ACTA ARITHMETICA
133.2 (2008)

Primes with preassigned digits II

by

Glyn Harman (Egham) and Imre Kátai (Budapest)

Dedicated to Professor W. M. Schmidt on his 75th birthday

1. Introduction. The purpose of this paper is to show that significant improvements can be made on the results given in [5] using the ideas contained in the second author's paper [8]. Indeed, the results stated in [8] anticipated Wolke's conjecture in [14] by some twenty years. The first named author was unaware of this work until it was pointed out to him by Cécile Dartyge. There are some oversights in the proofs of the results in [8] which mean that the full strength of the theorems in that paper was not established at the time. However, even after correcting those proofs, what remains is strong enough to improve the work in [5]. In this paper we shall give improved versions of those proofs which establish substantially stronger results.

We recall the history of this problem. In 1951 Sierpiński [11] began the investigation of prime numbers with preassigned digits. In [12] he showed that in any given base g and with given a, b with $1 \le a \le g-1$, $\gcd(b, g) = 1$, $1 \le b \le g-1$, one can find infinitely many primes p having a as its first digit and b as its last. As was shown in [5], these results are elementary deductions from deeper results on the distribution of primes. Wolke [14] considered the more difficult case where one preassigns at most two digits *anywhere* in the expansion of an integer with k digits, and obtained an asymptotic formula (valid for $k \to \infty$) in this case.

We will use here the notation of [8] to state the problem formally rather than what was used in [14, 5]. We suppose that an integer n is written in the standard form to base q:

$$n = \sum_{j=0}^{\infty} a_j(n) q^j, \qquad 0 \le a_j(n) \le q-1.$$

2000 *Mathematics Subject Classification*: Primary 11N05.
Key words and phrases: distribution of primes.

Given a set of integers $\{b_l : 1 \le l \le r\}$ with $0 \le b_l < q$, and a set of non-negative integers $\{j_1, \dots, j_r\}$, we then write

$$\Pi\left(x \,\middle|\, \begin{matrix} j_1, \dots, j_r \\ b_1, \dots, b_r \end{matrix}\right) = |\{p \le x : a_{j_l}(p) = b_l,\ 1 \le l \le r\}|.$$

Our aim is to give an asymptotic formula for this expression when $0 \le j_1 < j_2 < \cdots < j_r \le k-1$. Henceforth we suppose that the leftmost digit is non-zero, that is, $j_r = k-1 \Rightarrow b_r \ne 0$ with $q^{k-1} \le x \le q^k - 1$. We write $\mathbf{b} = (b_1, \dots, b_r)$ and $\mathbf{j} = (j_1, \dots, j_r)$. We may then abbreviate our notation to

$$\Pi\left(x \,\middle|\, \begin{matrix} j_1, \dots, j_r \\ b_1, \dots, b_r \end{matrix}\right) = \Pi\left(x \,\middle|\, \begin{matrix} \mathbf{j} \\ \mathbf{b} \end{matrix}\right).$$

Wolke conjectured that in the case $x = q^k - 1$ we have

$$\Pi\left(x \,\middle|\, \begin{matrix} \mathbf{j} \\ \mathbf{b} \end{matrix}\right) \sim \mathrm{Li}(x) f(\mathbf{b}, \mathbf{j}). \tag{1}$$

Here

$$\mathrm{Li}(x) = \int_2^x \frac{1}{\log u}\, du \quad \text{and} \quad f(\mathbf{b}, \mathbf{j}) = \begin{cases} q^{-r} & \text{if } j_1 > 0, \\ 0 & \text{if } j_1 = 0, (b_1, q) > 1, \\ q^{1-r}\phi(q)^{-1} & \text{if } j_1 = 0,\ (b_1, q) = 1. \end{cases}$$

In the above, $\phi(n)$ denotes, as usual, Euler's totient function. Wolke gave a proof of (1) for $r = 1$ and $r = 2$, but for general r his proof was dependent on the Generalized Riemann Hypothesis. To be precise, he established a result equivalent to the following.

THEOREM 1 (Wolke). *Assume the Riemann Hypothesis for the L-functions with characters (mod q^k). Then, for any $\varepsilon \in (0,1)$, $k \ge k_0(\varepsilon)$, and all possible $\mathbf{b}, \mathbf{j}$ with $1 \le r \le (1-\varepsilon)k^{1/2}$, we have*

$$\Pi\left(x \,\middle|\, \begin{matrix} \mathbf{j} \\ \mathbf{b} \end{matrix}\right) = \mathrm{Li}(x) f(\mathbf{b}, \mathbf{j}) + O\left(\frac{x}{q^r (\log x)^2}\right)$$

when $x = q^k - 1$.

In [5] the first named author proved the following result.

