

1. THE TWIN PRIME CONJECTURE

One of the most famous problems in mathematics is the Twin Prime Conjecture.

Conjecture 1 (Twin Prime Conjecture). *There are infinitely many pairs of primes which differ by 2.*

There is some debate as to how old the Twin Prime Conjecture is - it was certainly considered by de Polignac over 150 years ago [17], but there has been speculation that it could go back much further, potentially as far back as Euclid and the ancient Greeks over 2000 years ago.

Unfortunately the Twin Prime Conjecture remains unsolved, but it is strongly believed to be true. Hardy and Littlewood [32] made a more precise conjecture about the density of twin primes. In its most optimistic form, this is the following.

Conjecture 2 (Quantitative Twin Prime Conjecture). *For any fixed $\epsilon > 0$, the number of twin primes up to x , $\pi_2(x)$, is given by*

$$\pi_2(x) = \#\{p \leq x : p + 2 \text{ prime}\} = \mathfrak{S} \int_2^x \frac{dt}{\log^2 t} + O(x^{1/2+\epsilon}),$$

where

$$\mathfrak{S} = 2 \prod_{p \geq 3} \left(1 - \frac{1}{(p-1)^2}\right).$$

(Here, and throughout the paper, ϵ will denote a small positive quantity, and various implied constants in $O(\cdot)$ notation will be allowed to depend on ϵ without us explicitly saying so. In particular, the implied constant in the term $O(x^{1/2+\epsilon})$ may depend on ϵ .)

1.1. Numerical evidence. The quantitative variant of the Twin Prime Conjecture given by Conjecture 2 allows one to test it numerically, and the evidence is as compelling as one could possibly hope for. A huge amount of computation has been done. If we let $\text{Li}_2(x)$ denote the integral in Conjecture 2, then we have the following.

x	$\pi_2(x)$	$\mathfrak{S} \operatorname{Li}_2(x)$	$\pi_2(x) - \mathfrak{S} \operatorname{Li}_2(x)$
10	2	4.8...	-2.8...
10^2	8	13.5...	-5.5...
10^3	35	45.7...	-10.7...
10^4	205	214.2...	-9.2...
10^5	1224	1248.7...	-24.7...
10^6	8169	8248.0...	-79.0...
10^7	58980	58753.8...	226.1...
10^8	440312	440367.7...	-55.7...
10^9	3424506	3425308.1...	-802.1...
10^{10}	27412679	27411416.5...	1262.4...
10^{11}	224376048	224368864.7...	7183.2...
10^{12}	1870585220	1870559867.8...	25352.1...
10^{13}	15834664872	15834598305.0...	66566.9...
10^{14}	135780321665	135780264894.4...	56770.5...
10^{15}	1177209242304	1177208491860.6...	750443.3...

The point of this table is that as soon as x is moderately large, the approximation $\mathfrak{S} \operatorname{Li}_2(x)$ is exceptionally close to $\pi_2(x)$, and the difference consistently fits the prediction of being of size about \sqrt{x} . The twin prime counts are known up to at least 10^{18} ; we have stopped at 10^{15} in the interests of space alone!

1.2. Cramér's model. The Twin Prime Conjecture is supported by various heuristic arguments to justify the main term $\mathfrak{S} \operatorname{Li}_2(x)$. One such heuristic is based on a random model of the primes due to Harold Cramér.

If one chooses an integer n uniformly at random between x and $2x$, then the probability that n is prime is roughly $1/\log x$ (this is a consequence of the Prime Number Theorem). We might naively guess that the property of n being prime should be roughly independent of the property of $n+2$ being prime when x is large, and so the probability that both n and $n+2$ are prime would be about $1/(\log n) \cdot 1/\log(n+2) \approx 1/(\log n)^2$. This would lead one to guess that there are about $x/(\log x)^2$ twin primes up to x . This is the same order of magnitude as the prediction of Conjecture 2, but misses the constant \mathfrak{S} .

The reason for this discrepancy is because we also need to account for the fact that primes have no small prime factors - if $n > 2$ is prime, then we know that $n+2$ is odd, which makes it slightly more likely to be prime. If both n and $n+2$ are primes of size x , then they cannot be a multiple of any small prime p . If n and $n+2$ were independent, the chances of neither being a multiple of p would be about $(1-1/p)^2$. However, in reality they are not independent; neither is a multiple of p with probability $(1-2/p)$ if $p > 2$, since n needs to avoid the residue classes 0 and $-2 \pmod{p}$. Therefore we should refine the independence assumption, to take this into account. (If $p = 2$

then n just needs to avoid the residue class $0 \pmod{2}$, which occurs with probability about $1/2$.) This suggests we should have a correction factor of $(1 - 2/p)/(1 - 1/p)^2$ for each such small prime p , and so the total correction factor should be

$$\frac{1/2}{(1 - 1/2)^2} \prod_{p>2} \frac{1 - 2/p}{(1 - 1/p)^2} = \mathfrak{S}.$$

This leads to the guess of $\mathfrak{S}x/\log^2 x$ twin primes up to x , as in Conjecture 2.

1.3. Circle method heuristic. An alternative heuristic comes from a technique known as the Hardy-Littlewood circle method. One way to try to prove results like the Twin Prime Conjecture is to try to use the circle method to count the number of twin primes up to x . The idea is to first express the count as an integral:

$$\pi_2(x) = \int_0^1 \left| \sum_{p \leq x+2} e^{2\pi i p \theta} \right|^2 e^{4\pi i \theta} d\theta,$$

and then to try to understand the integral by separating parts when the integrand is large from the parts when it is small. This is the way we prove the three prime version of Goldbach's conjecture, for example. We know that the integrand is large whenever θ is close to a rational with small denominator. For these parts of the integral, we can understand the relevant exponential sums using results on primes in arithmetic progressions. This allows us to get an asymptotic estimate for the contribution from these θ , which turns out to be

$$(1 + o(1))\mathfrak{S} \int_2^x \frac{dt}{\log^2 t}.$$

This is exactly the expected main term! We would successfully prove the Twin Prime Conjecture if we could show the contribution amongst θ *not* close to a rational with small denominator were small. Unfortunately, we do not know how to do this, although we expect this to be the case. In similar simpler problems we can show these terms are negligible, but for the Twin Prime Problem we don't know how to show cancellation between different θ , and considering each θ individually is insufficient. However, since we still expect the contribution to be small, this gives an alternative justification for the guess of Conjecture 2.

2. SIEVE METHODS AND RELATED PROBLEMS

Although the Twin Prime Conjecture is still very much open, we have succeeded in making partial progress in various directions. The aim of this article is to outline some recent breakthroughs making dramatic new progress on questions related to the conjecture.

2.1. Early results from sieve methods. During the 20th century, a key development in Prime Number Theory was the invention of *sieve methods*. These are flexible tools for studying the primes, by relating the primes to simpler mathematical objects which can be analyzed effectively. These simpler objects are weighted integers, where the weights are chosen carefully so as to be heavily concentrated on the primes, but are simple enough to calculate with.

Viggo Brun, the pioneer of sieve methods, showed that the number of twin primes up to x is $O(x(\log \log x / \log x)^2)$, which is only slightly larger than the conjectured truth. Slight refinements allow us to show that Conjecture 2 cannot underestimate the number of twin primes by more than a bounded amount.

Theorem 1 (Sieve bound on Twin Primes). *There is a constant $C > 0$ such that for $x \geq 3$*

$$\pi_2(x) \leq C \mathfrak{S} \int_2^x \frac{dt}{(\log t)^2}.$$

The current record is that one can take C slightly below 3.4 if x is large enough ([74], building on the earlier work [14, 25, 22, 9, 23]). Even Brun's original result implied that the sum of reciprocals of twin primes converges, unlike the sum of reciprocals of the primes.

Corollary 2 (Brun's Constant). *The sum of reciprocals of all twin prime pairs converges to a constant*

$$B := \left(\frac{1}{3} + \frac{1}{5}\right) + \left(\frac{1}{5} + \frac{1}{7}\right) + \left(\frac{1}{11} + \frac{1}{13}\right) + \dots \approx 1.902\dots$$

As well as proving upper bounds for counts of primes with certain primes, sieve methods are also generally able to show one can solve many such problems if one replaces 'prime' with 'an integer with at most r of prime factors', for some fixed constant r . In the context of the Twin Prime Conjecture, the strongest result of this type is Chen's celebrated theorem [13].

Theorem 3 (Chen's Theorem). *There are infinitely many primes p such that $p + 2$ has at most 2 prime factors.*

Unfortunately there is a key limitation inherent in standard sieve methods which means very significant new ideas are required if one wants to prove the Twin Prime Conjecture itself. Indeed, it is known that sieve methods struggle to distinguish between integers with an even or an odd number of prime factors, and without overcoming this limitation one can't hope to prove a stronger result than Chen's theorem.

2.2. Bounded gaps between primes. An alternative approximation to the Twin Prime Conjecture is to try to show that there are pairs of primes which are close to each other, even if they do not exactly differ

by 2. This problem was studied by many authors [19, 61, 62, 8, 54, 71, 34, 35, 36, 44], but progress was slow. In 2005 a breakthrough due to Goldston, Pintz and Yıldırım [29] showed that there are pairs of primes arbitrarily close to each other compared with the average gap.

Theorem 4. *Let $\epsilon > 0$. Then there are infinitely many pairs (p_1, p_2) of distinct primes such that*

$$|p_1 - p_2| \leq \epsilon \log p_1$$

Moreover, the new method - now known as the ‘GPY method’ - could show that there are infinitely many such pairs (p_1, p_2) with $|p_1 - p_2|$ bounded, if one assumed a widely believed conjecture on the distribution on primes in arithmetic progressions. Despite many significant results in this direction [24, 9, 10, 11] this conjecture seemed out of reach. In 2013, Zhang [75] proved a suitable variant of this conjecture, which allowed him to use the GPY method to prove that there are infinitely many bounded gaps between primes.

Theorem 5 (Zhang’s Theorem). *There are infinitely many pairs (p_1, p_2) of distinct primes such that*

$$|p_1 - p_2| \leq 70\,000\,000.$$

After Zhang’s breakthrough, a new variant of the GPY method was discovered independently by the author [50] and Tao (unpublished). This gave an alternative way of proving bounded gaps between primes, but had several other consequences as well since it was flexible and could show the existence of clumps of many primes in bounded length intervals. Moreover, it turned out this new method produced stronger numerical results for the size of the gap. A joint collaborative project was set up to optimize the numerics in this bound. First Zhang’s arguments were optimized and improved in [60], and then the new method was also optimized, resulting in the current world record [59].

Theorem 6. *There are infinitely many pairs (p_1, p_2) of distinct primes such that*

$$|p_1 - p_2| \leq 246.$$

All of these approaches actually prove a weak approximation of a far-reaching generalization of the Twin Prime Conjecture, and then deduce results about gaps between primes from this result. The generalization is instead of asking for simultaneous prime values of the functions n and $n + 2$, it considers a more general set of many linear functions.

