ISRAEL JOURNAL OF MATHEMATICS **206** (2015), 165–182
DOI: 10.1007/s11856-014-1129-5

PRESCRIBING THE BINARY DIGITS OF PRIMES, II

BY

JEAN BOURGAIN*

Institute for Advanced Study, Princeton, NJ 08540, USA
e-mail: bourgain@math.ias.edu

ABSTRACT

We obtain the expected asymptotic formula for the number of primes $p < N = 2^n$ with r prescribed (arbitrary placed) binary digits, provided $r < cn$ for a suitable constant $c > 0$. This result improves on our earlier result where r was assumed to satisfy $r < c\big(\frac{n}{\log n}\big)^{4/7}$.

1. Summary

This paper is a follow up on [B1]. We establish the following stronger statement.

THEOREM: *Let $N = 2^n$, n large enough, and $A \subset \{1, \dots, n-1\}$ such that*

$$r = |A| < cn \tag{1.1}$$

(where c is an absolute constant). Then, considering binary expansions $x = \sum_{j<n} x_j 2^j$ ($x_0 = 1$ and $x_j = 0, 1$ for $1 \le j < n$) and assignments α_j for $j \in A$, we have

$$|\{p < N;\ \text{for}\ \ j \in A,\ \text{the } j\text{-digit of } p \text{ equals } \alpha_j\}| = (1 + o(1))2^{-r} \frac{N}{\log N}. \tag{1.2}$$

In [B1], the corresponding result was proven under the more restrictive condition

$$r < c\Big(\frac{n}{\log n}\Big)^{4/7}. \tag{1.3}$$

* This work was partially supported by NSF grants DMS-1301619 and DMS-0835373.

Received June 25 2013 and in revised form January 9, 2014

We also refer the reader to [B1], [B2] for some background and motivation. In particular, the paper of Harman and Katai [H-K] and complexity issues for the Moebius function and the primes, raised by G. Kalai, were seminal to those investigations.

The strategy followed here is roughly similar to the one in [B1], except for the fact that the additive Fourier spectrum (together with Vinogradov's estimate) is only used to bound the contribution of the minor arcs. Also, the estimates on the additive Fourier transform established in §2 are significantly stronger than those used in [B1]. In the treatment of the major arcs, we switch immediately to multiplicative characters (see (3.8) below) and are led to study correlations of both the von Mangoldt function Λ and the given function $f = 1_{[x<N; x_j=\alpha_j \text{ for } j\in A]}$ with multiplicative characters. This issue for Λ is classical and depends on Dirichlet L-function theory. We rely here on the same basic facts that were used in [B1]. As in [B1], we subdivide primitive characters $\mathcal{X}$ into two classes $\mathcal{G}$ and $\mathcal{B}$ ('good' and 'bad') depending on the zero-free region of $L(s,\mathcal{X})$. It turns out that non-trivial bounds on the multiplicative spectrum of f are only required if $\mathcal{X}\in\mathcal{B}$ (which is a small set of characters). In order to establish those bounds, we rely again on estimates on the additive Fourier transform of f. Also, as in [B1], we are invoking the Gallagher–Iwaniec estimate on the improved zero-free region of $L(s,\mathcal{X})$ for $\mathcal{X}(\operatorname{mod} q)$ with q a power of 2, though the precise quantitative form of [I] (which was responsible for the condition (1.3)) is no longer relevant here. Basically any statement that for q as above, $L(s,\mathcal{X})$ has a zero-free region $1-\sigma < c\frac{\log\log qT}{\log qT}$, $|\gamma|<T$, where $\rho=\sigma+i\gamma$, would suffice for our purpose. This fact ensures then that no character $\mathcal{X}\in\mathcal{B}$ has conductor which is a power of 2, which is essential to our analysis. Note that possible Siegel zeros in any case forces us to introduce the class $\mathcal{B}$, even if r were further reduced. As in [B1], one needs to evaluate sums of the form

$$\sum_{x\in I, q_0|x} f(x) \tag{1.4}$$

and

$$\sum_{x\in I, q_0|x} \mathcal{X}(x)f(x) \tag{1.5}$$

with $I\subset\{1,\ldots,N\}$ intervals of a certain size and $\mathcal{X}\in\mathcal{B}$, $\mathcal{X}(\operatorname{mod} q)$, $(q,q_0)=1$. The main technical innovation compared with [B1] is a more efficient strategy to estimate these sums, leading to the required information under a less restrictive hypothesis on r.

The above theorem may be seen as a relative of Linnik's result on the least prime in an arithmetic progression. One key difference is that possible Siegel zeros do not affect the final statement (though they technically play a role in the argument).

Our presentation is completely self-contained, apart from basic number theoretic results, and we will not refer to [B1].

2. Preliminaries

For $x \in \{1, 2, \dots, 2^n - 1\}$, write $x = \sum_{0 \le j < n} x_j 2^j$ with $x_j = 0, 1$.

Let

$$f(x) = 1_{[x<2^n; x_j = \alpha_j \text{ for } j \in A]} \quad \text{and} \quad N = 2^n$$

where

$$A = \{0 = j_0 < j_1 < \dots < j_r\} \subset \{0, 1, \dots, n-1\}.$$

We assume

$$r + 1 = |A| = \rho n \tag{2.1}$$

with $\rho > 0$ bounded by a sufficiently small constant $c > 0$.

