



# On the minimal distance of a polynomial code

Peter Pal Pach, Csaba Szabo

## ► To cite this version:

Peter Pal Pach, Csaba Szabo. On the minimal distance of a polynomial code. Discrete Mathematics and Theoretical Computer Science, 2011, Vol. 13 no. 4 (4), pp.33–43. 10.46298/dmtcs.556 . hal-00990488

**HAL Id: hal-00990488**

**<https://inria.hal.science/hal-00990488>**

Submitted on 13 May 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# On the minimal distance of a polynomial code

Péter Pál Pach<sup>†</sup> and Csaba Szabó<sup>‡</sup>

*Eötvös Loránd University, Department of Algebra and Number Theory, Budapest, Hungary*

*received 28<sup>th</sup> April 2010, revised 16<sup>th</sup> March 2011, accepted 17<sup>th</sup> March 2011.*

For a polynomial  $f(x) \in \mathbb{Z}_2[x]$  it is natural to consider the near-ring code generated by the polynomials  $f \circ x, f \circ x^2, \dots, f \circ x^k$  as a vectorspace. It is a 19 year old conjecture of Günter Pilz that for the polynomial  $f(x) = x^n + x^{n-1} + \dots + x$  the minimal distance of this code is  $n$ .

The conjecture is equivalent to the following purely number theoretical problem. Let  $\underline{m} = \{1, 2, \dots, m\}$  and  $A \subset \mathbb{N}$  be an arbitrary finite subset of  $\mathbb{N}$ . Show that the number of products that occur odd many times in  $\underline{n} \cdot A$  is at least  $n$ . Pilz also formulated the conjecture for the special case when  $A = \underline{k}$ . We show that for  $A = \underline{k}$  the conjecture holds and that the minimal distance of the code is at least  $n/(\log n)^{0.223}$ .

While proving the case  $A = \underline{k}$  we use different number theoretical methods depending on the size of  $k$  (respect to  $n$ ). Furthermore, we apply several estimates on the distribution of primes.

**Keywords:** near-ring code, minimal distance, prime

## 1 Introduction

For two finite subsets of the positive integers,  $A$  and  $B$  let  $A * B = \{ab \mid a \in A, b \in B \text{ and } ab \text{ occurs odd many times in } A \cdot B\}$ . In other words, if  $A = \{a_1, \dots, a_k\}$ , then  $A * B = a_1 B \Delta \dots \Delta a_k B$ , where  $\Delta$  denotes the symmetric difference. For a positive integer  $m$  let  $\underline{m} = \{1, 2, \dots, m\}$ .

**Conjecture 1** *If  $n, k$  are positive integers, then  $|\underline{n} * \underline{k}| \geq n$ .*

For an arbitrary finite subset  $A \subset \mathbb{N}$  it was proved that  $|\underline{m} * A| \geq \pi(m) + 1$ , where  $\pi(x)$  is the prime counting function, and the following conjecture was formulated (Pilz (1992)):

**Conjecture 2** *Let  $n$  be a positive integer and  $K \subset \mathbb{N}$  be a finite set of integers. Then  $|\underline{n} * K| \geq n$ .*

These purely number theoretical problems originate in the theory of near-ring codes. A near-ring can be described as a ring, where the addition is not necessarily commutative and only one of the distributive laws is required. A typical example is the near-ring of polynomials, where the addition is the usual polynomial addition, and multiplication is the composition of the polynomials. In this example the addition

<sup>†</sup>Email: ppp24@cs.elte.hu

<sup>‡</sup>Email: csaba@cs.elte.hu

is commutative and only the right distributive law holds. Near-rings play an important role in combinatorics: They are used to construct block designs that give rise to efficient error correcting codes. For more information on these codes see Eggetsberger (2011), Pilz (1983) and Pilz (2011). A special and very interesting near-ring code is defined in the following way: Let  $f \in \mathbb{Z}_2[x]$  be a polynomial and  $C(f, k)$  the code generated (as a subspace) by the polynomials  $f = f \circ x, f \circ x^2, \dots, f \circ x^k$ . For  $f = x + x^2 + \dots + x^n$  a typical codeword is

$$\sum_{i \in K} f \circ x^i = \sum_{j \in K * \underline{n}} x^j,$$

where  $K$  is a finite subset of  $\underline{k}$ . As  $C(f, k)$  is a linear code, its minimal distance is equal to the minimal weight of any nonzero codeword. Hence the minimum distance of  $C(f, k)$  is the minimal value of  $|\underline{n} * K|$  for some  $K \subseteq \underline{k}$ .