Definition (Admissibility). *Let $\mathcal{L} = \{L_1, \dots, L_k\}$ be a set of linear functions in one variable with integer coefficients and positive lead coefficient (i.e. $L_i(n) = a_i n + b_i$ for some $a_i, b_i \in \mathbb{Z}$ with $a_i > 0$).*

We say \mathcal{L} is admissible if $\prod_{i=1}^k L_i(n)$ has no fixed prime divisor (i.e. if for every prime p there is an integer n_p such that $\prod_{i=1}^k L_i(n_p)$ is coprime to p .)

Given a set of linear functions, it is easy to check whether they are admissible or not, so this should be thought of as a simple property. If a set of linear functions is not admissible (i.e. there is a fixed prime divisor of $\prod_{i=1}^k L_i(n)$) then it is straightforward to see that there can only be finitely many integers n such that all the $L_i(n)$ are prime. The far-reaching Prime k -tuples Conjecture claims that this simple necessary condition is in fact sufficient.

Conjecture 3 (Prime k -tuples Conjecture). *Let $\mathcal{L} = \{L_1, \dots, L_k\}$ be an admissible set of linear functions. Then there are infinitely many integers n such that all of $L_1(n), \dots, L_k(n)$ are primes.*

Conjecture 1 follows immediately from this by taking $\mathcal{L} = \{L_1, L_2\}$ with $L_1(n) = n$, $L_2(n) = n + 2$, but the Prime k -tuples Conjecture goes much further and would have a huge number of applications. It almost completely describes the ‘small scale’ structure of the primes. Since we do not know how to solve the Twin Prime Conjecture, we certainly do not know how to solve the Prime k -tuples Conjecture. Recent developments, however, have been based on establishing weak versions of the Prime k -tuples Conjecture, where *several* rather than *all* of the linear functions are simultaneously prime. One consequence is we now can prove the following Theorem [51].

Theorem 7 (Weak k -tuples). *There is a constant $c > 0$ such that the following holds.*

Let \mathcal{L} be an admissible set of k linear functions. Then there are infinitely many integers n such that at least $c \log k$ of $L_1(n), \dots, L_k(n)$ are primes.

The fact that there are infinitely many bounded gaps between primes follows immediately from Theorem 7 on choosing k such that $c \log k \geq 2$ and $\mathcal{L} = \{L_1, \dots, L_k\}$ with $L_i(n) = n + h_i$ for suitable constants h_i such that \mathcal{L} is admissible (choosing $h_i = i \times k!$ works, for example). The stronger numerical results such as Theorem 6 come from numerically optimizing the proof of Theorem 7 to determine what the smallest value of k is such that $c \log k$ can be replaced by 2.

We briefly overview some of the ideas behind the GPY method in Section 3, and then sketch some of the ideas in the proof of Theorem 7 in Section 4.

2.3. Chowla’s conjecture. Even though we can prove results such as Theorem 6, significant new ideas are required to prove the Twin Prime Conjecture itself. As mentioned before, sieve methods have a limitation in that they struggle to distinguish numbers with an even number of prime factors from numbers with an odd number of prime factors. Results such as Theorem 7 exploit the flexibility of only asking for some of the linear functions to be prime to sidestep this limitation,

without directly addressing it. To prove the Twin Prime Conjecture, however, it is likely one will need to address this limitation directly.

One way of seeing this is to consider variants of questions about the primes where instead one considers the Liouville λ function, given by

$$\lambda(n) = \begin{cases} -1, & n \text{ has an odd number of prime factors,} \\ 1, & \text{otherwise.} \end{cases}$$

Showing cancellation in $\lambda(n)$ was generally considered as difficult as studying the primes in any given problem. For example, Chowla [15] made a general conjecture in the same spirit as Conjecture 3. For the analogue of the Twin Prime Conjecture, this takes the following form.

Conjecture 4 (Weak form of Chowla's Conjecture). *We have*

$$\sum_{n \leq x} \lambda(n)\lambda(n+2) = o(x).$$

Until very recently it was believed that this conjecture was essentially as hard as the Twin Prime Conjecture, and the limitations of sieve methods meant that they were poorly suited to addressing this problem. It was therefore considered a major advance when Tao [66] showed that a slightly weaker version of Conjecture 4 is true, where we use a logarithmic average instead of a standard average.

Theorem 8 (Logarithmically Averaged Chowla).

$$\sum_{n \leq x} \frac{\lambda(n)\lambda(n+2)}{n} = o(\log x).$$

This result introduced several new ideas, but also relied heavily on a recent breakthrough of Matomäki and Radziwiłł [46], who had shown that for the question of almost all short intervals, one could prove results for $\lambda(n)$ which are much better than what is known for primes.

Theorem 9 (Liouville in short intervals). *For almost all intervals $[x, x+h] \subseteq [Y, 2Y]$, we have*

$$\sum_{n \in [x, x+h]} \lambda(n) = O\left(\frac{h}{(\log h)^{1/10}}\right).$$

The Matomäki-Radziwiłł Theorem is optimal in the sense that provided $h \rightarrow \infty$ there is cancellation in the Liouville function in almost all short intervals. By way of comparison, for primes we only know that there is the expected number of primes in almost all intervals $[x, x+h] \subseteq [Y, 2Y]$ when $h \geq Y^{1/6}$, a much weaker result. This therefore demonstrates that the previous belief that many questions about $\lambda(n)$ were as hard as the same question about the primes is not correct.

Unfortunately it currently appears that these breakthroughs do not directly address the Twin Prime Conjecture itself - they rely on showing

that the behaviour of small prime factors and large prime factors is roughly independent, and cancellation comes from small prime factors of integers. This means the methods do not appear to adapt to say anything about the primes (where there are no small prime factors), but the mere fact that something can be said about these questions which directly address the limitations of sieve techniques is a big advance.

Although new ideas appear to be required to prove the Twin Prime Conjecture, the breakthroughs of Theorems 7, 8 and 9 have already had several further applications, including new results on large gaps between primes [21, 20, 52], the resolution of the Erdős discrepancy problem [65], as well as many other results on the distribution of primes [67, 51, 69, 12, 6, 28, 57, 42, 2, 49, 3, 16, 72, 70, 55, 56, 33, 4, 5, 1, 37, 53, 58, 1, 45, 43] and correlations of multiplicative functions [68, 48, 39, 38, 30, 41, 27, 26].

We will sketch some of the key ideas behind Theorem 8 in Section 5, assuming a variant of Theorem 9. We then sketch some of the ideas behind Theorem 9 in Section 6.

3. THE GPY METHOD AND WEAK PRIME k -TUPLES

We aim to prove Theorem 7 by the ‘GPY method’, which can be interpreted as a first moment method which was introduced to study small gaps between primes by Goldston, Pintz and Yıldırım [29]. This takes the following basic steps, for some given set $\{L_1, \dots, L_k\}$ of distinct functions satisfying the hypotheses of Conjecture 3:

1. We choose a probability measure w supported on integers in $[X, 2X]$.
2. We calculate the expected number of the functions $L_i(n)$ which are prime at n , if n is chosen randomly with probability $w(n)$.
3. If this expectation is at least m , then there must be some $n \in [X, 2X]$ such that at least m of $L_1(n), \dots, L_k(n)$ are primes.
4. If this holds for all large X , then there are infinitely many such integers n .

This procedure only works if we can find a probability measure w which is suitably concentrated on integers n where many of the $L_i(n)$ are prime, but at the same time is simple enough that we can calculate this expectation unconditionally. We note that by linearity of expectation, it suffices to be able to calculate the probability that any one of the linear functions is prime. However, any slowly changing smooth function w is insufficient since the primes have density 0 in the integers, whereas any choice of w explicitly depending on the joint distribution of prime values of the L_i is likely to be too complicated to handle unconditionally.

Sieve methods provide natural choices for the probability measure w , since they produce arithmetic objects concentrated on primes but

simple enough to calculate with. The situation of simultaneous prime values of L_1, \dots, L_k is a ‘ k -dimensional’ sieve problem. For such problems when k is large but fixed, the Selberg sieve tends to be the type of sieve which performs best. The standard choice of Selberg sieve weights (which are essentially optimal in closely related situations) are

$$(3.1) \quad w(n) \propto \left(\sum_{\substack{d \mid \prod_{i=1}^k L_i(n) \\ d < R}} \mu(d) \left(\log \frac{R}{d} \right)^k \right)^2,$$

where R is a parameter which controls the complexity of the sieve weights, and w is normalized to sum to 1 on $[X, 2X]$.

To calculate the probability that $L_j(n)$ is prime with this choice of $w(n)$, we wish to estimate the sum of $w(n)$ over $n \in [X, 2X]$ such that $L_j(n)$ is prime. To do this we typically expand the divisor sum in the definition (3.1) and swap the order of summation. This reduces the problem to estimating the number of prime values of $L_j(n)$ for $n \in [X, 2X]$ in many different arithmetic progressions with moduli of size about R^2 . The Elliott-Halberstam conjecture [18] asserts that we should be able to do this when $R^2 < X^{1-\epsilon}$, but unconditionally we only know how to do this when $R^2 < X^{1/2-\epsilon}$, using the Bombieri-Vinogradov Theorem [7, 73]. After some computation one finds that, provided we do have suitable estimates for primes in arithmetic progressions, the choice (3.1) gives

$$(3.2) \quad \mathbb{E} \#\{1 \leq i \leq k : L_i(n) \text{ prime}\} = \left(2 - \frac{2}{k+1} + o(1) \right) \frac{\log R}{\log X}.$$

In particular, this is less than 1 for all large X , even if we optimistically assume the Elliott-Halberstam conjecture and take $R \approx X^{1/2-\epsilon}$, and so it appears that we will not be able to conclude anything about primes in this manner.

The groundbreaking work of Goldston-Pintz-Yıldırım [29] showed that a variant of this choice of weight actually performs much better. They considered

$$(3.3) \quad w(n) \propto \left(\sum_{\substack{d \mid \prod_{i=1}^k L_i(n) \\ d < R}} \mu(d) \left(\log \frac{R}{d} \right)^{k+\ell} \right)^2,$$

where ℓ is an additional parameter to be optimized over. With the choice (3.3), one finds that provided we have suitable estimates for primes in arithmetic progressions, we obtain

$$\mathbb{E} \#\{1 \leq i \leq k : L_i(n) \text{ prime}\} = \left(4 + O\left(\frac{1}{\ell}\right) + O\left(\frac{\ell}{k}\right) \right) \frac{\log R}{\log X}.$$

This improves upon (3.2) by a factor of about 2 when k is large and $\ell \approx k^{1/2}$. This falls just short of proving that two of the linear functions

are simultaneously prime when $R = X^{1/4-\epsilon}$ as allowed by the Bombieri-Vinogradov theorem, but any small improvement allowing $R = X^{1/4+\epsilon}$ would give bounded gaps between primes! By considering additional possible primes, Goldston Pintz and Yıldırım were to use this argument to prove Theorem 4.