For $\lambda \in \mathbb{R}$, denote

$$\hat{f}(\lambda) = 2^{-n} \sum_{x=0}^{2^n-1} e^{2\pi i \lambda x} f(x) = 2^{-|A|} \prod_{j \in A} e^{2\pi i \lambda \alpha_j 2^j} \prod_{\substack{1 \le j < n \\ j \notin A}} \frac{1 + e^{i\pi\lambda 2^{j+1}}}{2}.$$

Thus

$$|\hat{f}(\lambda)| = 2^{-|A|} \prod_{\substack{1 \le j < n \\ j \notin A}} |\cos \pi \lambda 2^j|. \tag{2.2}$$

Lemma 1:

$$2^{r+1} \sum_{k=0}^{2^n-1} \Big| \hat{f}\Big(\frac{k}{2^n}\Big) \Big| < 2^{C\rho(\log \frac{1}{\rho})n} \tag{2.3}$$

for some constant C.

Proof. By (2.2), the left side of (2.3) equals

$$\sum_{k=0}^{2^n-1} \prod_{\substack{1\le j<n\\ j\notin A}} \Big|\cos\pi\frac{k}{2^{n-j}}\Big|. \tag{2.4}$$

Writing $k = k' + 2^{n-j_1}\ell$, $0 \le k' < 2^{n-j_1}$, $0 \le \ell < 2^{j_1}$, we get

$$(2.4) = \sum_{k'<2^{n-j_1}} \sum_{\ell<2^{j_1}} \prod_{j=1}^{j_1-1} \Big|\cos\pi\Big(\frac{k'}{2^{n-j}} + \frac{\ell}{2^{j_1-j}}\Big)\Big| \prod_{\substack{j_1<j<n\\ j\notin A}} \Big|\cos\pi\frac{k'}{2^{n-j}}\Big|.$$

Evaluate the inner sum with fixed k' as

$$\begin{aligned}\sum_{\ell<2^{j_1}} \prod_{j=1}^{j_1-1} \Big|\cos\pi\Big(\frac{k'}{2^{n-j}} + \frac{\ell}{2^{j_1-j}}\Big)\Big| &\le \max_\theta \sum_{\ell<2^{j_1}} \prod_{j=1}^{j_1-1} \Big|\cos\pi\frac{2^j(\ell+\theta)}{2^{j_1}}\Big| \\ &= \max_\theta\Big\{\sum_{\ell<2^{j_1}} 2^{-j_1}\Big|\sum_{x=0}^{2^{j_1}-1} e^{2\pi i x 2^{-j_1}(\ell+\theta)}\Big|\Big\} \\ &\le \max_\theta\Big\{\sum_{\ell<2^{j_1}} \frac{4}{2^{j_1}\|\frac{\ell+\theta}{2^{j_1}}\|+1}\Big\} < Cj_1\end{aligned}$$

for some constant C. Hence

$$(2.4) < C(j_1-j_0) \sum_{k'<2^{n-j_1}} \prod_{\substack{j_1<j<n\\ j\notin A}} \Big|\cos\pi\frac{k'}{2^{n-j}}\Big|$$

and we repeat the process with the k'-sum, replacing n by $n-j_1$, j_1 by j_2-j_1 etc. It follows that

$$(2.4) < C^r \prod_{s=1}^{r}(j_{s+1}-j_s) \tag{2.5}$$

where we have set $j_{r+1} = n$. Since $u \le \frac{1}{\theta}2^{\theta u}$ for $u \ge 0$, $\theta \ge 0$,

$$(2.5) < \Big(\frac{C}{\theta}\Big)^r 2^{\theta n} = \Big(\Big(\frac{C}{\theta}\Big)^\rho 2^\theta\Big)^n < 2^{C\rho(\log\frac{1}{\rho})n}$$

for an appropriate choice of θ, proving (2.3). ∎

LEMMA 2:

$$2^r\int_0^1 |\hat f(\theta)|d\theta < 2^{C\rho(\log\frac{1}{\rho})n-n}. \tag{2.6}$$

Proof. Since $\hat{f}(\theta)$ is a trigonometric polynomial with spectrum in $\{0,1,\ldots,2^n-1\}$,

$$\hat{f}(\theta) = 2^{-n} \sum_{k=0}^{2^n-1} \hat{f}\Big(\frac{k}{2^n}\Big) D_n\Big(\theta - \frac{k}{2^n}\Big)$$

with $D_n(\theta) = \sum_{k=0}^{2^n-1} e^{2\pi ik\theta}$ the Dirichlet kernel. It follows from Lemma 1 that

$$2^r \int_0^1 |\hat{f}(\theta)| d\theta \le \|D_n\|_1 \, 2^{C\rho(\log \frac{1}{\rho})n-n} < Cn2^{C\rho(\log \frac{1}{\rho})n-n}$$

proving (2.6). ∎

Lemma 3: *Let $Q < 2^{n/100}$. Then*

$$2^r \sum_{\substack{q<Q, q \text{ odd} \\ 1\le a<q,(a,q)=1}} \Big|\hat{f}\Big(\frac{a}{q}\Big)\Big| < Q^{C\rho \log \frac{1}{\rho}}. \tag{2.7}$$

Proof. Take m such that

$$2^{m-1} \le Q^2 < 2^m.$$

Clearly there is $0 < j_* < n - 2m$ so that the interval $I = \{j_*, \ldots, j_* + m - 1\}$ satisfies

$$r' = |A'| = |A \cap I| < 2\rho m. \tag{2.8}$$

We note that

$$2^r \Big|\hat{f}\Big(\frac{a}{q}\Big)\Big| \le \prod_{j\in I\setminus A'} \Big|\cos \frac{\pi a 2^j}{q}\Big| = 2^{r'} \Big|\hat{g}\Big(\frac{a2^{j_*}}{q}\Big)\Big|$$

where

$$g = 1_{[x<2^m; x_j = \alpha_{j+j_*} \text{ for } j\in A'-j_*]}.$$

It follows then from (2.6), (2.8) that

$$2^{r'} \int_0^1 |\hat{g}(\theta)| d\theta < 2^{C\rho(\log \frac{1}{\rho})m-m}. \tag{2.9}$$

