxiliary functions f for use in this theorem, or how to optimize the choice of f, i.e., minimize r. In fact, the exact optimum is known only for $n = 1$, 8, and 24. However, one can perform a numerical optimization over a suitable space of functions, such as polynomials of fixed degree times a Gaussian, with the hope that it will converge to the global optimum as the degree tends to infinity. Figure 2.1 compares the resulting numerical bound with the density of the best packing known.

In most dimensions, the linear programming bound seems nowhere near sharp, but the upper and lower bounds appear to touch in eight and twenty-four dimensions. Cohn and Elkies conjectured that they were equal in those cases, and the solutions of the sphere packing problem in these dimensions come from proving this conjecture.[2]

The optimal auxiliary functions in eight and twenty-four dimensions have come to be known as magic functions, because obtaining an exact solution in these dimensions feels like a miracle. To see how this miracle comes about, we will examine a proof of Theorem 2.1 for the special case of lattice packings. It is based on the Poisson

[2]The linear programming bound also seems to be sharp in two dimensions, but no proof is known, despite the fact that the two-dimensional sphere packing problem itself can be solved by elementary means.

summation formula, which states that

$$\sum_{x\in\Lambda} f(x) = \frac{1}{\operatorname{vol}(\mathbb{R}^n/\Lambda)} \sum_{y\in\Lambda^*} \widehat{f}(y)$$

for every Schwartz function $f\colon \mathbb{R}^n \to \mathbb{C}$ and lattice Λ in $\mathbb{R}^n$. In this formula, $\operatorname{vol}(\mathbb{R}^n/\Lambda)$ is the volume of the quotient torus (i.e., the volume of a fundamental parallelotope for the lattice, or equivalently the absolute value of the determinant of a basis), and Λ^* is the dual lattice, which is spanned by the dual basis $v_1^*,\dots,v_n^*$ to any basis $v_1,\dots,v_n$ of Λ (i.e., $\langle v_i^*, v_j\rangle = \delta_{i,j}$). Poisson summation expresses a fundamental duality for Fourier analysis on $\mathbb{R}^n$, and we can apply it as follows.

Proof of Theorem 2.1 for lattice packings. Suppose our sphere packing consists of spheres centered at the points of a lattice Λ in $\mathbb{R}^n$. The sphere packing density is scaling-invariant, and so without loss of generality we can assume that the minimal nonzero vectors in Λ have length r. In other words, the sphere packing uses spheres of radius $r/2$, so that neighboring spheres are tangent to each other. Then the packing density is $\operatorname{vol}(B^n_{r/2})/\operatorname{vol}(\mathbb{R}^n/\Lambda)$, since there is one sphere for each fundamental cell of Λ.

We now apply Poisson summation to the auxiliary function f, to obtain

$$\sum_{x\in\Lambda} f(x) = \frac{1}{\operatorname{vol}(\mathbb{R}^n/\Lambda)} \sum_{y\in\Lambda^*} \widehat{f}(y).$$

The left side of this equation is bounded above by $f(0) = 1$, because $f(x) \le 0$ whenever $|x| \ge r$, and the right side is bounded below by $\widehat{f}(0)/\operatorname{vol}(\mathbb{R}^n/\Lambda) = 1/\operatorname{vol}(\mathbb{R}^n/\Lambda)$, since every summand is nonnegative. Thus, we conclude that $1/\operatorname{vol}(\mathbb{R}^n/\Lambda) \le 1$, and the sphere packing density satisfies $\operatorname{vol}(B^n_{r/2})/\operatorname{vol}(\mathbb{R}^n/\Lambda) \le \operatorname{vol}(B^n_{r/2})$, as desired. □

The proof for more general packings is similar in spirit, but it applies Poisson summation to periodic packings given by unions of translates of a lattice. See [10] or [8] for the details.

Note that the proof of Theorem 2.1 does not actually require f to be radial. However, the conditions on f are linear and rotation-invariant, and thus we can assume f is radial without loss of generality via rotational averaging.

What sort of function f could show that a lattice Λ is an optimal sphere packing? The proof given above drops the terms $f(x)$ with $x \in \Lambda\setminus\{0\}$ and $\widehat{f}(t)$ for $y \in \Lambda^*\setminus\{0\}$. Thus, we obtain a sharp bound if and only if all these omitted terms vanish. Because f and $\widehat{f}$ are radial functions, these conditions amount to saying that f vanishes on all the nonzero vector lengths in Λ, while $\widehat{f}$ vanishes on all the nonzero vector lengths in Λ^*. Furthermore, the last sign change should occur at the first root of f.

It turns out that the E_8 and Leech lattice are both self-dual, and their nonzero vector lengths are simply $\sqrt{2k}$ for integers $k \ge 1$ in E_8 and $k \ge 2$ in Λ_{24}. Thus, we know exactly what the roots of the magic functions should be. These roots are shown in Figure 2.2 for eight dimensions.

Now the whole problem comes down to constructing magic functions with these roots. That might not seem so difficult, but controlling the behavior of f and $\widehat{f}$ simultaneously is a subtle problem. Of course we can obtain any roots we'd like for f or $\widehat{f}$ in isolation, but not necessarily at the same time. This phenomenon is a form of uncertainty principle [7, 20, 11], much like the Heisenberg uncertainty principle.

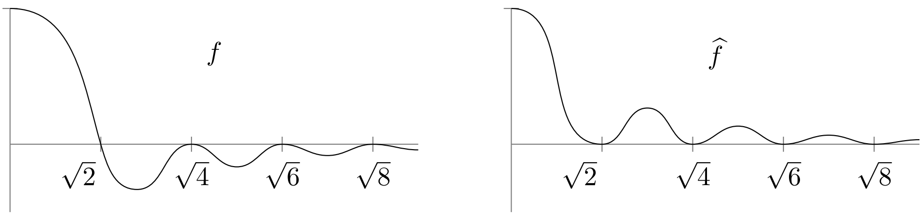


FIGURE 2.2. This diagram, which is taken from [8], shows the roots of the magic function f and its Fourier transform $\widehat{f}$ in eight dimensions. It is not an accurate plot, since these functions decrease very rapidly.

Viazovska gave a remarkable construction of the eight-dimensional magic function in terms of modular forms, which are a class of special functions defined on the upper half-plane $\mathbb{H} = \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}$ and satisfying certain transformation laws. The general theory of modular forms can feel somewhat forbidding to beginners, but Poisson summation gives us a simple way to get our hands on one example. The theta function $\theta \colon \mathbb{H} \to \mathbb{C}$ is defined by

$$\theta(z) = \sum_{n \in \mathbb{Z}} e^{\pi i n^2 z} = 1 + 2e^{\pi i z} + 2e^{4\pi i z} + 2e^{9\pi i z} + \cdots,$$

which converges because $z \in \mathbb{H}$ means $\mathrm{Im}(z) > 0$. This function satisfies two key identities,

$$\theta(z+2) = \theta(z) \qquad \text{and} \qquad \theta(-1/z) = (-iz)^{1/2}\theta(z). \tag{2.1}$$

The first identity follows immediately from the defining series, while the second is more subtle and will be proved below. In this equation, we have to choose the branch for $(-iz)^{1/2}$ carefully. Throughout this paper, fractional powers such as this one will be defined to be positive on the upper imaginary axis $(0, \infty)i$ in $\mathbb{H}$ and continuous on $\mathbb{H}$.

To prove that $\theta(-1/z) = (-iz)^{1/2}\theta(z)$, we will use Poisson summation for the one-dimensional lattice $\mathbb{Z}$ in $\mathbb{R}$. Consider the complex Gaussian $f \colon \mathbb{R} \to \mathbb{C}$ defined by

$$f(x) = e^{\pi i z x^2}$$

with $z \in \mathbb{H}$. When z is purely imaginary, this function is an ordinary Gaussian, and the other points in $\mathbb{H}$ behave much the same. In particular, one can check that

$$\widehat{f}(y) = (-iz)^{-1/2} e^{\pi i (-1/z) y^2}, \tag{2.2}$$

which is the complex generalization of the fact that the Fourier transform of a wide Gaussian is a narrow Gaussian and vice versa. Now Poisson summation says that

$$\sum_{x \in \mathbb{Z}} f(x) = \sum_{y \in \mathbb{Z}} \widehat{f}(y),$$

because $\mathbb{Z}$ is self-dual. This equation amounts to

$$\sum_{x \in \mathbb{Z}} e^{\pi i z x^2} = \sum_{y \in \mathbb{Z}} (-iz)^{-1/2} e^{\pi i (-1/z) y^2},$$

and thus $\theta(-1/z) = (-iz)^{1/2}\theta(z)$.

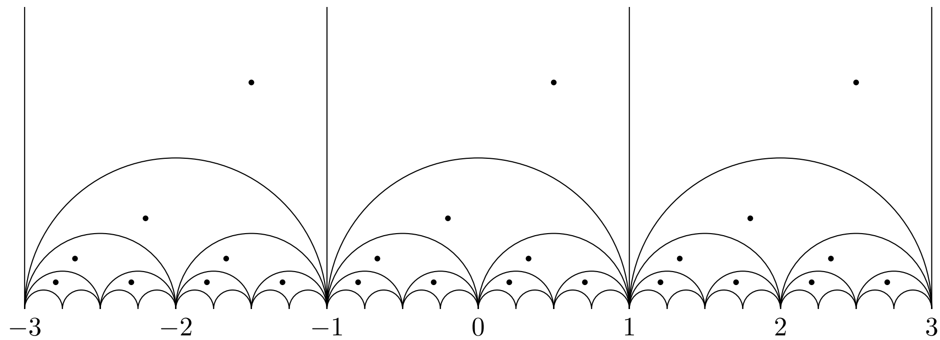


FIGURE 2.3. The regions shown here are ideal hyperbolic triangles (i.e., triangles in the hyperbolic plane with vertices at infinity), and they are fundamental domains for the action of Γ_θ on the upper half-plane. In particular, each Γ_θ-orbit intersects each triangle exactly once, unless it intersects the boundary of the triangle. The dots show a typical Γ_θ-orbit.

The functions $z \mapsto z + 2$ and $z \mapsto -1/z$ map $\mathbb{H}$ to itself, and they generate a group of linear fractional transformations of $\mathbb{H}$ called Γ_θ, in honor of the function θ. One can put a metric on $\mathbb{H}$ that turns it into the hyperbolic plane, at which point Γ_θ becomes a discrete group of isometries of $\mathbb{H}$, but we will not need this interpretation. See Figure 2.3 for a picture of a Γ_θ-orbit in $\mathbb{H}$.

Together with analyticity and some growth conditions, the identities (2.1) say that θ is a *modular form of weight* $1/2$ for the group Γ_θ. Viazovska's solution of the eight-dimensional sphere packing problem constructs the magic function using θ and a number of other modular forms, in a way that looks rather mysterious. What do modular forms have to do with radial Schwartz functions?

Instead of examining the details of her construction, let's think about a bigger picture. We know the eight-dimensional magic function f should satisfy

$$\begin{aligned} f\big(\sqrt{2k}\big) &= 0 \qquad \text{for } k \geq 1,\\ f'\big(\sqrt{2k}\big) &= 0 \qquad \text{for } k \geq 2,\\ \widehat{f}\big(\sqrt{2k}\big) &= 0 \qquad \text{for } k \geq 1, \text{ and}\\ \widehat{f}\,'\big(\sqrt{2k}\big) &= 0 \qquad \text{for } k \geq 1, \end{aligned}$$

as in Figure 2.2. Viazovska conjectured that this data, together with the nonzero value $f'\big(\sqrt{2}\big)$, would be enough to determine f uniquely. In fact, that turns out to be true:

Theorem 2.2 (Cohn, Kumar, Miller, Radchenko, and Viazovska [13])**.** *Let* (n, k_0) *be* $(8, 1)$ *or* $(24, 2)$*. Then every radial Schwartz function* $f\colon \mathbb{R}^n \to \mathbb{C}$ *is uniquely determined by the values* $f\big(\sqrt{2k}\big)$, $f'\big(\sqrt{2k}\big)$, $\widehat{f}\big(\sqrt{2k}\big)$, *and* $\widehat{f}\,'\big(\sqrt{2k}\big)$ *for integers* $k \geq k_0$*. Specifically, there exists an interpolation basis* $a_k, b_k, \widehat{a}_k, \widehat{b}_k$ *of radial Schwartz* *functions on* $\mathbb{R}^n$ *for* $k \geq k_0$ *such that for every* f *and* $x \in \mathbb{R}^n$*,*

$$f(x) = \sum_{k=k_0}^{\infty} f\big(\sqrt{2k}\big)\, a_k(x) + \sum_{k=k_0}^{\infty} f'\big(\sqrt{2k}\big)\, b_k(x) + \sum_{k=k_0}^{\infty} \widehat{f}\big(\sqrt{2k}\big)\, \widehat{a}_k(x) + \sum_{k=k_0}^{\infty} \widehat{f}\,'\big(\sqrt{2k}\big)\, \widehat{b}_k(x),$$

where these sums converge absolutely.

In particular, up to scaling the magic function is the interpolation basis function b_{k_0} in this theorem. One does not need this interpolation theorem to solve the sphere packing problem, but it is needed for analyzing ground states of more general particle systems in $\mathbb{R}^8$ and $\mathbb{R}^{24}$ (see [13]), and it provides a broader context for the magic functions.

Theorem 2.2 is similar in spirit to other interpolation theorems in mathematics. The simplest and most famous of these theorems is Lagrange interpolation, which says that a polynomial in one variable of degree less than n can be reconstructed from its values at any n distinct points. If the interpolation points are $x_1, \dots, x_n$, then we can write down an interpolation basis $p_1, \dots, p_n$ as

$$p_k(x) = \prod_{\substack{j=1 \\ j \neq k}}^{n} \frac{x - x_j}{x_k - x_j}, \tag{2.3}$$

so that every polynomial f of degree less than n is given by

$$f(x) = \sum_{j=1}^{n} f(x_j) p_j(x).$$

Lagrange interpolation can be generalized to Hermite interpolation, which takes into account derivatives along similar lines to Theorem 2.2: a polynomial f can be reconstructed from the values $f^{(j)}(x_k)$ with $0 \leq j < d_k$ and $1 \leq k \leq m$ if its degree is less than $\sum_{k=1}^{m} d_k$.

One important relative of Lagrange interpolation is Shannon sampling, which in the case of Schwartz functions $f\colon \mathbb{R} \to \mathbb{C}$ says that if $\widehat{f}$ vanishes outside the interval $[-r/2, r/2]$ for some r, then f is determined by its values on $r^{-1}\mathbb{Z}$ via

$$f(x) = \sum_{n \in \mathbb{Z}} f(n/r) \frac{\sin \pi(rx - n)}{\pi(rx - n)}.$$

This theorem plays a crucial role in information theory, since it says that a band-limited signal (i.e., one with a limited range of frequencies) is determined by periodic samples. It's worth noting that the product formula

$$\frac{\sin \pi x}{\pi x} = \prod_{j=1}^{\infty} \left(1 - \frac{x}{j^2}\right) \tag{2.4}$$

is analogous to the products (2.3) in the Lagrange interpolation basis. Much is known about Shannon sampling and its variations; see, for example, [27] and the references cited therein.