THEOREM 2 (Harman). *Given $r \ge 1$ there exists $k_0(r)$ such that for $k \ge k_0$ we have*

$$\Pi\left(x \,\middle|\, \begin{matrix} \mathbf{j} \\ \mathbf{b} \end{matrix}\right) \gg \mathrm{Li}(x) f(\mathbf{b}, \mathbf{j}) \tag{2}$$

when $f(\mathbf{b}, \mathbf{j}) \ne 0$ and $x = q^k - 1$.

It is stated at the end of that paper that (2) can be improved to (1) with more effort. The method used in [5] only gave a poor relation between k_0 and r, and the results were ineffective owing to an appeal to Siegel's theorem.

In [8] the following result was claimed by the second author (we alter the terminology there slightly).

THEOREM 3 (Kátai). *Let* $q^{k-1} < x < q^k$, $1 \le r < \sqrt{k}$, $0 = j_1 < j_2 < \cdots < j_r \le k$, $(b_1, q) = 1$. *Then*

$$\Pi\left(x \,\middle|\, \begin{matrix}\mathbf{j}\\ \mathbf{b}\end{matrix}\right) = \frac{\mathrm{Li}(x)}{q^{r-1}\phi(q)} + O\left(\frac{\mathrm{Li}(x)}{q^r}\left(\exp(-d(\log x)^{1/2}) + (\log x)^4 (q^{j_r}/x)^{1/2}\right)\right) \tag{3}$$

with a suitable positive constant d, *uniformly in* $r, \mathbf{j}, \mathbf{b}$. *Moreover, if* $2^r < k^{1/5}$ *then we have*

$$\Pi\left(x \,\middle|\, \begin{matrix}\mathbf{j}\\ \mathbf{b}\end{matrix}\right) = \frac{q}{\phi(q)\log x} A\left(x \,\middle|\, \begin{matrix}\mathbf{j}\\ \mathbf{b}\end{matrix}\right) + O\left(\frac{x}{q^r}(\log x)^{9/20-2}\right). \tag{4}$$

Here

$$A\left(x \,\middle|\, \begin{matrix}\mathbf{j}\\ \mathbf{b}\end{matrix}\right) = |\{n \in \mathbb{N} : a_{j_l} = b_l,\ 1 \le l \le r\}|.$$

This result would establish Wolke's conjecture with $k > 2^{5r}$ for all large r. Unfortunately, as indicated above, there are some oversights in the proof. Upon fixing these we find that (3) becomes

$$\Pi\left(x \,\middle|\, \begin{matrix}\mathbf{j}\\ \mathbf{b}\end{matrix}\right) = \frac{\mathrm{Li}(x)}{q^{r-1}\phi(q)} + O(x((\log x)^{-C} + (\log x)^4 (q^{j_r}/x)^{1/2})),$$

so long as

$$r < \frac{\log\log x}{\log(C+1) + \log\log\log x}.$$

The quality of (4) is likewise considerably diminished (note that (3) is used in the proof of (4) in [8]), so that

$$r < C \log\log\log x$$

is required to obtain an asymptotic formula. Arguing as in [2] one can take $r < C\log\log x$ for any fixed C, so long as

$$k^{1/3} \le j_2 < j_3 < \cdots < j_r \le k - k^{1/3}.$$

Before proceeding further we should remark the significance of the size of the error terms and the positions of the prescribed digits. In particular, if the leftmost prescribed digit is near the beginning of the number's expansion to base q (that is, j_r is near k), then one cannot hope to get an error term that is smaller than $(\log x)^{-1}$ times the main term. This is evident upon considering the probability that a number n is prime, namely $(\log n)^{-1}$. This can change in size by a factor $1 + O(1/\log x)$ as the leftmost prescribed digit is altered when j_r is near to k in size. It is therefore reasonable to

change our prime counting function to

$$\Psi\left(x \,\middle|\, \begin{matrix}\mathbf{j}\\ \mathbf{b}\end{matrix}\right) = \Psi\left(x \,\middle|\, \begin{matrix}j_1,\dots,j_r\\ b_1,\dots,b_r\end{matrix}\right) = \sum_p \log p,$$

where the summation is over

$$p \in \{p \le x : a_{j_l}(p) = b_l,\ 1 \le l \le r\}.$$

Indeed, at the expense of an error $O(x^{1/2})$ we may replace $\log p$ by $\Lambda(n)$, the von Mangoldt function. Our main result is then as follows.

THEOREM 4. *Let $\varepsilon > 0$, $1 \le r < \sqrt{k}$, $0 = j_1 < j_2 < \cdots < j_r \le k-1$, $(b_1, q) = 1$, $x > \exp(q^2)$. Suppose that $0 < \Delta < 1/(2q)$. Then there exists $\delta(\varepsilon) > 0$ such that*

$$\Psi\left(x \,\middle|\, \begin{matrix}\mathbf{j}\\ \mathbf{b}\end{matrix}\right) = \frac{q}{\phi(q)} A\left(x \,\middle|\, \begin{matrix}\mathbf{j}\\ \mathbf{b}\end{matrix}\right) + O\left(x\Delta r\,\frac{q}{\phi(q)}\right) + O(x(\log x)^3(3-\log\Delta)^r(\xi + \Delta^{-1}q^{-T/2})). \tag{5}$$

Here

$$\xi = \exp(-\delta(\varepsilon)(\log x)^{4/7-\varepsilon}), \qquad T = \max_{1\le t\le r}(j_{t+1} - j_t),$$

with the convention that $j_{r+1} = k$. All implied constants are absolute and effectively computable.