The set of points

$$\mathcal{F} = \Big\{\frac{a2^{j_*}}{q} (\operatorname{mod} 1); q < Q, q \text{ odd}, (a, q) = 1\Big\}$$

are clearly pairwise 2^{-m}-separated. Write for $\xi \in \mathcal{F}$

$$|\hat{g}(\xi)| \le 2^m \int_{|\theta-\xi|<2^{-m-1}} |\hat{g}(\theta)| d\theta + 2^m \int_{|\theta-\xi|<2^{-m-1}} |\hat{g}(\theta) - \hat{g}(\xi)| d\theta$$

and

$$\sum_{\xi\in\mathcal{F}}|\hat{g}(\xi)|\leq 2^m\int_0^1|\hat{g}|+\int_0^1|(\hat{g})'|,$$

$$2^{r'}\sum_{\xi\in\mathcal{F}}|\hat{g}(\xi)|\lesssim 2^{m+r'}\int_0^1|\hat{g}|<2^{C\rho(\log\frac{1}{\rho})m}<Q^{2C\rho\log\frac{1}{\rho}},$$

where we used Bernstein's inequality and (2.9). This proves (2.7). ∎

For small q, there is the following individual bond.

LEMMA 4: *Let $1<q<n^{\frac{1}{10\rho}}$ and odd, $(a,q)=1$. Then*

$$2^r\Big|\hat{f}\Big(\frac{a}{q}\Big)\Big|<2^{-\sqrt{n}}. \tag{2.10}$$

Proof. Clearly

$$2^r\Big|\hat{f}\Big(\frac{a}{q}\Big)\Big|=\prod_{\substack{1\leq j<n\\ j\notin A}}\Big|\cos\pi 2^j\frac{a}{q}\Big|\leq\gamma^{\frac{1}{2}\frac{n}{\ell}}\quad\text{with}\quad\ell=[\log_2 q]+1 \tag{2.11}$$

where γ is an upper bound on

$$\prod_{\substack{0\leq j<\ell\\ j\notin E}}\Big|\cos\pi 2^j\frac{a'}{q}\Big|, \tag{2.12}$$

where $(a',q)=1$ and $E\subset\{0,1,\dots,\ell-1\}$ satisfies $|E|<2\rho\ell$.

Take $0\leq j_*\leq 2\rho\ell$ such that $j_*\notin E$ and set $2^{-\ell'-1}\leq\Big\|2^{j_*}\frac{a'}{q}\Big\|<2^{-\ell'}$, $0\leq\ell'\leq\ell$.

Then

$$\Big\|2^j\frac{a'}{q}\Big\|\sim 2^{j-j_*-\ell'}\quad\text{for } 0\leq j-j_*<\ell'-1.$$

If $\ell'>10+2\rho\ell$, we can find j such that $\ell'-2\rho\ell\leq j-j_*<\ell'-1$ and $j\notin E$.

Then $\big\|2^j\frac{a'}{q}\big\|\gtrsim 2^{-2\rho\ell}$. Hence, in either case we find some $0\leq j<\ell$, $j\notin E$, such that $\|2^j\frac{a'}{q}\|>c2^{-2\rho\ell}>cq^{-2\rho}$. It follows that

$$(2.12)\leq\Big|\cos\pi 2^j\frac{a'}{q}\Big|<1-\frac{1}{2}\Big\|\frac{2^ja'}{q}\Big\|^2<1-cq^{-4\rho}.$$

Substituting in (2.11) gives the bound $e^{-\frac{cn}{q^{4\rho}\log q}}$ and the Lemma follows. ∎

3. Minor arcs contribution

Let $N = 2^n$. Write

$$\sum_{k\le N}\Lambda(k)f(k) = \int_0^1 S(\alpha)\overline{S}_f(\alpha)d\alpha, \tag{3.1}$$

denoting

$$S(\alpha) = \sum \Lambda(k)e(k\alpha) \tag{3.2}$$

and

$$S_f(\alpha) = \sum f(k)e(k\alpha) = N\hat{f}(\alpha). \tag{3.3}$$

We assume $f(k) = 0$ for k even, since obviously $k \equiv 1 (\mathrm{mod}\, 2)$ is a necessary condition for f to capture primes.

We fix a parameter $B = B(n)$ which will be specified later, B at most a small power of N.

The major arcs are defined by

$$\mathcal{M}(q,a) = \Big[\Big|\alpha - \frac{a}{q}\Big| < \frac{B}{qN}\Big] \quad \text{where } q < B. \tag{3.4}$$

Given α, there is $q < \frac{N}{B}$ such that

$$\Big|\alpha - \frac{a}{q}\Big| < \frac{B}{qN} < \frac{1}{q^2}.$$

From Vinogradov's estimate (Theorem 13.6 in [I-K])

$$\begin{aligned}|S(\alpha)| &< (q^{\frac12}N^{\frac12} + q^{-\frac12}N + N^{\frac45})(\log N)^3\\ &< C\Big(\frac{N}{\sqrt{B}} + \frac{N}{\sqrt{q}} + N^{4/5}\Big)(\log N)^3.\end{aligned} \tag{3.5}$$

Hence if $q \ge B$,

$$|S(\alpha)| < C\frac{N}{\sqrt{B}}(\log N)^3. \tag{3.6}$$

Thus the minor arcs contribution in (3.1) is at most

$$C\frac{N}{\sqrt{B}}(\log N)^3\|S_f\|_1. \tag{3.7}$$

Since by (2.16)

$$\|S_f\|_1 < 2^{-r}N^{C\rho\log\frac{1}{\rho}}, \tag{3.8}$$

we take

$$\log B > 3C\rho\Big(\log\frac{1}{\rho}\Big)n \tag{3.9}$$

which takes care of the minor arcs contribution.