Both Lagrange interpolation and Shannon sampling rely on a notion of size. We measure the size of a polynomial by its degree, and the size of a bandlimited function by its bandwidth, the smallest r such that $\operatorname{supp}(\widehat{f}) \subseteq [-r/2, r/2]$. Then the larger a function is, the more interpolation points are required to reconstruct it, with "more" referring to density in the bandlimited case. Here the intuition is that size controls how many roots a function can have.[3]

Puzzlingly, Theorem 2.2 shows no sign of a similar notion of size. It is reminiscent of Shannon sampling, in that it takes into account both f and $\widehat{f}$, but it treats them symmetrically. In particular, there is little hope of a product formula along the lines of (2.3) or (2.4), because specifying the roots of f will not yield the roots of $\widehat{f}$. There seems to be a fundamental difference between these interpolation formulas, and neither Lagrange interpolation nor Shannon sampling offers a clue as to how to prove Theorem 2.2.

3. First-order Fourier interpolation

How does one prove an interpolation theorem like Theorem 2.2? We'll examine a technically simpler variant due to Radchenko and Viazovska, which is important in its own right and a beautiful illustration of Fourier interpolation. It deals with functions of one variable (so "radial" becomes "even"), and it studies interpolation to first order, without derivatives. This first-order interpolation theorem does not seem to have any applications to sphere packing, but it's a fundamental fact about Fourier analysis, and it is remarkable that it was not known until well into the 21st century.

Theorem 3.1 (Radchenko and Viazovska [33]). *There exist even Schwartz functions $a_n\colon \mathbb{R} \to \mathbb{R}$ for $n = 0, 1, 2, \dots$ such that every even Schwartz function $f\colon \mathbb{R} \to \mathbb{R}$ satisfies*

$$f(x) = \sum_{n=0}^{\infty} f\big(\sqrt{n}\big)\, a_n(x) + \sum_{n=0}^{\infty} \widehat{f}\big(\sqrt{n}\big)\, \widehat{a}_n(x)$$

for all $x \in \mathbb{R}$, and these sums converge absolutely.

There is also a corresponding theorem about odd functions (Theorem 7 in [33]), which can be proved in almost the same way. We'll focus on even functions here for simplicity. Note also that the root spacing has changed from $\sqrt{2n}$ to $\sqrt{n}$ in comparison with Theorem 2.2, which reflects the change in the order of interpolation.

As a consequence of this interpolation theorem, if an even Schwartz function $f\colon \mathbb{R} \to \mathbb{R}$ satisfies $f\big(\sqrt{n}\big) = \widehat{f}\big(\sqrt{n}\big) = 0$ for $n = 0, 1, 2, \dots$, then f vanishes identically. It's not so surprising that constructing an explicit interpolation basis $a_0, a_1, \dots$ would require special functions, such as modular forms, but it's noteworthy that even this corollary about vanishing does not seem easy to prove directly.

In the remainder of this section, we'll sketch a proof of Theorem 3.1. The sketch will omit a number of analytic details, but it will outline the techniques and explain where additional work is required.

The central question is where the interpolation basis $a_0, a_1, \dots$ comes from. We need to characterize these functions and prove that they have the desired properties. A first observation is that the interpolation basis is not quite unique, because Poisson summation over $\mathbb{Z}$ implies that every even Schwartz function f satisfies

$$f(0) + 2f(1) + 2f(2) + \cdots = \widehat{f}(0) + 2\widehat{f}(1) + 2\widehat{f}(2) + \cdots .$$

[3]Furthermore, size is related to growth at infinity. For degrees of polynomials this is clear, while a bandlimited function of bandwidth r can be analytically continued to the entire complex plane and satisfies $|f(z)| = O(e^{\pi r|z|})$. In other words, it is an entire function of exponential type πr.

In particular, $\widehat{f}(0)$ is determined by the values $f(0), f(1), f(2), \dots$ and $\widehat{f}(1), \widehat{f}(2), \dots$. To account for this redundancy, we will impose the constraint $\widehat{a}_0 = a_0$, so that the interpolation formula becomes

$$f(x) = (f(0) + \widehat{f}(0))a_0(x) + \sum_{n=1}^{\infty} f\big(\sqrt{n}\big)\, a_n(x) + \sum_{n=1}^{\infty} \widehat{f}\big(\sqrt{n}\big)\, \widehat{a}_n(x).$$

It turns out that this formula is now irredundant, with no additional linear relations between the values $f\big(\sqrt{n}\big)$ and $\widehat{f}\big(\sqrt{n}\big)$, and the interpolation basis is uniquely determined. Substituting $f = a_n$ shows that we can characterize a_n by its values at the points $\sqrt{m}$ with $m = 0, 1, 2, \dots$. Specifically, for $n, m \geq 1$, we must have

$$a_n\big(\sqrt{m}\big) = \begin{cases} 1 & \text{if } m = n, \text{ and} \\ 0 & \text{otherwise,} \end{cases}$$

$\widehat{a}_n\big(\sqrt{m}\big) = 0$, and $a_n(0) + \widehat{a}_n(0) = 0$, while a_0 must satisfy $\widehat{a}_0 = a_0$, $a_0(0) = 1/2$, and $a_0\big(\sqrt{m}\big) = 0$ for all $m \geq 1$.

These constraints let us get a handle on a_n, and we can use them to compute numerical approximations to a_n. More dramatically, they allow us to use Viazovska's modular form techniques from [39] to construct a_n explicitly. For example, we can write down a_0 as follows:

Lemma 3.2. *Let $a_0 \colon \mathbb{R} \to \mathbb{C}$ be defined by*

$$a_0(x) = \frac{1}{4} \int_{-1}^{1} \theta(z)^3 e^{\pi i z x^2}\, dz,$$

where we integrate over a semicircle in the upper half-plane $\mathbb{H}$. Then a_0 is an even Schwartz function with Fourier transform $\widehat{a}_0 = a_0$, and it satisfies $a_0(0) = 1/2$ and $a_0\big(\sqrt{m}\big) = 0$ for all positive integers m.

We'll use the same semicircular contour of integration in all integrals from -1 to 1 below. Recall that the theta function in this integral is defined for $z \in \mathbb{H}$ by

$$\theta(z) = \sum_{n \in \mathbb{Z}} e^{\pi i n^2 z}$$

and satisfies the functional equations $\theta(z+2) = \theta(z)$ and $\theta(-1/z) = (-iz)^{1/2}\theta(z)$.

Sketch of proof. The function a_0 is manifestly even, and we can prove that it is a Schwartz function by analyzing the behavior of $\theta(z)$ as z tends to ± 1. Specifically, if we remove small neighborhoods of ± 1 from the contour, then we obtain a smooth function of x. One can show that this function and its derivatives are rapidly decreasing as $x \to \infty$, essentially because the complex phases interfere destructively. To show that a_0 itself is a Schwartz function, we just have to check that the behavior as $z \to \pm 1$ is not bad enough to ruin this analysis.

To show that $\widehat{a}_0 = a_0$, we can take the Fourier transform of the complex Gaussian under the integral sign using (2.2) and change variables to $u = -1/z$, to obtain

$$\begin{aligned}\widehat{a}_0(x) &= \frac{1}{4}\int_{-1}^{1} \theta(z)^3 (-iz)^{-1/2} e^{\pi i(-1/z)x^2}\, dz\\ &= \frac{1}{4}\int_{1}^{-1} \theta(-1/u)^3 (i/u)^{-1/2} e^{\pi i u x^2} u^{-2}\, du\\ &= \frac{1}{4}\int_{-1}^{1} \theta(u)^3 e^{\pi i u x^2}\, du\\ &= a_0(x),\end{aligned}$$

where the third line follows from $\theta(-1/u)^3 = (-iu)^{3/2}\theta(u)^3$ and

$$-(-iu)^{3/2}(i/u)^{-1/2}u^{-2} = 1$$

for $u \in \mathbb{H}$. (To check this last identity, note that the left side is always ± 1, it is continuous for $u \in \mathbb{H}$, and it equals 1 when $u = i$.)

Finally, we can compute $a_0\big(\sqrt{m}\big)$ for nonnegative integers m using the identity

$$\begin{aligned}a_0\big(\sqrt{m}\big) &= \frac{1}{4}\int_{-1}^{1} \theta(z)^3 e^{m\pi i z}\, dz\\ &= \frac{1}{4}\int_{-1+i}^{1+i} \theta(z)^3 e^{m\pi i z}\, dz,\end{aligned}$$

where we have deformed the contour to a straight line from $-1+i$ to $1+i$, which is possible because the integrals between 0 and $-1+i$ and between $1+i$ and 1 cancel due to $\theta(z+2) = \theta(z)$. Now we write

$$\theta(z) = 1 + 2e^{\pi i z} + 2e^{4\pi i z} + 2e^{9\pi i z} + \cdots$$

and expand $\theta(z)^3$ as a series in powers of $e^{\pi i z}$. By orthogonality, the value

$$a_0\big(\sqrt{m}\big) = \frac{1}{4}\int_{-1+i}^{1+i} \theta(z)^3 e^{m\pi i z}\, dz$$

is $1/2$ times the coefficient of $e^{-m\pi i z}$ in this expansion of $\theta(z)^3$. In particular, $a_0(0) = 1/2$ and $a_0\big(\sqrt{m}\big) = 0$ for positive integers m, as desired, since there are no negative powers of $e^{\pi i z}$ in this series. □

What made this proof work is that the identity $\theta(-1/z)^3 = (-iz)^{3/2}\theta(z)^3$ gave us $\widehat{a}_0 = a_0$, while the identity $\theta(z+2)^3 = \theta(z)^3$ let us compute the values $a_0\big(\sqrt{m}\big)$ as Fourier series coefficients. One can obtain each basis function a_n using similar constructions, which require increasingly elaborate replacements for $\theta(z)^3$ as n grows, and it is not immediately clear how to describe or analyze them systematically. Furthermore, obtaining the basis functions individually does not explain why the interpolation formula actually holds: these functions could in principle exist yet not suffice to reconstruct an arbitrary even Schwartz function in Theorem 3.1.

To give a uniform account of these functions, we will construct generating functions for the interpolation basis. For $\tau \in \mathbb{H}$, let

$$F(\tau, x) = \sum_{n=0}^{\infty} a_n(x) e^{n\pi i \tau},$$

and denote its Fourier transform in x by

$$\widehat{F}(\tau, x) = \sum_{n=0}^{\infty} \widehat{a}_n(x) e^{n\pi i\tau}.$$

Being Fourier series, these functions satisfy the functional equations

$$F(\tau+2, x) = F(\tau, x) \qquad \text{and} \qquad \widehat{F}(\tau+2, x) = \widehat{F}(\tau, x).$$

Furthermore, the formula (2.2) for the Fourier transform of a complex Gaussian implies that the interpolation formula from Theorem 3.1 for the function $f(x) = e^{\pi i\tau x^2}$ is equivalent to

$$F(\tau, x) + (-i\tau)^{-1/2}\widehat{F}(-1/\tau, x) = e^{\pi i\tau x^2},$$

and thus F and $\widehat{F}$ must satisfy this functional equation as well.

In fact, these functional equations turn out to be almost all we need to obtain a working interpolation basis. The following lemma is stated somewhat informally, but it can be made precise.

Lemma 3.3. *If there exists a function F such that F and $\widehat{F}$ satisfy these three functional equations and certain analyticity and growth bounds, then Theorem 3.1 follows.*

Sketch of proof. The idea behind the proof is surprisingly simple. If F and $\widehat{F}$ are sufficiently well-behaved, then the functional equations $F(\tau+2, x) = F(\tau, x)$ and $\widehat{F}(\tau+2, x) = \widehat{F}(\tau, x)$ imply that they can be expanded as Fourier series. We can define the functions a_n to be the Fourier coefficients of $F(\tau, x)$, and $\widehat{a}_n$ must be the corresponding coefficient of $\widehat{F}(\tau, x)$, as in the original definitions of F and $\widehat{F}$ above. The fact that there are no terms with $n < 0$ amounts to boundedness as $\operatorname{Im}(\tau) \to \infty$, and the constraint that $a_0 = \widehat{a}_0$ can be phrased similarly (namely that $F(\tau, x) - \widehat{F}(\tau, x)$ decays as $\operatorname{Im}(\tau) \to \infty$).

Now the third functional equation says that

$$\sum_{n=0}^{\infty} a_n(x) e^{n\pi i\tau} + \sum_{n=0}^{\infty} \widehat{a}_n(x)(-i\tau)^{-1/2} e^{n\pi i(-1/\tau)} = e^{\pi i\tau x^2},$$

which becomes

$$\sum_{n=0}^{\infty} a_n(x) f\big(\sqrt{n}\big) + \sum_{n=0}^{\infty} \widehat{a}_n(x) \widehat{f}\big(\sqrt{n}\big) = f(x)$$

if we set $f(x) = e^{\pi i\tau x^2}$. In other words, it states that the interpolation theorem holds when f is a complex Gaussian.

One can show that complex Gaussians span a dense subspace of the even Schwartz functions. To complete the proof, all we need to show is that for each $x \in \mathbb{R}$, the functional Λ_x that takes an even Schwartz function f to

$$\Lambda_x(f) = f(x) - \sum_{n=0}^{\infty} f\big(\sqrt{n}\big)\, a_n(x) - \sum_{n=0}^{\infty} \widehat{f}\big(\sqrt{n}\big)\, \widehat{a}_n(x)$$

is continuous, so that vanishing on a dense subspace implies vanishing everywhere. The topology on the space of Schwartz functions is defined by a family of seminorms, and proving that Λ_x is continuous requires proving that the seminorms of a_n and $\widehat{a}_n$ grow at most polynomially as $n \to \infty$. To prove the required bounds, we can use Fourier orthogonality to write $a_n(x)$ and $\widehat{a}_n(x)$ as integrals in τ of $F(\tau, x)$ and

$\widehat{F}(\tau, x)$, respectively, and then use suitable growth bounds for F and $\widehat{F}$ to bound the seminorms of these integrals. □

We can now imitate the construction of a_0 from $\theta(z)^3$ in Lemma 3.2 to obtain the generating functions F and $\widehat{F}$ explicitly. To do so, we will replace $\theta(z)^3$ with the functions K and $\widehat{K}$ from the following proposition, which is again stated informally. Note that $\widehat{K}$ is not a Fourier transform of K; instead, this notation is simply mnemonic, since $\widehat{K}$ will be used to construct $\widehat{F}$.