Building on earlier work, Zhang [75] succeeded in establishing an extended version of the Bombieri-Vinogradov Theorem for moduli with no large prime factors allowing $R = X^{1/4+\epsilon}$, and ultimately this allowed him to show

$$\mathbb{E} \#\{1 \leq i \leq k : L_i(n) \text{ prime}\} > 1,$$

if $k > 3\,500\,000$ and X is sufficiently large. The key breakthrough in Zhang's work was this result on primes in arithmetic progressions of modulus slightly larger than $X^{1/2}$. By choosing suitable linear functions, this then established Theorem 5.

An alternative approach to extending the work of Goldston, Pintz and Yıldırım was developed independently by the author [50] and Tao (unpublished). The key difference was to consider the multidimensional generalization

$$(3.4) \quad w(n) \propto \left(\sum_{\substack{d_1, \dots, d_k \\ d_i | L_i(n) \\ \prod_{i=1}^k d_i < R}} \mu(d_1 \cdots d_k) F\left(\frac{\log d_1}{\log R}, \dots, \frac{\log d_k}{\log R}\right) \right)^2,$$

for suitable smooth functions $F : \mathbb{R}^k \rightarrow \mathbb{R}$ supported on $[0, \infty)^k$. The flexibility of allowing the function F to depend on each divisor d_1, \dots, d_k of $L_1(n), \dots, L_k(n)$ allows us to make $w(n)$ more concentrated on integers n when many of the $L_i(n)$ are prime (at the cost of making it less concentrated on integers n when *all* the linear functions are prime). This is the key new idea allowing for the proof of Theorem 7.

For further reading on the GPY method, we recommend the excellent survey article of Soundararajan [63]. For further reading about the more recent developments surrounding bounded gaps between primes, we recommend the survey articles of Granville [31] and Kowalski [40].

4. A MODIFIED GPY SIEVE METHOD

For a given fixed admissible set $\{L_1, \dots, L_k\}$, we wish to estimate

$$\mathbb{E} \#\{1 \leq i \leq k : L_i(n) \text{ prime}\} = \sum_{i=1}^k \mathbb{P}(L_i(n) \text{ prime}),$$

when $n \in [x, 2x]$ is chosen randomly with probability according to (3.4). Let

$$\nu(n) = \left(\sum_{\substack{d_1, \dots, d_k \\ d_i | L_i(n) \\ \prod_{i=1}^k d_i < R}} \mu(d_1 \cdots d_k) F\left(\frac{\log d_1}{\log R}, \dots, \frac{\log d_k}{\log R}\right) \right)^2,$$

for some piecewise smooth function F , independent of x . Thus $w(n) = \nu(n) / (\sum_{m \in [x, 2x]} \nu(m))$, and we have

$$\mathbb{P}(L_i(n) \text{ prime}) = \frac{\sum_{n \in [x, 2x]} \nu(n) \mathbf{1}_{\text{prime}}(L_i(n))}{\sum_{n \in [x, 2x]} \nu(n)}.$$

Therefore it suffices to get asymptotic estimates for the numerator and denominator, and show that this is large when x is large and $F = F_k$ is chosen suitably.

4.1. First sieve calculations. We first try to estimate $\sum_{n \in [x, 2x]} \nu(n)$. Expanding the square in the definition of $\nu(n)$ and swapping the order of summation, we find that this is given by

$$(4.1) \quad \sum_{\substack{d_1, \dots, d_k \\ \prod_{i=1}^k d_i < R}} \sum_{\substack{e_1, \dots, e_k \\ \prod_{i=1}^k e_i < R}} \mu(\mathbf{d}) \mu(\mathbf{e}) F\left(\frac{\log \mathbf{d}}{\log R}\right) F\left(\frac{\log \mathbf{e}}{\log R}\right) \sum_{\substack{n \in [x, 2x] \\ d_i | L_i(n) \forall i \\ e_i | L_i(n) \forall i}} 1,$$

where we have written $\mu(\mathbf{d}) = \mu(d_1 \cdots d_k)$ and

$$F\left(\frac{\log \mathbf{d}}{\log R}\right) = F\left(\frac{\log d_1}{\log R}, \dots, \frac{\log d_k}{\log R}\right)$$

for notational simplicity.

We see that the conditions $d_i | L_i(n)$ and $e_i | L_i(n)$ mean that $\text{lcm}(d_i, e_i) | L_i(n)$. Imagine for simplicity that $L_i(n) = a_i n + b_i$ are such that $\prod_{i < j} (b_i - b_j)$ divides a_ℓ for all $1 \leq \ell \leq k$ (this avoids a few minor issues with small primes). Let $A = \text{lcm}(a_1, \dots, a_k)$. Then, by the Chinese Remainder Theorem, we can rewrite the conditions

$$\text{lcm}(d_i, e_i) | L_i(n) \quad \text{for } 1 \leq i \leq k$$

as

$$n \equiv n_0 \pmod{A \prod_{i=1}^k \text{lcm}(d_i, e_i)}$$

for some n_0 , provided $A, \text{lcm}(d_1, e_1), \dots, \text{lcm}(d_{k-1}, e_{k-1})$ and $\text{lcm}(d_k, e_k)$ are all pairwise coprime. If they are not pairwise coprime, then there are no solutions to the conditions.

Therefore, we find that the inner sum of (4.1) simply counts integers in an arithmetic progression whenever it is non-empty. The number of

$n \in [x, 2x]$ with $n \equiv a \pmod{q}$ is $x/q + O(1)$. Thus (4.1) simplifies to

$$\frac{x}{A} \sum_{\substack{d_1, \dots, d_k \\ \prod_{i=1}^k d_i < R}}^* \sum_{\substack{e_1, \dots, e_k \\ \prod_{i=1}^k e_i < R}}^* \frac{\mu(\mathbf{d})\mu(\mathbf{e})}{\prod_{i=1}^k \text{lcm}(d_i, e_i)} F\left(\frac{\log \mathbf{d}}{\log R}\right) F\left(\frac{\log \mathbf{e}}{\log R}\right) + O(R^2).$$

Here the asterisk indicates we have the restriction that $a_1 \dots a_k, d_1 e_1, \dots, d_k e_k$ are pairwise coprime.

This is now a smoothed sum of multiplicative functions, which are standard sums to evaluate in analytic number theory, typically via complex analysis.

Lemma 10 (Estimate of smoothed sum of multiplicative functions).
Let

$$S = \sum_{\substack{d_1, \dots, d_k \\ \prod_{i=1}^k d_i < R}}^* \sum_{\substack{e_1, \dots, e_k \\ \prod_{i=1}^k e_i < R}}^* \frac{\mu(\mathbf{d})\mu(\mathbf{e})}{\prod_{i=1}^k \text{lcm}(d_i, e_i)} F\left(\frac{\log \mathbf{d}}{\log R}\right) F\left(\frac{\log \mathbf{e}}{\log R}\right).$$

Then

$$S = (1 + o(1)) \frac{A^k}{\varphi(A)^k (\log R)^k} \int_{\substack{t_1, \dots, t_k \geq 0 \\ t_1 + \dots + t_k \leq 1}} \frac{\partial^k F(t_1, \dots, t_k)^2}{\partial t_1 \dots \partial t_k} dt_1 \dots dt_k.$$

Therefore, letting $\tilde{I}(F)$ be the integral above, we see that

$$(4.2) \quad \sum_{n \in [x, 2x]} \nu(n) = (1 + o(1)) \frac{x A^{k-1}}{\varphi(A)^k (\log R)^k} \tilde{I}(F).$$

This gives our asymptotic evaluation of $\sum_n \nu(n)$.

4.2. Sieve calculations with primes. We now want to calculate $\sum_{n \in [x, 2x]} \nu(n) \mathbf{1}_{\text{prime}}(L_j(n))$ for each j . For notational simplicity we will consider the case $j = k$, the other cases being essentially identical. As before, expanding out we find this is given by

$$(4.3) \quad \sum_{\substack{d_1, \dots, d_k \\ \prod_{i=1}^k d_i < R}} \sum_{\substack{e_1, \dots, e_k \\ \prod_{i=1}^k e_i < R}} \mu(\mathbf{d})\mu(\mathbf{e}) F\left(\frac{\log \mathbf{d}}{\log R}\right) F\left(\frac{\log \mathbf{e}}{\log R}\right) \sum_{\substack{n \in [x, 2x] \\ d_i | L_i(n) \forall i \\ e_i | L_i(n) \forall i}} \mathbf{1}_{\text{prime}}(L_k(n)).$$

Since we have restricted $L_k(n)$ to be prime, we must have $d_k = e_k = 1$ since $R < x$. As before, we can combine the remaining conditions $d_i, e_i | L_i(n)$ into a single congruence via the Chinese Remainder Theorem

$$n \equiv n'_0 \pmod{A \prod_{i=1}^{k-1} \text{lcm}(d_i, e_i)},$$

for some n'_0 provided $A, d_1 e_1, \dots, d_{k-1} e_{k-1}$ are pairwise coprime (and if they are not the inner sum is empty). Moreover, we can check that n'_0 is coprime to the modulus. Therefore we wish to estimate primes

in arithmetic progressions. Dirichlet's famous theorem (and its refinement by Siegel and Walfisz) allows us to count primes in arithmetic progressions, but only when the modulus is very small. Unfortunately here the modulus can be as large as R^2 , which we want to be a power of x , which is much too large to handle without making significant progress on the issue of Siegel zeros. Fortunately, we only need an average form of this result, and the celebrated Bombieri-Vinogradov Theorem [7, 73] provides an adequate substitute for the Generalized Riemann Hypothesis here.