4. Major arcs analysis

Next, we analyze the major arcs contributions ($q < B$)

$$\sum_{(a,q)=1}\ \int_{|\alpha-\frac{a}{q}|<\frac{B}{qN}} S(\alpha)\overline{S}_f(\alpha)d\alpha. \tag{4.1}$$

Write $\alpha = \frac{a}{q}+\beta$. Defining

$$\tau(\chi) = \sum_{m=1}^{q}\chi(m)e_q(m)$$

we have (see [D], p. 147)

$$S(\alpha) = \frac{1}{\phi(q)}\sum_{\chi}\tau(\overline{\chi})\chi(a)\Big[\sum_{k\le N}\chi(k)\Lambda(k)e(k\beta)\Big] + O((\log N)^2). \tag{4.2}$$

Assume χ is inducted by χ_1 which is primitive $(\bmod\, q_1), q_1|q$. Then from [D], p. 67

$$\tau(\bar\chi) = \mu\Big(\frac{q}{q_1}\Big)\bar\chi_1\Big(\frac{q}{q_1}\Big)\tau(\bar\chi_1) \tag{4.3}$$

which vanishes, unless $q_2 = \frac{q}{q_1}$ is square free with $(q_1,q_2)=1$.

The contribution of χ in (4.1) equals

(4.4)

$$\frac{\tau(\overline{\chi})}{\phi(q)}\int_{|\beta|<\frac{B}{qN}}\Big[\sum_{k\le N}\chi(k)\Lambda(k)e(k\beta)\Big]\Big[\sum_{k<N}f(k)\Big(\sum_{a=1}^{q}\chi(a)e_q(-ak)\Big)e(-k\beta)\Big]d\beta.$$

We have

$$\begin{aligned}\sum_{a=1}^{q}e_q(ak)\chi(a) &= \sum_{(a,q)=1}e_q(ak)\chi_1(a)\\ &= \Big[\sum_{(a_1,q_1)=1}e_{q_1}(a_1k)\chi_1(a_1)\Big]\Big[\sum_{(a_2,q_2)=1}e_{q_2}(a_2k)\Big]\mathcal{X}_1(q_2)\\ &= \overline{\chi_1}(k)\tau(\chi_1)c_{q_2}(k)\mathcal{X}_1(q_2),\end{aligned} \tag{4.5}$$

where

$$c_{q_2}(k) = \frac{\mu\Big(\frac{q_2}{(q_2,k)}\Big)\phi(q_2)}{\phi\Big(\frac{q_2}{(q_2,k)}\Big)}. \tag{4.6}$$

From (4.3), (4.5), (4.6)

$$\begin{aligned} \frac{\tau(\overline{\chi})}{\phi(q)}\Big[\sum_{a=1}^{q}\chi(a)e_q(-ak)\Big] &= \frac{|\tau(\chi_1)|^2}{\phi(q_1)}\frac{1}{\phi\Big(\frac{q_2}{(q_2,k)}\Big)}\mu((q_2,k))\overline{\chi_1}(-k) \\ &= \frac{q_1}{\phi(q_1)}\ \frac{1}{\phi(\frac{q_2}{(q_2,k)})}\mu((q_2,k))\overline{\chi_1}(-k). \end{aligned} \tag{4.7}$$

Returning to (4.2), rather than integrating in β over the interval $|\beta| < \frac{B}{qN}$, we introduce a weight function

$$w\Big(\frac{qN}{B}\beta\Big)$$

where $0 \le w \le 1$ is a smooth bumpfunction on $\mathbb{R}$ such that $w = 1$ on $[-1, 1]$, supp $w \subset [-2, 2]$ and

$$|\hat{w}(y)| < Ce^{-|y|^{1/2}}.$$

See [I] for the existence of such function.

(Note that this operation creates in (4.1) an error term that is captured by the minor arcs contribution (3.7).)

Hence, substituting (4.7), (4.4) becomes

$$\frac{q_1}{\phi(q_1)}\frac{B}{qN}\sum_{k_1,k_2<N}\hat{w}\Big(\frac{B}{qN}(k_1-k_2)\Big)\mathcal{X}(k_1)\Lambda(k_1)f(k_2)\frac{\mu((q_2,k_2))}{\phi(\frac{q_2}{(q_2,k_2)})}\overline{\mathcal{X}_1}(k_2) \tag{4.8}$$

and we observe that by our assumption on $\hat{w}$ the k_1, k_2 summation in (4.8) is restricted to $|k_1 - k_2| < \frac{qN}{B}n^3$, up to a negligible error.

We first examine the contribution of the principal characters.