Proposition 3.4. *There exist meromorphic functions K and $\widehat{K}$ on $\mathbb{H} \times \mathbb{H}$ that satisfy the following conditions for all $\tau, z \in \mathbb{H}$:*

(1) $K(\tau + 2, z) = K(\tau, z)$ *and* $\widehat{K}(\tau + 2, z) = \widehat{K}(\tau, z)$,
(2) $K(\tau, z + 2) = K(\tau, z)$ *and* $\widehat{K}(\tau, z + 2) = \widehat{K}(\tau, z)$,
(3) $K(-1/\tau, z) = -(-i\tau)^{1/2}\widehat{K}(\tau, z)$,
(4) $K(\tau, -1/z) = (-iz)^{3/2}\widehat{K}(\tau, z)$,
(5) $z \mapsto K(\tau, z)$ *and* $z \mapsto \widehat{K}(\tau, z)$ *have poles only when z is in the Γ_θ-orbit of τ,*
(6) *all their poles are simple poles,*
(7) *the residue of $z \mapsto K(\tau, z)$ at $z = \tau$ is $1/(2\pi i)$ and at $z = -1/\tau$ is 0 (in other words, there is no pole there),*
(8) *the residue of $z \mapsto \widehat{K}(\tau, z)$ at $z = \tau$ is 0, and*
(9) *K and $\widehat{K}$ satisfy certain growth bounds, which we will not discuss here.*

The motivation behind the transformation laws in Proposition 3.4 is that they generalize how $\theta(z)^3$ transforms, and we'll see that they perfectly describe what we need to obtain F and $\widehat{F}$ as integrals of K and $\widehat{K}$. At first glance the most mysterious aspect may be the poles, which did not occur for $\theta(z)^3$. We'll see below that the poles lead to the inhomogeneous term $e^{\pi i \tau x^2}$ in the functional equation

$$F(\tau, x) + (-i\tau)^{-1/2}\widehat{F}(-1/\tau, x) = e^{\pi i \tau x^2}.$$

Before we examine how to use K and $\widehat{K}$ to construct F and $\widehat{F}$, we will take a look at how Proposition 3.4 is proved.

Sketch of proof. The functions K and $\widehat{K}$ can be described explicitly in terms of modular forms, using three ingredients: the theta function θ, the modular function λ, and a Hauptmodul (principal modular function) J for the group Γ_θ.

We have already been using θ, and λ is a similar analytic function on $\mathbb{H}$ that dates back to the 19th century. For our purposes, its key properties will be how Γ_θ acts on it, namely

$$\lambda(z + 2) = \lambda(z) \qquad \text{and} \qquad \lambda(-1/z) = 1 - \lambda(z).$$

Note that it is not quite invariant under Γ_θ. We define $J(z)$ to be $\lambda(z)(1 - \lambda(z))/16$, so that $J(z)$ is invariant under both generators of Γ_θ; i.e.,

$$J(z + 2) = J(z) \qquad \text{and} \qquad J(-1/z) = J(z).$$

Then it turns out that J generates the function field of the quotient of $\mathbb{H}$ by the action of Γ_θ (this quotient has genus 0), and $J(z) = J(\tau)$ if and only if z and τ are in the same orbit of Γ_θ.

Using these tools, we can guess much of what $K(\tau, z)$ and $\widehat{K}(\tau, z)$ should look like. Conditions (3) and (4) suggest that these functions should have factors of $\theta(\tau)\theta(z)^3$, to get the correct weights for the transformation laws. Conditions (4) and (5) imply that they should be given by $1/(J(z) - J(\tau))$ times something holomorphic, and the signs in (3) and (4) can be obtained using $1 - 2\lambda(-1/z) = -(1 - 2\lambda(z))$.

In fact, we can take

$$K(\tau, z) = \theta(\tau)\theta(z)^3 \frac{J(z)(1 - 2\lambda(\tau)) + J(\tau)(1 - 2\lambda(z))}{4(J(z) - J(\tau))}$$

and

$$\widehat{K}(\tau, z) = \theta(\tau)\theta(z)^3 \frac{J(z)(1 - 2\lambda(\tau)) - J(\tau)(1 - 2\lambda(z))}{4(J(z) - J(\tau))},$$

and fairly routine computations show that (1) through (9) hold. The functions K and $\widehat{K}$ turn out to be uniquely determined by these conditions, but we will not verify that here, to avoid having to state the conditions more carefully and deal with residues and growth bounds.

It's worth noting that one can simplify some of the verification by writing K and $\widehat{K}$ in terms of the function $h := 1 - 2\lambda$ via

$$K(\tau, z) = \theta(\tau)\theta(z)^3 \frac{1 - h(\tau)h(z)}{4(h(\tau) - h(z))}$$

and

$$\widehat{K}(\tau, z) = \theta(\tau)\theta(z)^3 \frac{1 + h(\tau)h(z)}{4(h(\tau) + h(z))}.$$

For example, h is a Hauptmodul for a subgroup of Γ_θ called $\Gamma(2)$, and these formulas show that the poles of $z \mapsto K(\tau, z)$ and $z \mapsto \widehat{K}(\tau, z)$ occur only on the $\Gamma(2)$-orbits of τ and $-1/\tau$, respectively. $\square$

All that remains is to use K and $\widehat{K}$ to construct functions F and $\widehat{F}$ for use in Lemma 3.3. To do so, we can imitate Lemma 3.2. As a first attempt to produce F from K, we could try setting

$$F(\tau, x) = \int_{-1}^{1} K(\tau, z) e^{\pi i z x^2} \, dz. \tag{3.1}$$

However, this formula can't possibly hold for all τ, because the integrand has poles on the Γ_θ-orbit of τ, and as one varies τ, sometimes these poles cross the contour of integration. Instead, we can use this definition only on subsets of $\mathbb{H}$ for which the poles avoid the contour of integration. As shown in Figure 2.3, one such subset consists of all the points $\tau \in \mathbb{H}$ such that τ has distance strictly greater than 1 from $2\mathbb{Z}$. For such τ, we define $F(\tau, x)$ by (3.1); we will deal with other values of τ via analytic continuation in Lemma 3.5.

To obtain $\widehat{F}(\tau, x)$ we can take the Fourier transform of $F(\tau, x)$ in x. For τ strictly further than distance 1 from $2\mathbb{Z}$, we can use the semicircular contour, and almost exactly the same computation as in the proof of Lemma 3.2 shows that

$$\begin{aligned}\widehat{F}(\tau,x) &= \int_{-1}^{1} K(\tau,z)(-iz)^{-1/2}e^{\pi i(-1/z)x^2}\,dz\\ &= -\int_{-1}^{1} K(\tau,-1/z)(i/z)^{-1/2}e^{\pi izx^2}z^{-2}\,dz\\ &= -\int_{-1}^{1} \widehat{K}(\tau,z)(-iz)^{3/2}(i/z)^{-1/2}z^{-2}e^{\pi izx^2}\,dz\\ &= \int_{-1}^{1} \widehat{K}(\tau,z)e^{\pi izx^2}\,dz.\end{aligned}$$

Lemma 3.5. *The functions $\tau \mapsto F(\tau,x)$ and $\tau \mapsto \widehat{F}(\tau,x)$ can be analytically continued to all of $\mathbb{H}$, and they satisfy the functional equations $F(\tau+2,x) = F(\tau,x)$, $\widehat{F}(\tau+2,x) = \widehat{F}(\tau,x)$, and $F(\tau,x) + (-i\tau)^{-1/2}\widehat{F}(-1/\tau,x) = e^{\pi i\tau x^2}$.*

Sketch of proof. Let $S = \{\tau \in \mathbb{H} : |\tau - 2n| > 1 \text{ for all } n \in \mathbb{Z}\}$. We have defined $F(\tau,x)$ and $\widehat{F}(\tau,x)$ for $\tau \in S$, and the functional equations

$$F(\tau+2,x) = F(\tau,x) \qquad \text{and} \qquad \widehat{F}(\tau+2,x) = \widehat{F}(\tau,x)$$

for $\tau \in S$ are immediate consequences of

$$K(\tau+2,z) = K(\tau,z) \qquad \text{and} \qquad \widehat{K}(\tau+2,z) = \widehat{K}(\tau,z).$$

To prove the lemma, it will suffice to analytically continue $\tau \mapsto F(\tau,x)$ and $\tau \mapsto \widehat{F}(\tau,x)$ to some open neighborhood of the closure of S in $\mathbb{H}$, such that the continuations satisfy

$$F(\tau,x) + (-i\tau)^{-1/2}\widehat{F}(-1/\tau,x) = e^{\pi i\tau x^2}$$

whenever τ and $-1/\tau$ are both in this neighborhood. Then we can use the functional equations to extend these functions to all the hyperbolic triangles in Figure 2.3.[4]

We can now use the information about poles and residues in Proposition 3.4. When we analytically continue $F(\tau,x)$ to τ just below the semicircle from -1 to 1, the only relevant pole of $z \mapsto K(\tau,z)$ is at $z = \tau$, since $-1/\tau$ is the only other nearby point in the Γ_θ-orbit of τ, and there is no pole at that point. We can set

$$F(\tau,x) = \int_{\mathcal{C}_\tau} K(\tau,z)e^{\pi izx^2}\,dz, \tag{3.2}$$

where $\mathcal{C}_\tau$ is a deformation of the semicircle to form a contour from -1 to 1 that passes below τ, so that τ never lies on the contour.

Similarly, we can analytically continue $\widehat{F}(\tau,x)$ to just below the semicircle via

$$\widehat{F}(\tau,x) = \int_{\mathcal{C}'_\tau} \widehat{K}(\tau,z)e^{\pi izx^2}\,dz,$$

where this time there is no pole at $x = \tau$, and the condition is that the contour $\mathcal{C}'_\tau$ stays above the pole of $z \mapsto \widehat{K}(\tau,z)$ at $z = -1/\tau$.

Now we can prove the functional equation

$$F(\tau,x) + (-i\tau)^{-1/2}\widehat{F}(-1/\tau,x) = e^{\pi i\tau x^2}$$

[4]Note that as we pass from a triangle to the adjacent triangles, we can never reach the same triangle via two different paths of adjacencies, and thus we don't need to worry about inadvertently defining a multivalued function of τ.

as follows when τ is just below the semicircle. The identity

$$K(-1/\tau, z) = -(-i\tau)^{1/2}\widehat{K}(\tau, z),$$

or equivalently

$$\widehat{K}(-1/\tau, z) = -(-i\tau)^{1/2}K(\tau, z),$$

shows that

$$\begin{aligned}(-i\tau)^{-1/2}\widehat{F}(-1/\tau, x) &= \int_{\mathcal{C}'_{-1/\tau}} (-i\tau)^{-1/2}\widehat{K}(-1/\tau, z)e^{\pi izx^2}\,dz \\ &= \int_{\mathcal{C}'_{-1/\tau}} -(-i\tau)^{-1/2}(-i\tau)^{1/2}K(\tau, z)e^{\pi izx^2}\,dz \\ &= -\int_{\mathcal{C}'_{-1/\tau}} K(\tau, z)e^{\pi izx^2}\,dz.\end{aligned} \tag{3.3}$$

Combining (3.2) and (3.3) with the residue theorem implies that

$$F(\tau, x) + (-i\tau)^{-1/2}\widehat{F}(-1/\tau, x)$$

is $2\pi i$ times the sum of the residues of all the poles of $z \mapsto K(\tau, z)e^{\pi izx^2}$ between $\mathcal{C}_\tau$ and $\mathcal{C}'_{-1/\tau}$. The only pole that could lie between these contours is at $z = \tau$, since $z \mapsto K(\tau, z)$ has no pole at $z = -1/\tau$, and by construction it does lie between them. The residue of $z \mapsto K(\tau, z)e^{\pi izx^2}$ at $z = \tau$ is $e^{\pi i\tau x^2}/(2\pi i)$, and so

$$F(\tau, x) + (-i\tau)^{-1/2}\widehat{F}(-1/\tau, x) = e^{\pi i\tau x^2},$$

as desired. □

Lemma 3.5 shows that $F(\tau, x)$ and $\widehat{F}(\tau, x)$ can be analytically continued to all $\tau \in \mathbb{H}$ in such a way that they satisfy the three functional equations. That is almost everything we need to prove Theorem 3.1 using Lemma 3.3. However, to apply this lemma we need to verify certain growth conditions for $F(\tau, x)$ and $\widehat{F}(\tau, x)$ as τ approaches the real line. Verifying these conditions is the most technical part of the proof of the interpolation theorem, and we will not examine it here. In short, the verification combines bounds on K and $\widehat{K}$ with careful accounting of how quickly the inhomogeneous terms from the third functional equation can accumulate during the analytic continuation. Once this is done, the proof of Theorem 3.1 is complete.

This proof is satisfyingly thorough, in that it not only proves the interpolation formula, but also provides plenty of additional information. For example, we can obtain explicit formulas for the interpolation basis $a_0, a_1, \dots$ by using the identity $K(\tau + 2, z) = K(\tau, z)$ to write K as a Fourier series

$$K(\tau, z) = \sum_{n=0}^{\infty} \varphi_n(z)e^{n\pi i\tau}$$

when $\mathrm{Im}(\tau)$ is large. Then

$$a_n(x) = \int_{-1}^{1} \varphi_n(z)e^{\pi izx^2}\,dz,$$

which generalizes Lemma 3.2. Similarly, the Fourier coefficients of $\widehat{K}$ yield formulas for $\widehat{a}_n$.

On the other hand, some aspects of the proof are quite delicate. For example, it is very sensitive to the form $\sqrt{n}$ of the interpolation points. Specifically, the proof of the functional equation

$$F(\tau, x) + (-i\tau)^{-1/2}\widehat{F}(-1/\tau, x) = e^{\pi i \tau x^2},$$

depends on the fact that the complex Gaussian $x \mapsto e^{\pi i \tau x^2}$ equals $e^{n\pi i \tau}$ when evaluated at the interpolation point $x = \sqrt{n}$. If we replaced $\sqrt{n}$ with other interpolation points r_n, then the Fourier series for $F(\tau, x)$ would have to be replaced with

$$\sum_{n=0}^{\infty} a_n(x) e^{r_n^2 \pi i \tau},$$

and it would no longer satisfy $F(\tau + 2, x) = F(\tau, x)$ if the values r_n^2 are not integers. That would disrupt the algebraic mechanism behind the proof.

Much remains to be understood regarding generalizations of the Radchenko-Viazovska theorem and how Fourier interpolation fits into a broader picture. One significant line of work [2, 3] connects Fourier interpolation to uniqueness theory for the Klein-Gordon equation [24, 25, 26]. Other noteworthy papers examine the density of possible interpolation points [28, 34] and whether they can be perturbed [30], interpolation formulas using zeros of zeta and L-functions [5], and extensions to non-radial functions [35, 31, 32]. Perhaps the most surprising development so far has been a paper on sphere packing and quantum gravity [23], which shows the equivalence of linear programming bounds with the spinless modular bootstrap bound for free bosons in conformal field theory, and which furthermore shows that certain bases of special functions constructed by Mazáč and Paulos [29] for the conformal bootstrap can be transformed into Fourier interpolation bases.

Acknowledgements

I am grateful to Danylo Radchenko for helpful feedback on a draft of this article.