Theorem 11 (Bombieri-Vinogradov Theorem). *Let $\pi(x, q, a)$ denote the number of primes $p \leq x$ with $p \equiv a \pmod{q}$. Then for any $A, \epsilon > 0$ there is a constant $C(A, \epsilon) > 0$ such that*

$$\sum_{q \leq x^{1/2-\epsilon}} \sup_{(a,q)=1} \left| \pi(x, q, a) - \frac{\pi(x)}{\varphi(q)} \right| \leq \frac{C(A, \epsilon)x}{(\log x)^A}.$$

This gives a good result provided all our moduli are bounded by $x^{1/2-\epsilon}$, which means we require $R^2 \leq x^{1/2-\epsilon}$. Using this, we find that (4.3) is essentially given by

$$\frac{\pi(x)}{\varphi(A)} \sum_{\substack{d_1, \dots, d_{k-1} \\ \prod_{i=1}^k d_i < R \\ d_k=1}}^* \sum_{\substack{e_1, \dots, e_{k-1} \\ \prod_{i=1}^k e_i < R \\ e_k=1}}^* \frac{\mu(\mathbf{d})\mu(\mathbf{e})}{\prod_{i=1}^{k-1} \varphi(\text{lcm}(d_i, e_i))} F\left(\frac{\log \mathbf{d}}{\log R}\right) F\left(\frac{\log \mathbf{e}}{\log R}\right),$$

where the asterisks indicate the summation is subject to $A, d_1e_1, d_2e_2, \dots, d_{k-1}e_{k-1}$ being coprime. For the purposes of the summation, there is essentially nothing lost if we replace $\varphi(\text{lcm}(d_i, e_i))$ with $\text{lcm}(d_i, e_i)$. Thus, a variant of Lemma 10 gives

$$\begin{aligned} & \sum_{\substack{d_1, \dots, d_{k-1} \\ \prod_{i=1}^k d_i < R \\ d_k=1}}^* \sum_{\substack{e_1, \dots, e_{k-1} \\ \prod_{i=1}^k e_i < R \\ e_k=1}}^* \frac{\mu(\mathbf{d})\mu(\mathbf{e})}{\prod_{i=1}^{k-1} \varphi(\text{lcm}(d_i, e_i))} F\left(\frac{\log \mathbf{d}}{\log R}\right) F\left(\frac{\log \mathbf{e}}{\log R}\right) \\ &= \frac{(1 + o(1))A^{k-1}}{\varphi(A)^{k-1}(\log R)^{k-1}} \int_{\substack{t_1, \dots, t_{k-1} \geq 0 \\ t_1 + \dots + t_{k-1} \leq 1}} \frac{\partial^{k-1} F(t_1, \dots, t_{k-1}, 0)^2}{\partial t_1 \dots \partial t_{k-1}} dt_1 \dots dt_{k-1}. \end{aligned}$$

Letting $\tilde{J}_k(F)$ denote the integral above, and using the Prime Number Theorem $\pi(x) = (1 + o(1))x/\log x$, we see this gives

$$(4.4) \quad \sum_{n \in [x, 2x]} \nu(n) \mathbf{1}_{\text{prime}}(L_k(n)) = \frac{(1 + o(1))A^{k-1}x}{\varphi(A)^k (\log R)^{k-1} \log x} \tilde{J}_k(F).$$

Thus, combining (4.2) and (4.4), we find that

$$\begin{aligned}\mathbb{P}(L_k(n) \text{ prime}) &= \frac{\sum_{n \in [x, 2x]} \nu(n) \mathbf{1}_{\text{prime}}(L_k(n))}{\sum_{n \in [x, 2x]} \nu(n)} \\ &= (1 + o(1)) \frac{\log R}{\log x} \frac{\tilde{J}_k(F)}{\tilde{I}(F)}.\end{aligned}$$

4.3. Choice of F . It will be convenient to work with

$$\tilde{F} = \frac{\partial^k}{\partial t_1 \dots \partial t_k} F.$$

With this transform, the previous calculations show that for $R \leq x^{1/4-\epsilon}$

$$(4.5) \quad \mathbb{E} \#\{1 \leq i \leq k : L_i(n) \text{ prime}\} = \left(\frac{\sum_{i=1}^k J_i(\tilde{F})}{I(\tilde{F})} + o(1) \right) \frac{\log R}{\log X},$$

where the $J_i(\tilde{F})$ and $I(\tilde{F})$ are given by

$$\begin{aligned}J_\ell(\tilde{F}) &= \int \dots \int \left(\int_0^{1-\sum_{i \neq \ell} t_i} \tilde{F}(t_1, \dots, t_k) dt_\ell \right)^2 dt_1 \dots dt_{\ell-1} dt_{\ell+1} \dots dt_k, \\ I(\tilde{F}) &= \int \dots \int \tilde{F}(t_1, \dots, t_k)^2 dt_1 \dots dt_k.\end{aligned}$$

In particular, we can show that many of the L_i are simultaneously prime infinitely often, if we can show that $\sup_{\tilde{F}} (\sum_{i=1}^k J_i(\tilde{F})/I(\tilde{F})) \rightarrow \infty$ as $k \rightarrow \infty$, where the supremum is over all piecewise smooth \tilde{F} .

A key advantage of this generalization is that we can make use of high dimensional phenomena such as concentration of measure. We concentrate on functions \tilde{F} the form

$$\tilde{F}(t_1, \dots, t_k) = \begin{cases} \prod_{i=1}^k G(t_i), & \sum_{i=1}^k t_i < 1, \\ 0, & \text{otherwise,} \end{cases}$$

for some smooth function $G(t)$ (depending on k). This allows us to make a probabilistic interpretation of the integrals $J_i(\tilde{F})$ and $I(\tilde{F})$. Let Z_1, \dots, Z_k be i.i.d. random variables (depending on k) on $[0, \infty]$ with probability density function G^2 , expectation $\mu = \int_0^\infty tG(t)^2 dt$ and variance $\sigma^2 = \int_0^\infty (t - \mu)^2 G(t)^2 dt$. Then

$$\begin{aligned}I(\tilde{F}) &= \mathbb{P}\left(\sum_{i=1}^k Z_i < 1\right), \\ J_\ell(\tilde{F}) &\geq \left(\int_0^{1/2} G(t) dt\right)^2 \mathbb{P}\left(\sum_{i=1}^k Z_i < \frac{1}{2}\right).\end{aligned}$$

The random variable $\sum_{i=1}^k Z_i$ has mean $k\mu$ and variance $k\sigma^2$, and so it becomes concentrated on $k\mu$ when k is large provided $\sigma/k^{1/2}\mu \rightarrow 0$

as $k \rightarrow \infty$. In particular, if $\mu < 1/3k$ and $\sigma^2 k \rightarrow 0$, then $\mathbb{P}(\sum_{i=1}^k Z_i < 1/2)$ approaches 1 as $k \rightarrow \infty$. Therefore, to show that $\sum_{\ell=1}^k J_\ell(\tilde{F})/I(\tilde{F}) \rightarrow \infty$ as $k \rightarrow \infty$, it suffices to find a function G satisfying:

$$\begin{aligned} \int_0^\infty tG(t)^2 &< \frac{1}{3k}, \\ \int_0^\infty G(t)^2 dt &= 1, \\ k \int_0^\infty t^2 G(t)^2 dt &\rightarrow 0, \\ k \left(\int_0^\infty G(t) dt \right)^2 &\rightarrow \infty \quad \text{as } k \rightarrow \infty. \end{aligned}$$

We find that choosing

$$(4.6) \quad G(t) \approx \begin{cases} \frac{\sqrt{k \log k}}{1+tk \log k}, & t < k^{-3/4}, \\ 0, & \text{otherwise,} \end{cases}$$

gives a function G satisfying these constraints. (This choice can be found via the calculus of variations, and further calculations show that this choice is essentially optimal.) Putting this all together, we find that for some constant $c > 0$

$$\mathbb{E} \#\{i : L_i(n) \text{ prime}\} \geq (c \log k + o(1)) \frac{\log R}{\log X}.$$

In particular, taking $R = X^{1/4-o(1)}$ (as allowed by the Bombieri-Vinogradov theorem) and letting k be sufficiently large, we find that there are infinitely many integers n such that $(c/4 + o(1)) \log k$ of the $L_i(n)$ are simultaneously prime. Performing these calculations carefully allows one to take $c \approx 1$ when k is large.

Morally, the effect of such a choice of function \tilde{F} is to make ‘typical’ divisors (d_1, \dots, d_k) occurring in (3.4) to have $\prod_{i=1}^k d_i$ smaller, but for it to be more common that some of the components of (d_1, \dots, d_k) are unusually large when compared with (3.1) or (3.3). This correspondingly causes the random integer n chosen with probability $w(n)$ to be such that the $L_i(n)$ are more likely to have slightly smaller prime factors, but it is also more likely that some of the $L_i(n)$ are prime.

4.4. Consequences for small gaps between primes. The above argument shows that there is a constant $c > 0$ such that for any set $\{L_1, \dots, L_k\}$ of integral linear functions satisfying the hypotheses of Conjecture 3, at least $c \log k$ of the linear functions L_i are simultaneously prime infinitely often. To show that there are primes close together, we simply take the linear functions to be of the form $L_i(n) = n + h_i$ for some integers $h_1 < \dots < h_k$ chosen to make $\prod_{i=1}^k L_i(n)$ have no fixed prime divisor, and so that $h_k - h_1$ is small.

In general, a good choice of the h_i is to take h_i to be the i^{th} prime after the integer k . With this choice, $\prod_{i=1}^k L_i(n)$ has no fixed prime divisor and $h_k - h_1 \approx k \log k$, so we can find intervals of length $k \log k$ containing $c \log k$ primes infinitely often. By working out the best possible implied constants, the argument we have sketched allows us to show that there are m primes in an interval of length $O(m^3 e^{4m})$ infinitely often. The fact that we can find many primes rather than just 2, and we can do so taking $R = x^\epsilon$ has important other applications, such as for large gaps between primes [52, 21, 20].

4.5. Numerical optimization. If we are only interested in just how small a *single* gap can be, then we can improve the analysis and get an explicit bound on the size of the gap by adopting a numerical analysis perspective. The key issue is to find the smallest value of k such that for all large X

$$\mathbb{E} \#\{1 \leq i \leq k : L_i(n) \text{ prime}\} > 1,$$

since this immediately implies that two of the linear functions are simultaneously prime. Recalling that we can choose $R = x^{1/4-\epsilon}$ for any $\epsilon > 0$, using (4.5) we reduce the problem to finding a value of k as small as possible, such that we can find a function $\tilde{F} : [0, \infty)^k \rightarrow \mathbb{R}$ with

$$\frac{\sum_{i=1}^k J_i(\tilde{F})}{I(\tilde{F})} > 4.$$

We fix some basis functions $g_1, \dots, g_r : [0, \infty)^k \rightarrow \mathbb{R}$ which are supported on (t_1, \dots, t_k) satisfying $\sum_{i=1}^k t_i \leq 1$, and restrict our attention to functions \tilde{F} in the linear span of the g_i ; that is functions \tilde{F} of the form

$$\tilde{F}(t_1, \dots, t_k) = \sum_{i=1}^r f_i g_i(t_1, \dots, t_k),$$

for some coefficients $\mathbf{f} = (f_1, \dots, f_r) \in \mathbb{R}^r$ which we think of as variables we will optimize over. For such a choice, we find that $\sum_{i=1}^k J_i(\tilde{F})$ and $I(\tilde{F})$ are both quadratic forms in the variables f_1, \dots, f_r , and the coefficients of these quadratic forms are explicit integrals in terms of g_1, \dots, g_r . If we choose a suitably nice basis g_1, \dots, g_r , then these integrals are explicitly computable, and so we obtain explicit $r \times r$ real symmetric matrices M_1 and M_2 such that

$$\frac{\sum_{i=1}^k J_i(\tilde{F})}{I(\tilde{F})} = \frac{\mathbf{f}^T M_1 \mathbf{f}}{\mathbf{f}^T M_2 \mathbf{f}}.$$

We then find that the choice of coefficients \mathbf{f} which maximizes this ratio is the eigenvector of $M_2^{-1} M_1$ corresponding to the largest eigenvalue, and the value of the ratio is given by this largest eigenvalue. Thus the existence of a good function \tilde{F} can be reduced to checking whether

the largest eigenvalue of a finite matrix is larger than 4, which can be performed numerically by a computer.