For $\mathcal{X} = \mathcal{X}_0, q_1 = 1$ and (4.8) becomes

$$\frac{B}{qN}\sum_{k_1,k_2<N}\hat{w}\Big(\frac{B}{qN}(k_1-k_2)\Big)\Lambda(k_1)f(k_2)\frac{\mu((q,k_2))}{\phi(\frac{q}{(q,k_2)})}. \tag{4.9}$$

Fixing k_2, perform the k_1-summation in (4.9). Writing

$$\psi(x) = x - \sum_{\substack{\zeta(\rho)=0\\ |\gamma|<B^2}}\frac{x^\rho}{\rho} + O\Big(\frac{x}{B^2}(\log x)^2\Big) \tag{4.10}$$

for $x > \frac{N}{B}$ and assuming also

$$\log B < \frac{n}{1000}, \tag{4.11}$$

partial summation, together with the usual zero-density and zero-free region estimate, give

$$(4.9) = \sum_{k\leq N} f(k) \frac{\mu((q,k))}{\phi(\frac{q}{(q,k)})} \tag{4.12}$$

$$+ O\Big\{\Big[\sum_{k\leq N} \frac{f(k)}{\phi(\frac{q}{(q,k)})}\Big] \exp(-(\log N)^{\frac{1}{2}})\Big\}. \tag{4.13}$$

Let $\kappa(q)$ be a function satisfying the following:

Assumption A: *Let $q_0 < B$ be odd and square free. Then*

$$\sum_{q_0|k, k<N} f(k) = \mathbb{E}[f] \frac{N}{q_0} + O(\kappa(q_0) N \mathbb{E}[f]), \tag{4.14}$$

where $\mathbb{E}[f]$ denotes the normalized average $2^{-r} \sum_{1\leq x\leq 2^r} f(x)$.

Assuming q square-free (sf) and odd, (4.12) equals

$$\sum_{q'|q} \frac{\mu(q')}{\phi(\frac{q}{q'})} \Big[\sum_{(q,k)=q'} f(k)\Big] = \sum_{q'|q} \frac{\phi(q')}{\phi(q)} \mu(q') \sum_{q''|\frac{q}{q'}} \mu(q'') \Big[\sum_{\substack{q'q''|k\\ k<N}} f(k)\Big],$$

and substituting (4.14) we obtain

$$N\,\mathbb{E}[f] \sum_{q'|q} \frac{\mu(q')}{\phi(q/q')} \sum_{q''|\frac{q}{q'}} \frac{\mu(q'')}{q'q''} \tag{4.15}$$

$$+ O\Big(N\, 2^{-r} \sum_{q'|q} \frac{\phi(q')}{\phi(q)} \sum_{q''|\frac{q}{q'}} \kappa(q'q'')\Big). \tag{4.16}$$

Next,

$$\begin{aligned}
(4.15) &= N\,\mathbb{E}[f] \sum_{q'|q} \frac{\mu(q')}{\phi(q/q')} \frac{1}{q'} \prod_{p|\frac{q}{q'}} \Big(1 - \frac{1}{p}\Big) \\
&= 2N\mathbb{E}[f] \sum_{q'|q} \frac{\mu(q')}{q} \qquad (4.17)\\
&= \begin{cases} N\,\mathbb{E}[f] & \text{if } \ q = 1, \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}$$

Summing (4.16) over $q < B$ sf and odd, we have the estimate (setting $q_1 = q'q''$)

$$\begin{aligned} 2^{-r}N \sum_{\substack{q_1<B\\ q_1 \text{ sf}}} \kappa(q_1)\Bigg[\sum_{\substack{q''|q_1,(q_1,q_2)=1\\ q_2<B \text{ sf}}} \frac{1}{\phi(q'')\phi(q_2)}\Bigg] \\ < 2^{-r}N(\log B)^2\Bigg[\sum_{\substack{q<B\\ q \text{ sf, odd}}} \kappa(q)\Bigg]. \end{aligned} \tag{4.18}$$

For q sf and even, set $q = 2q_1$ and note that (4.12) equals

$$\sum_{q'|q_1} \frac{\mu(q')}{\phi\big(q_1/q'\big)}\Bigg[\sum_{(q_1,k)=q'} f(k)\Bigg],$$

and we proceed similarly as above with the same conclusion and q replaced by q_1.

The first factor in (4.13) contributes for

$$\begin{aligned} \sum_{q<B,q \text{ sf}} \sum_{q'|q,q' \text{ odd}} \frac{1}{\phi(\frac{q}{q'})} \sum_{\substack{k\leq N\\ q'|k}} f(k) \overset{(4.14)}{\leq} N2^{-r} \sum_{\substack{q',q''<B\\ q',q'' \text{ sf}, q' \text{ odd}}} \frac{1}{\phi(q'')}\Big(\frac{1}{q'}+\kappa(q')\Big) \\ < N2^{-r}\Bigg[(\log B)^2 + (\log B)\Bigg(\sum_{\substack{q<B\\ q \text{ sf, odd}}} \kappa(q)\Bigg)\Bigg]. \end{aligned} \tag{4.19}$$

Thus, from the preceding, the contribution of the principal characters equals

$$2\mathbb{E}[f]N + CN2^{-r}n^2\Bigg\{\sum_{\substack{q<B\\ q \text{ sf, odd}}} \kappa(q)\Bigg\} + CN2^{-r}n^2\exp(-n^{\frac{1}{2}}). \tag{4.20}$$

Next, consider non-principal characters, i.e., $q_1 > 1$.

Estimate (4.8) by

$$\frac{q_1}{\phi(q_1)}\,\frac{B^3}{N}(4.21).(4.22) + O\Big(\frac{N}{qB^2}\Big) \tag{4.23}$$

with

$$(4.21) = \max_{|I|\sim \frac{N}{B^3}} \Big|\sum_{k\in I} \mathcal{X}(k)\Lambda(k)\Big|$$

and

$$(4.22) = \sum_{I}\Bigg|\sum_{k\in I} f(k)\,\frac{\mu((q_2,k))}{\phi(\frac{q_2}{(q_2,k)})}\bar{\mathcal{X}}_1(k)\Bigg|,$$

where I runs over a partition in intervals of size $\sim \frac{N}{B^3}$.

Obviously

$$(4.22) \leq \sum_{k \leq N} \frac{f(k)}{\phi(\frac{q_2}{(q_2,k)})}$$

and summing over $q_2 < B, q_2$ sf, gives the estimate (4.19).