References

[1] N. Afkhami-Jeddi, H. Cohn, T. Hartman, D. de Laat, and A. Tajdini, *High-dimensional sphere packing and the modular bootstrap*, J. High Energy Phys. **2020** (2020), no. 12, Paper No. 066, 44 pp. arXiv:2006.02560 doi:10.1007/jhep12(2020)066

[2] A. Bakan, H. Hedenmalm, A. Montes-Rodríguez, D. Radchenko, and M. Viazovska, *Fourier uniqueness in even dimensions*, Proc. Natl. Acad. Sci. USA **118** (2021), no. 15, Paper No. 2023227118, 4 pp. doi:10.1073/pnas.2023227118

[3] A. Bakan, H. Hedenmalm, A. Montes-Rodríguez, D. Radchenko, and M. Viazovska, *Hyperbolic Fourier series*, preprint, 2021. arXiv:2110.00148

[4] M. R. Best, *Binary codes with a minimum distance of four*, IEEE Trans. Inform. Theory **26** (1980), no. 6, 738–742. doi:10.1109/TIT.1980.1056269

[5] A. Bondarenko, D. Radchenko, and K. Seip, *Fourier interpolation with zeros of zeta and L-functions*, preprint, 2020. arXiv:2005.02996

[6] S. V. Borodachov, D. P. Hardin, and E. B. Saff, *Asymptotics of best-packing on rectifiable sets*, Proc. Amer. Math. Soc. **135** (2007), no. 8, 2369–2380. arXivmath-ph/0605021 doi:10.1090/S0002-9939-07-08975-7

[7] J. Bourgain, L. Clozel, and J.-P. Kahane, *Principe d'Heisenberg et fonctions positives*, Ann. Inst. Fourier (Grenoble) **60** (2010), no. 4, 1215–1232. arXiv:0811.4360 doi:10.5802/aif.2552

[8] H. Cohn, *A conceptual breakthrough in sphere packing*, Notices Amer. Math. Soc. **64** (2017), no. 2, 102–115. arXiv:1611.01685 doi:10.1090/noti1474

[9] H. Cohn, *The work of Maryna Viazovska*, Fields medal laudatio, 2022. doi:10.4171/ICM2022/213

[10] H. Cohn and N. Elkies, *New upper bounds on sphere packings I*, Ann. of Math. (2) **157** (2003), no. 2, 689–714. arXiv:math/0110009 doi:10.4007/annals.2003.157.689

[11] H. Cohn and F. Gonçalves, *An optimal uncertainty principle in twelve dimensions via modular forms*, Invent. Math. **217** (2019), no. 3, 799–831. arXiv:1712.04438 doi:10.1007/s00222-019-00875-4

[12] H. Cohn, A. Kumar, S. D. Miller, D. Radchenko, and M. Viazovska, *The sphere packing problem in dimension* 24, Ann. of Math. (2) **185** (2017), no. 3, 1017–1033. arXiv:1603.06518 doi:10.4007/annals.2017.185.3.8

[13] H. Cohn, A. Kumar, S. D. Miller, D. Radchenko, and M. Viazovska, *Universal optimality of the E_8 and Leech lattices and interpolation formulas*, Annals of Mathematics, to appear. arXiv:1902.05438 doi:10.4007/annals.2022.196.3.3

[14] J. H. Conway and N. J. A. Sloane, *What are all the best sphere packings in low dimensions?*, Discrete Comput. Geom. **13** (1995), no. 3–4, 383–403. doi:10.1007/BF02574051

[15] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*, third edition, Grundlehren der Mathematischen Wissenschaften **290**, Springer, New York, 1999. doi:10.1007/978-1-4757-6568-7

[16] D. de Laat and F. Vallentin, *A breakthrough in sphere packing: the search for magic functions*, Nieuw Arch. Wiskd. (5) **17** (2016), no. 3, 184–192. arXiv:1607.02111

[17] P. Delsarte, *Bounds for unrestricted codes, by linear programming*, Philips Res. Rep. **27** (1972), 272–289.

[18] W. Ebeling, *Lattices and Codes: A Course Partially Based on Lectures by Friedrich Hirzebruch*, third edition, Advanced Lectures in Mathematics, Springer Spektrum, Wiesbaden, 2013. doi:10.1007/978-3-658-00360-9

[19] L. Fejes, *Über einen geometrischen Satz*, Math. Z. **46** (1940), 83–85. doi:10.1007/BF01181430

[20] F. Gonçalves, D. Oliveira e Silva, and S. Steinerberger, *Hermite polynomials, linear flows on the torus, and an uncertainty principle for roots*, J. Math. Anal. Appl. **451** (2017), no. 2, 678–711. arXiv:1602.03366 doi:10.1016/j.jmaa.2017.02.030

[21] T. C. Hales, *A proof of the Kepler conjecture*, Ann. of Math. (2) **162** (2005), no. 3, 1065–1185. doi:10.4007/annals.2005.162.1065

[22] T. Hales, M. Adams, G. Bauer, T. D. Dang, J. Harrison, L. T. Hoang, C. Kaliszyk, V. Magron, S. McLaughlin, T. T. Nguyen, Q. T. Nguyen, T. Nipkow, S. Obua, J. Pleso, J. Rute, A. Solovyev, T. H. A. Ta, N. T. Tran, T. D. Trieu, J. Urban, K. Vu, and R. Zumkeller, *A formal proof of the Kepler conjecture*, Forum Math. Pi **5** (2017), e2, 29 pp. arXiv:1501.02155 doi:10.1017/fmp.2017.1

[23] T. Hartman, D. Mazáč, and L. Rastelli, *Sphere packing and quantum gravity*, J. High Energy Phys. **2019** (2019), no. 12, 048, 66 pp. arXiv:1905.01319 doi:10.1007/jhep12(2019)048

[24] H. Hedenmalm and A. Montes-Rodríguez, *Heisenberg uniqueness pairs and the Klein-Gordon equation*, Ann. of Math. (2) **173** (2011), no. 3, 1507–1527. arXiv:0910.1246 doi:10.4007/annals.2011.173.3.6

[25] H. Hedenmalm and A. Montes-Rodríguez, *The Klein-Gordon equation, the Hilbert transform, and dynamics of Gauss-type maps*, J. Eur. Math. Soc. (JEMS) **22** (2020), no. 6, 1703–1757. arXiv:1503.04038 doi:10.4171/jems/954

[26] H. Hedenmalm and A. Montes-Rodríguez, *The Klein-Gordon equation, the Hilbert transform and Gauss-type maps: H^∞ approximation*, J. Anal. Math. **144** (2021), no. 1, 119–190. doi:10.1007/s11854-021-0173-4

[27] J. R. Higgins, *Five short stories about the cardinal series*, Bull. Amer. Math. Soc. (N.S.) **12** (1985), no. 1, 45–89. doi:10.1090/S0273-0979-1985-15293-0

[28] A. Kulikov, *Fourier interpolation and time-frequency localization*, J. Fourier Anal. Appl. **27** (2021), no. 3, Paper No. 58, 8 pp. arXiv:2005.12836 doi:10.1007/s00041-021-09861-y

[29] D. Mazáč and M. F. Paulos, *The analytic functional bootstrap. Part II. Natural bases for the crossing equation*, J. High Energy Phys. **2019** (2019), no. 2, 163, 53 pp. arXiv:1811.10646 doi:10.1007/jhep02(2019)163

[30] J. P. G. Ramos and M. Sousa, *Perturbed interpolation formulae and applications*, preprint, 2020. arXiv:2005.10337

[31] J. P. G. Ramos and M. Stoller, *Perturbed Fourier uniqueness and interpolation results in higher dimensions*, J. Funct. Anal. **282** (2022), no. 12, Paper No. 109448, 34 pp. arXiv:2103.12015 doi:10.1016/j.jfa.2022.109448

[32] D. Radchenko and M. Stoller, *Fourier non-uniqueness sets from totally real number fields*, Comment. Math. Helv. **97** (2022), no. 3, 513–553. arXiv:2108.11828 doi:10.4171/cmh/538

[33] D. Radchenko and M. Viazovska, *Fourier interpolation on the real line*, Publ. Math. Inst. Hautes Études Sci. **129** (2019), 51–81. arXiv:1701.00265 doi:10.1007/s10240-018-0101-z

[34] N. T. Sardari, *Higher Fourier interpolation on the plane*, preprint, 2021. arXiv:2102.08753

[35] M. Stoller, *Fourier interpolation from spheres*, Trans. Amer. Math. Soc. **374** (2021), no. 11, 8045–8079. arXiv:2002.11627 doi:10.1090/tran/8440

[36] T. M. Thompson, *From Error-correcting Codes through Sphere Packings to Simple Groups*, Carus Mathematical Monographs **21**, Mathematical Association of America, Washington, DC, 1983.

[37] A. Thue, *Om nogle geometrisk-taltheoretiske Theoremer*, Forhdl. Skand. Naturforsk. **14** (1892), 352–353.

[38] A. Thue, *Über die dichteste Zusammenstellung von kongruenten Kreisen in der Ebene*, Norske Vid. Selsk. Skr. **1** (1910), 1–9.

[39] M. S. Viazovska, *The sphere packing problem in dimension* 8, Ann. of Math. (2) **185** (2017), no. 3, 991–1015. arXiv:1603.04246 doi:10.4007/annals.2017.185.3.7

[40] M. Viazovska, *Sharp sphere packings*, Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018. Vol. II. Invited lectures, 455–466, World Sci. Publ., Hackensack, NJ, 2018. doi:10.1142/9789813272880_0063

[41] M. Viazovska, *Almost impossible E_8 and Leech lattices*, Eur. Math. Soc. Mag. No. **121** (2021), 4–8. doi:10.4171/mag-47

Microsoft Research New England, One Memorial Drive, Cambridge, MA 02142, USA
Email address: `cohn@microsoft.com`

A STROLL AROUND THE CRITICAL POTTS MODEL

MARTIN HAIRER

Abstract. Over the past decade or so, a broad research programme spearheaded by H. Duminil-Copin and his collaborators has vastly increased our understanding of a number of critical or near-critical statistical mechanics models. Most prominently, these include the q-state Potts models and, essentially equivalently, the FK cluster models. In this short review, we present a small selection of recent results from this research area.

Keywords: Ising model, Potts model, percolation

Contents

1. Introduction

One of the simplest, yet extremely rich, models of statistical mechanics is the Ising model which has historically been introduced as a toy model for the behaviour of ferromagnets. (This model was actually first invented by Wilhelm Lenz in 1920, who then gave it to his student Ernst Ising to study.) The definition of the model goes as follows. Given a finite graph G, identified here with its set of vertices, we consider the configuration space $\Omega = \{-1, 1\}^G$ and define on Ω an energy functional $E(\sigma) = -\frac{1}{2}\sum_{x\sim y}\sigma_x\sigma_y$, where $x \sim y$ if and only if the vertices x and y are connected by an edge in G. One should think here of the vertices of G as indexing spatial locations, for example of individual atoms in a metallic solid, of the graph structure as indicating which locations are neighbours in space, and of σ_x as denoting a spin variable associated to such a location. The energy is then defined in such a way that states of low energy are those where many pairs of neighbouring spins are aligned.

Given an inverse temperature β, one then defines a probability measure μ_β on Ω by setting $\mu_\beta(\{\sigma\}) = Z^{-1}\exp(-\beta E(\sigma))$, where Z is such that $\mu_\beta(\Omega) = 1$. For definiteness, when we talk about "the Ising model on G at inverse temperature β", we mean the measure μ_β as just described. The interpretation of the model in terms of spins and atoms suggests that an interesting special case is that where G is a large

2020 *Mathematics Subject Classification.* 82B26, 82B43, 82B20.

piece of a lattice, for example $G = \Lambda_N = \{-N, \ldots, N\}^d$ or $G = \mathbf{Z}^d \cap N\mathcal{O}$ for some open set $\mathcal{O} \subset \mathbf{R}^d$ with smooth boundary, with edges between nearest neighbours. Writing μ_β^N for the Ising model on G_N, it turns out that the limit $\mu_\beta = \lim_{N\to\infty} \mu_\beta^N$ exists and can therefore be interpreted as the Ising model on $\mathbf{Z}^d$.

One very interesting qualitative feature of this model is that it exhibits a *phase transition* in every dimension $d \geq 2$: there exists a critical (dimension-dependent) value β_c which delineates two regimes in which the measure μ_β behaves very differently. At "high temperature", namely for $\beta < \beta_c$, the *spontaneous magnetisation*, namely the random quantity $M = N^{-d} \sum_{i\in\Lambda_N} \sigma_i$, converges to 0 in probability as $N \to \infty$. For $\beta > \beta_c$ on the other hand, it converges in probability to a limiting random variable that can take exactly two possible values $\pm h_\beta \neq 0$ with equal probabilities. The actual value of β_c is only known in dimension 2 where it equals $\beta_c = \log\bigl(1 + \sqrt{2}\bigr)$ [Ons44]. (There is no phase transition at all in dimension 1 and the spontaneous magnetisation M always vanishes, so in some sense $\beta_c = +\infty$ there.)

The expression and just mentioned result for the spontaneous magnetisation M has the flavour of a "law of large numbers", so it is natural to ask whether there is an associated "central limit theorem" describing the fluctuations of the magnetisation. In other words, does the law of the quantity $N^{-d/2} \sum_{i\in\Lambda_N} (\sigma_i - M)$ converge to that of a normal distribution? This is indeed the case when $\beta \neq \beta_c$, but the corresponding variance diverges as $\beta \to \beta_c$. The behaviour *at* the critical temperature is highly non-trivial and it is not even clear at first sight how such an expression should be normalised. In other words, does there exist a value α such that the law of

$$N^{-\alpha} \sum_{i\in\Lambda_N} (\sigma_i - M)$$

admits a non-degenerate limit distribution as $N \to \infty$ when $\beta = \beta_c$? It was shown in a recent series of works [CGN15, CGN16] that if one chooses $\alpha = 15/8$ in dimension $d = 2$, then this is indeed the case. Actually even more was shown there, namely one can consider the joint distribution of finitely many quantities of the form

$$I_\phi^N(\sigma) = N^{-\alpha} \sum_{x\in\Lambda_N} \phi(x/N)\sigma_x \;, \tag{1.1}$$

for ϕ a smooth test function supported on $[-1, 1]^d$, and these all converge. One way of interpreting this is that there exists a random distribution ζ on the hypercube such that the quantities $I_\phi^N(\sigma)$ all converge jointly in law to the quantities $\zeta(\phi)$.