COROLLARY. *Suppose that $1 \le r < C\sqrt{k}/\log k$, $0 = j_1 < j_2 < \cdots < j_r \le k-1$, $(b_1, q) = 1$, $x > \exp(q^2)$. Then*

$$\Psi\left(x \,\middle|\, \begin{matrix}\mathbf{j}\\ \mathbf{b}\end{matrix}\right) = \frac{q}{\phi(q)}\, A\left(x \,\middle|\, \begin{matrix}\mathbf{j}\\ \mathbf{b}\end{matrix}\right) + O\left(\frac{x}{q^{r-1}\phi(q)}\exp(-(\log x)^{1/2})\right). \tag{6}$$

In particular, if $x = q^k - 1$ we have

$$\Psi\left(x \,\middle|\, \begin{matrix}\mathbf{j}\\ \mathbf{b}\end{matrix}\right) = \frac{x}{q^{r-1}\phi(q)}\,(1 + O(\exp(-(\log x)^{1/2}))). \tag{7}$$

The implied constant here is effectively computable and depends only on q and C.

The corollary follows quickly from the theorem with $\varepsilon = 10^{-2}$, say, upon putting $\Delta = (rq^r)^{-1}\exp(-(\log x)^{1/2})$ and noting that $T > C^{-1}\sqrt{k}\log k$.

COROLLARY. *Suppose that $1 \le r < C\sqrt{k}/\log k$, $0 \le j_1 < j_2 < \cdots < j_r \le k-1$, $x > \exp(q^2)$. Then*

$$\Pi\left(x \,\middle|\, \begin{matrix}\mathbf{j}\\ \mathbf{b}\end{matrix}\right) = \frac{q^r f(\mathbf{b},\mathbf{j})}{\log x}\, A\left(x \,\middle|\, \begin{matrix}\mathbf{j}\\ \mathbf{b}\end{matrix}\right) + O\left(\frac{x\log\log x}{\phi(q)\, q^{r-1}(\log x)^2}\right). \tag{8}$$

In particular, if $x = q^k - 1$ we have, in the case $f(\mathbf{b},\mathbf{j}) \ne 0$,

$$\Pi\left(x \,\middle|\, \begin{matrix}\mathbf{j}\\ \mathbf{b}\end{matrix}\right) = f(\mathbf{b},\mathbf{j})\,\frac{x}{\log x}\left(1 + O\left(\frac{\log\log x}{\log x}\right)\right). \tag{9}$$

The implied constant here is effectively computable and depends only on C and q.

We note that this is just a factor $\log\log x$ weaker than Wolke's result on the Generalized Riemann Hypothesis. We have already noted that the error term cannot be better than $(\log x)^{-1}$ times the main term (even with $x/\log x$ replaced by $\mathrm{Li}(x)$), so we have come very close to the best possible result.

Proof. Clearly, we can suppose that $j_1 = 0$, since otherwise we can replace r by $r+1$ and sum over the $\phi(q)$ possible values for b_1. Write

$$z = \frac{x}{(\log x)^2}.$$

Then

$$\log n = \log x\left(1 + O\left(\frac{\log(x/z)}{\log x}\right)\right)$$

for $z \le n \le x$. This clearly gives

$$\Pi\left(x \,\middle|\, \begin{matrix}\mathbf{j}\\ \mathbf{b}\end{matrix}\right) = \frac{1}{\log x}\,\Psi\left(x \,\middle|\, \begin{matrix}\mathbf{j}\\ \mathbf{b}\end{matrix}\right)\left(1 + O\left(\frac{\log(x/z)}{\log x}\right)\right) + O\left(\Pi\left(z \,\middle|\, \begin{matrix}\mathbf{j}\\ \mathbf{b}\end{matrix}\right)\right). \tag{10}$$

The last term on the right side of (10) is trivially

$$\ll \frac{z}{q^r}.$$

The result then follows from the previous corollary. ∎

The basic idea of the proof is the same as in [8, 5], namely we find the hth digit of n by considering

$$\left\{\frac{n}{q^h}\right\},$$

where $\{\cdot\}$ denotes fractional part, which reduces the problem to Fourier analysis and the estimation of exponential sums. We then have a "major arc" technique for fractions with small denominator, a "minor arc" technique for fractions with large denominator, but not too near x in size. Finally, we have a method using primes in short intervals for exponential sums that cannot be considered by other means. The reader should note that all constants will be effective here. We are able to avoid any appeal to Siegel's theorem in view of the special nature of the moduli (equivalently, fraction denominator) under consideration and our demand that $x > \exp(q^2)$.