In this paper we settle Conjecture 1, and prove that for arbitrary  $n \in \mathbb{N}$  and finite set  $K \subset \mathbb{N}$  we have  $|\underline{n} * K| \geq c \cdot \frac{n}{\log^{0.223} n}$  for some  $c > 0$ . Note that the minimal distance in  $C(f, k)$  depends heavily on  $f$ .

If, for example, we start with  $f(x) = x + x^2 + x^4 + \dots + x^{2^k}$ , then  $f \circ x + f \circ x^2 = x + x^{2^{k+1}}$ , hence the minimal distance of the corresponding code is 2.

The natural logarithm will be denoted by  $\log$  through the whole paper.

## 2 The general case

Let us denote by  $g(n)$  the minimal size of the set  $\underline{n} * K$ , where  $K$  is a finite subset of the positive integers. In Pilz (1992) it is proved that  $g(n) \geq \pi(n) + 1$ . In this section we improve this lower bound and prove that  $g(n) \geq c \cdot \frac{n}{\log^{0.223} n}$  for some  $c > 0$ . The proof is based on the following lemma:

**Proposition 1** *For every positive integer  $n$*

$$g(n) \geq \sum_{p \leq n} g(\lfloor n/p^{\alpha_p} \rfloor),$$

where the sum goes over the primes less than  $n$ , and  $\alpha_p$  is the largest integer such that  $p^{\alpha_p} \leq n$ .

**Proof:** Let  $p \leq n$  be a prime and  $K_p \subseteq K$  the subset of  $K$  containing the elements that are divisible by the largest power of  $p$  occurring as divisor of some element of  $K$  (possibly  $p^0 = 1$ ). Similarly, let  $\underline{n}_p \subseteq \underline{n}$  be the set of elements of  $\underline{n}$  that are divisible by  $p^{\alpha_p}$ . Note that  $\underline{n}_p$  is never empty. By the maximality of the exponents of  $p$  in  $K_p$  and  $\underline{n}_p$ , for any  $a \in \underline{n}_p$ ,  $b \in K_p$  and  $c \in \underline{n}$ ,  $d \in K$  if  $ab = cd$ , then  $c \in \underline{n}_p$  and  $d \in K_p$  hold. We prove that for  $p < q \leq n$  different primes  $\underline{n}_p \cdot K_p$  and  $\underline{n}_q \cdot K_q$  are disjoint. If for some  $a \in \underline{n}$  and  $b \in K$  we have  $ab \in \underline{n}_p \cdot K_p \cap \underline{n}_q \cdot K_q$ , then  $a \in \underline{n}_p \cap \underline{n}_q$ . Thus  $a = pqd'$ , and  $\bar{a} = p^2 d' < a$  is in  $\underline{n}$ . The exponent of  $p$  in  $\bar{a}$  is larger than the one in  $a$ , which is contradiction. Hence,  $\underline{n} * K$  contains the disjoint union of the sets  $\underline{n}_p \cdot K_p$  for  $p \leq n$ , so

$$|\underline{n} * K| \geq \sum_{p \leq n} |\underline{n}_p * K_p|. \quad (1)$$

As  $p^{\alpha_p} \leq n < p^{\alpha_p+1}$ , clearly,  $\underline{n}_p = \{p^{\alpha_p}, 2p^{\alpha_p}, \dots, \lfloor n/p^{\alpha_p} \rfloor p^{\alpha_p}\}$ , where  $\lfloor n/p^{\alpha_p} \rfloor < p$ . Dividing by  $p^{\alpha_p}$ , we obtain that  $|\underline{n}_p * K_p| = |\lfloor n/p^{\alpha_p} \rfloor * K_p|$ , thus by the definition of  $g$  we get

$$|\underline{n}_p * K_p| = |\lfloor n/p^{\alpha_p} \rfloor * K_p| \geq g(\lfloor n/p^{\alpha_p} \rfloor).$$

By (1) we have

$$g(n) \geq \sum_{p \leq n} g(\lfloor n/p^{\alpha_p} \rfloor),$$

and this is what we wanted to prove.  $\square$

**Theorem 2** For every  $\lambda > \lambda_0$  there exists a  $c = c(\lambda) > 0$  such that for every  $n > 1$

$$g(n) \geq c \cdot \frac{n}{\log^\lambda n},$$

where  $\lambda_0$  satisfies  $\int_0^1 \left(\frac{2}{y}\right)^{\lambda_0} \frac{1}{2-y} dy = 1$ . Note that  $\lambda_0 \sim 0.2223\dots$

