Serdica J. Computing **9** (2015), No. 3–4, 257–268

Serdica
Journal of Computing

Bulgarian Academy of Sciences
Institute of Mathematics and Informatics

COMBINATORIAL COMPUTATIONS ON AN EXTENSION OF A PROBLEM BY PÁL TURÁN

Petar Gaydarov*, Konstantin Delchev

Abstract. Turán's problem asks what is the maximal distance from a polynomial to the set of all irreducible polynomials over $\mathbb{Z}$. It turns out it is sufficient to consider the problem in the setting of $\mathbb{F}_2$. Even though it is conjectured that there exists an absolute constant C such that the distance $L(f-g) \leq C$, the problem remains open. Thus it attracts different approaches, one of which belongs to Lee, Ruskey and Williams, who study what the probability is for a set of polynomials 'resembling' the irreducibles to satisfy this conjecture. In the following article we strive to provide more precision and detail to their method, and propose a table with better numeric results.

1. Introduction. In 1962 P. Turán [9] asked whether every polynomial with integer coefficients is close to an irreducible polynomial with an equal or smaller degree. For a polynomial $f(x) = \sum_{k=0}^{n} a_k x^k$, let $L(f)$ denote its length,

ACM Computing Classification System (1998): H.1.1.

Key words: irreducible polynomials, distance sets, finite fields.

*This author is partially supported by the High School Students Institute of Mathematics and Informatics.

defined by

$$L(f) = \sum_{k=0}^{n} |a_k|.$$

Thus we can rephrase the question: Is there an absolute constant C such that for every polynomial $f \in \mathbb{Z}[x]$ there exists an irreducible polynomial $g \in \mathbb{Z}[x]$ with $\deg(g) \leq \deg(f)$ and $L(f - g) \leq C$.

In general the problem remains open, but a number of partial results are known. In 1970, Schinzel [8] proved that if we remove the restriction about the degree, then $C = 3$ is such a constant. Moreover, he showed that for a polynomial f of degree n, there is an irreducible polynomial g with $L(f - g) \leq 3$ and

$$\deg(g) \leq \exp\bigl((5n + 7)(\|f\|^2 + 3)\bigr),$$

where $\|f\|^2$ stands for the sum of the squares of the coefficients of f. More recently, Banerjee and Filaseta [1] improved this by showing that the bound on the degree of the irreducible polynomial g depends only linearly on that of f (though exponentially on $\|f\|^2$). More precisely, the bound on the degree of g satisfies

$$\deg(g) \leq 8 \max\{n + 3, n_0\} 5^{8\|f\|^2+9},$$

where n_0 is an effectively computable constant.

Turán's problem has been tested and verified for polynomials of small degree n by explicit computations. In 1997, Bérczes and Hajdu [2] showed that $C = 5$ suffices for polynomials of degree $n \leq 22$, and in 1998 [3] demonstrated that this bound suffices for $n \leq 24$. In 2008, Ruskey, Lee, and Williams [6] established that $C = 5$ is sufficient for $n \leq 32$ by using an algorithm developed in [4]. More recently, Mossinghoff [7] extended this result to $n \leq 34$. This was again bettered by Filaseta and Mossinghoff [5] as they proved that constant to be sufficient for degree at most 40.

The results were proven by showing that $C = 3$ suffices for polynomials with leading and constant terms which are both odd. For such a polynomial f with degree n, by Eisenstein's criterion with prime $p = 2$, there exists an irreducible polynomial $g(x)$ with $\deg(g) = n$ and $L(f - g) \leq n$. For a positive integer n, we denote by C_n the smallest positive integer having the following property: For every $f \in \mathbb{Z}[x]$ with degree n and odd leading and trailing terms, there exists an irreducible polynomial $g \in \mathbb{Z}[x]$ with $\deg(g) \leq \deg(f)$ and $L(f - g) \leq C_n$.