As far as our method is concerned, the worst case scenario is to have the prescribed digits equally spaced out among the digits of our potential primes. As remarked in [5] one can preassign up to $0.472k$ digits so long as they appear in at most two blocks, one at the beginning, the other at the end of the expansion. A long block in the middle can also be considered by using

techniques from Diophantine approximation with primes [13], leading to the possibility of preassigning a block of length nearly $k/4$ if one is looking for an asymptotic formula, and even longer if one only desires a lower bound (using [6], for example). We note the interest of some authors in prescribing the digits in base 2 of products of two primes [10].

2. Preliminary lemmas. First we state the lemma converting the problem from Diophantine approximation to exponential sum form.

LEMMA 1. *Let* $0 < \Delta < 1/(2q), 0 \le b \le q-1$. *Then there is a function* $\chi_b(x)$ *such that*

$$0 \le \chi_b(x) \le 1,$$

$$\chi_b(x) = \sum_{h=-\infty}^{\infty} c_h e(x),$$

$$c_0 = \frac{1}{q}, \quad c_{jq} = 0 \quad \textit{when } j \ne 0,$$

$$|c_m| \le \min\left(\frac{1}{q}, \frac{1}{\pi m}, \frac{1}{\Delta \pi^2 m^2}\right),$$

$$\chi_b(x) = \begin{cases} 1 & \textit{if } \Delta < \{x - b/q\} < 1/q - \Delta, \\ 0 & \textit{if } 1/q + \Delta < \{x - b/q\} < 1 - \Delta. \end{cases}$$

Proof. This is a familiar type of result, and is the corrected form of what appears in [8, p. 344]. ∎

We require the following result to deal with the cases of fractional parts falling in the narrow region where our test function lies strictly between 0 and 1. We will need to prove results for both the "prime counting" and the "all integer counting" cases.

LEMMA 2. *Let* $x^{-1/2} < \xi < 1/(2q)$, $0 \le b \le q$, $q^j \le x$, $j \ge 2$. *Write*

$$\mathcal{A} = \{n \le x : b/q - \xi < \{n/q^j\} < b/q + \xi\}.$$

Then

$$\sum_{p \in \mathcal{A}} \log p \le \log q + 16 \xi x \frac{q}{\phi(q)}. \tag{11}$$

Also, if $(a, q) = 1$ *we have*

$$\sum_{\substack{n \in \mathcal{A} \\ n \equiv a \,(\mathrm{mod}\, q)}} 1 \le 4x\xi. \tag{12}$$

Proof. For (11) we need to count solutions to

$$-\xi q^j < p - q^j m - b q^{j-1} < \xi q^j, \quad 0 \le m < x q^{-j}.$$

First we note that if $\xi q^j < 1$ there is at most one solution to the above (and that only when $j = 2$ and q is a prime). We recall the Brun–Titchmarsh inequality in the form given by Montgomery and Vaughan [9], namely that

$$\sum_{\substack{z-y<p<z\\ p\equiv a\,(\mathrm{mod}\,t)}} 1 \le \frac{2y}{\phi(t)\log(y/t)}.$$

If $q^j \le x^{3/4}$ then we apply this at most $2\xi q^j$ times with $t = q^j$, $z = y = x$, to give a bound

$$< 4\xi q^j \frac{x}{\phi(q^j)\log(x/q^j)} \le 16x\xi \frac{q}{\phi(q)}.$$

If, on the other hand, $q^j > x^{3/4}$ then we can apply the inequality $\le xq^{-j}$ times with $t = 1$, $y = 2\xi q^j$ to give a bound

$$< 4xq^{-j} \frac{x}{\log(\xi q^j)} \le 16x\xi.$$

The bound (11) then follows.

To deal with the case (12) we note there are no solutions if $\xi q^j < 1$, and we can simply count the number of solutions to be no more than $2\xi q^j(xq^{-j}+1)$ otherwise. ∎

Our "minor arc" estimate for dealing with the sums which arise from Lemma 1 is as follows. This also covers the major arcs estimate for the unrestricted sum.

LEMMA 3. *Suppose that* $x > 2$, $\gcd(u,v) = \gcd(b,d) = 1$, $\gcd(d,v) = h$. *Then, if* $|\alpha - u/v| < 2/x$, *we have*

$$\sum_{\substack{n<x\\ n\equiv b\,(\mathrm{mod}\,d)}} \Lambda(n)e(n\alpha) \ll x(\log x)^3\left(\frac{h}{dv^{1/2}} + \left(\frac{v}{xh}\right)^{1/2} + \frac{1}{x^{1/5}d^{2/5}}\right). \tag{13}$$

Also, if $h < v$ *and* $d|\alpha - u/v| < 1/(2v)$, *then*

$$\sum_{\substack{n<x\\ n\equiv b\,(\mathrm{mod}\,d)}} e(n\alpha) \ll v. \tag{14}$$

Proof. The bound (13) is given in [1]. The estimate (14) is elementary. ∎

Now, when dealing with primes in arithmetic progressions (which is relevant to the problem when there are prescribed digits near the right hand end of the expansion), it is usual for difficulties to arise from possible Siegel zeros. The next result shows that the special nature of the moduli under consideration precludes the existence of such nuisances.