**Proof:** Fix  $1 > \lambda > \lambda_0$ . We claim that there exists some  $c > 0$  such that the inequality

$$g(n) \geq c \cdot \frac{n}{\log^\lambda n} \quad (2)$$

holds for every  $n > 1$ . The proof is by induction on  $n$ . First we discuss the induction step. Assume that (2) holds for  $n < m$ . Now, we show that it holds for  $n = m$ , as well. The value of  $c$  will be chosen later. By Proposition 1 and the induction hypothesis:

$$\begin{aligned} g(m) &\geq \sum_{\sqrt{m} < p \leq m} g(\lfloor m/p \rfloor) \geq \sum_{\sqrt{m} < p < m/2} c \cdot \frac{\lfloor m/p \rfloor}{\log^\lambda(\lfloor m/p \rfloor)} \geq \\ &\geq \sum_{\sqrt{m} < p < m/2} c \cdot \frac{\lfloor m/p \rfloor}{\log^\lambda(\lfloor m/p \rfloor)} \geq \sum_{\sqrt{m} < p < m/2} c \cdot \frac{m/p - 1}{\log^\lambda(\lfloor m/p \rfloor)} = \\ &= \sum_{\sqrt{m} < p < m/2} c \cdot \frac{m/p}{\log^\lambda(\lfloor m/p \rfloor)} - \sum_{\sqrt{m} < p < m/2} c \cdot \frac{1}{\log^\lambda(\lfloor m/p \rfloor)}. \quad (3) \end{aligned}$$

In Rosser and Schoenfeld (1962) it is proved that  $\pi(m) < \frac{1.25506m}{\log m}$  for every  $m > 1$ , hence  $\pi(m/2) - \pi(\sqrt{m}) \leq \pi(m) < 1.5 \cdot \frac{m}{\log m}$ . For the second term of the last line of (3) we obtain:

$$\sum_{\sqrt{m} < p < m/2} c \cdot \frac{1}{\log^\lambda(\lfloor m/p \rfloor)} \leq \sum_{\sqrt{m} < p < m/2} c \cdot \frac{1}{(\log 2)^\lambda} \leq 1.5 \cdot \frac{m}{\log m} \cdot \frac{c}{\log 2} = o\left(\frac{m}{\log^\lambda m}\right), \quad (4)$$

since  $\lambda < 1$ .

Now we estimate the main term. By Mertens' theorem, there exists a constant  $M$  such that  $\sum_{p \leq x} \frac{1}{p} = \log \log x + M + o(1)$ . Hence, for every  $\varepsilon > 0$  there exists  $B = B(\varepsilon)$  such that for  $B \leq a \leq b$

$$\left| \sum_{a < p < b} \frac{1}{p} - \log \log b + \log \log a \right| < \varepsilon \quad (5)$$

holds. For  $m > 2^{2K}$  we have  $m^{\frac{1}{2} + \frac{K-1}{2K}} < m/2$ . Applying (5) to the interval  $I_h = (m^{\frac{1}{2} + \frac{h-1}{2K}}, m^{\frac{1}{2} + \frac{h}{2K}}]$ , where  $h$  is an integer satisfying  $1 \leq h \leq K-1$  we obtain that

$$\sum_{p \in I_h} \frac{1}{p} > \log \frac{K+h}{K+h-1} - \varepsilon. \quad (6)$$

If  $p \in I_h$ , then  $\log^\lambda(m/p) \leq \log^\lambda(m) \left(\frac{K-h+1}{2K}\right)^\lambda$ . Substituting into the main term of the last line of (3), omitting the integer parts and rearranging we get that

$$\begin{aligned} \sum_{\sqrt{m} < p < m/2} c \cdot \frac{m/p}{\log^\lambda(\lfloor m/p \rfloor)} &\geq cm \sum_{\sqrt{m} < p < m/2} \frac{1/p}{\log^\lambda(m/p)} \geq \\ &\geq \frac{cm}{\log^\lambda m} \sum_{h=1}^{K-1} \sum_{p \in I_h} \left( \frac{2K}{K-h+1} \right)^\lambda \cdot \frac{1}{p} \geq \\ &\geq \frac{cm}{\log^\lambda m} \left( \sum_{h=1}^{K-1} \left( \frac{2K}{K-h+1} \right)^\lambda \log \frac{K+h}{K+h-1} - \varepsilon \sum_{h=1}^{K-1} \left( \frac{2K}{K-h+1} \right)^\lambda \right). \end{aligned} \quad (7)$$

Now we show that there exists some  $K$  such that

$$S_K = \sum_{h=1}^{K-1} \left( \frac{2K}{K-h+1} \right)^\lambda \log \frac{K+h}{K+h-1} > 1. \quad (8)$$

Let  $f_K(y) = \left(\frac{2}{y}\right)^\lambda K \cdot \log \left(1 + \frac{1}{K(2-y)}\right)$  and  $f(y) = \left(\frac{2}{y}\right)^\lambda \cdot \frac{1}{2-y}$ . The sequence of functions  $f_K$  converges to  $f$ . Then

$$S_K = \frac{f_K(\frac{1}{K}) + f_K(\frac{2}{K}) + \cdots + f_K(\frac{K}{K})}{K} - \frac{f_K(\frac{1}{K})}{K}.$$