The factor (4.21) is bounded by

$$\max_{\frac{N}{B^2} < x < N} |\psi(x+h, \mathcal{X}) - \psi(x, \mathcal{X})| \quad \text{with } h \sim \frac{N}{B^3}. \tag{4.24}$$

Choose a parameter $B < T < N^{\frac{1}{100}}$ and denote by $N(\alpha, T; \mathcal{X})$ the number of zeros of $L(s, \mathcal{X})$ such that

$$\alpha \leq \sigma \leq 1, \ |t| \leq T \quad (s = \sigma + it).$$

Then (see [Bom], Theorem 14)

$$N(\alpha) = \sum_{q \leq Q} \ \sideset{}{^*}\sum_{\chi (\mathrm{mod}\, q)} N(\alpha, T; \chi) < C(TQ)^{8(1-\alpha)}, \tag{4.25}$$

where $\sum^*$ refers to summation over primitive characters.

Let χ be a non-principal character. From Proposition 5.25 in [I-K], for $T \leq x$

$$\psi(x, \chi) = -\sum_{\substack{L(\rho,\chi)=0 \\ |\gamma| \leq T}} \frac{x^\rho - 1}{\rho} + O\Big(\frac{x}{T}(\log xq)^2\Big), \tag{4.26}$$

where $\rho = \beta + i\gamma$. We denote

$$\eta = \eta(\chi) = \min(1 - \beta)$$

with min taken over all zeros ρ of $L(s, \chi)$ with $|\gamma| \leq T$.

Taking $T = B^5$, we get from (4.26) that

$$(4.24) \leq h \sum_{\substack{L(\rho,\mathcal{X}_1)=0 \\ |\gamma|<B^5}} \frac{1}{x^{1-\beta}} + O\Big(\frac{Nn^2}{B^5}\Big). \tag{4.27}$$

At this point, we fix some $\eta^* = \eta^*(n)$ and subdivide the primitive characters $\mathcal{X}_1$ in classes $\mathcal{G}$ and $\mathcal{B}$ depending on whether $\eta \geq \eta^*$ or $\eta < \eta^*$.

Recall that $q \leq B$.

Summing (4.27) over q_1, $\mathcal{X}_1(\bmod q_1)$ primitive and $\mathcal{X}_1 \in \mathcal{G}$, we obtain from the density bound (4.25).

(4.28)
$$\sum_{\mathcal{X}_1\in\mathcal{G}} \frac{1}{x^{1-\beta}} = -2\int_{\frac{1}{2}}^{1-\eta_*} \frac{1}{x^{1-\alpha}} dN(\alpha) \le 2x^{-\frac{1}{2}} N\Big(\frac{1}{2}\Big) + 2\log x \int_{\eta_*}^{\frac{1}{2}} \Big(\frac{B^{48}}{x}\Big)^\tau d\tau$$
$$< \frac{B^{21}}{N^{\frac{1}{2}}} + N^{-\frac{1}{2}\eta_*} < O(N^{-\frac{1}{2}\eta_*}).$$

Hence, the contribution to (4.23) of the $\mathcal{X}_1 \in \mathcal{G}$ may be estimated by

$$(\log\log B)\frac{B^3}{N}\Big[\frac{N}{B^3}N^{-\frac{1}{2}\eta_*} + \frac{Nn^2}{B^4}\Big].(4.19)$$
$$< n^3(N^{-\frac{1}{2}\eta_*} + n^2B^{-1})\Big(1 + \sum_{q<B}\kappa(q)\Big)2^{-r}N. \tag{4.29}$$

Next, we consider the contribution of the $\mathcal{X}_1 \in \mathcal{B}$. Again from (4.25)

$$|\mathcal{B}| \ll (TB)^{8\eta_*} \le B^{48\eta_*}. \tag{4.30}$$

Use the trivial bound $\frac{N}{B^3}$ on (4.21) for $\mathcal{X}_1 \in \mathcal{B}$. We get the following estimate for the $\mathcal{B}$-contribution to the first term of (4.23):

$$\sum_{\mathcal{X}_1\in\mathcal{B}} \sum_{\substack{q_2<B\\ q_2 \text{ sf}}} (4.22). \tag{4.31}$$

Introduce another parameter $\alpha(q_1, q_0)$ satisfying the condition.

Assumption B: *Given $q_0, q_1 < B, (q_0, q_1) = 1$ with q_0 sf and odd, $\chi_1(\bmod q_1)$ primitive, $\chi_1 \in \mathcal{B}$,*

$$\Big|\sum_{k\in I, q_0|k} f(k)\chi_1(k)\Big| < \alpha(q_1, q_0)\Big[\sum_{k\in I} f(k) + |I|2^{-r}\Big] \tag{4.32}$$

holds, whenever $I \subset [1, N]$ is an interval of size $\sim \frac{N}{B^3}$.