This time however, unlike in the central limit theorem, the limiting distributions are not Gaussian (the random variables $\zeta(\phi)$ actually exhibit an even faster decaying tail behaviour) and no nice closed form expression exists for them (but there *does* exist a closed form expression for their joint moments, which was first derived heuristically in the physics literature [BPZ84, Car84, BG93] and recently made rigorous in [CHI15]). Note that the exponent α is closely related to the behaviour of $\mathbf{E}_c\sigma_u\sigma_v$ (where $\mathbf{E}_c$ denotes the expectation under μ_{β_c}) since, assuming that $\mathbf{E}_c\sigma_u\sigma_v \approx |u - v|^{-2\delta}$, one finds that

$$\begin{aligned}\mathbf{E}_c\bigl(I_\phi^N(\sigma)\bigr)^2 &= N^{-2\alpha} \sum_{u,v} \phi(u/N)\phi(v/N)\mathbf{E}_c\sigma_u\sigma_v \\ &\lesssim N^{-2\alpha} \sum_{u,v} |u - v|^{-2\delta} \approx N^{2d-(2\delta\wedge d)-2\alpha} \;,\end{aligned}$$

so that one expects the relation $\alpha = d - (\delta \wedge d/2)$, which (correctly) leads to the prediction $\delta = \frac{1}{8}$. Interestingly, the limiting distribution ζ exhibits a form of covariance under the action of the conformal group(oid) in the following sense. Given any smooth simply connected domain $D \subset \mathbf{R}^2$, one can consider expressions like (1.1), but this time with $\Lambda_N = ND \cap \mathbf{Z}^2$. It turns out that these do again converge, this time to a random distribution ζ_D on the domain D. Given two such domains D and $\bar{D}$ and a bijective conformal map $\psi\colon D \to \bar{D}$, the pushforward η of $\zeta_{\bar{D}}$ to D given by

$$\eta(\phi) = \zeta_{\bar{D}}(\phi \circ \psi^{-1}) , \tag{1.2a}$$

is equal in law to the random distribution $\bar{\eta}$ given by

$$\bar{\eta}(\phi) = \zeta_D(|\psi'|^{15/8}\phi) = \zeta_D(|\psi'|^{\alpha}\phi) , \tag{1.3}$$

where $\alpha = 2 - \delta$ is as above. This and a number of other properties of the Ising model at criticality allows to associate it to the conformal field theory with central charge $c = \frac{1}{2}$.

The picture in dimensions greater than 2 is less clear. For $d \geq 5$, it was shown in [Aiz81, Aiz82, Frö82] that the correct scaling exponent to use in (1.1) at $\beta = \beta_c$ is $\alpha = 1 + \frac{d}{2}$ and that the limit is a Gaussian Free Field, namely the Gaussian random distribution with correlation function given by the Green's function of the Laplacian (with Neuman boundary conditions on the square). In dimension $d = 3$, virtually nothing is known rigorously about the critical Ising model, not even the value of its scaling exponents, although much progress has been made at a non-rigorous (but very well supported) level with the development of the "conformal bootstrap" [ESPP$^+$12, ESPP$^+$14]. Regarding the case $d = 4$, it was somewhat unclear until very recently whether the Ising model at criticality should be "trivial" (i.e. described by Gaussian distributions) or not. This was eventually settled by Aizenman and Duminil-Copin in the work [ADC21] where they show that any subsequential limit for expressions of the form (1.1) as $N \to \infty$ (and $\beta \to \beta_c$) must necessarily be Gaussian.

1.1. **A general picture.** The general picture that has been emerging over the past half century or so regarding the behaviour of many statistical mechanics systems can be summarised as follows:

(1) Many of the simplest local equilibrium systems in dimension 2 or higher do exhibit a phase transition, namely there exists a critical value β_c at which the qualitative large scale behaviour of the system changes abruptly. In general, a system may depend on additional parameters in which case one may see a more complicated *phase diagram* with several regions in parameter space where the global behaviour of the system displays qualitatively different behaviour. In any case, the "high temperature / small β phase" is expected to behave in such a way that what happens in well separated regions of space is very close to independent.

(2) In dimension 2, many of these systems appear to exhibit a form of conformal invariance at criticality, even though no rotation symmetry is built a priori into their description. When this happens, the link to $2d$ conformal field theories (and the associated probabilistic objects like SLE [Sch00], QLE [MS16], etc) provides a hugely powerful machinery to predict – and in a number of cases also rigorously prove – their behaviour. In the case of the

Potts model (see below for its definition), these links are on a strong rigorous footing for $q \in \{0, 2\}$, but much needs to be done for other values of q.

(3) The universe of local statistical mechanics models can be subdivided into broad classes of models that exhibit a shared large-scale behaviour at criticality. These are called "universality classes" and, in the $2d$ equilibrium case, they are expected to come in families parametrised by a real parameter, the central charge. (For certain values of the central charge, one expects to have several "subclasses", but we will not discuss this kind of subtlety here.)

(4) Although one still expects conformal invariance at criticality in higher dimensions, this is a much smaller symmetry there and therefore appears to provide somewhat less insight[1]. One also expects the situation there to be more rigid than in two dimensions, with fewer universality classes. (Possibly only a discrete family.)

(5) Models that have "obvious" variants in every dimension typically have a critical dimension above which their behaviour at criticality is "trivial" in the sense that it exhibits Gaussian behaviour. (Typically with correlation function given by the Green's function of the Laplacian.) In the case of the Ising universality class, this critical dimension is 4, while in the case of Bernoulli percolation it is 6.

One important branch of modern probability theory aims to put this general picture onto rigorous mathematical footing. The remainder of this article is devoted to a short overview of some of the recent contributions to this vast programme, mainly focusing around the example of the critical Potts model where much recent progress was made by Hugo Duminil-Copin and his collaborators.

Acknowledgements. This work was supported by the Royal Society through a research professorship, grant number RP/R1/191065. Sections 3 and 4 of this review were previously published in my 2022 laudatio for Hugo Duminil-Copin's Fields medal.

2. The Potts and random cluster models

The Potts model is a natural generalisation of the Ising model: this time the configuration space is given by $\Omega = \{1, \ldots, q\}^G$ and the corresponding energy functional is given by $E_q(\sigma) = -\sum_{x\sim y} \mathbf{1}\{\sigma_x = \sigma_y\}$. We denote by

$$\mathbf{P}_{\beta,q}(\sigma) \propto \exp\bigl(-E_q(\sigma)\bigr) ,$$

the corresponding Gibs measure. Note that the case $q = 2$ yields the Ising model, modulo a recentering of the energy (which doesn't affect the measures $\mu_\beta = \mathbf{P}_{\beta,2}$). For $q \neq 2$, the Potts model does not exhibit the kind of exact solvability that the Ising model does in two dimensions (as discovered by Onsager [Ons44] in his famous computation of its partition function), so that it is one of the simplest possible models of statistical mechanics that isn't known to be exactly solvable.

One important feature of the Potts model is that it is very closely related to a different model, the random cluster model, introduced by Fortuin and Kasteleyn [FK72], which however makes sense for all $q > 0$, not just integer values. This model is usually interpreted as a percolation model, i.e. its state space is given by $\bar\Omega = \{0, 1\}^E$ where E denotes the set of edges of the graph G and, given a configuration $\omega \in \bar\Omega$, we say that the edge e is "open"

[1]See however the recent breakthrough made in the approximation of the critical exponents of the $3d$ Ising model using the "conformal bootstrap" [ESPP+12, ESPP+14] already mentioned above.

if and only if $\omega_e = 1$. Given two fixed parameters $p \in (0,1)$ and $q > 0$, the probability of a configuration ω is then proportional to

$$\mathbf{Q}_{p,q}(\omega) \propto p^{|\omega|}(1-p)^{|1-\omega|}q^{|K_\omega|}\;,$$

where $|\omega| = \sum_{e\in E}\omega_e$ and K_ω denotes the set of connected components (also called "clusters" in this context) of the subgraph G_ω of G given by replacing the edge set E with the set $E_\omega = \{e : \omega_e = 1\}$ of "open" edges. (Here an isolated vertex counts as a connected component.)

It turns out, see for example [Gri06, Thm 1.13] that given any finite graph G and provided that β and p are related by the identity

$$p = 1 - e^{-\beta}\;, \tag{2.1}$$

one can find a probability measure $\mathbf{P}$ on $\Omega \times \bar\Omega$ with the following properties:

- The marginal of $\mathbf{P}$ on Ω coincides with the Potts model, namely $\mathbf{P}(A\times\bar\Omega) = \mathbf{P}_{\beta,q}(A)$.
- The marginal of $\mathbf{P}$ on $\bar\Omega$ coincides with the random cluster model, namely $\mathbf{P}(\Omega\times A) = \mathbf{Q}_{p,q}(A)$.
- Under $\mathbf{P}$, almost every configuration (σ,ω) is such that for every open edge xy (i.e. such that $\omega_{xy} = 1$), one has $\sigma_x = \sigma_y$.
- Conditional on a configuration σ, the law of ω under $\mathbf{P}$ is obtained by setting the values $\{\omega_{xy} : \sigma_x = \sigma_y\}$ to be i.i.d. Bernoulli random variables with parameter p.
- Conditional on a configuration ω, the law of σ under $\mathbf{P}$ is obtained by assigning to every cluster $A \in K_\omega$ independently a "colour" $\sigma_A \in \{1,\ldots,q\}$, and then setting $\sigma_x = \sigma_A$ for all $x \in A$.

The advantage of the random cluster model is that it exhibits a nice duality in the case when G is a connected planar graph (for example a chunk of the two-dimensional lattice). In that case, one can define a dual graph (G^*, E^*) whose vertex set G^* consists of the faces of the original graph G and such that there a bijection between E and E^* mapping any edge $e \in E$ to an edge e^* connecting the two faces adjacent to E. (This may generate self-loops.)

Every configuration ω on E then determines a dual configuration ω^* on E^* by setting $\omega^*_{e^*} = 1 - \omega_e$, where e and e^* are related as just described. See Figure 1 for an example of a configuration ω on a chunk of the square lattice, as well as the corresponding dual configuration. Write $\mathbf{Q}^*_{p,q}$ for the pushforward of the measure $\mathbf{Q}_{p,q}$ under the map $\omega \mapsto \omega^*$. One then has the following result.

Proposition 2.1. *The measure $\mathbf{Q}^*_{p,q}$ coincides with the random cluster model on G^* with parameters (p^*, q) where p^* is given by*

$$p^* = \frac{q - pq}{p + q - pq}\;.$$

Proof. Recall that, given any configuration ω, G_ω is the (planar) subgraph of G obtained by only retaining the "open edges" $E_\omega = \{e : \omega_e = 1\}$. The proof is then based on two remarks. First, writing F_ω for the set of faces of G_ω (with the usual convention that there is an infinite outer face) and K_ω for the set of its connected components, we note that one has the identity

$$|G| + |F_\omega| = 1 + |E_\omega| + |K_\omega|\;.$$

(This variant of the Euler characteristic formula is true for any planar graph and can easily be shown by induction over the number of vertices and edges. The reason why we have G appearing there is to emphasise that the vertex set of the graph G_ω is independent of the configuration ω, which will be important in the sequel.) The second remark relates the graph G_ω to the subgraph $G^*_{\omega^*}$ of G^* generated by the configuration dual to ω.[2] One can

[2]Note that $G^*_{\omega^*}$ is very different from the dual graph of G_ω.

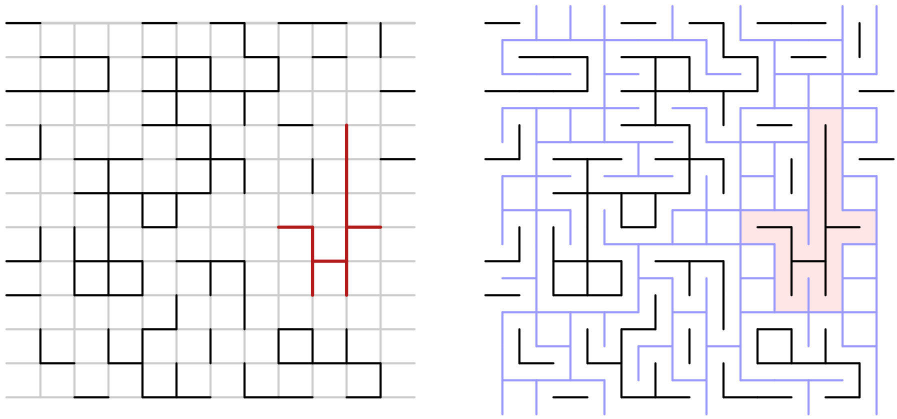


FIGURE 1. On the left, we draw a configuration ω for the random cluster model with $N = 11$, with one of the clusters highlighted in red. On the right, the same configuration is drawn together with its dual configuration in light blue. The face of the dual configuration corresponding to the cluster is shaded in light red.

see that connected components of G_ω are then in one-to-one correspondence with faces of $G^*_{\omega^*}$, see Figure 1 for an illustration of this fact. In other words, one has the identity $|K_\omega| = |F^*_{\omega^*}|$.

Using this correspondence and the fact that $|E_\omega| + |E^*_{\omega^*}| = |E|$ by definition of the dual configuration, it then follows that

$$\begin{aligned}\mathbf{Q}^*_{p,q}(\omega^*) &\propto p^{|\omega|}(1-p)^{|1-\omega|}q^{k(\omega)} \propto \big(p/(1-p)\big)^{|E_\omega|}q^{|K_\omega|} \\ &= \big(p/(1-p)\big)^{|E|-|E^*_{\omega^*}|}q^{|F^*_{\omega^*}|} \propto \big((1-p)/p\big)^{|E^*_{\omega^*}|}q^{1+|K^*_{\omega^*}|+|E^*_{\omega^*}|-|G^*|} \\ &\propto \big(q(1-p)/p\big)^{|E^*_{\omega^*}|}q^{|K^*_{\omega^*}|} = \big(p^*/(1-p^*)\big)^{|E^*_{\omega^*}|}q^{|K^*_{\omega^*}|} \propto \mathbf{Q}_{p^*,q}\;,\end{aligned}$$

which is precisely the desired claim. □

Since the square lattice is self-dual, this leads to the natural conjecture that the critical value of p for the random cluster model on $\mathbf{Z}^2$ is given by the (unique) value such that $p^* = p$, namely

$$p^2(q-1) - 2pq + q = 0 \qquad \Rightarrow \qquad p = \frac{q-\sqrt{q}}{q-1} = 1 - \frac{1}{1+\sqrt{q}}\;.$$

Thanks to (2.1) and the close link between the random cluster model and the Potts model, this motivates the following recent result [BDC12].

Theorem 2.2. *The critical inverse temperature for the q-colour Potts model is given by* $\beta_c = \log(1+\sqrt{q})$.

In the remainder of this article, we describe several recent results for the random cluster and Potts models at criticality. Our main focus is on the two-dimensional case, but we'll see that one important result is the continuity of the phase transition in dimension 3.