LEMMA 4. *There is an absolute constant* $c > 0$ *such that if* $Q \mid q^m$ *and* $x > \exp(q^2)$, *then no zero of an L-function for a real character* (*mod* Q)

has a real zero ϱ with

$$\varrho > 1 - \frac{c}{(\log x)^{3/4}}.$$

Proof. We recall that the only real characters $\pmod Q$ are induced by primitive characters to the modulus g where $g = 2^b h$ with $b = 0$, 2, or 3 (and if Q is odd then $b \equiv 0$), and h is square-free with $h \mid Q$. Hence $g \le 4q$. If $L(s,\chi)$ is an L-function with a real character $\chi \pmod Q$ then its zeros in the critical strip are identical to the zeros of the L-function for the character $\pmod g$ inducing it. By [3, Chapter 14] any real zero ϱ of this L-function has

$$\varrho < 1 - \frac{c'}{g^{1/2}(\log g)^2} < 1 - \frac{c}{(\log x)^{3/4}}.$$

This completes the proof. ∎

Next we need a zero-free region given by Iwaniec that generalizes the work of Gallagher [4].

LEMMA 5. *Let $Q \mid q^m$, $T > 0$. There exists at most one character χ $(\mathrm{mod}\ Q)$ such that $L(s,\chi)$ has a zero ϱ satisfying*

$$\operatorname{Re}\varrho > 1 - \theta, \qquad |\operatorname{Im}\theta| < T,$$

where

$$\theta^{-1} = 4 \cdot 10^4 (\log q + (\ell(\log \ell))^{3/4}), \qquad \ell = \log(Q(T+3)).$$

If there does exist such a character, then it is real and the corresponding zero is real.

Proof. This follows immediately from [7, Theorem 2]. ∎

We can now combine the last two results to obtain the following.

LEMMA 6. *Given $\varepsilon > 0$ there is a constant $c = c(\varepsilon) > 0$ such that for $x > \exp(q^2)$, $Q \mid q^m$ for some $m \in \mathbb{N}$, and all a with $(a,q) = 1$, we have*

$$\sum_{\substack{n \le x \\ n \equiv a \,(\mathrm{mod}\, Q)}} \Lambda(n) = \frac{x}{\phi(Q)} + O(x \exp(-c(\log x)^{4/7-\varepsilon})), \tag{15}$$

uniformly for $Q \le \exp(c(\log x)^{4/7-\varepsilon})$.

Proof. This follows from the work in Chapters 14 and 20 of [3] with the zero-free regions given above. ∎

We can use this to establish the following "major arc" result, which is an improvement of [8, Lemma 2]. Henceforth in this paper we write

$$E = \exp\left(\tfrac{1}{2} c(\log x)^{4/7-\varepsilon}\right), \qquad \eta = E^{-1},$$

where $c = c(\varepsilon)$ is the constant in (15).

LEMMA 7. *Let $H \mid q^m$ for some $m \in \mathbb{N}$, $H \nmid q$, $(A,H)=1$, $1 \le b \le q-1$, $(b,q)=1$. Let $\varepsilon > 0$ and $x > \exp(q^2)$. Then, when $H \le E$, we have*

$$\sum_{\substack{n \le x \\ n \equiv b \,(\mathrm{mod}\, q)}} \Lambda(n) e\left(\frac{nA}{H}\right) \ll x\eta. \tag{16}$$

Proof. We follow the proof on pages 345–346 of [8], but we can make use of our improved results on the distribution of primes in arithmetic progressions given above. Write $d = \mathrm{lcm}[H,q]$. The left hand side of (16) is

$$\sum_{\substack{u \,(\mathrm{mod}\, d) \\ u \equiv b \,(\mathrm{mod}\, q)}} \Psi(x,d,u) e\left(\frac{uA}{H}\right) = S(x) \quad \text{say.}$$

Here

$$\Psi(x,d,u) = \sum_{\substack{n \le x \\ n \equiv u \,(\mathrm{mod}\, d)}} \Lambda(n) = \frac{x}{\phi(d)} + O(x\eta^2),$$

by (15). Thus

$$S(x) = \frac{x}{\phi(d)} \sum_{\substack{u \,(\mathrm{mod}\, d) \\ u \equiv b \,(\mathrm{mod}\, q)}} e\left(\frac{uA}{H}\right) + O(xH\eta^2).$$

As shown in [8], it is elementary that

$$\sum_{\substack{u \,(\mathrm{mod}\, d) \\ u \equiv b \,(\mathrm{mod}\, q)}} e\left(\frac{uA}{H}\right) = 0.$$

We thus have

$$S(x) = O(xH\eta^2) = O(x\eta)$$

as desired. ∎

Our final lemma on exponential sums is in fact a reinterpretation of a result on primes in short intervals. It provides an important link between the sum weighted by $\Lambda(n)$ and the unweighted sum.