Let

$$T_K = \frac{f(\frac{1}{K}) + f(\frac{2}{K}) + \cdots + f(\frac{K}{K})}{K}.$$

As  $1 > \lambda > \lambda_0$ , the Riemann-sum  $T_k$  converges to  $\int_0^1 f > 1$ . As  $f_K(\frac{1}{K})/K$  converges to 0, it is easy to see that  $S_K - T_K$  converges to 0. Hence we can fix a  $K$  such that  $S_K > 1$ . Now, we can choose some

$\varepsilon > 0$  such that

$$\eta = \sum_{h=1}^{K-1} \left( \frac{2K}{K-h+1} \right)^\lambda \log \frac{K+h}{K+h-1} - 1 - \varepsilon \sum_{h=1}^{K-1} \left( \frac{2K}{K-h+1} \right)^\lambda > 0.$$

According to (4) there exists some  $R$  such that if  $R < m$ , then

$$\sum_{\sqrt{m} < p < m/2} c \cdot \frac{1}{\log^\lambda(\lfloor m/p \rfloor)} \leq \eta \cdot c \cdot \frac{m}{\log^\lambda m}.$$

By (3) and (7) we obtain that  $g(m) \geq c \cdot \frac{m}{\log^\lambda m}$  holds. If we choose  $c > 0$  such that (2) holds for  $n \leq \max(2^{2K}, B^2(\varepsilon), R)$ , then (3) is gained.  $\square$

### 3 The case $K = \underline{k}$

In this section we prove Conjecture 1. We distinguish cases according to how large is  $k$  according to  $n$ . The conjecture is true for  $k \leq 8$ . (Pilz (1992))

*Case 1:*

$$9 \leq k \leq 1.34 \cdot \log n$$

We show that in this case the number of elements that occur exactly once in the product  $\underline{n} \cdot \underline{k}$  is at least  $n$ . We shall need the following two observations.

**Lemma 3** *Let  $n/2 < a \leq n$  and  $b \in \underline{k}$  such that  $a$  is relatively prime to every number less than  $k$ . Then  $ab$  occurs once in  $\underline{n} \cdot \underline{k}$ .*

**Proof:** Let us assume that  $a_1, a_2 \in \underline{n}$  and  $b_1, b_2 \in \underline{k}$  satisfy the conditions of the lemma, and  $a_1 b_1 = a_2 b_2$ . Now,  $a_1 | a_2 b_2$  and  $a_1$  and  $b_2$  are relatively prime, hence  $a_1 | a_2$ . As  $a_1 > n/2$  we have  $2a_1 > n \geq a_2$ , thus  $a_1 = a_2$ , which implies  $b_1 = b_2$ .  $\square$

**Lemma 4** *If  $k \geq 14$ , then  $\prod_{p \leq k} \left(1 - \frac{1}{p}\right) \geq \frac{0.5}{\log k}$ .*

**Proof:** In Rosser and Schoenfeld (1962) it is shown that for  $k > 1$

$$\frac{e^{-\gamma}}{\log k} \left(1 - \frac{1}{\log^2 k}\right) \leq \prod_{p \leq k} \left(1 - \frac{1}{p}\right),$$

where  $\gamma$  is the Euler constant. For  $k > 21$  by using the monotonicity of the logarithm function and  $e^{-\gamma} > 0.56$  we get that

$$\frac{e^{-\gamma}}{\log k} \left(1 - \frac{1}{\log^2 k}\right) \geq \frac{0.56}{\log k} \left(1 - \frac{1}{\log^2 22}\right) > \frac{0.5}{\log k}.$$

For  $14 \leq k \leq 21$  it is enough to check the statement when  $k = 14, 17$  and  $19$ . For these numbers the values of  $(\log k) \cdot \prod_{p \leq k} \left(1 - \frac{1}{p}\right)$  are 0.506, 0.511 and 0.503, respectively, hence the statement holds.  $\square$