If we choose the g_i to be symmetric polynomials of low degree, then this provides a nice basis since the corresponding integrals have a closed form solution, and allows one to make arbitrarily accurate numerical approximations to the optimal function \tilde{F} with enough computation. For the problem at hand, these numerical calculations are large but feasible. This approach ultimately allows one to show that if $k = 54$ there is a function \tilde{F} such that $\sum_{i=1}^{54} J_i(\tilde{F})/I(\tilde{F}) > 4$.

To turn this into small gaps between primes we need to choose the shifts h_i in our linear functions $L_i(n) = n + h_i$ so that $\prod_{i=1}^{54} L_i(n)$ has no fixed prime divisor, and the h_i are in as short an interval as possible. This is a feasible exhaustive numerical optimization problem, with an optimal choice of the $\{h_1, \dots, h_{54}\}$ given by ¹

$$\{0, 2, 6, 12, 20, 26, 30, 32, 42, 56, 60, 62, 72, 74, 84, 86, 90, 96, \\ 104, 110, 114, 116, 120, 126, 132, 134, 140, 144, 152, 156, \\ 162, 170, 174, 176, 182, 186, 194, 200, 204, 210, 216, 222, \\ 224, 230, 236, 240, 242, 246, 252, 254, 260, 264, 266, 270\}.$$

Putting this together, we find that there are infinitely many pairs of primes $p \neq q$ such that

$$|p - q| \leq 270.$$

This result is not quite the current record - in [59] we make some further technical refinements to the sieve, which corresponds to modifying the expressions $J_i(\tilde{F})$ and $I(\tilde{F})$ slightly. This ultimately allows us to improve $k = 54$ to $k = 50$ and correspondingly improve from gaps of size at most 270 to gaps of at most 246, giving Theorem 6. The main ideas are the same as above.

5. LOGARITHMIC CHOWLA CONJECTURE

We wish to sketch the proof of the following result due to Tao. For minor technical reasons, we will consider the following variant of Theorem 8, where we replace $n + 2$ with $n + 1$:

Theorem 12 (Variant of logarithmically averaged 2-point Chowla conjecture). *As $x \rightarrow \infty$, we have*

$$\sum_{n < x} \frac{\lambda(n)\lambda(n+1)}{n} = o(\log x)$$

This will rely heavily on the Matomäki-Radziwiłł Theorem (Theorem 9), which we talk about in the next section. In fact, for simplicity

¹Such computations were first performed by Engelsma - see <http://www.opertech.com/primes/k-tuples.html>

we will use a slight variant of their theorem, which is the following, established in [47].

Theorem 13 (Extended Matomäki-Radziwiłł). *Let $\theta \in \mathbb{R}$. Then for all but at most $X/(\log h)^{1/10}$ integers $n \leq X$ we have*

$$\left| \sum_{j \leq h} \lambda(n+j)e(j\theta) \right| \ll \frac{h}{(\log h)^{1/10}}.$$

5.1. Reduce to a ternary problem. First we use the multiplicativity of λ . Since $\lambda(n) = -\lambda(pn)$ for any prime p , we have

$$\begin{aligned} \sum_{n < x} \frac{\lambda(n)\lambda(n+1)}{n} &= \sum_{n < x} \frac{\lambda(pn)\lambda(pn+p)}{n} \\ &= \sum_{m < px} \frac{\lambda(m)\lambda(m+p)}{m} \mathbf{1}_{p|m} \\ &= \sum_{m < x} \frac{\lambda(m)\lambda(m+p)}{m} \mathbf{1}_{p|m} + O(\log p). \end{aligned}$$

Provided p is small enough compared with x , we see that the error term is negligible (here we make important use of the logarithmic weighting). We now average this expression over $p \in [P, 2P]$. This gives

$$\sum_{n < x} \frac{\lambda(n)\lambda(n+1)}{n} = \frac{\log P}{P} \sum_{m < x} \frac{\lambda(m)}{m} \sum_{p \in [P, 2P]} \lambda(m+p) \mathbf{1}_{p|m} + o(\log x).$$

The key point of these manipulations is that we have gone from a binary problem where we want to understand the simultaneous factorizations of n and $n+1$ but only have one variable at play, to a ternary problem where we want to understand the factorization of m and $m+p$ and we have two variables at play. Such estimates are much more amenable to analytic estimates (this is roughly why we can solve the ternary Goldbach problem, but not the binary one). If we ignored the indicator function $\mathbf{1}_{p|m}$ and relaxed $p \in [P, 2P]$ to $p \leq x$ then we can estimate

$$\sum_{m, p < x} \lambda(m)\lambda(m+p)$$

using the circle method. Similarly, if we ignored the indicator function $\mathbf{1}_{p|m}$ and the fact p is prime, we can evaluate

$$\sum_m \left| \sum_{n \in [P, 2P]} \lambda(m+n) \right|$$

using the Matomäki-Radziwiłł Theorem. Thus our main difficulty is to replace the indicator function with something simpler, and then to control the fact we have primes in short intervals.

Since $1/m \approx 1/(m+j)$ if $|j|$ is smaller than m , we see that

$$\begin{aligned} & \sum_{m < x} \frac{\lambda(m)}{m} \sum_{p \in [P, 2P]} \lambda(m+p) \mathbf{1}_{p|m} \\ & \approx \frac{1}{J} \sum_{m < x} \frac{1}{m} \sum_{j \leq J} \sum_{p \in [P, 2P]} \lambda(m+j) \lambda(m+j+p) \mathbf{1}_{p|m+j}. \end{aligned}$$

Thus the proof boils down to showing that for ‘most’ $m \leq x$ we have

$$\sum_{j \leq J} \sum_{p \in [P, 2P]} \lambda(m+j) \lambda(m+j+p) \mathbf{1}_{p|m+j} = o\left(\frac{J}{\log P}\right),$$

for a suitable choice of J and P of similar size, both very small compared with x .

5.2. Correlation between sign patterns and residue classes. We see that the values of $\lambda(m+j)\lambda(m+j+p)$ appearing above are determined by the vector $\mathbf{x}_m = (\lambda(m+1), \dots, \lambda(m+2P+J)) \in \{-1, 1\}^{2P+J}$. Similarly, the values of $\mathbf{1}_{p|m+j}$ are determined by $m \pmod{Q}$, where $Q = \prod_{p \in [P, 2P]} p$. Thus we can rewrite our expression as

$$\sum_{\mathbf{a} \in \{-1, 1\}^{2P+J}} \sum_{h \pmod{Q}} \sum_{j \leq J} \sum_{p \in [P, 2P]} a_j a_{j+p} \mathbf{1}_{p|h+j} \sum_{\substack{m \leq x \\ m \equiv h \pmod{Q} \\ \mathbf{x}_m = \mathbf{a}}} \frac{1}{m}.$$

In particular, if for any sign pattern $\mathbf{a} \in \{-1, 1\}^{2P+J}$ we knew that the values of m such that $\mathbf{x}_m = \mathbf{a}$ were equidistributed \pmod{Q} , then we could replace $\mathbf{1}_{p|m+j}$ with $1/p$.

Unfortunately we do not know that the values are equidistributed. However, if we temporarily fix a sign pattern \mathbf{a} , we see that we are left to consider

$$\left(\sum_{\substack{m \leq x \\ m \equiv h \pmod{Q} \\ \mathbf{x}_m = \mathbf{a}}} \frac{1}{m} \right) \cdot \sum_{p \in [P, 2P]} \left(\sum_{j \leq J} a_j a_{j+p} \mathbf{1}_{p|h+j} \right).$$

The second expression in parentheses depends only on $h \pmod{p}$, which is independently uniformly distributed as h varies \pmod{Q} . Therefore the sum over p is a sum of many independent expressions, so we can use concentration of measure to show that for each \mathbf{a} there can only be a few residue classes which cause problems.

Lemma 14 (Application of Chernoff Bound). *Let $\mathbf{a} \in \{-1, 1\}^{2P+J}$, and let $\mathcal{E}_{\mathbf{a}}$ be the set of residue classes $y \pmod{Q}$ such that*

$$\left| \sum_{p \in [P, 2P]} \left(\sum_{j \leq J} a_j a_{j+p} \mathbf{1}_{p|y+j} \right) - \sum_{p \in [P, 2P]} \frac{1}{p} \left(\sum_{j \leq J} a_j a_{j+p} \right) \right| \geq \frac{\epsilon J}{\log P}.$$

Then $\mathcal{E}_{\mathbf{a}}$ has size

$$\#\mathcal{E}_{\mathbf{a}} \ll \exp\left(-\frac{10P^3}{J^2 \log P}\right)Q.$$

This means that the situation we have to rule out is that the set of $m \leq x$ with $\mathbf{x}_m = \mathbf{a}$ are highly concentrated in a very small number of residue classes $(\bmod Q)$, and that for each residue class $b \pmod{Q}$ the set of $m \leq x$ with $m \equiv b \pmod{Q}$ are concentrated on a very small number of sign patterns.

5.3. The entropy decrement argument. Probably the most novel part of Tao's argument is handling the possibility that the residue classes are concentrated on a small set. This is the 'entropy decrement argument'.

We wish to handle this possibility by showing there exists a scale (i.e. a choice of J and P) such that the set of m with $\mathbf{x}_m = \mathbf{a}$ is not concentrated in a small number of residue classes $(\bmod Q)$, for almost all relevant choices of \mathbf{a} . By Lemma 14, we see that this means we can effectively replace $\mathbf{1}_{p|h+j}$ with $1/p$, since the bad residue classes from $\mathcal{E}_{\mathbf{a}}$ have a negligible effect.

This will be done by showing that if at some scale (J_1, P_1) this is not the case, then there is a larger scale (J_2, P_2) where \mathbf{x}_m is concentrated in a relatively smaller set. We can repeat this observation at most a bounded number of times, and so there must be some scale (J', P') when the residue classes were essentially equidistributed.