3. (Dis)continuity of phase transitions

One very natural question in statistical mechanics is whether one can take the limit $N \to \infty$ for the finite volume Gibbs measures. At this stage, we note that there are actually several inequivalent natural ways in which one can define the Ising or Potts model

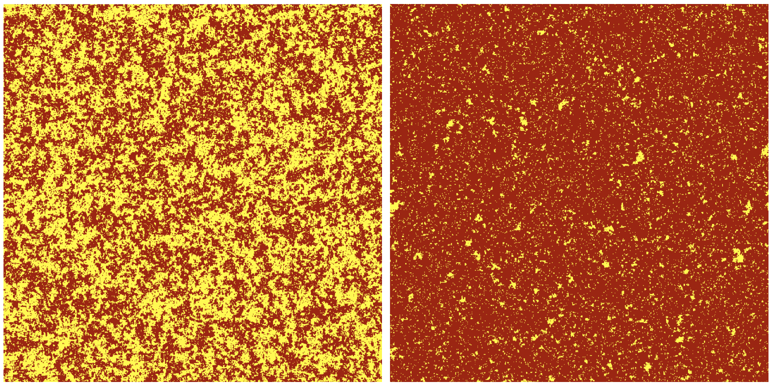


FIGURE 2. Typical Ising configurations for $\beta < \beta_c$ (left) and $\beta > \beta_c$ (right).

in a region of size N of $\mathbf{Z}^d$. One possibility is to simply consider $\Lambda_N = \{-N, \dots, N\}^d$ as a subgraph of the lattice $\mathbf{Z}^d$, as we have done so far. However, one could also extend configurations $\sigma \in \{1, \dots, q\}^{\Lambda_N}$ to all of $\{1, \dots, q\}^{\mathbf{Z}^d}$ by fixing a reference configuration $\bar\sigma \in \{1, \dots, q\}^{\mathbf{Z}^d}$ and postulating that $\sigma_x = \bar\sigma_x$ for $x \notin \Lambda_N$. (A natural choice is to take $\bar\sigma$ constant and we will mainly consider such a situation here.) Finally, one could identify $-N$ with N in Λ_N and consider the Potts model on larger and larger discrete tori. In this way, we have different choices of "boundary conditions" yielding different definitions for the finite volume measures $\mu_{\beta,N}$.

In many examples of interest (including the case of the Potts models), the measure $\mu_\beta = \lim_{N\to\infty} \mu_{\beta,N}$ is well-defined (i.e. independent of the choice of boundary condition) for $\beta < \beta_c$ while one can obtain several distinct limits in the case $\beta > \beta_c$. Figure 2 shows typical samples drawn from μ_β for the Ising model with $\bar\sigma \equiv 1$. In the case $\beta > \beta_c$, the resulting sample clearly "remembers" the bias introduced by $\bar\sigma$ in the sense that a typical configuration consists of a "sea" of spins taking the dominant value $+1$ (brown) with small "islands" of spins taking the value -1 (yellow). Had we set $\bar\sigma \equiv -1$, we would have obtained a sample with the opposite behaviour, which illustrates the non-uniqueness of the infinite-volume measure μ_β in this case. In the case $\beta < \beta_c$ on the other hand, each one of the two possible spin values is about equally represented and the measure is symmetric under the substitution $1 \leftrightarrow -1$, which illustrates the uniqueness of μ_β. It is in fact a theorem in the case of the Ising model that for $\beta > \beta_c$ there exist exactly two translation invariant infinite volume measures $\mu_\beta^\pm$ corresponding to boundary conditions $\bar\sigma \equiv \pm 1$ and that every accumulation point of $\mu_{\beta,N}$ for any sufficiently homogeneous boundary condition as $N \to \infty$ is a convex combination of μ_β^+ and μ_β^-. (In fact a similar statement holds for the Potts model with q states, where one has exactly q distinct infinite volume Gibbs measures when $\beta > \beta_c$.)

This raises the question of the uniqueness of μ_β at $\beta = \beta_c$. If it is, then we say that the phase transition is *continuous*, otherwise it is said to be *discontinuous*. The reason for this terminology is that continuity in this sense turns out to be equivalent to the continuity of the maps $\beta \mapsto \mu_\beta^\pm$ at $\beta = \beta_c$. It has been known for quite some time [Yan52, AF86] that the phase transition for the Ising model is continuous in dimensions $d = 1, 2$ as well as $d \geq 4$. The reason why dimensions 1 and 2 are typically much better understood is that the Ising model is "solvable" in these dimensions in the sense that explicit expressions can be derived for the expectation of a large number of observables under $\mu_{\beta,N}$ (this solution is straightforward in $d = 1$ [Isi25] where no phase transition is present, but it was a major breakthrough when Onsager obtained his exact solution for $d = 2$ [Ons44]).

Dimension $d = 4$ on the other hand is the "upper critical dimension" beyond which the model is expected to be "trivial" (i.e. described by Gaussian random variables in the scaling limit) which allows to use a number of powerful techniques, including for example the *lace expansion* [HS94, Sak07].

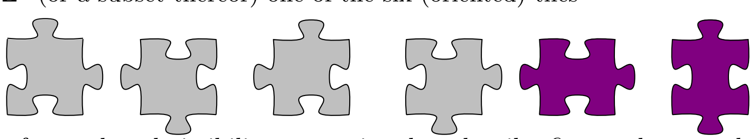
This leaves the case $d = 3$ which is of course the physically most interesting one since the Ising model is a toy model of ferromagnetism and its dimensions represent the usual spatial dimensions. Heuristic considerations suggest that the phase transition is also continuous there, and this is consistent with physical experiments, assuming that the Ising model belongs to the same universality class as that of a genuine physical magnet. In the article [ADCS15], Duminil-Copin et al. gave the first rigorous proof that this is indeed the case. The proof relies on the introduction of the quantity

$$M(\beta) = \inf_{B\subset \mathbf{Z}^3} \frac{1}{|B|^2} \sum_{x,y\in B} \int \sigma_x \sigma_y \, \mu_\beta^0(d\sigma) \;,$$

where μ_β^0 denotes the infinite volume limit obtained from using "free" conditions, as well as three main steps. First, they rely on results of [FSS76, FILS78] to argue that the Fourier transform of $x \mapsto \int \sigma_0\sigma_x \, \mu_\beta^0(d\sigma)$ belongs to L^1 at $\beta = \beta_c$, which implies that $M(\beta_c) = 0$. Then, and this is the main step, they show that having $M(\beta) = 0$ implies that a certain percolation model with long-range correlations constructed from the Ising model admits no infinite clusters. Finally, they use a variant of the "switching lemma" [GHS70] to show that the quantity $\int \sigma_0\sigma_x \, \mu_\beta^+(d\sigma) - \int \sigma_0\sigma_x \, \mu_\beta^0(d\sigma)$ is dominated by an explicit function times the probability of the origin belonging to an infinite cluster in the above mentioned model and therefore has to vanish at $\beta = \beta_c$. Once this is known, it is not too difficult to show that the spontaneous magnetisation of the Ising model at criticality must vanish (namely one has $\int \sigma_0 \, \mu_{\beta_c}^+(d\sigma) = 0$), which in turn yields the desired uniqueness statement.

Considering more general values of q for the Potts model illustrates in a rather striking way the fact that continuity of the phase transition, whatever the dimension, is a rather non-trivial property that isn't necessarily expected in general. Indeed, it was conjectured by Baxter in the 70's [Bax71, Bax73] that the Potts model on $\mathbf{Z}^2$ exhibits a continuous phase transition if and only if $q \le 4$. The pair of articles [DCST17, DCGH$^+$21] by Duminil-Copin et al. provides proofs of both directions of this conjecture. For the sake of brevity we will not comment on the proofs in any detail, but we note that the proof of continuity of the phase transition for $q \le 4$ is almost completely disjoint from that in the case of the $3d$ Ising model. A milestone is again to show that the model at criticality with boundary condition set to one fixed element of S admits no infinite cluster. However both the proof of this fact (exploiting a form of discrete holomorphicity of certain cleverly chosen observables) and the proof of its equivalence with the uniqueness of the infinite-volume measure at criticality (actually they show equivalence of a list of 5 quite distinct properties which are of independent interest for the study of the critical Potts model) are completely different.

Regarding the proof of *dis*continuity when $q > 4$, the main tool is a close relation, first discovered by Temperley–Lieb [TL71] in a restricted context and then by Baxter et al. [BKW76] in more generality, between the FK model on $\mathbf{Z}^2$ and the so-called six-vertex model. Configurations of the latter can be visualised as jigsaws where one assigns to each vertex of $\mathbf{Z}^2$ (or a subset thereof) one of the six (oriented) tiles



and one enforces the admissibility constraint that the tiles fit together seamlessly. One further postulates that the probability of seeing a given admissible configuration is proportional to $c^{\#p}$, where $\#p$ denotes the number of purple tiles in the configuration and c is some fixed constant. The relation between the six-vertex model and the critical FK model holds for the specific choice $c = \sqrt{2+\sqrt{q}}$. The advantage gained from this relation is that the six-vertex model is "solvable" in a certain sense using the transfer matrix formalism. This doesn't get one out of the woods since the transfer matrices V_N involved are very large: they act on a vector space of dimension 2^N, but are block diagonal with each block $V_N^{[n]}$ acting on a subspace of dimension $\binom{n}{N}$. Each of these blocks is irreducible with positive entries and therefore admits a Perron–Frobenius vector. The main technical result of [DCST17] is a very sharp asymptotic for the Perron–Frobenius eigenvalues of $V_N^{[N/2-r]}$ for fixed r as $N \to \infty$. Interestingly, the authors are able to prove that the ratios between these values converge to finite (and explicit, at least as explicit convergent series) limits as $N \to \infty$ and that the values themselves diverge exponentially in N with known exponent, but the common lower-order behaviour of that divergence is not known. This asymptotic is however sufficient to obtain good control over the partition function of the six vertex model and to exploit it to compute an explicit expression for the inverse correlation length of the critical Potts model with free boundary conditions when $q > 4$. The finiteness of that expression finally allows to deduce the discontinuity of the phase transition.

To conclude this section, I would like to mention the beautiful article [DCRT19] which, although not quite dealing with the question of continuity of the phase transition, does have a related flavour. The question there is that of the "sharpness" of the phase transition which in this particular case is couched as the question whether it is really true that the measure μ_β has exponentially decaying correlations (in the sense that the covariance between $f(\sigma_0)$ and $f(\sigma_x)$ decays exponentially fast as $|x| \to \infty$ for any "nice enough" function $f\colon S \to \mathbf{R}$) for *every* $\beta < \beta_c$ and not just for small enough values where a perturbation argument around $\beta = 0$ (where $f(\sigma_0)$ and $f(\sigma_x)$ are independent under μ_0 as soon as $x \neq 0$) may apply. One difficulty with this type of statements is that one will in general not know any closed-form expression for β_c: in the case of the FK model on the square lattice such an expression can be derived by a duality argument [BDC12], but it is not known for more general situations. The main result of [DCRT19] is that the phase transition of the FK model on *any* vertex-transitive infinite graph is sharp.

The main tool in their proof is a novel and far-reaching generalisation of the OSSS inequality [OSSS05]. The context here is that of increasing random variables $f\colon \{0,1\}^E \to [0,1]$ (for a finite set E and for the natural coordinate-wise partial order on $\{0,1\}^E$) where $\{0,1\}^E$ is furthermore equipped with a probability measure $\mathbf{P}$ that is itself *monotonic* in the sense that for every $F \subset E$ and every $e \in E \setminus F$, the conditional probabilities $\mathbf{P}(w_e = 1 \,|\, \mathcal{F}_F)$ are increasing functions. (Here $\mathcal{F}_F$ denotes the σ-algebra generated by the evaluations $w \mapsto w_e$ for $e \in F$.) One then considers *any* algorithm that reveals one by one the values of an input $w \in \{0,1\}^E$ in such a way that the coordinate to be revealed next depends in a deterministic way on the information gleaned from the revealement up to that point. (In particular, the first coordinate to be revealed is always the same since no information has been obtained yet at that point.) The algorithm stops once the revealed values are sufficient to determine the value of $f(w)$, thus yielding a random set $\hat{E} \subset E$ of revealed values. The result of [DCRT19] is then that one has the inequality

$$\mathrm{Var}(f) \leq \sum_{e \in E} \mathbf{P}(e \in \hat{E})\ \mathrm{Cov}(f, w_e)\ , \tag{3.1}$$

which looks formally the same as the result of [OSSS05], but the assumption there was that the measure $\mathbf{P}$ is simply the uniform measure. Since the latter is clearly monotonic (it is such that $\mathbf{P}(w_e = 1 \,|\, \mathcal{F}_F)$ is constant), the results of [OSSS05] follow as a special case.

Using this result, [DCRT19] then obtain the following dichotomy which yields the desired sharpness statement.

Theorem 3.1. *Let G be any transitive graph and let $\mathbf{P}_{\beta,n}$ be the FK measure on the ball Λ_n of radius n in G. Then, there exists $\beta_c \in \mathbf{R}$ such that, for every $\beta < \beta_c$ there exists $c_\beta > 0$ such that $\mathbf{P}_{\beta,n}(0 \leftrightarrow \partial\Lambda_n) \lesssim e^{-c_\beta n}$, uniformly in n. For $\beta > \beta_c$ on the other hand, there exists $c > 0$ such that $\mathbf{P}_{\beta,n}(0 \leftrightarrow \partial\Lambda_n) \geq c \min\{1, \beta - \beta_c\}$.*

Once (3.1) is known, the proof is surprisingly simple and relies on two ingredients. First, one can show that the measures $\mathbf{P}_{\beta,n}$ and the function $\mathbf{1}_{0\leftrightarrow\partial\Lambda_n}$ satisfy the assumptions of (3.1). Setting $\theta_n(\beta) = \mathbf{P}_{\beta,n}(0 \leftrightarrow \partial\Lambda_n)$, a clever choice of search algorithm for the (potential) cluster connecting the origin 0 to $\partial\Lambda_n$ then allows to show that one has the bound

$$\theta_n'(\beta) \gtrsim \sum_{e\in E} \mathrm{Cov}_\beta(\mathbf{1}_{0\leftrightarrow\partial\Lambda_n}, w_e) \geq \frac{n}{8\Sigma_n(\beta)}\theta_n(\beta)(1-\theta_n(\beta)) \;. \tag{3.2}$$

where $\Sigma_n = \sum_{k=0}^{n-1}\theta_n$. The fact that the first inequality holds is known and can be checked in an elementary way. The second fact is that *any* sequence of functions $\beta \mapsto \theta_n(\beta)$ satisfying a differential inequality of the form (3.2) necessarily satisfies a dichotomy of the type appearing in the statement of Theorem 3.1. Since we are not interested in the regime where θ_n is large, we can rewrite (3.2) as $\theta_n' \geq \frac{cn}{\Sigma_n}\theta_n$. The fact that the θ_n then should satisfy such a dichotomy is quite clear: if β is such that they converge to a non-vanishing limit θ, then $\Sigma_n/n \sim \theta$ and one must have $\theta' \geq c$. If on the other hand they converge to 0 on a whole interval $[a,b]$, then that convergence must take place sufficiently fast so that $\Sigma_n/n \gg \theta_n$ (since otherwise the previous argument applies). Since $\Sigma_n/n \sim \theta_n$ for $\theta_n \sim n^{-\alpha}$ as soon as $\alpha < 1$, it is then plausible that for any $c < b$ one has $\theta_n \ll n^{-1/2}$ (say), implying $\theta_n' \gtrsim \sqrt{n}\theta_n$ and therefore $\theta_n \lesssim e^{-\sqrt{n}(c-\beta)}$ for $\beta < c$. This shows that Σ_n is bounded for $\beta < c$, leading to $\theta_n' \gtrsim n\theta_n$ and therefore an exponentially (in n) small bound as claimed.