Hence

$$(4.22) = \sum_I \sum_{\substack{q_2'|q_2\\ q_2' \text{ odd}}} \frac{1}{\phi(\frac{q_2}{q_2'})}\Big|\sum_{k\in I,(k,q_2)=q_2'} f(k)\chi_1(k)\Big|$$
$$\le N2^{-r} \sum_{\substack{q_2'|q_2\\ q_2' \text{ odd}}} \frac{1}{\phi(\frac{q_2}{q_2'})} \sum_{q_2''|\frac{q_2}{q_2'}, q_2'' \text{ odd}} \alpha(q_1, q_2'q_2'')$$

and summation over sf $q_2 < B$ gives

$$N2^{-r} \sum_{\substack{q_3<B \\ q_3 \text{ sf, odd}}} \alpha(q_1, q_3) \sum_{\substack{q_2''|q_3, q_2'''<B \\ q_2''' \text{ sf}, (q_2''', q_3)=1}} \frac{1}{\phi(q_2'')\phi(q_2''')} \tag{4.33}$$
$$< Cn^2 N 2^{-r} \Big[\sum_{\substack{q_3<B \\ q_3 \text{ sf, odd}}} \alpha(q_1, q_3) \Big].$$

By (4.30), this gives the following bound on (4.31):

$$n^2 B^{48\eta_*} N 2^{-r} \max_{\mathcal{X}\in\mathcal{B}} \Big[\sum_{q_0<B, q_0 \text{ sf, odd}} \alpha(q_1, q_0) \Big]. \tag{4.34}$$

In the next section, we will establish bounds on

$$\sum_{\substack{q<B \\ q \text{ sf, odd}}} \kappa(q) \tag{4.35}$$

and

$$\sum_{\substack{q_0<B \\ q_0 \text{ sf, odd}}} \alpha(q_1, q_0). \tag{4.36}$$

In particular, (4.35) $< O(1)$ so that a choice

$$\eta_* = O\Big(\frac{\log n}{n}\Big) \tag{4.37}$$

suffices for (4.29) to be conclusive.

It is important to note that for this choice of η_*, no $\mathcal{X}_1 \in \mathcal{B}$ has a conductor q_1 which is a power of 2. Indeed, recalling the Gallagher–Iwaniec result (see [H-K], Lemma 5), if $\mathcal{X}_1$ is primitive $(\operatorname{mod} 2^m)$, we obtain the following improved zero-free region:

$$\eta(\mathcal{X}_1) > c[(\log 2^m T)(\log\log 2^m T)]^{-\frac{3}{4}} > c(\log B \log\log B)^{-\frac{3}{4}} > \eta_* \tag{4.38}$$

(recall also that Siegel zeros are not a concern).

5. Further estimates

It remains to obtain suitable bounds on $\kappa(q)$ and $\alpha(q_1, q_0)$ introduced in Assumptions A and B.

Write for $q < B$, q sf and odd

$$\sum_{q|k} f(k) = \frac{N}{q}\mathbb{E}[f] + \frac{1}{q}\sum_{a=1}^{q-1}\sum_{k} f(k) e\Big(\frac{ak}{q}\Big)$$

and hence

$$\kappa(q) \leq \frac{2^r}{q}\sum_{a=1}^{q-1}\Big|\hat{f}\Big(\frac{a}{q}\Big)\Big|. \tag{5.1}$$

It follows that (4.35) may be bounded by

$$2^r \sum_{\substack{1<q<B\\ q\,\text{sf, odd}}} \frac{1}{q}\sum_{a=1}^{q-1}\Big|\hat{f}\Big(\frac{a}{q}\Big)\Big| \leq \log B \sum_{\substack{1<q<B\\ q\text{ sf, odd}\\ (a,q)=1}} \frac{2^r}{q}\Big|\hat{f}\Big(\frac{a}{q}\Big)\Big|. \tag{5.2}$$

Consider dyadic ranges $q \sim Q < B$. Lemma 3 provides an estimate $Q^{C\rho(\log\frac{1}{\rho})-1} < Q^{-\frac{1}{2}}$ (for ρ small enough) for the corresponding contribution to the sum in the r.h.s. of (5.2), while, for $Q < n^{\frac{1}{10\rho}}$, Lemma 4 gives an estimate $Q2^{-\sqrt{n}}$. It follows that

$$(5.2) < n.n^{-\frac{1}{20\rho}} + n^{\frac{1}{10\rho}+1}e^{-\sqrt{n}} < 2.n^{-\frac{1}{20\rho}+1}. \tag{5.3}$$

Next, consider Assumption B. Observe first that we can assume (by subdivision) I to be of the form $[0, 2^m - 1] + u2^m$ with $m = [\frac{n}{2}]$ say.

Fix $u \in \{0, 1, \dots, 2^{n-m} - 1\}$ such that $u_j = \alpha_{j+m}$ for $j + m \in A$ and define

$$f_1(x) = f(x + u2^m) \quad \text{for } x \in \{0, \dots, 2^m - 1\}. \tag{5.4}$$

Thus

$$f_1 = 1_{[x<2^m; x_j=\alpha_j \text{ for } j\in A\cap[1,m-1]]}.$$

It clearly suffices to establish inequalities (4.32) with $f|_I$ replaced by f_1, provided $\mathcal{X}_1(k)$ is replaced by $\mathcal{X}_1(k + u2^m)$. This basically leads to evaluate

$$\sum_{k<N, k+b\equiv 0 (\mathrm{mod}\, q_0)} f(k)\mathcal{X}_1(k+b) \tag{5.5}$$

without taking the restriction $k \in I$ into consideration.

Let us first assume $q_1 > 1$ is odd. Write (5.5) as

$$\frac{1}{q_0}\sum_{a_0=0}^{q_0-1}\sum_k f(k)e\Big(\frac{a_0}{q_0}(k+b)\Big)\mathcal{X}_1(k+b) = \frac{N}{q_0}\sum_{a_0=0}^{q_0-1}\sum_{(a_1,q_1)=1}\hat{\mathcal{X}}_1(a_1)e\Big(b\Big(\frac{a_0}{q_0}+\frac{a_1}{q_1}\Big)\Big)\hat{f}\Big(\frac{a_0}{q_0}+\frac{a_1}{q_1}\Big) \tag{5.6}$$

with

$$\hat{\mathcal{X}}_1(a_1) = \frac{1}{q_1}\sum_{x=0}^{q_1-1}\mathcal{X}_1(x)e\Big(-\frac{xa_1}{q}\Big).$$