**Proposition 5** *Let  $9 \leq k \leq 1.34 \cdot \log n$ . Then  $|\underline{n} * \underline{k}| \geq n$ .*

**Proof:** We show that there are at least  $n$  products satisfying the conditions of Lemma 3. For this we need to estimate the number of integers between  $n/2$  and  $n$  that are not divisible by a prime less than  $k$ . This number will be denoted by  $D$ . By the inclusion-exclusion principle

$$D = n - \lfloor n/2 \rfloor + \sum_{h=1}^r (-1)^h \sum_{1 \leq i_1 < \dots < i_h \leq r} \left( \left\lfloor \frac{n}{p_{i_1} \dots p_{i_h}} \right\rfloor - \left\lfloor \frac{n/2}{p_{i_1} \dots p_{i_h}} \right\rfloor \right), \quad (9)$$

where  $\pi(k) = r$  and  $p_1, \dots, p_r$  are the primes up to  $k$ . Applying  $x - 1 < \lfloor x \rfloor \leq x$  to all  $2^{r+1}$  terms of the right side we get that

$$\begin{aligned} D &\geq n - n/2 + \sum_{h=1}^r (-1)^h \sum_{1 \leq i_1 < \dots < i_h \leq r} \left( \frac{n}{p_{i_1} \dots p_{i_h}} - \frac{n/2}{p_{i_1} \dots p_{i_h}} \right) - 2^r = \\ &= \frac{n}{2} \prod_{p \leq k} \left(1 - \frac{1}{p}\right) - 2^r. \end{aligned} \quad (10)$$

If  $k \geq 14$ , Lemma 4 applies, and

$$D \geq \frac{n}{2} \prod_{p \leq k} \left(1 - \frac{1}{p}\right) - 2^r \geq \frac{0.25n}{\log k} - 2^r$$

As  $k \leq 1.34 \log n$ , for  $k \geq 14$  we have the estimation

$$2^r = 2^{\pi(k)} \leq 2^{k/2} \leq \frac{1}{100 \log k} \cdot e^{\frac{k}{1.34}} \leq \frac{n}{100 \log k}.$$

Hence,  $D \geq \frac{0.24n}{\log k}$ . Using Lemma 3 we obtain  $|\underline{n} * \underline{k}| \geq Dk$ . The function  $x/\log x$  is monotone increasing on  $[1, \infty)$ , thus

$$|\underline{n} * \underline{k}| \geq Dk \geq \frac{0.24k}{\log k} n \geq \frac{0.24 \cdot 14}{\log 14} n > n.$$

For  $9 \leq k \leq 13$  we have

$$|\underline{n} * \underline{k}| \geq Dk \geq \left( \frac{n}{2} \prod_{p \leq k} \left(1 - \frac{1}{p}\right) - 2^{\pi(k)} \right) k.$$

For  $10 \leq k \leq 13$  it is obtained by calculation that the right hand side is greater than  $n$  if  $n \geq e^{k/1.34}$ . For  $k = 9$  the inequality holds if  $n > 5040$ . By brute force the statement can be checked for  $k = 9$  and  $n \leq 5040$ . Thus we obtained  $|\underline{n} * \underline{k}| > n$ .  $\square$

**Case 2:**  $1.34 \cdot \log n \leq k \leq n - \frac{0.22 \cdot n}{\log n}$  and  $n \geq 1410$ .

Let  $k_1 = \max(k, n/7)$  and  $k_1 < p \leq n$  a prime. As  $k < p$ , the set of elements of  $\underline{n} * \underline{k}$ , which are divisible by  $p$  is  $\{p, 2p, \dots, \lfloor n/p \rfloor p\} * \underline{k}$ . This set has the same cardinality as the set  $\lfloor n/p \rfloor * \underline{k}$ . Now,  $\lfloor n/p \rfloor \leq 6$ , hence  $|\lfloor n/p \rfloor * \underline{k}| \geq k$ . It is easy to see that for  $p > q > n/7$  an element of  $\underline{n} * \underline{k}$  cannot be divisible by both  $p$  and  $q$ . Hence,  $|\underline{n} * \underline{k}| \geq (\pi(n) - \pi(k_1))k$ .

At first, suppose that  $k \leq n/7$ . By a theorem of Dusart (1999) for  $x \geq 17$

$$\frac{x}{\log x} \leq \pi(x) \leq \frac{x}{\log x} \left(1 + \frac{1.2762}{\log x}\right)$$

holds. Hence,  $\pi(n) - \pi(n/7) \geq 0.749 \cdot \frac{n}{\log n}$  for  $n \geq 1410$ . As  $1.34 \cdot \log n \leq k$ , we have

$$|\underline{n} * \underline{k}| \geq 1.34 \cdot 0.749 \cdot n > n.$$

Secondly, let us consider the case when  $n/7 < k \leq n/2$ . As  $\pi(n) - \pi(n/2) \geq 7$ ,

$$|\underline{n} * \underline{k}| \geq (\pi(n) - \pi(k_1))k > 7 \cdot n/7 = n.$$

Finally, let  $n/2 < k < n - \frac{0.22 \cdot n}{\log n}$ . Then by the estimates in Dusart (1999) and Robin (1983) there are at least two primes between  $k$  and  $n$  if  $n > 90000$ . It can be checked that this also holds for  $n > 1410$ . Thus

$$|\underline{n} * \underline{k}| \geq (\pi(n) - \pi(k))k \geq 2(n/2) = n.$$

We continue with the case when  $k$  is "large", that is,  $n - \frac{0.4 \cdot n}{\log n + 1.02} \leq k$ . By calculation we have  $n - \frac{0.4 \cdot n}{\log n + 1.02} \leq n - \frac{0.22 \cdot n}{\log n}$  for  $n \geq 4$ .