4. Rotational invariance for the critical FK models

As already mentioned a number of times, a crucial feature of $2d$ equilibrium statistical mechanics is the fact that most models (at least those with sufficiently "local" interactions) are expected to obey a form of conformal invariance, or equivariance as in (1.2), when considering large-scale observables (crossing probabilities, averages, etc) at the critical temperature. This expectation and the resulting link to the well understood world of $2d$ conformal field theories allows to generate a plethora of conjectures regarding the large-scale behaviour of these models, but these are in many cases extremely hard to prove. Consider for example the N-step $2d$ self-avoiding random walk which is simply the uniform measure on all functions $h\colon \{0,\ldots,N\} \to \mathbf{Z}^2$ such that $h(0) = 0$ and such that $|h(i+1) - h(i)| = 1$ for all $i < N$. Exploiting the expected conformal invariance of its suitably rescaled large-N limit, one expects the size of $h(N)$ to be of order $N^{3/4}$ and its rescaling by $N^{3/4}$ to converge to a specific continuous random curve, namely $\mathrm{SLE}_{8/3}$ [LSW04]. Rigorously, almost *nothing* non-trivial is known: although the diameter of the range of h trivially has to be at least $\sqrt{N/\pi}$, the current best lower bound on the endpoint does not even match that! Instead, one only knows the bound $(\mathbf{E}|h(N)|^p)^{1/p} \geq \frac{1}{6}N^{p/(2p+2)}$ that was recently obtained by Madras [Mad14]. Similarly, while one trivially has $|h(N)| \leq N$, the best non-trivial upper bound is pretty much the weakest possible improvement, namely that for every $p \geq 1$ one has $\lim_{N\to\infty} N^{-1}(\mathbf{E}|h(N)|^p)^{1/p} = 0$, obtained around the same time by Duminil-Copin and Hammond [DCH13]. One main obstruction is that there is at the moment no proof showing that the self-avoiding random walk is conformally invariant at large scales.

While this illustrates the importance of showing that statistical models are conformally invariant (or at least rotationally invariant as a crucial first step) at criticality, the strategy of proof for such claims has so far mostly relied on finding a large enough collection of observables that already satisfy a discrete analogue of conformal invariance, typically by solving a discrete analogue of the Cauchy–Riemann equations. See for example Chelkak and Smirnov's proof of conformal invariance for the Ising model on isoradial graphs [CS12] and Smirnov's proof of conformal invariance for critical percolation [SS11]. The two-dimensional FK model with $q \leq 4$ already mentioned in Section 3 is one of the simplest models where conformal invariance at criticality is expected, but where it is not known how to obtain this from a suitable discrete conformal invariance. In the recent work

[DCKK+20], Duminil-Copin et al. show that the large-scale behaviour of these models is indeed rotationally invariant.

To define the notion of "large-scale behaviour", we recall that the configuration space of the FK model is the same as that for regular percolation, see Figure 1. Such a configuration can alternatively be described as a collection of non self-intersecting loops separating the percolation clusters from the clusters of the dual configuration. (Actually it naturally yields *two* collections of loops, depending on whether the loop encloses a percolation cluster of the primary or of the dual configuration, but we will ignore this detail for the sake of our exposition.) Given two collections $\mathcal{F}$ and $\bar{\mathcal{F}}$ of non self-intersecting loops in the plane, one then defines a distance between them in the following way. Given (small) $\eta > 0$, write $\mathcal{B}_\eta \subset \mathbf{R}^2$ for a large chunk of a fine lattice in $\mathbf{R}^2$, for example $\mathcal{B}_\eta = \eta\mathbf{Z}^2 \cap [-\eta^{-1}, \eta^{-1}]^2$. Given a loop γ and assuming that its image doesn't intersect the set $\mathcal{B}_\eta$, one then denotes by $[\eta]_\gamma$ its homotopy class in $\mathbf{R}^2 \setminus \mathcal{B}_\eta$. One then postulates that $d_H(\mathcal{F}, \bar{\mathcal{F}}) \leq \eta$ if and only if, for every $\gamma \in \mathcal{F}$ that encloses at least two elements of $\mathcal{B}_\eta$ but not all of it, there exists $\bar{\gamma} \in \bar{\mathcal{F}}$ such that $[\gamma]_\eta = [\bar{\gamma}]_\eta$ and vice-versa. (The H here stands for 'homotopy'.)

Given a metric space (M, d), the metric d lifts naturally to a metric on the space of probability measures on M which metrises the topology of weak convergence (at least when M is "nice", for example Polish). This is done by considering the Wasserstein (also sometimes called Kantorovich–Rubinstein or Monge–Kantorovich) distance

$$d(\mu, \nu) = \inf_{\mathbf{P} \in \mathcal{C}(\mu,\nu)} \int d(x, y)\, \mathbf{P}(dx, dy)\ ,$$

where $\mathcal{C}(\mu_1, \mu_2)$ denotes the set of all couplings between μ_1 and μ_2, that is probability measures on M^2 with ith marginal equal to μ_i. Note that with this definition, the map that assigns to x the probability measure δ_x concentrated at x is an isometry.

Fix now once and for all $q \in [1, 4]$ and consider a smooth bounded simply connected domain $\Omega \subset \mathbf{R}^2$. For $\varepsilon > 0$, write $\mathbf{P}_{\varepsilon,\Omega}$ for the critical FK measure (viewed as a measure on collections of loops) on $\varepsilon\mathbf{Z}^2 \cap \Omega$ with free boundary conditions. We also write $\mathbf{P}_\varepsilon$ for the limit of $\mathbf{P}_{\varepsilon,\Omega}$ as $\Omega \to \mathbf{R}^2$. Given an angle $\theta \in \mathbf{R}$, we also write R_θ for the rotation by θ, which naturally acts on loops in $\mathbf{R}^2$. The large-scale rotational invariance of the critical FK model can then be formulated as follows.

Theorem 4.1. *For every domain $\Omega \subset \mathbf{R}^2$ as above and every angle θ one has*

$$\lim_{\varepsilon \to 0} d_H\big(R_\theta^* \mathbf{P}_{\varepsilon,\Omega}, \mathbf{P}_{\varepsilon,R_\theta\Omega}\big) = 0\ .$$

Furthermore, one has $\lim_{\varepsilon\to 0} d_H(R_\theta^* \mathbf{P}_\varepsilon, \mathbf{P}_\varepsilon) = 0$.

We only focus on the second statement since it turns out that the first one can be deduced from it without too much effort. In fact, the authors of [DCKK+20] show a type of universality statement for the FK model on rectangular lattices, but its formulation requires some preparation. We start by defining a specific class of isoradial embeddings of the two-dimensional square lattice into the plane. Recall that a planar graph embedded in the plane is isoradial if, for each face f, there exists a circle of radius 1 containing all the vertices of f. (For example, the canonical embedding of the square lattice is isoradial.)

Given a bi-infinite sequence $\alpha\colon \mathbf{Z} \to (-\frac{\pi}{2}, \frac{\pi}{2})$, we consider the map $\iota_\alpha\colon \mathbf{Z}^2 \to \mathbf{R}^2$ given by

$$\iota_\alpha\colon (x, y) \mapsto \big(x + s_y, c_y\big)\ , \qquad s_y = \sum_{k\in(0,y]} \sin(\alpha_k)\ , \quad c_y = \sum_{k\in(0,y]} \cos(\alpha_k)\ ,$$

with the convention that for $y < 0$, $\sum_{(0,y]} = -\sum_{(y,0]}$. This defines an isoradial graph $L(\alpha)$ by considering the embedding of $\{(x, y) : x + y \,\text{even}\}$ (joined by diagonal edges) under ι_α (see Figure 3). The dual graph $L^*(\alpha)$ of $L(\alpha)$ is then given by the embedding of $\{(x, y) : x + y \,\text{odd}\}$. The associated "diamond graph" has as its vertices both the vertices of $L(\alpha)$ and the centres of its faces, and its edges are given by all pairs (v, f) with v a

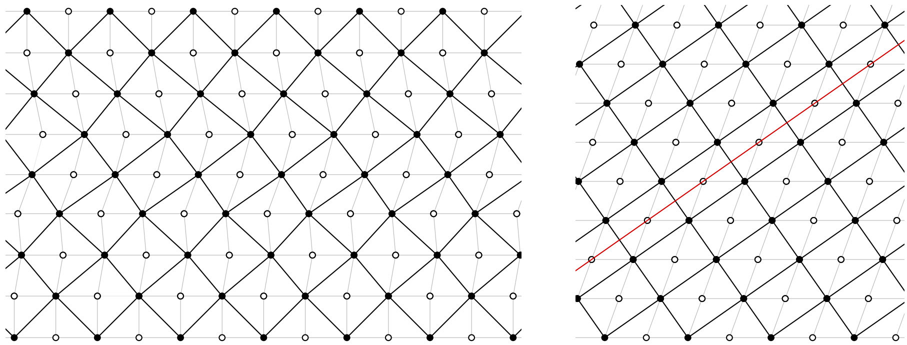


FIGURE 3. Examples of graphs $L(\alpha)$. On the left is a generic α while on the right α is constant but non-zero. The graph itself is drawn in black, the vertices of its dual graph are drawn in white, and the associated diamond graph is light gray. In red, we draw one of the symmetry axes of the second graph.

vertex and f a face such that $v \in f$. The diamond graph is simply given by the embedding of the usual lattice $\mathbf{Z}^2$ with nearest-neighbour edges under ι_α.

It is crucial at this stage to note that the critical FK model on $L(\alpha)$ is *not* given by simply pushing forward the critical FK model on $\mathbf{Z}^2$ under the map ι_α. Instead, one reweighs each edge of the graph in a very specific way that depends on the length of the edge. More specifically, viewing a configuration of the FK model as a subset $\omega \subset E$ of the set of edges of the (finite) graph on which the model is considered, the probability of seeing a given configuration ω is proportional to

$$\Big(\prod_{e\in\omega} p_e\Big)\Big(\prod_{e\in E\setminus\omega} (1-p_e)\Big)q^{k(\omega)}\;, \tag{4.1}$$

where $k(\omega)$ denotes the number of connected components of the subgraph ω. The formula for p_e as a function of q and the length of the edge e is explicit but not relevant for the sake of this discussion.

The most important step in the proof is to show that the large-scale connectivity properties of the critical FK model on $L(\alpha)$ are very close to those of the model on $L(T_j\alpha)$, where T_j swaps the jth and $(j+1)$th component:

$$(T_j\alpha)_k = \begin{cases} \alpha_{j+1} & \text{if } k=j,\\ \alpha_j & \text{if } k=j+1,\\ \alpha_k & \text{otherwise.}\end{cases}$$

Furthermore, there exists a natural coupling between the FK measures on the two lattices which implements this "closedness". This part of the proof exploits the link to the six vertex model and its "solvability" using the transfer matrix formalism. One then deduces from this that the model on the standard lattice $L(0)$ is very close to that on a rotated rectangular lattice $L(\alpha)$ with $k \mapsto \alpha_k$ constant (see the right half of Figure 3). This works by fixing some large $N > 0$ (which is then eventually sent to infinity) and starting from $\alpha_k^{(i)} = \alpha \mathbf{1}_{k\ge N}$ and then swapping components in such a way as to move some of the non-zero components down until une ends up with $\alpha_k^{(f)} = \alpha(\mathbf{1}_{|k|\le N} + \mathbf{1}_{k>3N})$. Since one has $L(0) \approx L(\alpha^{(i)})$ and $L(\alpha) \approx L(\alpha^{(f)})$, the desired statement follows if one can control the error made at each step of the argument. This turns out to be extremely delicate and one has to exploit subtle stochastic cancellations along the way. One trick is to allow the vertices of the set $\mathcal{B}_\eta$ around which the homotopy classes are computed to move a little bit with each application of a swapping operator T_j and to show that this motion ends up being diffusive (and therefore "slow") rather than ballistic.

Once one knows that $\lim_{\varepsilon\to 0} d_H(\mathbf{P}_{\varepsilon,L(0)}, \mathbf{P}_{\varepsilon,L(\alpha)}) = 0$, the second part of Theorem 4.1 follows at once. The idea is simply to note that $L(\alpha)$ is invariant under reflection along a line with angle $\frac{\pi}{4} - \frac{\alpha}{2}$, but that the effect of this reflection on $L(0)$ is the same as that of a rotation by angle α (since it is itself invariant under reflection along a line with angle $\frac{\pi}{4}$), so that

$$d_H(\mathbf{P}_\varepsilon, R_\alpha^*\mathbf{P}_\varepsilon) \le d_H(\mathbf{P}_{\varepsilon,L(0)}, \mathbf{P}_{\varepsilon,L(\alpha)}) + d_H(\mathbf{P}_{\varepsilon,L(\alpha)}, R_\alpha^*\mathbf{P}_{\varepsilon,L(0)}) = 2d_H(\mathbf{P}_{\varepsilon,L(0)}, \mathbf{P}_{\varepsilon,L(\alpha)})\ ,$$

and the claim follows.

References

[ADC21] M. Aizenman and H. Duminil-Copin. Marginal triviality of the scaling limits of critical 4D Ising and ϕ_4^4 models. *Ann. of Math. (2)* **194**, no. 1, (2021), 163–235. arXiv:1912.07973. doi:10.4007/annals.2021.194.1.3.

[ADCS15] M. Aizenman, H. Duminil-Copin, and V. Sidoravicius. Random currents and continuity of Ising model's spontaneous magnetization. *Comm. Math. Phys.* **334**, no. 2, (2015), 719–742. arXiv:1311.1937. doi:10.1007/s00220-014-2093-y.

[AF86] M. Aizenman and R. Fernández. On the critical behavior of the magnetization in high-dimensional Ising models. *J. Statist. Phys.* **44**, no. 3-4, (1986), 393–454. doi:10.1007/BF01011304.

[Aiz81] M. Aizenman. Proof of the triviality of φ_d^4 field theory and some mean-field features of Ising models for $d > 4$. *Phys. Rev. Lett.* **47**, no. 1, (1981), 1–4. doi:10.1103/PhysRevLett.47.1.

[Aiz82] M. Aizenman. Geometric analysis of φ^4 fields and Ising models. I, II. *Comm. Math. Phys.* **86**, no. 1, (1982), 1–48.

[Bax71] R. J. Baxter. Generalized ferroelectric model on a square lattice. *Studies in Appl. Math.* **50**, (1971), 51–69. doi:10.1002/sapm197150151.

[Bax73] R. J. Baxter. Potts model at the critical temperature. *Journal of Physics C: Solid State Physics* **6**, no. 23, (1973), L445–L448. doi:10.1088/0022-3719/6/23/005.

[BDC12] V. Beffara and H. Duminil-Copin. The self-dual point of the two-dimensional random-cluster model is critical for $q \ge 1$. *Probab. Theory Related Fields* **153**, no. 3-4, (2012), 511–542. arXiv:1006.5073. doi:10.1007/s00440-011-0353-8.