We can also further simplify Turán's problem by considering a local version. For a polynomial $f \in \mathbb{F}_2[x]$, we define its length $L_2(f)$ as its number of monomials. Let $C_n(2)$ be the smallest positive integer with the property that for every $f \in \mathbb{F}_2[x]$ with degree n and constant term 1, there exists an irreducible polynomial $g \in \mathbb{F}_2[x]$ with the same degree and $L_2(f - g) \leq C_n(2)$. We know that any polynomial $g \in \mathbb{Z}[x]$ with an odd leading coefficient is necessarily irreducible in $\mathbb{Z}[x]$ if its reduction modulo 2 is irreducible in $\mathbb{F}_2[x]$. Then it follows that $C_n \leq C_n(2)$. Thus, to establish the result $C \leq 5$, we only need to prove that $C_n(2) \leq 3$. An additional result by Filaseta and Mossinghoff [5] states that for $n \geq 246$ the distance of a positive proportion of polynomials in $\mathbb{F}_2[x]$ to every irreducible polynomial is greater or equal to 4, i. e., $C_n(2) \geq 3$.

Lee, Ruskey and Williams [6] study the Hamming distance from polynomials to classes of polynomials that share certain properties of irreducible polynomials over $\mathbb{F}_2$. The results give some insight into whether or not irreducible polynomials can be effectively modeled by these more general classes of polynomials. The properties they examine are the number of elements which is $\lfloor 2^n/n \rfloor$, the non-zero constant term, the odd density and reciprocal-closeness.

At first they choose uniformly randomly sets S satisfying the first three properties. They derive a formula for the expected number of polynomials at a certain distance from these sets S (a full proof of the formula is presented by Mossinghoff [7]). However, these results turn out not to be close enough to the actual data about irreducible polynomials over $\mathbb{F}_2$. The authors then examine uniformly randomly chosen sets R which satisfy all four properties. Ruskey et al. first examine those sets for an odd degree n of the polynomials. They derive a formula for the expected number of polynomials in the *neighbourhood* (distance 1) of those uniformly randomly chosen sets R. For the case for polynomials of even degree n they only conjecture a formula for the number of polynomials in the neighbourhood of the sets R.

In this paper we strive to clarify and expand on them. In Section 2 we discuss the prior work of Ruskey et al. In Section 3 we examine the uniformly randomly chosen sets R of polynomials of even degree n and find a formula for the expected number of polynomials in the neighbourhood of those sets which is different from the one conjectured by Ruskey et al.

2. Prior work. It will be convenient to restate the results in terms of bitstrings. We identify the bitstring $b_1 b_2 \ldots b_N$ with the polynomial $x^{N+1} + b_N x^N + \cdots + b_1 x + 1$.

Hamming distances to an odd density set. By $S(N, M)$ we denote a set of M odd density bitstrings each of length N, chosen uniformly at random from the set of all 2^N odd density bitstrings of length N. We say that a bitstring is *odd* if it has odd density; otherwise, it is *even.* The Hamming distance $H(b, c)$ between two bitstrings b and c of length N is the number of positions in which the corresponding bits differ. We say that two bitstrings are *adjacent* if their Hamming distance is one; i. e., they are adjacent in the hypercube. We extend the notation to sets S of bitstrings by defining $H(b, S) = \min\{H(b, s) | s \in S\}$.

Given a set of length N bitstrings S, the *neighbourhood*, $H_1(S)$, of S is the set $\{b \in \{0,1\}^N | H(b, s) = 1 \text{ for some } s \in S\}$; in other words it is exactly the same as the open neighbourhood of S in the hypercube Q_N, in the graph-theoretic sense.

Theorem 1 ([6]). *Asymptotically the expectation*

$$\mathbb{E}|\{b \in \{0,1\}^N | H(b, S(N, M)) = d\}|$$

is equal to

$$\begin{cases} 2^n/n & \textit{when } d = 0; \\ 2^{N-1}(1 - e^{-4}) & \textit{when } d = 1; \\ 2^{N-1} - 2^n/n & \textit{when } d = 2; \\ 2^{N-1}e^{-4} & \textit{when } d = 3; \\ 0 & \textit{when } d > 3. \end{cases}$$

When we compare these results with the actual data about irreducible polynomials they do not match. Since the current model does not explain the data, we need to refine it. We restrict the random strings to be reverse-closed. Given a binary string b, let b^R represent the reversal of b. A set of binary strings S is reverse-closed if $b^R \in S$ whenever $b \in S$.