LEMMA 8. *Suppose that $x > \exp(q^2)$. Let $(b,q)=1$ and $\alpha = A/H$, where $1 < H \mid q^k$, $(A,H)=1$ and*

$$\|\alpha\| < Ex^{-1}.$$

Then

$$\sum_{\substack{n \le x \\ n \equiv b \,(\mathrm{mod}\, q)}} \Lambda(n) e(n\alpha) = \frac{q}{\phi(q)} \sum_{\substack{n \le x \\ n \equiv b \,(\mathrm{mod}\, q)}} e(n\alpha) + O(x\eta). \tag{17}$$

Proof. We note that by Lemma 6, for $y \le x$, we have

$$\Psi(y,q,b) = \frac{y}{\phi(q)} + O(x\eta^2).$$

It is elementary that

$$\sum_{\substack{n<y \\ n\equiv b\,(\mathrm{mod}\,q)}} 1 = \frac{y}{q} + O(q)$$

and, for any m,

$$|e(m\alpha) - e((m+1)\alpha)| \le 2\pi\|\alpha\| \ll Ex^{-1}.$$

Hence, by partial summation,

$$\begin{aligned}
&\sum_{\substack{n\le x \\ n\equiv b\,(\mathrm{mod}\,q)}} (\Lambda(n) - q/\phi(q))e(n\alpha) \\
&\quad = \sum_{m\le x} (e(m\alpha) - e((m+1)\alpha)) \sum_{\substack{n\le m \\ n\equiv b\,(\mathrm{mod}\,q)}} (\Lambda(n) - q/\phi(q)) \\
&\qquad + e((x+1)\alpha) \sum_{\substack{n\le x \\ n\equiv b\,(\mathrm{mod}\,q)}} (\Lambda(n) - q/\phi(q)) \\
&\quad \ll \sum_{m\le x} |e(m\alpha) - e((m+1)\alpha)|(x\eta^2 + q) + x\eta^2 + q \\
&\quad \ll x(Ex^{-1})(x\eta^2 + q) + x\eta^2 + q \ll x\eta.
\end{aligned}$$

This completes the proof. ∎

3. Proof of Theorem 4. First we note that (5) is trivial for $r\Delta > 1$ or $\Delta < x^{-1/2}$, so we can assume henceforth that

$$r^{-1} \ge \Delta \ge x^{-1/2}.$$

Let $\chi^*(n)$ denote the characteristic function of the set

$$\{n \le x : a_{j_l}(n) = b_l,\ 2 \le l \le r\}.$$

Write $\Theta(n) = \Theta(n,q) = \Lambda(n) - q/\phi(q)$. To prove (5) we need only show that

$$\begin{aligned}
(18)\qquad &\sum_{\substack{n\le x \\ n\equiv b_1\,(\mathrm{mod}\,q)}} \chi^*(n)\Theta(n) \\
&\qquad \ll x\Delta r\,\frac{q}{\phi(q)} + x(\log x)^3(3 - \log\Delta)^r(\xi + \Delta^{-1}q^{-T/2}).
\end{aligned}$$

Now, using the notation of Lemma 1, put

$$\chi(n) = \prod_{l=2}^{r} \chi_{b_l}\left(\frac{n}{q^{j_l}}\right).$$

Then we have $\chi^*(n) = \chi(n)$, unless $\{nq^{-j_l}\}$ falls into the region where χ_{b_l} is strictly between zero and one. By Lemma 2 the error this introduces is $O(\Delta rxq/\phi(q))$.

Now write

$$\chi_{b_l}(x) = f_l(x) + g_l(x),$$

where

$$g_l(x) = \sum_{|h|\Delta^2>1} c_h e(x).$$

Then write

$$\prod_{l=2}^{r} \chi_{b_l}\left(\frac{n}{q^{j_l}}\right) = F(n) + G(n), \quad \text{where} \quad F(n) = \prod_{l=2}^{r} f_l\left(\frac{n}{q^{j_l}}\right).$$

Now, for every l, we have

$$|g_l(x)| \le 2 \sum_{h\Delta^2>1} \frac{1}{\pi^2 h^2 \Delta} < \Delta.$$

Hence $|f_l(x) - \chi_{b_l}(x)| < \Delta$, and so

$$G(n) \le (1+\Delta)^{r-1} - 1 < 2r\Delta.$$

We then have

$$\sum_{n\le x} \Theta(n)G(n) = O(r\Delta xq/\phi(q)),$$

which is a suitable error.