Let us first choose (J_1, P_1) as our initial scales, and let

$$\mathbf{x}_m^{(1)} = (\lambda(m+1), \dots, \lambda(m+J_1))$$

and $Q_1 = \prod_{P_1 \leq p \leq 2P_1} p$. Imagine for a simplification that for each residue class $b \pmod{Q_1}$, all $m \equiv b \pmod{Q_1}$ have $\mathbf{x}_m^{(1)} \in \mathcal{E}_b$ for some set \mathcal{E}_b of size $\leq 2^{\alpha_1 J_1}$. Let

$$\mathbf{x}_m^{(2)} = (\lambda(m+1), \dots, \lambda(m+rJ_1))$$

be the concatenation of $\mathbf{x}_m, \mathbf{x}_{m+J_1}, \dots, \mathbf{x}_{m+(r-1)J_1}$ to form a vector at scale $J_2 = rJ_1$. There are $\leq Q_1$ choices of a residue class for m and given such a choice there are $\leq 2^{\alpha_1 J_1}$ choices for each \mathbf{x}_j , and so $\leq Q_1 2^{\alpha_1 J_2}$ choices of $\mathbf{x}_m^{(2)}$. For r large, this is $2^{(\alpha_1 + o(1))J_2}$.

We now let $P_2 = rP_1$, and consider the relevance of residue classes modulo $Q_2 = \prod_{P \leq p \leq 2P} P$. First imagine that for most residue classes $b \pmod{Q_2}$ the set of possible $\mathbf{x}_m^{(2)}$ with $m \equiv b \pmod{Q_2}$ is close to the maximal size $2^{\alpha_1 J_2}$ - say at least $2^{(\alpha_1 - \epsilon)J_2}$. Then by double counting we see that typically at least $2^{-\epsilon J_2} Q_2$ different residue classes appear in the set of m such that $\mathbf{x}_m = \mathbf{a}$ for all of the relevant choices of \mathbf{a} . In particular, for ϵ small enough we see this set of residue classes isn't highly concentrated. This means that at this scale we avoid the obstruction, and so we're good.

Alternatively, it could be the case that for each residue class there are many fewer possible sign patterns - say $\leq 2^{\alpha_2 J_2}$ for some $\alpha_2 < \alpha_1$. Then we can consider a larger scale $J_3 = r_2 J_2$, and the concatenation $\mathbf{x}_m^{(3)}$ of $\mathbf{x}_m^{(2)}, \mathbf{x}_{m+J_2}^{(2)}, \dots, \mathbf{x}_{m+(r_2-1)J_2}^{(2)}$. There are $\leq Q_2$ choices of residue class for $m \pmod{Q_2}$, and given such a choice $\leq 2^{\alpha_2 J_2}$ choices for each $\mathbf{x}_j^{(2)}$. Thus there are $\leq Q_2 2^{\alpha_2 J_3} = 2^{(\alpha_2 + o(1))J_3}$ choices.

Continuing in this way (be aware that we are oversimplifying details), we either find a good scale, or a decreasing sequence $\alpha_1 > \alpha_2 > \dots$ such that there are $\leq 2^{(\alpha_\ell + o(1))J_\ell}$ choices of $\mathbf{x}^{(\ell)}$. By carefully controlling the quantitative aspects, we find that the α_j must decrease enough at each stage such that this process terminates, meaning that we are in one of the good situations above for some choice of scale (J', P') , although we only have very weak control over the size of J' and P' .

The outcome of this is that we have found J' and P' such that for most $m \leq x$

$$\sum_{j \leq J'} \sum_{p \in [P', 2P']} \lambda(m+j)\lambda(m+j+p) \mathbf{1}_{p|m+j} \approx \sum_{j \leq J'} \sum_{p \in [P', 2P']} \frac{\lambda(m+j)\lambda(m+j+p)}{p}.$$

5.4. The circle method. It suffices to show for most $m \leq x$ that

$$\left| \sum_{j \leq J'} \sum_{p \in [P', 2P']} \lambda(m+j)\lambda(m+j+p) \right| = o\left(\frac{J'P'}{\log P'}\right).$$

As mentioned before, this can be thought of a version of the Ternary Goldbach problem in short intervals. Let $H = J' + 2P'$. We see that by Fourier expansion

$$\begin{aligned} & \sum_{j \leq J'} \sum_{p \in [P', 2P']} \lambda(m+j)\lambda(m+j+p) \\ &= \frac{1}{H} \sum_{b \pmod{H}} \left| \sum_{j \leq H} \lambda(m+j) e\left(\frac{bj}{H}\right) \right|^2 \sum_{p \in [P', 2P']} e\left(\frac{-bp}{H}\right), \end{aligned}$$

where $e(x) = e^{2\pi i x}$. The set of b for which

$$\left| \sum_{p \in [P', 2P']} e\left(\frac{-bp}{H}\right) \right| = o\left(\frac{P'}{\log P'}\right)$$

contribute a total of $o(HP'/\log P')$. However, the set \mathcal{B} of b for which the right hand side is $\gg P'/\log P'$ has size $O(1)$, by considering a 4th power mean value. Thus these bad frequencies contribute

$$\ll \frac{P'}{H \log P'} \sup_{b \in \mathcal{B}} \left| \sum_{j \leq H} \lambda(m+j) e\left(\frac{bj}{H}\right) \right|^2.$$

But this is $o(HP'/\log P')$ by the Theorem 13.

6. LIOUVILLE IN SHORT INTERVALS

Rather than sketch the proof of Theorem 13 itself, we will give a sketch of just the hardest case $\theta = 0$. This constitutes the main new ideas in the work of Matomäki and Radziwiłł, and is essentially just a qualitative version of Theorem 9.

Theorem 15 (Liouville function in almost all short intervals). *Let $\epsilon > 0$. Then there is a constant $c(\epsilon)$ such that the following holds.*

If $\delta X \geq c(\epsilon)$ and $0 < \delta < 1$, then for all but at most ϵX values of $y \in [X, 2X]$ we have

$$\left| \frac{1}{\delta y} \sum_{y \leq n \leq y(1+\delta)} \lambda(n) \right| \leq \epsilon.$$

Informally, this shows that ‘almost all’ intervals $[y, y(1 + \delta)]$ have $\sum_{n \in [y, y(1+\delta)]} \lambda(n) = o(\delta X)$, even when δX is going to infinity arbitrarily slowly with X .

6.1. Reduction to Dirichlet Polynomials. To show such a result, it suffices to show

$$\int_X^{2X} \left| \frac{1}{\delta y} \sum_{y \leq n \leq y(1+\delta)} \lambda(n) \right|^2 = o(X)$$

as $\delta X \rightarrow \infty$. Indeed, let \mathcal{S}_ϵ be the set of $y \in [X, 2X]$ for which

$$\left| \frac{1}{\delta y} \sum_{n \in [y, y(1+\delta)]} \lambda(n) \right| \geq \epsilon.$$

Then we see that \mathcal{S}_ϵ contributes at least $\epsilon^2 m(\mathcal{S}_\epsilon)$ to the integral, where $m(\mathcal{S}_\epsilon)$ is the measure of \mathcal{S}_ϵ . Therefore $m(\mathcal{S}_\epsilon) = o(X)$ for any fixed $\epsilon > 0$, as required.

The standard way to attempt to estimate such integrals is to relate them to large values of Dirichlet polynomials. This can be achieved via the following lemma.

Lemma 16 (Counting using Mellin transforms). *Let $|a_n| \leq 1$ be a real sequence. Then for any $C > 0$ we have*

$$\int_X^{2X} \left| \frac{1}{\delta y} \sum_{y \leq n \leq y(1+\delta)} a_n \right|^2 dy \ll \frac{X}{C} + \frac{1}{X} \int_0^{C/\delta} \left| \sum_{X \leq n \leq 3X} \frac{a_n}{n^{it}} \right|^2 dt.$$

This reduces the task to showing that $\sum_n a_n/n^{it}$ cannot be large very often. In particular, by choosing C as a large constant, we would be done if we could show

$$(6.1) \quad \int_0^{C/\delta} \left| \sum_{n \leq X} \frac{\lambda(n)}{n^{it}} \right|^2 dt = o(X^2).$$

We expect (6.1) to hold for $\delta X \rightarrow \infty$, but we only know how to show this when $\delta \geq X^{-5/12+\epsilon}$. This is very frustrating, since one easily gets a bound which is almost good enough from the well-known mean value theorem.

Lemma 17 (Mean value Theorem for Dirichlet polynomials). *Let $|a_n| \leq 1$ be a complex sequence. Then*

$$\int_T^{2T} \left| \sum_{n \leq N} \frac{a_n}{n^{it}} \right|^2 dt \ll TN + N^2.$$

This result is essentially sharp; if $a_n = n^{it_0}$ for some fixed $t_0 \in [T, 2T]$ then the contribution near t_0 would contribute about N^2 , whereas a random choice of a_n would have a total contribution of about TN . The mean value estimate can be improved slightly if one is integrating over a smaller set.

Lemma 18 (Montgomery-Halasz Mean Value Estimate). *Let $|a_n| \leq 1$ be a complex sequence, and $\mathcal{U} \subseteq [T, 2T]$ be a set of 1-separated points. Then*

$$\sum_{t \in \mathcal{U}} \left| \sum_{n \leq N} \frac{a_n}{n^{it}} \right|^2 \ll N^2 + \#\mathcal{U}T^{1/2}N.$$

From this, we see that one key roadblock is that it is necessary to show that $\lambda(n)$ cannot behave like n^{it_0} for any $t_0 \ll 1/\delta$, since this would genuinely cause problems. Fortunately, since $\lambda(n)$ is closely connected to the zeta function, we can show that $\sum_n \lambda(n)n^{-it}$ is always a bit smaller than the trivial bound, which at least removes this obstruction.

Lemma 19 (Consequences of a zero-free region for $\zeta(s)$). *Let $|t| \leq \exp((\log x)^{5/4})$. Then there is a constant $c > 0$ such that*

$$\left| \sum_{p < x} \frac{1}{p^{it}} \right| \ll \frac{x}{t \log x} + x \exp(-c(\log x)^{1/7}),$$

$$\left| \sum_{n < x} \frac{\lambda(n)}{n^{it}} \right| \ll x \exp(-c(\log x)^{1/7}).$$

If one happens to have coefficients which factor nicely, then the above L^∞ bounds can be combined with Lemma 17 to give a suitable result. For example, we see that for $(\log P)^{100} \leq T \leq \exp((\log P)^{5/4})$ we have

$$\begin{aligned} \int_T^{2T} \left| \sum_{P \leq p_1 \leq 2P} \frac{1}{p_1^{it}} \right|^2 \left| \sum_{X/P \leq p_2 \leq 2X/P} \frac{1}{p_2^{it}} \right|^2 dt &\ll \frac{P^2}{(\log P)^{100}} \int_T^{2T} \left| \sum_{X/P \leq p_2 \leq 2X/P} \frac{1}{p_2^{it}} \right|^2 dt \\ &\ll \frac{X^2}{(\log P)^{100}} \left(1 + \frac{TP}{X} \right), \end{aligned}$$

which gives a good bound if $T < X/P$. Unfortunately one doesn't necessarily have a nice factorization like this in general. Building on

the pioneering work of Vinogradov, the strongest previous work on primes and Liouville was on using combinatorial formulae to express such expressions with terms with nice factorizations, or with terms with very well-controlled coefficients.