Even length odd density reverse-closed sets. Let O represent the binary strings of length N with odd density, and let E represent the binary strings of length N with even density. Since N is even, there is no $b \in O$ where $b = b^R$. Therefore, we can partition O into $O^>$ and $O^<$ where $O^>$ and $O^<$ are defined as:

$$O^> = \{b \in O : b > b^R\}$$
$$O^< = \{b \in O : b < b^R\}.$$

Let $R = R(N, M)$ be a reverse-closed set of M odd density bitstrings of length N, chosen uniformly randomly. Our objective is to calculate the expected size of $H_1(R(N, M))$ as a function of M and N.

Theorem 2 ([6]). *When N is even the expected size of the set $H_1(R(N,M))$ is*

$$\mathbb{E}|H_1(R(N,M))| = 2^{N/2}\left(2^{N/2-1} - \frac{\binom{2^{N-2}-N/2}{M/2} + \left(2^{N/2-1}-1\right)\binom{2^{N-2}-N}{M/2}}{\binom{2^{N-2}}{M/2}}\right).$$

From Theorem 2 it follows that

Theorem 3 ([6]). *When N is even, then*

$$\mathbb{E}|H_1(R(n-1, 2^n/n))| \sim 2^{n-2}(1-e^{-4}).$$

3. Odd length odd density reverse-closed sets.

When N is odd the computation becomes more complicated because some bitstrings satisfy $b = b^R$. Again let $R = R(N,M) \subseteq O$ be a uniformly randomly chosen reverse-closed set of M bitstrings. Our goal is yet again to compute the expectation for $H_1(R)$ depending on M and N. We define $O^>$ and $O^<$ the same way and let $O^=$ be

$$O^= = \{b \in O : b = b^R\}.$$

We can partition E into E^0, E^1 and E^2 where:

$$E^2 = \{b \in E : b \text{ and } b^R \text{ differ in exactly two positions}\};$$

$$E^1 = \{b \in E : b = b^R\};$$

$$E^0 = \{b \in E : b \notin E^2 \cup E^1\}.$$

To illustrate these sets, let us consider an example for $N = 5$.

The string $b = 01100 \in E^2$ since $H_1(\{b\}) \cup O^= = \{00100, 01110\}$. In general, b is in E^2 if $b_n = 1$, and $b_x \neq b_{N-x}$ has a unique solution for $0 \le x \le (N-1)/2$. Then $H_1(\{b\}) \cup O^=$ contains the result of changing the x^{th} or $(N-x)^{\text{th}}$ bit of b.

The string $01010 \in E^1$ since $H_1(\{b\}) \cup O^= = \{01110\}$. In general, E^1 contains binary strings b in E such that $b_n = 0$, and $b_x = b_{N-x}$ for all $0 \le x \le N$. Then $H_1(\{b\}) \cup O^=$ contains the result of changing the middle bit of b to one.

The string $01001 \in E^0$ since $H_1(\{b\}) \cup O^= = \varnothing$. The set E^0 contains the binary strings in E that are not in E^2 or E^1.

We now prove the following proposition.

Proposition 1. *Let $b(i)$ denote the bitstring which differs from b in the i^{th} position. The equation $b(M) = b^R(L)$ has a solution if and only if $b = b^R$.*

Proof. Let b have an even length. Let b_i denote the i^{th} position in b. There are two possibilities for the equality $b(M) = b^R(L)$ to hold true.

I Let $M \neq N - L$. Then the following must be true:

$$\begin{aligned}
&b_1 = b_N;\\
&b_2 = b_{N-1};\\
&\dots\\
&b_M \neq b_{N-M};\\
&\dots\\
&b_L = b_{N-L};\\
&\dots\\
&b_{N/2} = b_{N/2+1};\\
&\dots\\
&b_{N-L} \neq b_L;\\
&\dots\\
&b_{N-M} = b_M.
\end{aligned}$$

Obviously it is impossible for $b_M \neq b_{N-M}$ and $b_M = b_{N-M}$ to hold true simultaneously.