Hence

$$|(5.6)| \le \frac{N}{q_0\sqrt{q_1}}\sum_{a_0=0}^{q_0-1}\sum_{(a_1,q_1)=1}\Big|\hat{f}\Big(\frac{a_0}{q_0}+\frac{a_1}{q_1}\Big)\Big|,$$

and summing over $1 \le q_0 < B, q_0$ sf, odd, $(q_0,q_1)=1$, we obtain a bound

$$\frac{N}{\sqrt{q_1}}\log B \sum_{\substack{1\le q_0<B, q_0 \text{ sf, odd}, (q_0,q_1)=1\\ (a_0,q_0)=1,(a_1,q_1)=1}}\frac{1}{q_0}\Big|\hat{f}\Big(\frac{a_0}{q_0}+\frac{a_1}{q_1}\Big)\Big| \le N.n\sum_{\substack{1<q<B^2, q \text{ sf, odd}\\ (a,q)=1}}\frac{1}{\sqrt{q}}\Big|\hat{f}\Big(\frac{a}{q}\Big)\Big|. \tag{5.7}$$

By a similar estimate as used for (5.2), we get for ρ small enough

$$(5.3) < n^{-\frac{1}{30\rho}+1}N.2^{-r}$$

and a bound

$$(4.36) < n^{-\frac{1}{30\rho}+1}. \tag{5.8}$$

If q_1 is even, write $q_1 = 2^\nu q_1'$, $(q_1',2)=1$ and $q_1' > 1$ since q_1 is not a power of 2. Let $\mathcal{X}_1 = \mathcal{X}_0\mathcal{X}_1'$ with $\mathcal{X}_0(\operatorname{mod} 2^\nu)$ and $\mathcal{X}_1'$ primitive $(\operatorname{mod} q_1')$. Write $k = z + 2^\nu x$ with $z \in \{0,1,\dots,2^\nu-1\}$, $x < 2^{n-\nu}$ and

$$(5.5) = \sum_{z=0}^{2^\nu-1}\mathcal{X}_0(b+z)\sum_{\substack{x<2^{n-\nu}\\ b+z+2^\nu x\equiv 0(\operatorname{mod} q_0)}}\mathcal{X}_1'(b+z+2^\nu x)f_z(x) \tag{5.9}$$

denoting

$$f_z(x) = f(z+2^\nu x). \tag{5.10}$$

Thus

$$|(5.9)| \leq \sum_{z=0}^{2^\nu-1} \max_{b'} \Big| \sum_{\substack{x<2^{n-\nu} \\ x+b'\equiv 0 (\mathrm{mod}\, q_0)}} \mathcal{X}_1(b'+x) f_z(x) \Big|. \tag{5.11}$$

Estimate the inner sum in (5.11) similarly to (5.5), with f replaced by f_z, q_1 by q_1', N by $2^{-\nu}N$. This gives a bound

$$2^{n-\nu}\mathbb{E}[f_z].(n-\nu)^{-\frac{1}{30\rho}+1}. \tag{5.12}$$

Summation of (5.12) over $z < 2^\nu$ implies that

$$(5.11) < Cn^{-\frac{1}{30\rho}+1} N 2^{-r}$$

so that (5.8) holds in general.

Summarizing, we have proved that

$$\sum_{\substack{q<B \\ q \text{ sf, odd}}} \kappa(q) < Cn^{-\frac{1}{20\rho}+1} \tag{5.13}$$

and also

$$\sum_{\substack{q_0<B, q_0 \text{ sf, odd} \\ (q_0,q_1)=1}} \alpha(q_1, q_0) < Cn^{-\frac{1}{30\rho}+1}. \tag{5.14}$$

6. Conclusion

Recalling (4.20), (4.23), (4.29), (4.34) and inserting the estimates (5.13), (5.14), we find that

(6.1)
$$\sum_{x<N} \Lambda(x) f(x) = 2\mathbb{E}[f]N + N\mathbb{E}[f] O(n^{-\frac{1}{20\rho}+3} + n^2 e^{-\sqrt{n}} + 2^r B^{-1} + n^3 N^{-\frac{1}{2}\eta_*} + n^5 B^{-1} + n^{-\frac{1}{30\rho}+3} B^{48\eta_*}).$$

Recall also conditions (3.9), (4.11) on B, i.e. $B = N^u$ for some sufficiently small $u > 0$.

It remains to choose $\eta_* \sim \frac{\log n}{n}$ appropriately and let ρ be small enough to conclude the Theorem.

References

[Bom] E. Bombieri, *Le Grand Crible dans la Théorie Analytique des Nombres*, Astérisque **18** (1987).

[B1] J. Bourgain, *Prescribing the binary digits of primes*, Israel J. Math. **194** (2013), 935–955.

[B2] J. Bourgain, *Monotone Boolean functions capture their primes*, J. Anal. Math., to appear.

[D] H. Davenport, *Multiplicative Number Theory*, Springer-Verlag, New York–Berlin, 1980.

[Ga] P. X. Gallagher, *Primes in progressions to prime-power modulus*, Invent. Math. **16** (1972), 191–201.

[I] H. Iwaniec, *On zeros of Dirichlet L-series*, Invent. Math. **23** (1974), 97–104.

[I-K] H. Iwaniec and E. Kowalski, *Analytic Number Theory*, American Mathematical Society, Providence, RI, 2004.

[H-K] G. Harman and I. Katai, *Primes with preassigned digits. II*, Acta Arith. **133** (2008), 171–184.

[I] A. E. Ingham, *A note on Fourier transforms*, J. London Math. Soc. **9** (1934), 29–32.