**Case 3:**  $n - \frac{0.4 \cdot n}{\log n + 1.02} \leq k \leq n$  and  $n > 5000$ .

If  $k = n$ , then  $\underline{k} \cdot \underline{n} = \{1, \dots, n\} \cdot \{1, \dots, n\}$ . If  $a \neq b$ , then pairing  $ab$  with  $ba$  only the products of the form  $a \cdot a$  are left, hence  $\underline{n} * \underline{k} = \{1^2, 2^2, \dots, n^2\}$ . Thus

$$|\underline{n} * \underline{k}| = n.$$

Assume now that  $k < n$ . Then

$$|\underline{n} * \underline{k}| = |(\underline{k} * \underline{k}) \Delta ((\underline{n} \setminus \underline{k}) * \underline{k})| = |\underline{k} * \underline{k}| + |(\underline{n} \setminus \underline{k}) * \underline{k}| - 2|(\underline{k} * \underline{k}) \cap ((\underline{n} \setminus \underline{k}) * \underline{k})|. \quad (11)$$

For the first term on the right side of (11) we have

$$|\underline{k} * \underline{k}| = |\{1^2, 2^2, \dots, k^2\}| = k. \quad (12)$$



**Lemma 6** For the second term of (11) we have

$$|(\underline{n} \setminus \underline{k}) * \underline{k}| \geq 2k - n. \quad (13)$$

**Proof:** We use the following observation: If

$$i \leq \frac{k}{n-k} \quad \text{and} \quad k+1 \leq j \leq n,$$

then  $ij$  appears exactly once in  $(\underline{n} \setminus \underline{k}) \cdot \underline{k}$ , so  $ij \in (\underline{n} \setminus \underline{k}) * \underline{k}$ . Let us assume that  $ij = i'j'$  such that  $1 \leq i' \leq k$  and  $k+1 \leq j' \leq n$ . If  $i = i'$ , then  $j = j'$ . If  $i' < i$ , then  $1 \leq i' \leq \frac{k}{n-k}$  and  $k+1 \leq j' \leq n$ . Now, changing the roles of  $(i, j)$  and  $(i', j')$  we may assume that  $i < i'$ . As  $ij = i'j'$ , we have  $\frac{i}{i'} = \frac{j'}{j}$  and

$$\frac{i}{i'} \leq \frac{i}{i+1} \leq \frac{\frac{k}{n-k}}{\frac{k}{n-k} + 1} = \frac{k}{n} < \frac{k+1}{n} \leq \frac{j'}{j},$$

which is a contradiction. For  $(\underline{n} \setminus \underline{k}) * \underline{k}$  we obtain that

$$|(\underline{n} \setminus \underline{k}) * \underline{k}| \geq \left\lfloor \frac{k}{n-k} \right\rfloor (n-k) \geq \left( \frac{k}{n-k} - 1 \right) (n-k) = k - (n-k) = 2k - n. \quad (14)$$

□

Now, we focus on the third term of (11).

**Lemma 7** For the third second term of (11)

$$|(\underline{k} * \underline{k}) \cap ((\underline{n} \setminus \underline{k}) * \underline{k})| \leq 0.431 \cdot k. \quad (15)$$

holds.

**Proof:** It is enough to show that among the numbers  $1^2, 2^2, \dots, k^2$  at most  $0.431k$  many has a divisor in the interval  $[k+1, n]$ . Let  $k+1 \leq m \leq n$  and  $m = a_m b_m^2$ , where  $b_m^2$  is the largest square divisor of  $m$ . Since  $a_m$  is squarefree,  $m|i^2$  if and only if  $a_m b_m|i$ . Let  $S$  denote the following upper bound of the number of elements of the set  $\{1^2, 2^2, \dots, k^2\}$  which have a divisor in  $[k+1, n]$ :

$$S = \sum_{m=k+1}^n \left\lfloor \frac{k}{a_m b_m} \right\rfloor \leq \sum_{m=k+1}^n \frac{k}{a_m b_m} = k \sum_{m=k+1}^n \frac{b_m}{m}.$$