We are then left to estimate

$$\sum_{\substack{n\equiv b_1 \,(\mathrm{mod}\, q)\\ n\le x}} \Theta(n)F(n) = \frac{1}{q^{r-1}} \sum_{\substack{n\equiv b_1 \,(\mathrm{mod}\, q)\\ n\le x}} \Theta(n) + O(X),$$

where

$$X = \sum_{0<|\mathbf{h}|<\Delta^{-2}} \prod_{l=2}^{r} \min\left(\frac{1}{q}, \frac{1}{\pi|h_l|}, \frac{1}{\Delta\pi^2 h_l^2}\right) \Big| \sum_{\substack{n\equiv b_1 \,(\mathrm{mod}\, q)\\ n\le x}} \Theta(n) e(\mu(\mathbf{h})n)\Big|.$$

Here $\mathbf{h} = (h_2, \dots, h_r)$ and

$$|\mathbf{h}| = \max_l |h_l|, \qquad \mu(\mathbf{h}) = \left(\frac{h_2}{q^{j_2}} + \cdots + \frac{h_r}{q^{l_r}}\right).$$

We have already remarked that

$$\sum_{\substack{n\equiv b_1 \,(\mathrm{mod}\, q)\\ n\le x}} \Lambda(n) = \frac{x}{\phi(q)} + O(x\eta^2) = \sum_{\substack{n\equiv b_1 \,(\mathrm{mod}\, q)\\ n\le x}} \frac{q}{\phi(q)} + O(x\eta^2).$$

It therefore remains to bound X. First we note that

$$X \le \left(\frac{1}{q} + 2 \sum_{0<h<\Delta^{-2}} \min\left(\frac{1}{q}, \frac{1}{\pi h}, \frac{1}{\Delta \pi^2 h^2}\right)\right)^r S \le (3 - \log \Delta)^r S$$

with

$$S = \max_{0<|\mathbf{h}|\le\Delta^{-2}}{}^{*} \Big| \sum_{\substack{n \equiv b_1 \pmod q \\ n \le x}} \Theta(n) e(\mu(\mathbf{h}) n) \Big|.$$

Here the $*$ indicates that the maximum excludes those $\mathbf{h}$ for which any non-zero $h_l \equiv 0 \pmod q$ in view of the properties of the Fourier coefficients given in Lemma 1. It now remains to estimate this exponential sum for all the values of $\mu(\mathbf{h})$ that can arise.

We know that $\mu(\mathbf{h})$ is a rational with denominator $Q \mid q^k$. It is also important that in lowest form the denominator is not 1. Indeed, writing this fraction as A/H we must have $H \nmid q$. This vital fact follows from our knowledge that no $h_l \equiv 0 \pmod q$ and $j_2 \ge 2$. The first-named author used a different test function in [5] and this caused problems with the possibility that the expression corresponding to $\mu(\mathbf{h})$ could have been an integer (and so no saving would have been possible on the exponential sum). Paraphrasing the argument on page 345 of [8], if y is the largest integer with $h_{j_y} \ne 0$ then

$$\frac{A}{H} q^{j_y} = h_{j_y} + h_{j_{y-1}} q^{j_y - j_{y-1}} + \cdots .$$

Since $j_y \ge 2$, if $H \mid q$ then the left hand side is divisible by q and this forces $h_{j_y} \equiv 0 \pmod q$, and this value has been excluded since the corresponding Fourier coefficient is zero. We can then apply Lemma 7 and (14) immediately when $H \le E$ to obtain a satisfactory estimate.

Now suppose that $E < H < x\eta$. Our minor arc estimate, Lemma 3, then provides us with a satisfactory upper bound.

We are then left with the case when $H \ge x\eta$. If

$$\|A/H\| < Ex^{-1}, \tag{19}$$

we can use Lemma 8 to bound this sum. The reader should note that both sides of (17) can be significantly large when $x \ne q^k - 1$ and so this part of the argument is a genuine "major arc" connecting the number of solutions weighted with $\Lambda(n)$ with those weighted with the constant $q/\phi(q)$.

If (19) fails then we need to approximate A/H by a rational with smaller denominator. The principle we now follow is very simple, but the details might look unclear, so we outline the basic idea first. The fraction A/H is a finite length decimal to base q, which contains long blocks of $q-1$ and/or 0. This is ensured by the relative sizes of k and r and, in particular, by the size of T, along with our restriction $|h_j| < \Delta^{-2}$. The blocks then have a length of size at least $q^T \Delta^2$. It follows that A/H cannot be approximated

very well by a fraction with small denominator (that is, less than around $\Delta^2 q^T$), unless that denominator divides q^{k-1}. We are then able to appeal either to Lemma 3 or 7 as above.