II Let $M = N - L$. Then the following must hold true:

$$\begin{aligned}
&b_1 = b_N;\\
&b_2 = b_{N-1};\\
&\dots\\
&b_M \neq b_{N-M} = b_{N-L} \neq b_L;\\
&\dots
\end{aligned}$$

However, this means $b_M = b_L$, i. e., the bitstring b must be a palindrome. The proof is analogous when b has on odd length. □

The following lemma was first presented by Lee et al. [6], however, the full proof was omitted from their paper.

Lemma 1. *Using this partition we obtain:*

$$|\{c \in O^{=} : c \in R \Rightarrow b \in H_1(R)\}| = \begin{cases} 0, & \textit{when } b \in E^0; \\ 1, & \textit{when } b \in E^1; \\ 2, & \textit{when } b \in E^2 \end{cases}$$

and

$$|\{c \in O^{>} : c \in R \Rightarrow b \in H_1(R)\}| = \begin{cases} N, & \textit{when } b \in E^0; \\ (N-1)/2, & \textit{when } b \in E^1; \\ N-2, & \textit{when } b \in E^2. \end{cases}$$

Proof. The first supposition follows from the definition of the sets. We need to prove the second one. Since there are two bitstrings in $O^{=}$ which are neighbours of $b \in E^2$, then the other $N-2$ different bitstrings neighbouring b are in either $O^{>}$ or $O^{<}$. Let us denote them by $O_b^{>}$ and $O_b^{<}$, respectively. Since the set R is reverse-closed, then every bitstring in $O^{<}$ corresponds to one which belongs to both $O^{>}$ and R. Neither one of the reverses of the bitstrings in $O_b^{<}$ concurs with any of the bitstrings in $O_b^{>}$ as proven in Proposition 1. The same holds for a bitstring $b \in E^0$. When $b \in E^1$ the difference is that every bitstring in $O_b^{<}$ concurs with one of the reverses of the ones in $O_b^{>}$. □

The above-mentioned sets have the following cardinalities:

$$\begin{aligned} |O| = |E| &= 2^{N-1}; \\ |O^{=}| &= 2^{(N-1)/2}; \\ |O^{>}| = |O^{<}| &= 2^{N-2} - 2^{(N-3)/2}; \\ |E^2| &= (N-1)2^{(N-3)/2}; \\ |E^1| &= 2^{(N-1)/2}; \\ |E^0| &= 2^{N-1} - (N+1)2^{(N-3)/2}. \end{aligned}$$

Theorem 4. *The probability for a fixed bitstring b to be in the* neighbourhood *of R is equal to*

$$\mathbb{P}(b \notin H_1(S) | b \in E) = \left(\frac{\sum \binom{|O^{=}|-2}{i} \binom{|O^{>}|-(N-2)}{(M-i)/2}}{\sum \binom{|O^{=}|}{i} \binom{|O^{>}|}{(M-i)/2}} \right) \frac{N-1}{2^{(N+1)/2}} +$$

$$+\left(\frac{\sum\binom{|O^=|-1}{i}\binom{|O^>|-(N-1)/2}{(M-i)/2}}{\sum\binom{|O^=|}{i}\binom{|O^>|}{(M-i)/2}}\right)\frac{1}{2^{(N-1)/2}}+$$

$$+\left(\frac{\sum\binom{|O^=|}{i}\binom{|O^>|-N}{(M-i)/2}}{\sum\binom{|O^=|}{i}\binom{|O^>|}{(M-i)/2}}\right)\left(1-\frac{N+1}{2^{(N+1)/2}}\right),$$

where we sum over i from 0 to $|O^=|$ such that $2|(M-i)$.