[BG93] T. W. Burkhardt and I. Guim. Conformal theory of the two-dimensional ising model with homogeneous boundary conditions and with disordred boundary fields. *Phys. Rev. B* **47**, (1993), 14306–14311. doi:10.1103/PhysRevB.47.14306.

[BKW76] R. J. Baxter, S. B. Kelland, and F. Y. Wu. Equivalence of the Potts model or Whitney polynomial with an ice-type model. *Journal of Physics A: Mathematical and General* **9**, no. 3, (1976), 397–406. doi:10.1088/0305-4470/9/3/009.

[BPZ84] A. Belavin, A. Polyakov, and A. Zamolodchikov. Infinite conformal symmetry in two-dimensional quantum field theory. *Nuclear Physics B* **241**, no. 2, (1984), 333–380. doi:https://doi.org/10.1016/0550-3213(84)90052-X.

[Car84] J. L. Cardy. Conformal invariance and surface critical behavior. *Nuclear Physics B* **240**, no. 4, (1984), 514–532. doi:https://doi.org/10.1016/0550-3213(84)90241-4.

[CGN15] F. Camia, C. Garban, and C. M. Newman. Planar Ising magnetization field I. Uniqueness of the critical scaling limit. *Ann. Probab.* **43**, no. 2, (2015), 528–571. arXiv:1205.6610. doi:10.1214/13-AOP881.

[CGN16] F. Camia, C. Garban, and C. M. Newman. Planar Ising magnetization field II. Properties of the critical and near-critical scaling limits. *Ann. Inst. Henri Poincaré Probab. Stat.* **52**, no. 1, (2016), 146–161. arXiv:1307.3926. doi:10.1214/14-AIHP643.

[CHI15] D. Chelkak, C. Hongler, and K. Izyurov. Conformal invariance of spin correlations in the planar Ising model. *Ann. of Math. (2)* **181**, no. 3, (2015), 1087–1138. doi:10.4007/annals.2015.181.3.5.

[CS12] D. Chelkak and S. Smirnov. Universality in the 2D Ising model and conformal invariance of fermionic observables. *Invent. Math.* **189**, no. 3, (2012), 515–580. arXiv:0910.2045. doi:10.1007/s00222-011-0371-2.

[DCGH+21] H. Duminil-Copin, M. Gagnebin, M. Harel, I. Manolescu, and V. Tassion. Discontinuity of the phase transition for the planar random-cluster and Potts models with $q > 4$. *Ann. Sci. Éc. Norm. Supér. (4)* **54**, no. 6, (2021), 1363–1413. arXiv:1611.09877. doi:10.24033/asens.2485.

[DCH13] H. Duminil-Copin and A. Hammond. Self-avoiding walk is sub-ballistic. *Comm. Math. Phys.* **324**, no. 2, (2013), 401–423. arXiv:1205.0401. doi:10.1007/s00220-013-1811-1.

[DCKK+20] H. Duminil-Copin, K. K. Kozlowski, D. Krachun, I. Manolescu, and M. Oulamara. Rotational invariance in critical planar lattice models. *arXiv preprint* (2020). arXiv:2012.11672.

[DCRT19] H. Duminil-Copin, A. Raoufi, and V. Tassion. Sharp phase transition for the random-cluster and Potts models via decision trees. *Ann. of Math. (2)* **189**, no. 1, (2019), 75–99. arXiv:1705.03104. doi:10.4007/annals.2019.189.1.2.

[DCST17] H. Duminil-Copin, V. Sidoravicius, and V. Tassion. Continuity of the phase transition for planar random-cluster and Potts models with $1 \leq q \leq 4$. *Comm. Math. Phys.* **349**, no. 1, (2017), 47–107. arXiv:1505.04159. doi:10.1007/s00220-016-2759-8.

[ESPP+12] S. El-Showk, M. F. Paulos, D. Poland, S. Rychkov, D. Simmons-Duffin, and A. Vichi. Solving the 3d Ising model with the conformal bootstrap. *Phys. Rev. D* **86**, (2012), 025022. arXiv:1203.6064. doi:10.1103/PhysRevD.86.025022.

[ESPP+14] S. El-Showk, M. F. Paulos, D. Poland, S. Rychkov, D. Simmons-Duffin, and A. Vichi. Solving the 3d Ising model with the conformal bootstrap II. c-minimization and precise critical exponents. *J. Stat. Phys.* **157**, no. 4-5, (2014), 869–914. arXiv:1403.4545. doi:10.1007/s10955-014-1042-7.

[FILS78] J. Fröhlich, R. Israel, E. H. Lieb, and B. Simon. Phase transitions and reflection positivity. I. General theory and long range lattice models. *Comm. Math. Phys.* **62**, no. 1, (1978), 1–34.

[FK72] C. M. Fortuin and P. W. Kasteleyn. On the random-cluster model. I. Introduction and relation to other models. *Physica* **57**, (1972), 536–564.

[Frö82] J. Fröhlich. On the triviality of $\lambda\varphi_d^4$ theories and the approach to the critical point in $d > 4$ dimensions. *Nuclear Phys. B* **200**, no. 2, (1982), 281–296. doi:10.1016/0550-3213(82)90088-8.

[FSS76] J. Fröhlich, B. Simon, and T. Spencer. Infrared bounds, phase transitions and continuous symmetry breaking. *Comm. Math. Phys.* **50**, no. 1, (1976), 79–95.

[GHS70] R. B. Griffiths, C. A. Hurst, and S. Sherman. Concavity of magnetization of an Ising ferromagnet in a positive external field. *J. Mathematical Phys.* **11**, (1970), 790–795. doi:10.1063/1.1665211.

[Gri06] G. Grimmett. *The random-cluster model*, vol. 333 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2006, xiv+377. doi:10.1007/978-3-540-32891-9.

[HS94] T. Hara and G. Slade. Mean-field behaviour and the lace expansion. In *Probability and phase transition (Cambridge, 1993)*, vol. 420 of *NATO Adv. Sci. Inst. Ser. C: Math. Phys. Sci.*, 87–122. Kluwer Acad. Publ., Dordrecht, 1994.

[Isi25] E. Ising. Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift fur Physik* **31**, no. 1, (1925), 253–258. doi:10.1007/BF02980577.

[LSW04] G. F. Lawler, O. Schramm, and W. Werner. On the scaling limit of planar self-avoiding walk. In *Fractal geometry and applications: a jubilee of Benoît Mandelbrot, Part 2*, vol. 72 of *Proc. Sympos. Pure Math.*, 339–364. Amer. Math. Soc., Providence, RI, 2004. doi:10.1214/aop/1079021469.

[Mad14] N. Madras. A lower bound for the end-to-end distance of the self-avoiding walk. *Canad. Math. Bull.* **57**, no. 1, (2014), 113–118. doi:10.4153/CMB-2012-022-6.

[MS16] J. Miller and S. Sheffield. Quantum Loewner evolution. *Duke Math. J.* **165**, no. 17, (2016), 3241–3378. arXiv:1312.5745. doi:10.1215/00127094-3627096.

[Ons44] L. Onsager. Crystal statistics. I. A two-dimensional model with an order-disorder transition. *Phys. Rev. (2)* **65**, (1944), 117–149.

[OSSS05] R. O'Donnell, M. Saks, O. Schramm, and R. Servedio. Every decision tree has an influential variable. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05)*, 31–39. 2005. arXiv:cs/0508071. doi:10.1109/SFCS.2005.34.

[Sak07] A. Sakai. Lace expansion for the Ising model. *Comm. Math. Phys.* **272**, no. 2, (2007), 283–344. doi:10.1007/s00220-007-0227-1.

[Sch00] O. Schramm. Scaling limits of loop-erased random walks and uniform spanning trees. *Israel J. Math.* **118**, (2000), 221–288. doi:10.1007/BF02803524.

[SS11] O. Schramm and S. Smirnov. On the scaling limits of planar percolation. *Ann. Probab.* **39**, no. 5, (2011), 1768–1814. With an appendix by Christophe Garban. arXiv:1101.5820. doi:10.1214/11-AOP659.

[TL71] H. N. V. Temperley and E. H. Lieb. Relations between the "percolation" and "colouring" problem and other graph-theoretical problems associated with regular planar lattices: some exact results for the "percolation" problem. *Proc. Roy. Soc. London Ser. A* **322**, no. 1549, (1971), 251–280. doi:10.1098/rspa.1971.0067.

[Yan52] C. N. Yang. The spontaneous magnetization of a two-dimensional Ising model. *Phys. Rev. (2)* **85**, (1952), 808–816.

EPFL, 1015 Lausanne, Switzerland, Imperial College London, London SW7 2AZ, UK
Email address: m.hairer@epfl.ch, m.hairer@imperial.ac.uk

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.

Friday, January 7, 2022 (Virtual)

Tom Scanlon, University of California, Berkeley
Tame geometry for Hodge theory

Elena Giorgi, Columbia University
The stability of black holes with matter

Anup Rao, University of Washington
Sunflowers: from soil to oil

Elamin Elbasha, Merck & Co., Inc.
Mathematics and the quest for vaccination-induced herd immunity threshold

January 18, 2021 (Virtual)

Abba Gumel, Arizona State University
Mathematics of the Dynamics and Control of the COVID-19 Pandemic

Ana Caraiani, Imperial College London
An excursion through the land of shtukas

Jennifer Hom, Georgia Institute of Technology
Getting a handle on the Conway knot

Richard Evan Schwartz, Brown University
Rectangles, Curves, and Klein Bottles

January 17, 2020 (Denver, CO)

Jordan S. Ellenberg, University of Wisconsin-Madison
Geometry, Inference, and Democracy

Bjorn Poonen, Massachusetts Institute of Technology
A p-adic approach to rational points on curves

Suncica Canic, University of California, Berkeley
Recent Progress on Moving Boundary Problems

Vlad C. Vicol, Courant Institute of Mathematical Sciences, New York University
Convex integration and fluid turbulence

January 18, 2019 (Baltimore, MD)

Bhargav Bhatt, University of Michigan
Perfectoid geometry and its applications

Thomas Vidick, California Institute of Technology
Verifying quantum computations at scale: a cryptographic leash on quantum devices

Stephanie van Willigenburg, University of British Columbia
The shuffle conjecture

Robert Lazarsfeld, Stony Brook University
Tangent Developable Surfaces and the Equations Defining Algebraic Curves

January 12, 2018 (San Diego, CA)

Richard D. James, University of Minnesota
Materials from mathematics

Craig L. Huneke, University of Virginia
How complicated are polynomials in many variables?

Isabelle Gallagher, Université Paris Diderot
From Newton to Navier-Stokes, or how to connect fluid mechanics equations from microscopic to macroscopic scales

Joshua A. Grochow, University of Colorado, Boulder
The Cap Set Conjecture, the polynomial method, and applications
(after Croot-Lev-Pach, Ellenberg-Gijswijt, and others)

January 6, 2017 (Atlanta, GA)

Lydia Bieri, University of Michigan
Black hole formation and stability: a mathematical investigation.

Matt Baker, Georgia Tech
Hodge Theory in Combinatorics.

Kannan Soundararajan, Stanford University
Tao's work on the Erdos Discrepancy Problem.

Susan Holmes, Stanford University
Statistical proof and the problem of irreproducibility.

January 8, 2016 (Seattle, WA)

Carina Curto, Pennsylvania State University
What can topology tell us about the neural code?

Lionel Levine, Cornell University and *Yuval Peres, Microsoft Research and University of California, Berkeley
Laplacian growth, sandpiles and scaling limits.

Timothy Gowers, Cambridge University
Probabilistic combinatorics and the recent work of Peter Keevash.

Amie Wilkinson, University of Chicago
What are Lyapunov exponents, and why are they interesting?

January 12, 2015 (San Antonio, TX)

Jared S. Weinstein, Boston University
Exploring the Galois group of the rational numbers: Recent breakthroughs.

Andrea R. Nahmod, University of Massachusetts, Amherst
The nonlinear Schrödinger equation on tori: Integrating harmonic analysis, geometry, and probability.

Mina Aganagic, University of California, Berkeley
String theory and math: Why this marriage may last.

Alex Wright, Stanford University
From rational billiards to dynamics on moduli spaces.

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
C*ontact 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

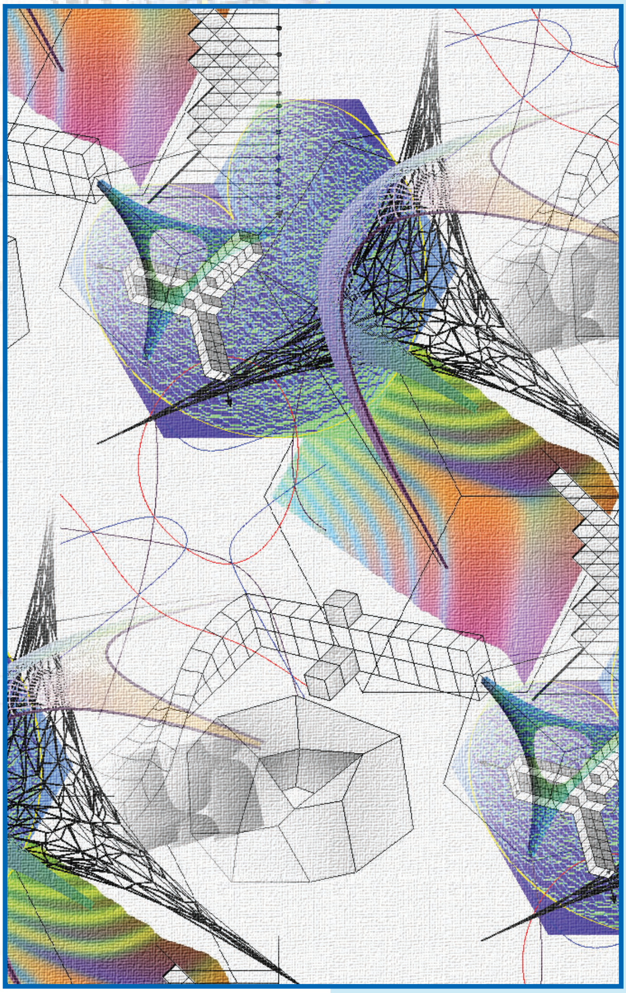
Ingrid Daubechies, Princeton University
Sublinear algorithms for sparse approximations with excellent odds

Edward Frenkel, University of California, Berkeley
Recent advances in the Langlands Program

Daniel Tataru, University of California, Berkeley
The wave maps equation

2023 CURRENT EVENTS BULLETIN

Committee



The back cover graphic is reprinted courtesy of Andrei Okounkov.

Cover graphic associated with Andrew Granville's talk courtesy of The American Mathematical Society. Guitar image courtesy of CaoChunhai / iStock / Getty Images.

Cover graphic associated with Christopher Eur's talk courtesy of The American Mathematical Society. Real points of the Cayley surface image courtesy of Salix alba/Wikimedia Commons.

Cover graphic associated with Henry Cohn's talk courtesy of Baloncici / iStock / Getty Images Plus.

Cover graphic associated with Martin Hairer's talk courtesy of The American Mathematical Society. Photography by Wolfgang Volz.