To fulfil this programme, write

$$W = \min\bigl(E^{1/2}, \tfrac{1}{8}\Delta^2 q^T\bigr), \qquad V = x/W.$$

By Dirichlet's theorem in Diophantine approximation, we can approximate A/H with a fraction u/v with $v > 1$, $(u, v) = 1$, and

$$|Av/H - u| \le V^{-1}, \qquad v \le V.$$

If $v \ge W$ we can appeal to Lemma 3 again. Note that in this case $|a/H - u/v| < x^{-1}$ as required in the hypothesis of that lemma. If it were possible for $v < W$ then we would have

$$\frac{h_r}{q^{j_r}} + \cdots + \frac{h_2}{q^{j_2}} = \frac{u}{v} + \theta$$

with

$$v|\theta| \le x^{-1}E^{1/2}.$$

Now $q^{j_r}v\theta$ is a non-zero integer, say C, and

$$v(h_r + \cdots + h_2 q^{j_r - j_2}) = uq^{j_r} + C. \tag{20}$$

We can suppose that $T = j_{y+1} - j_y$ for some $y < r$ since the case $y = r$ instantly forces

$$H \le xq^{-T}$$

leading to a satisfactory bound from Lemma 3. Let $Q = q^{j_r - j_y}$ and consider (20) (mod Q):

$$v(h_r + \cdots + h_{y+1}q^{j_r - j_{y+1}}) \equiv C \pmod{Q}.$$

For the moment we will assume that

$$v(h_r + \cdots + h_{y+1}q^{j_r - j_{y+1}}) \ne C. \tag{21}$$

It then follows that

$$v|h_r + \cdots + h_{y+1}q^{j_r - j_{y+1}}| > \tfrac{1}{2}Q.$$

However, the left hand side of the above is

$$\le \tfrac{1}{8}\Delta^2 q^T(2\Delta^{-2}q^{j_r - j_{y+1}}) = \tfrac{1}{4}Q.$$

This contradiction shows that we must have equality and not inequality in (21). It then follows that

$$v(h_y q^{j_r - j_y} + \cdots + h_2 q^{j_r - j_2}) = uq^{j_r},$$

and so

$$v(h_y + \cdots + h_2 q^{j_y - j_2}) = uq^{j_y}.$$

We must then have $v \mid q^{j_y}$ and $v \nmid q$. By partial summation we obtain

$$\Big| \sum_{\substack{n \equiv b_1 \,(\mathrm{mod}\, q) \\ n \le x}} \Theta(n) e(\mu(\mathbf{h}) n) \Big| \le (1 + x|\theta|) \max_{w \le x} \Big| \sum_{\substack{n \equiv b_1 \,(\mathrm{mod}\, q) \\ n \le w}} \Theta(n) e\Big(\frac{nu}{v}\Big) \Big|.$$

We can then apply Lemma 7 and (14) to obtain the bound

$$\ll (1 + x|\theta|)\eta \ll x\eta^{1/2}.$$

This completes the discussion of all cases and so completes the proof. The reader will notice that the value of $\delta(\varepsilon)$ in (5) is $\frac{1}{8}c(\varepsilon)$, where $c(\varepsilon)$ originates in Lemma 6. (In the interests of clarity we have not optimized the relationship between these two quantities.)

References

[1] A. Balog and A. Perelli, *Exponential sums over primes in an arithmetic progression*, Proc. Amer. Math. Soc. 93 (1985), 578–582.

[2] N. L. Bassily and I. Kátai, *Distribution of the values of q-additive functions on polynomial sequences*, Acta Math. Hungar. 68 (1995), 353–361.

[3] H. Davenport, *Multiplicative Number Theory*, 2nd ed., Springer, New York, 1980.

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

[5] G. Harman, *Primes with preassigned digits*, Acta Arith. 125 (2006), 179–185.

[6] D. R. Heath-Brown and C. H. Jia, *The distribution of αp modulo one*, Proc. London Math. Soc. (3) 84 (2002), 79–104.

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

[8] I. Kátai, *Distribution of digits of primes in q-ary canonical form*, Acta Math. Hungar. 47 (1986), 341–359.

[9] H. L. Montgomery and R. C. Vaughan, *On the large sieve*, Mathematika 20 (1973), 119–134.

[10] I. E. Shparlinski, *On RSA moduli with prescribed bit patterns*, Des. Codes Cryptogr. 39 (2006), 113–122.

[11] W. Sierpiński, *Sur l'existence des nombres premiers avec une suite arbitraire de chiffres initiaux*, Matematiche (Catania) 6 (1951), 135–137.

[12] —, *Sur les nombres premiers ayant des chiffres initiaux et finals donnés*, Acta Arith. 5 (1959), 265–266.

[13] R. C. Vaughan, *On the distribution of αp modulo 1*, Mathematika 24 (1977), 135–141.

[14] D. Wolke, *Primes with preassigned digits*, Acta Arith. 119 (2005), 201–209.

Department of Mathematics
Royal Holloway, University of London
Egham, Surrey TW20 0EX, UK
E-mail: G.Harman@rhul.ac.uk

Department of Computer Algebra
Eötvös Loránd University
Pázmány Péter s. 1/C
H-1117 Budapest, Hungary
E-mail: katai@compalg.inf.elte.hu

Received on 14.11.2007
and in revised form on 14.2.2008 (5573)