Proof. We seek to find the number of all possible reverse-closed sets of bitstrings with odd density. First, we find the number of ways to choose i bitstrings from $O^=$ and $(M-i)/2$ bitstrings from $O^>$ (the other $(M-i)/2$ bitstrings are the corresponding ones in $O^<$). We multiply the two numbers in order to compute the number of ways to uniformly randomly choose the set R. We sum this with respect to the values of i that have the same parity as M (the number of bitstrings from $|O^>|$ and $|O^<|$ is equal, hence the sum is an even number). This shows that the number $N(S)$ of possible sets S is

$$N(S)=\sum_{\substack{i=0\\2|(M-i)}}^{|O^=|}\binom{|O^=|}{i}\binom{|O^>|}{(M-i)/2}.$$

The number of sets which are not in the neighbourhood of R depends on whether b is in E^2, E^1 or E^0. We denote the numbers of these sets by $N(S^2)$, $N(S^1)$ and $N(S^0)$, respectively. We use the same method as when we computed $N(S)$. However, this time we need to subtract the number of bitstrings in the neighbourhood of S. We know these numbers thanks to Lemma 1. Therefore, we obtain the result:

$$N(S^0)=\sum_{\substack{i=0\\2|(M-i)}}^{|O^=|-2}\binom{|O^=|-2}{i}\binom{|O^>|-(N-2)}{(M-i)/2}$$

$$N(S^1)=\sum_{\substack{i=0\\2|(M-i)}}^{|O^=|-1}\binom{|O^=|-1}{i}\binom{|O^>|-(N-1)/2}{(M-i)/2}$$

$$N(S^2)=\sum_{\substack{i=0\\2|(M-i)}}^{|O^=|}\binom{|O^=|}{i}\binom{|O^>|-N}{(M-i)/2}.$$

Using the basic properties of probability we have

$$\mathbb{P}(b \notin H_1(S)|b \in E) = \sum_{i=0}^{2} \mathbb{P}(b \notin H_1(R)|b \in E^i)\mathbb{P}(b \in E^i|b \in E).$$

We then divide the number of favorable outcomes by the total number of possible outcomes and obtain

$$\mathbb{P}(b \notin H_1(R)|b \in E^i) = N(S^i)/N(S)$$

and

$$\mathbb{P}(b \in E^i|b \in E) = |E^i|/|E|.$$

Combining the formulas for $N(S^i)$ with the above equations, we achieve the desired result. □

This formula is different from the one conjectured by Lee et al. [6]. They have suggested the following formula:

$$\begin{aligned}\mathbb{P}(b \notin H_1(S)|b \in E) = {} & \left(\sum \frac{\binom{|O^=|-2}{i}}{\binom{O^=}{i}} \frac{\binom{|O^>|-(N-2)}{(M-i)/2}}{\binom{|O^>|}{(M-i)/2}} \right) \frac{N-1}{2^{(N+1)/2}} + \\ & + \left(\sum \frac{\binom{|O^=|-1}{i}}{\binom{O^=}{i}} \frac{\binom{|O^>|-(N-2)/2}{(M-i)/2}}{\binom{|O^>|}{(M-i)/2}} \right) \frac{1}{2^{(N-1)/2}} + \\ & + \left(\sum \frac{\binom{|O^>|-N}{(M-i)/2}}{\binom{|O^>|}{(M-i)/2}} \right) \left(1 - \frac{N+1}{2^{(N+1)/2}} \right),\end{aligned}$$

where in each of the sums the summation is over all $i = 0, 1, \dots, M$ such that $(M-i)/2$ is an integer.

In the following tables one can see that this formula is significantly different from the one in [6]. Moreover, computational results show that the original proposition, provided in [6] without proof, becomes negative for values greater than 16.