Recall that  $m = a_m b_m^2$ , where  $a_m$  is squarefree. Now, summing by  $j = b_m \leq \sqrt{m}$ :

$$S = k \sum_{j=1}^{\lfloor \sqrt{n} \rfloor} \sum_{\substack{j^2|m, \\ k+1 \leq m \leq n, \\ |\mu(m/j^2)|=1}} \frac{j}{m} \leq k \sum_{j=1}^{\lfloor \sqrt{n} \rfloor} j \sum_{\substack{j^2|m, \\ k+1 \leq m \leq n}} \frac{1}{m}.$$

Rewrite  $S = k(S_1 + S_2)$ , where

$$S_1 := \sum_{j=1}^{\lfloor \sqrt{n}/2 \rfloor} j \sum_{\substack{j^2 | m, \\ k+1 \leq m \leq n}} \frac{1}{m} \quad \text{and} \quad S_2 := \sum_{j=\lfloor \sqrt{n}/2 \rfloor + 1}^{\lfloor \sqrt{n} \rfloor} j \sum_{\substack{j^2 | m, \\ k+1 \leq m \leq n}} \frac{1}{m}.$$

First, we give an upper bound for  $S_1$ .

**Lemma 8**

$$S_1 \leq \left( \frac{\log n}{2} + 0.31 \right) (\log n - \log k) + \frac{n + 2\sqrt{n}}{8k}. \quad (16)$$

**Proof:** Let  $r_j = \left\lceil \frac{k+1}{j^2} \right\rceil$  and  $s_j = \left\lfloor \frac{n}{j^2} \right\rfloor$ . Then

$$S_1 = \sum_{j=1}^{\lfloor \sqrt{n}/2 \rfloor} j \sum_{l=r_j}^{s_j} \frac{1}{lj^2} = \sum_{j=1}^{\lfloor \sqrt{n}/2 \rfloor} \frac{1}{j} \sum_{l=r_j}^{s_j} \frac{1}{l}. \quad (17)$$

The function  $\frac{1}{x}$  is a nonnegative decreasing function on  $(0, \infty)$ , hence we can estimate the inside sum by

$$\sum_{l=r_j}^{s_j} \frac{1}{l} \leq \int_{r_j}^{s_j} \frac{1}{x} dx + \frac{1}{r_j} = \log s_j - \log r_j + \frac{1}{r_j}.$$

As  $\frac{k}{j^2} \leq r_j$  and  $s_j \leq \frac{n}{j^2}$  we have

$$\log s_j - \log r_j = \log \frac{s_j}{r_j} \leq \log \frac{n/j^2}{k/j^2} = \log n - \log k.$$

Substituting into (17) we obtain

$$S_1 \leq \sum_{j=1}^{\lfloor \sqrt{n}/2 \rfloor} \frac{1}{j} \left( \log s_j - \log r_j + \frac{1}{r_j} \right) \leq \sum_{j=1}^{\lfloor \sqrt{n}/2 \rfloor} \frac{1}{j} \left( \log n - \log k + \frac{j^2}{k} \right). \quad (18)$$

Since

$$\sum_{j=1}^{\lfloor \sqrt{n}/2 \rfloor} \frac{1}{j} \leq \log \lfloor \sqrt{n}/2 \rfloor + 1 \leq \frac{\log n}{2} - \log 2 + 1 \leq \frac{\log n}{2} + 0.31. \quad (19)$$

and

$$\sum_{j=1}^{\lfloor \sqrt{n}/2 \rfloor} j = \frac{\lfloor \sqrt{n}/2 \rfloor \cdot (\lfloor \sqrt{n}/2 \rfloor + 1)}{2} \leq \frac{n + 2\sqrt{n}}{8}, \quad (20)$$

from the inequalities (18), (19), (20) we get (16).  $\square$

Now we give an upper bound for  $S_2$ .

**Lemma 9**

$$S_2 \leq \left(1 + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}}\right) \cdot \frac{n-k}{2\sqrt{k}} \cdot \frac{\sqrt{n}}{k} + \frac{3\sqrt{n}}{k} < 1.15 \cdot \frac{(n-k)\sqrt{n}}{k^{3/2}} + \frac{3\sqrt{n}}{k}. \quad (21)$$

**Proof:**

$$S_2 = \sum_{j=\lfloor \sqrt{n}/2 \rfloor + 1}^{\lfloor \sqrt{n} \rfloor} \sum_{\substack{j^2 | m, \\ k+1 \leq m \leq n}} \frac{j}{m} \quad (22)$$