6.2. Factorization of typical integers. The key idea which Matomäki and Radziwiłł exploit is that unlike for primes, the Dirichlet polynomial associated to the Liouville function naturally obtains many convenient factorizations because of the fact that most integers have ‘small’ prime factors.

Lemma 20 (Almost all integers have many small prime factors). *Let $|a_n| \leq 1$ be a complex sequence. Let $\mathcal{I}_h = [\exp((\log h)^{9/10}), h]$ with $h \leq X^{1/2}$. Then for all but $O(y/(\log \log h)^{1/4})$ values of $y \in [X, 2X]$ we have*

$$\sum_{y \leq n \leq y(1+\delta)} a_n = \frac{10}{\log \log h} \sum_{y \leq n \leq y(1+\delta)} a_n \sum_{\substack{p|n \\ p \in \mathcal{I}_h}} 1 + O\left(\frac{\delta X}{(\log \log h)^{1/4}}\right).$$

Thus we can work with $a_n = \lambda(n) \sum_{p|n, p \in \mathcal{I}_h} 1 = -\sum_{mp=n} \lambda(m) \mathbf{1}_{p \in \mathcal{I}_h}$ instead of $\lambda(n)$. This might look like a more complicated function, but the key advantage is that now our Dirichlet polynomial can essentially be factored into two different Dirichlet polynomials, and we have control over the lengths of the two factors. After splitting the range of primes into intervals $[P, 2P]$, with this trick we find it is sufficient to show that for any $P \in \mathcal{I}_h$ we have

$$(6.2) \quad \int_0^{C/\delta} \left| \sum_{P \leq p \leq 2P} \frac{1}{p^{it}} \right|^2 \left| \sum_{X/2P \leq m \leq 3X/P} \frac{\lambda(m)}{m^{it}} \right|^2 dt = o\left(\frac{X^2}{(\log P)^2}\right).$$

By Lemma 19, we have that $\sum_m \lambda(m)/m^{it} \ll X/(P \log^{10} X)$ if $|t| \leq \log^3 X$, which gives the result in this case.

We now consider $\log^3 X \leq |t| \leq X$, so $|t| \geq \log P$. Provided $h \geq \exp((\log X)^{9/10})$, we see that $P \in \mathcal{I}_h$ implies $|t| \leq X \leq \exp((\log h)^{10/9}) \leq \exp((\log P)^{5/4})$. Thus, by Lemma 19, we have

$$\left| \sum_{P \leq p \leq 2P} \frac{1}{p^{it}} \right|^2 \ll \frac{P}{(\log P)^2}.$$

Combining this with the mean value estimate for $\sum_m \lambda(m)/m^{it}$ then shows this total contribution is $O(X/(\log P)^4)$ provided that $\delta X \geq h$. Choosing $h = \exp((\log X)^{9/10})$ gives the result for $\delta X \geq \exp((\log X)^{9/10})$. This is already well beyond what was viewed as within reach before the Matomäki-Radziwiłł work, and demonstrates the power of the new idea of using such factorizations to improve on mean-value estimates. Their result goes much further, however.

6.3. Iterative factorization. We can use Lemma 20 several times for different choices of h rather than just once. This gives that for a suitable choice of $h_1 \leq h_2 \leq \dots \leq h_J$

$$\sum_{y \leq n \leq y(1+\delta)} a_n = \prod_{i=1}^J \left(\frac{10}{\log \log h_j} \right) \sum_{y \leq n \leq y(1+\delta)} a_n \sum_{\substack{p_1, \dots, p_J | n \\ p_j \in \mathcal{I}_{h_j}}} 1 + o(\delta X).$$

As before, we are left to show something like

$$\int_{\log X}^{C/\delta} \left| \sum_{\substack{X/P_1 \leq mp_2 \dots p_J \leq 2X/P_1 \\ p_2 \in \mathcal{I}_{h_2}, \dots, p_J \in \mathcal{I}_{h_J}}} \frac{\lambda(m)}{m^{it} p_2^{it} \dots p_J^{it}} \right|^2 \sum_{P_1 \leq p_1 \leq 2P_1} \frac{1}{p_1^{it}} dt = o\left(\frac{X^2}{(\log P_1)^2}\right).$$

The set \mathcal{T}_1 of $t \in [\log X, C/\delta]$ for which $|\sum_{p_1} 1/p_1^{it}| \leq P_1^2/(\log P_1)^2$ cause no problems provided $\delta X \geq h_1$. However, for the terms when $|\sum_{p_1} 1/p_1^{it}|$ is large, now we cannot assume that h_1 is large enough that Lemma 19 can apply. We therefore have to study the exceptional set $\mathcal{E}_1 = [\log X, C/\delta] \setminus \mathcal{T}_1$. After splitting according to the size of p_2 we wish to show

$$\int_{\mathcal{E}_1} \left| \sum_{\substack{X/P_1 P_2 \leq mp_3 \dots p_J \leq 2X/P_1 P_2 \\ p_3 \in \mathcal{I}_{h_3}, \dots, p_J \in \mathcal{I}_{h_J}}} \frac{\lambda(m)}{m^{it} p_3^{it} \dots p_J^{it}} \right|^2 \sum_{P_2 \leq p_2 \leq 2P_2} \frac{1}{p_2^{it}} dt = o\left(\frac{X^2}{P_1^2 (\log P_2)^2}\right).$$

We first consider the contribution from $t \in \mathcal{T}_2$, those $t \in \mathcal{E}_1$ such that $|\sum_{p_2} p_2^{-it}| \leq P_2^{0.9}$. We would still lose too much from applying the mean value theorem directly, but instead we re-insert the sum $|\sum_{p_1} p_1^{-it}| \geq P_1/(\log P_1)^2$, to find for any $\ell > 0$ the above integral is bounded by

$$P_2^{1.8} \left(\frac{P_1}{(\log P_1)^2} \right)^{-2\ell} \int_{\mathcal{E}_1} \left| \sum_{\substack{X/P_1 P_2 \leq mp_3 \dots p_J \leq 2X/P_1 P_2 \\ p_3 \in \mathcal{I}_{h_3}, \dots, p_J \in \mathcal{I}_{h_J}}} \frac{\lambda(m)}{m^{it} p_3^{it} \dots p_J^{it}} \right|^2 \sum_{P_1 \leq p_1 \leq 2P_1} \frac{1}{p_1^{it}} dt.$$

If we choose ℓ such that $P_1^\ell \in [P_1 P_2, P_1^2 P_2]$, then Lemma 17 gives a bound

$$\ll P_1^{-2\ell + o(\ell)} P_2^{1.8} \left(\frac{X}{P_1 P_2} \cdot (2\ell P_1)^\ell \right)^2 \ll \ell^\ell \frac{X}{P_1^{2+o(1)} P_2^{0.2}}.$$

This is acceptable if ℓ is not too large, which happens if say $\log \log P_2 \leq (\log P_1)/100$.

We are therefore left to consider the remaining set $\mathcal{E}_2 = \mathcal{E}_1 \setminus \mathcal{T}_2$, where both the p_1 -sum and the p_2 -sum are large. Continuing in this manner, we consider in turn the p_3 -sum, the p_4 -sum etc until the p_J -sum, continually splitting the integral according to whether the sum is small (in which case we can bound it) or large (in which case we consider the next sum). At the j^{th} stage, we can handle the contribution from

parts of the integral where

$$\left| \sum_{P_j \leq p_j \leq 2P_j} p_j^{-it} \right| \leq P_j^{1-\epsilon_j}$$

for a suitable sequence $\epsilon_2 \leq \dots \leq \epsilon_J \leq 1/5$ (the choice $\epsilon_j = 1/5 - 1/(5j)$ works), and provided that the h_j are also chosen to increase suitably ($h_j = \exp(h_{j-1}^{1/100j^2})$ works). Thus we are left to consider the part of the integral when all the sums are large. If J is large enough, we find that $h_J \geq \exp((\log X)^{9/10})$, and so by Lemma 19 we have

$$(6.3) \quad \left| \sum_{P_j \leq p_j \leq 2P_j} p_j^{-it} \right| \leq \frac{P_j}{(\log X)^{100}}.$$

On the other hand, we may assume $h_J \leq \exp((\log X)^{99/100})$. By considering a large power of $\sum_{p_j} p_j^{-it}$ Lemma 17 shows that the set of $t \in [0, 1/\delta]$ for which $|\sum_{p_j} p_j^{-it}| > P_j^{4/5}$ has cardinality $O(X^{1/2-\epsilon})$. Thus we can use (6.3) and then apply Lemma 18 to deduce

$$\begin{aligned} & \int_{\mathcal{E}_J} \left| \sum_{P_j \leq p_j \leq 2P_j} \frac{1}{p_j^{it}} \right|^2 \left| \sum_{X/P_1 \dots P_J \leq m \leq 2X/P_1 \dots P_J} \frac{\lambda(m)}{m^{it}} \right|^2 dt \\ & \ll \frac{P_J^2}{(\log X)^{200}} \left(\frac{X^2}{P_1^2 \dots P_J^2} + X^{1-\epsilon} \frac{X}{P_1 \dots P_J} \right) \\ & \ll \frac{X^2}{P_1^2 \dots P_{J-1}^2 (\log X)^{200}}. \end{aligned}$$

Finally, this has bounded the last bit of the integral, giving the result.

For more details of the Matomäki-Radziwiłł Theorem, we refer the reader to the survey of Soundararajan [64].

REFERENCES

- [1] R. C. Baker and T. Freiberg. Limit points and long gaps between primes. *Q. J. Math.*, 67(2):233–260, 2016.
- [2] R. C. Baker and P. Pollack. Bounded gaps between primes with a given primitive root, II. *Forum Math.*, 28(4):675–687, 2016.
- [3] R. C. Baker and L. Zhao. Gaps between primes in Beatty sequences. *Acta Arith.*, 172(3):207–242, 2016.
- [4] R. C. Baker and L. Zhao. Gaps of smallest possible order between primes in an arithmetic progression. *Int. Math. Res. Not. IMRN*, (23):7341–7368, 2016.
- [5] W. D. Banks, T. Freiberg, and J. Maynard. On limit points of the sequence of normalized prime gaps. *Proc. Lond. Math. Soc. (3)*, 113(4):515–539, 2016.
- [6] W. D. Banks, T. Freiberg, and C. Turnage-Butterbaugh. Consecutive primes in tuples. *Acta Arith.*, 167(3):261–266, 2015.
- [7] E. Bombieri. On the large sieve. *Mathematika*, 12:201–225, 1965.
- [8] E. Bombieri and H. Davenport. Small differences between prime numbers. *Proc. Roy. Soc. Ser. A*, 293:1–18, 1966.
- [9] E. Bombieri, J. B. Friedlander, and H. Iwaniec. Primes in arithmetic progressions to large moduli. *Acta Math.*, 156(3-4):203–251, 1986.