N	# irreducible polynomials over $\mathbb{F}_2$	M	probability with Φ_1	probability with Φ_2
5	9	10	0.008333	0.006466
7	30	32	0.045619	0.009787
9	99	102	0.111236	0.012428
11	335	341	0.224294	0.013156
13	1161	1170	0.449207	0.013559
15	4080	4096	0.906438	0.0139
17	14532	14563	1.841681	0.014242
19	52377	52428	3.749562	0.014566

The expectation is calculated using the formula

$$\begin{aligned}\mathbb{E}|H_1(R)| &= |E|\mathbb{P}(b \in H_1(R)) \\ &= 2^{N-1}(1 - \mathbb{P}(b \notin H_1(R))).\end{aligned}$$

N	M	expectation with Φ_1	expectation with Φ_2	# polynomials with distance 1 from the irreducible
5	10	28.25824	15.86667	16
7	32	61.08037	63.37365	63
9	102	227.5236	252.8185	255
11	341	794.3226	1010.529	1020
13	1170	2256.048	4040.46	4048
15	4096	15329.13	16156.27	16216
17	14563	−55160.4	64602.6	64731
19	52428	−720781	258325	258718

For odd values of N the formula conjectured in the original paper does not always give feasible results as seen from the probability greater than 1 and the negative expectation in some cases. In comparison the formula derived in this paper provides results which seem closer to the actual data about the set of irreducible polynomials.

4. Conclusion. While we provide formulas for the expectation that a polynomial is at distance 1 from the sets R, several important questions remain. An exhaustive computational approach is needed to compare this to the actual data for the irreducibles which can possibly indicate whether the chosen properties are sufficient, or new ones are needed. The most likely such candidate is *non-divisibility by second degree irreducible polynomials.* Formulas for greater distances can also be of interest.

Acknowledgements. The authors would like to extend their gratitude to EUCYS for the opportunity to present the initial form of this project, and to Prof. P. Boyvalenkov and Dr. Y. Ollivier for their useful remarks and advice. Special thanks are due to Prof. I. Landjev for his extended review.

REFERENCES

[1] Banerjee P., M. Filaseta. On a polynomial conjecture of Pál Turán. *Acta Arith.*, **143** (2010), No 3, 239–255. MR2652578.

[2] Bérczes A., L. Hajdu. Computational experiences on the distances of polynomials to irreducible polynomials. *Math. Comp.*, **66** (1997), No 217, 391–398. MR1377660 (97c:11035).

[3] Bérczes A., L. Hajdu. On a problem of P. Turán concerning irreducible polynomials. In: Number Theory: Diophantine, Computational and Algebraic Aspects. Proc. of the Int. Conf. held in Eger, Hungary, July 29–August 2, 1996 (Eds K. Győry, A. Pethő, V. T. Sós). Berlin, Walter de Gruyter, 1998, 95–100. MR1628834 (99f:11032).

[4] Cattell K., F. Ruskey, J. Sawada, M. Serra, C. R. Miers. Fast algorithms to generate necklaces, unlabeled necklaces, and irreducible polynomials over GF(2). *J. Algorithms*, **37** (2000), No 2, 267–282. MR1788836 (2002a:05006).

[5] Filaseta M., M. Mossinghoff. The distance to an irreducible polynomial, II. *Mathematics of Computation*, **81** (2012), No 279, 1571–1585.

[6] Lee G., F. Ruskey, A. Williams. Hamming distance from irreducible polynomials over $\mathbb{F}_2$. In: Discrete Math. Theor. Comput. Sci. Proc., **AH** (2007), 169–180. MR2509520 (2010i:94124).

[7] Mossinghoff M. J. The distance to an irreducible polynomial. In: Gems in Experimental Mathematics (Eds T. Amdeberhan, L. A. Medina, V. H. Moll). *Contemp. Math.*, Amer. Math. Soc., Providence, RI, **517** (2010), 275–288. MR2731083.

[8] Schinzel A. Reducibility of lacunary polynomials, II. *Acta Arith.*, **16** (1970), 371–392. MR0265323 (42:233).

[9] Schinzel A. Reducibility of polynomials and covering systems of congruences. *Acta Arith.*, **13** (1967), 91–101. MR0219515 (36:2596).

Konstantin Delchev
Institute of Mathematics and Informatics
Bulgarian Academy of Sciences
Acad. G Bonchev St, bl. 8
1113 Sofia, Bulgaria
e-mail: `math_k_delchev@yahoo.com`

Petar Gaydarov
University of Cambridge
St John's College, CB2 1TP
Cambridge, UK
e-mail: `peter.gaydaro@gmail.com`