In (22) for every  $j$  we have

$$n \geq j^2 \geq (\lfloor \sqrt{n}/2 \rfloor + 1)^2 > \frac{n}{4}.$$

Hence  $m = j^2$  or  $2j^2$  or  $3j^2$ . As  $k < m \leq n$ , for  $m = ij^2$  ( $i = 1, 2, 3$ ) we get

$$\sqrt{\frac{k}{i}} < j \leq \sqrt{\frac{n}{i}} \quad \text{and} \quad \frac{j}{m} \leq \frac{\sqrt{n}}{k}.$$

For fixed  $i$ , the number of  $j$  such that  $m = ij^2$  is at most:

$$\left\lceil \frac{\sqrt{n} - \sqrt{k}}{\sqrt{i}} \right\rceil = \left\lceil \frac{1}{\sqrt{i}} \cdot \frac{n-k}{\sqrt{n} + \sqrt{k}} \right\rceil \leq \frac{1}{\sqrt{i}} \cdot \frac{n-k}{2\sqrt{k}} + 1,$$

thus

$$S_2 \leq \left(1 + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}}\right) \cdot \frac{n-k}{2\sqrt{k}} \cdot \frac{\sqrt{n}}{k} + \frac{3\sqrt{n}}{k} < 1.15 \cdot \frac{(n-k)\sqrt{n}}{k^{3/2}} + \frac{3\sqrt{n}}{k},$$

and this is what we wanted to show.  $\square$

Summarizing the results, from (16) and (21) we obtain:

$$\begin{aligned} S &= k(S_1 + S_2) \leq \\ &\leq k \left\{ \left( \frac{\log n}{2} + 0.31 \right) (\log n - \log k) + \frac{n + 2\sqrt{n}}{8k} + 1.15 \cdot \frac{(n-k)\sqrt{n}}{k^{3/2}} + \frac{3\sqrt{n}}{k} \right\}. \end{aligned} \quad (23)$$

We assumed that  $n - \frac{0.4 \cdot n}{\log n + 1.02} \leq k$  and  $n \geq 5000$ . By using the inequality  $e^{-x} < \frac{1}{1+x}$  we obtain that  $ne^{-\frac{0.2}{\frac{\log n}{2} + 0.31}} < n \cdot \frac{1}{1 + \frac{0.2}{\frac{\log n}{2} + 0.31}} = n - \frac{0.4 \cdot n}{\log n + 1.02} \leq k$ . As  $n \geq 5000$ , we have that  $\frac{k}{n} > 0.958$ .

By easy calculation from these inequalities the following ones can be deduced:

$$\left(\frac{\log n}{2} + 0.31\right)(\log n - \log k) < 0.2, \quad (24)$$

$$\frac{n + 2\sqrt{n}}{8k} < 0.135, \quad (25)$$

$$1.15 \cdot \frac{(n-k)\sqrt{n}}{k^{3/2}} + \frac{3\sqrt{n}}{k} < 0.096. \quad (26)$$

Adding (24), (25) and (26) using (23) we arrive at:

$$S \leq k(0.2 + 0.135 + 0.096) = 0.431 \cdot k. \quad (27)$$

Then from inequalities (12), (13) and (15) in case  $k/n > 0.958$  we get

$$|\underline{k} * \underline{n}| \geq k + 2k - n - 2S \geq 2.138 \cdot k - n > n,$$

thus we proved the statement in Case 3 as well.  $\square$

We proved the statement for all pairs  $n, k$  where  $n \geq 5000$ . Cases  $k \leq n \leq 5000$  can be checked by brute force.

## References

- P. Dusart. The  $k$ th prime is greater than  $k(\ln k + \ln \ln k - 1)$  for  $k \geq 2$ . *Math. Comp.*, 68(225):411–415, 1999.
- R. Eggetsberger. On constructing codes from planar nearrings. <http://www.algebra.uni-linz.ac.at/Nearrings/nrcodes.html>, 2011.
- G. Pilz. *Near-rings*, volume 23 of *North-Holland Mathematics Studies*. North-Holland Publishing Co., Amsterdam, second edition, 1983. ISBN 0-7204-0566-1. The theory and its applications.
- G. Pilz. On polynomial near-ring codes. In *Contributions to general algebra, 8 (Linz, 1991)*, pages 233–238. Hölder-Pichler-Tempsky, Vienna, 1992.
- G. Pilz. Near-rings: What they are and what they are good for. <http://www.algebra.uni-linz.ac.at/Nearrings/what-are.html>, 2011.
- G. Robin. Estimation de la fonction de Tchebychef  $\theta$  sur le  $k$ -ième nombre premier et grandes valeurs de la fonction  $\omega(n)$  nombre de diviseurs premiers de  $n$ . *Acta Arith.*, 42(4):367–389, 1983.
- J. B. Rosser and L. Schoenfeld. Approximate formulas for some functions of prime numbers. *Illinois J. Math.*, 6:64–94, 1962.