- [10] E. Bombieri, J. B. Friedlander, and H. Iwaniec. Primes in arithmetic progressions to large moduli. II. *Math. Ann.*, 277(3):361–393, 1987.
- [11] E. Bombieri, J. B. Friedlander, and H. Iwaniec. Primes in arithmetic progressions to large moduli. III. *J. Amer. Math. Soc.*, 2(2):215–224, 1989.
- [12] A. Castillo, C. Hall, R. Lemke Oliver, P. Pollack, and L. Thompson. Bounded gaps between primes in number fields and function fields. *Proc. Amer. Math. Soc.*, 143(7):2841–2856, 2015.
- [13] J. R. Chen. On the representation of a larger even integer as the sum of a prime and the product of at most two primes. *Sci. Sinica*, 16:157–176, 1973.
- [14] J. R. Chen. On the Goldbach’s problem and the sieve methods. *Sci. Sinica*, 21(6):701–739, 1978.
- [15] S. Chowla. *The Riemann hypothesis and Hilbert’s tenth problem*. Mathematics and Its Applications, Vol. 4. Gordon and Breach Science Publishers, New York-London-Paris, 1965.
- [16] L. Chua, S. Park, and G. D. Smith. Bounded gaps between primes in special sequences. *Proc. Amer. Math. Soc.*, 143(11):4597–4611, 2015.
- [17] A. de Polignac. Recherches nouvelles sur les nombres premiers. *Comptes Rendus Acad. Sci.*, (29):397–401, 1849.
- [18] P. D. T. A. Elliott and H. Halberstam. A conjecture in prime number theory. In *Symposia Mathematica, Vol. IV (INDAM, Rome, 1968/69)*, pages 59–72. Academic Press, London, 1970.
- [19] P. Erdős. The difference of consecutive primes. *Duke Math. J.*, 6:438–441, 1940.
- [20] K. Ford, B. Green, S. Konyagin, J. Maynard, and T. Tao. Long gaps between primes. *J. Amer. Math. Soc.*, 31(1):65–105, 2018.
- [21] K. Ford, B. Green, S. Konyagin, and T. Tao. Large gaps between consecutive prime numbers. *Ann. of Math. (2)*, 183(3):935–974, 2016.
- [22] É. Fouvry. Autour du théorème de Bombieri-Vinogradov. *Acta Math.*, 152(3-4):219–244, 1984.
- [23] É. Fouvry and F. Grupp. On the switching principle in sieve theory. *J. Reine Angew. Math.*, 370:101–126, 1986.
- [24] E. Fouvry and H. Iwaniec. On a theorem of Bombieri-Vinogradov type. *Mathematika*, 27(2):135–152 (1981), 1980.
- [25] E. Fouvry and H. Iwaniec. Primes in arithmetic progressions. *Acta Arith.*, 42(2):197–218, 1983.
- [26] N. Frantzikinakis. An averaged Chowla and Elliott conjecture along independent polynomials. *Int. Math. Res. Not. IMRN*, (12):3721–3743, 2018.
- [27] N. Frantzikinakis and B. Host. The logarithmic Sarnak conjecture for ergodic weights. *Ann. of Math. (2)*, 187(3):869–931, 2018.
- [28] T. Freiberg. Short intervals with a given number of primes. *J. Number Theory*, 163:159–171, 2016.
- [29] D. A. Goldston, J. Pintz, and C. Y. Yıldırım. Primes in tuples. I. *Ann. of Math. (2)*, 170(2):819–862, 2009.
- [30] É. Goudout. Théorème d’Erdős–Kac dans presque tous les petits intervalles. *Acta Arith.*, 182(2):101–116, 2018.
- [31] A. Granville. Primes in intervals of bounded length. *Bull. Amer. Math. Soc. (N.S.)*, 52(2):171–222, 2015.
- [32] G. H. Hardy and J. E. Littlewood. Some problems of ‘Partitio numerorum’; III: On the expression of a number as a sum of primes. *Acta Math.*, 44(1):1–70, 1923.
- [33] W. Huang and X. Wu. On the set of the difference of primes. *Proc. Amer. Math. Soc.*, 145(9):3787–3793, 2017.

- [34] M. N. Huxley. Small differences between consecutive primes. *Mathematika*, 20:229–232, 1973.
- [35] M. N. Huxley. Small differences between consecutive primes. II. *Mathematika*, 24(2):142–152, 1977.
- [36] M. N. Huxley. An application of the Fouvry-Iwaniec theorem. *Acta Arith.*, 43(4):441–443, 1984.
- [37] D. A. Kaptan. A note on small gaps between primes in arithmetic progressions. *Acta Arith.*, 172(4):351–375, 2016.
- [38] O. Klurman. Correlations of multiplicative functions and applications. *Compos. Math.*, 153(8):1622–1657, 2017.
- [39] O. Klurman and A. Mangerel. Rigidity theorems for multiplicative functions. *Math. Ann.*, 372(1-2):651–697, 2018.
- [40] E. Kowalski. Gaps between prime numbers and primes in arithmetic progressions [after Y. Zhang and J. Maynard]. *Astérisque*, (367-368):Exp. No. 1084, ix, 327–366, 2015.
- [41] S. Lester, K. Matomäki, and M. Radziwiłł. Small scale distribution of zeros and mass of modular forms. *J. Eur. Math. Soc. (JEMS)*, 20(7):1595–1627, 2018.
- [42] H. Li and H. Pan. Bounded gaps between primes of a special form. *Int. Math. Res. Not. IMRN*, (23):12345–12365, 2015.
- [43] J. Li, K. Pratt, and G. Shakan. A lower bound for the least prime in an arithmetic progression. *Q. J. Math.*, 68(3):729–758, 2017.
- [44] H. Maier. Small differences between prime numbers. *Michigan Math. J.*, 35(3):323–344, 1988.
- [45] H. Maier and M. T. Rassias. Large gaps between consecutive prime numbers containing perfect k -th powers of prime numbers. *J. Funct. Anal.*, 272(6):2659–2696, 2017.
- [46] K. Matomäki and M. Radziwiłł. Multiplicative functions in short intervals. *Ann. of Math. (2)*, 183(3):1015–1056, 2016.
- [47] K. Matomäki, M. Radziwiłł, and T. Tao. An averaged form of Chowla’s conjecture. *Algebra Number Theory*, 9(9):2167–2196, 2015.
- [48] K. Matomäki, M. Radziwiłł, and T. Tao. Sign patterns of the Liouville and Möbius functions. *Forum Math. Sigma*, 4:e14, 44, 2016.
- [49] K. Matomäki and X. Shao. Vinogradov’s three primes theorem with almost twin primes. *Compos. Math.*, 153(6):1220–1256, 2017.
- [50] J. Maynard. Small gaps between primes. *Ann. of Math. (2)*, 181(1):383–413, 2015.
- [51] J. Maynard. Dense clusters of primes in subsets. *Compos. Math.*, 152(7):1517–1554, 2016.
- [52] J. Maynard. Large gaps between primes. *Ann. of Math. (2)*, 183(3):915–933, 2016.
- [53] H. Parshall. Small gaps between configurations of prime polynomials. *J. Number Theory*, 162:35–53, 2016.
- [54] G. Z. Pil’tjai. The magnitude of the difference between consecutive primes. In *Studies in number theory, No. 4 (Russian)*, pages 73–79. Izdat. Saratov. Univ., Saratov, 1972.
- [55] J. Pintz. On the ratio of consecutive gaps between primes. In *Analytic number theory*, pages 285–304. Springer, Cham, 2015.
- [56] J. Pintz. Patterns of primes in arithmetic progressions. In *Number theory—Diophantine problems, uniform distribution and applications*, pages 369–379. Springer, Cham, 2017.
- [57] P. Pollack. Bounded gaps between primes with a given primitive root. *Algebra Number Theory*, 8(7):1769–1786, 2014.

- [58] P. Pollack and L. Thompson. Arithmetic functions at consecutive shifted primes. *Int. J. Number Theory*, 11(5):1477–1498, 2015.
- [59] D. H. J. Polymath. Variants of the Selberg sieve, and bounded intervals containing many primes. *Res. Math. Sci.*, 1:Art. 12, 83, 2014.
- [60] D. H. J. Polymath, W. Castryck, E. Fouvry, G. Harcos, E. Kowalski, P. Michel, P. Nelson, E. Paldi, J. Pintz, A. V. Sutherland, T. Tao, and X.-F. Xie. New equidistribution estimates of Zhang type. *Algebra Number Theory*, 8(9):2067–2199, 2014.
- [61] R. A. Rankin. The difference between consecutive prime numbers. IV. *Proc. Amer. Math. Soc.*, 1:143–150, 1950.
- [62] G. Ricci. Sull’andamento della differenza di numeri primi consecutivi. *Riv. Mat. Univ. Parma*, 5:3–54, 1954.
- [63] K. Soundararajan. Small gaps between prime numbers: the work of Goldston-Pintz-Yıldırım. *Bull. Amer. Math. Soc. (N.S.)*, 44(1):1–18, 2007.
- [64] K. Soundararajan. The Liouville function in short intervals. *Astérisque*, (390):Exp. No. 1119, 453–479, 2017. Séminaire Bourbaki. Vol. 2015/2016. Exposés 1104–1119.
- [65] T. Tao. The Erdős discrepancy problem. *Discrete Anal.*, pages Paper No. 1, 29, 2016.
- [66] T. Tao. The logarithmically averaged Chowla and Elliott conjectures for two-point correlations. *Forum Math. Pi*, 4:e8, 36, 2016.
- [67] J. Teräväinen. Almost primes in almost all short intervals. *Math. Proc. Cambridge Philos. Soc.*, 161(2):247–281, 2016.
- [68] J. Teräväinen. On binary correlations of multiplicative functions. *Forum Math. Sigma*, 6:e10, 41, 2018.
- [69] J. Thorner. Bounded gaps between primes in Chebotarev sets. *Res. Math. Sci.*, 1:1:4, 2014.
- [70] L. Troupe. Bounded gaps between prime polynomials with a given primitive root. *Finite Fields Appl.*, 37:295–310, 2016.
- [71] S. Uchiyama. On the difference between consecutive prime numbers. *Acta Arith.*, 27:153–157, 1975. Collection of articles in memory of Juriĭ Vladimirovič Linnik.
- [72] A. Vatwani. Bounded gaps between Gaussian primes. *J. Number Theory*, 171:449–473, 2017.
- [73] A. I. Vinogradov. The density hypothesis for Dirichet L -series. *Izv. Akad. Nauk SSSR Ser. Mat.*, 29:903–934, 1965.
- [74] J. Wu. Chen’s double sieve, Goldbach’s conjecture and the twin prime problem. *Acta Arith.*, 114(3):215–273, 2004.
- [75] Y. Zhang. Bounded gaps between primes. *Ann. of Math. (2)*, 179(3):1121–1174, 2014.