Patterson-Wiedemann Type Functions on 21 Variables

Selçuk Kavut and Subhamoy Maitra

Abstract

Nonlinearity is one of the most challenging combinatorial property in the domain of Boolean function research. Obtaining nonlinearity greater than the bent concatenation bound for odd number of variables continues to be one of the most sought after combinatorial research problems. The pioneering result in this direction has been discovered by Patterson and Wiedemann in 1983 (IEEE-IT), which considered Boolean functions on $5 \times 3 = 15$ variables that are invariant under the actions of the cyclic group $GF(2^5)^* \cdot GF(2^3)^*$ as well as the group of Frobenius authomorphisms. Such Boolean functions posses nonlinearity greater than the bent concatenation bound, namely $2^{n-1} - 2^{\frac{n-1}{2}}$. The next possible option for obtaining such functions is on $7 \times 3 = 21$ variables. However, obtaining such functions remained elusive for more than three decades even after substantial efforts as evident in the literature. In this paper, we exploit combinatorial arguments together with heuristic search to demonstrate such functions for the first time.

Index Terms

Covering Radius, First Order Reed-Muller Code, Nonlinearity, Patterson-Wiedemann Type Functions.

I. Introduction

CONSTRUCTING Boolean functions on odd number of variables n having nonlinearity greater than the bent concatenation bound $2^{n-1}-2^{\frac{n-1}{2}}$ is one of the most difficult problems in the area of cryptography, coding theory, and combinatorics. In 1983, Patterson and Wiedemann discovered [12] for the first time such Boolean functions for $n = 15$ using some combinatorial techniques and search methods together. The search space could be reduced substantially in [12] by considering several group actions as described later. The crux of the observation was that 15 can be written as 5×3, product of two primes. The next attempt in this direction should have been for $21 = 7 \times 3$, but that could not be achieved till that. In this paper, for the first time, we demonstrate such functions for 21 variables that could not be achieved for more than 32 years.

At this point, let us refer to the most important results in the field of maximum nonlinearity of Boolean functions on odd number of variables with the time-line. The time-line is mentioned here to highlight that this problem is indeed challenging as only a few results appeared in a substantially long duration even after a lot of attention to these problems. The main challenge is to check whether it is possible to overcome the bent concatenation bound. The bent functions are also combinatorially challenging and well studied class with application in coding theory and cryptography. These functions exist for even number of variables m having the provably maximum possible nonlinearity $2^{m-1} - 2^{\frac{m}{2}-1}$. Consider two m-variable bent functions f_0, f_1 and then construct an $n = m+1$ variable Boolean function F as $(1 \oplus x_{m+1})f_0 \oplus x_{m+1}f_1$. It can be easily checked that the nonlinearity of F is $2^{n-1} + 2^{\frac{n-1}{2}}$. Since the 2^n bit long truth table of F is concatenation of the two 2^m length truth tables of the bent functions f_0, f_1, this nonlinearity of F is called the bent concatenation nonlinearity.

- **1972:** In [2], it has been shown that for $n = 5$, the maximum nonlinearity of n-variable Boolean functions is the bent concatenation nonlinearity, which is 12.
- **1980:** In [11], the question for $n = 7$ could be solved and it has been noted that here also the maximum nonlinearity is the bent concatenation nonlinearity which is 56.
- **1983:** In [12], the seminal positive result has been presented by Patterson and Wiedemann showing that one can construct a 15-variable Boolean function f with nonlinearity $2^{15-1} - 2^{\frac{15-1}{2}} + 20 = 16276$. It is well known that using this function, one can construct any n-variable Boolean function F with nonlinearity

S. Kavut is with the Computer Engineering Department, Balıkesir University, Balıkesir, Turkey. Email: skavut@balikesir.edu.tr
S. Maitra is with the Applied Statistics Unit, Indian Statistical Institute, 203 B. T. Road, Kolkata 700108, India. E-mail: subho@isical.ac.in

$2^{n-1} - 2^{\frac{n-1}{2}} + 20 \cdot 2^{\frac{n-15}{2}}$ for $n > 15$. In fact, F can be written as $f \oplus g$, where g is an $(n-15)$-variable bent function.

- **2006-2010:** The 9-variable Boolean functions having nonlinearity 241 were identified [7] in the rotation-symmetric class and subsequently this result is improved [8] to 242 by considering the k-rotation-symmetric class. Thus, for $n = 9, 11, 13$, one can obtain Boolean functions having nonlinearity $2^{n-1} - 2^{\frac{n-1}{2}} + 2 \cdot 2^{\frac{n-9}{2}}$.

OUR CONTRIBUTION: The kind of constraints considered for constructing the 15-variable functions in [12] finally reduced to solving an integer programming problem on 11 binary variables, which was easy to solve using exhaustive search. However, the authors [12] pointed out the following in their paper:

> "We have not succeeded in understanding algebraically the choice of orbits made in (6) and thus have not succeeded in generalizing our construction to other dimensions although we suspect there is a construction for all $\mathcal{R}_m$ when m is not a prime power."

In fact, the situation is significantly harder for the 21-variable case as in a similar manner of [12] it reduces to an integer programming problem with 115 binary variables. An attempt has been made towards studying this class almost a decade back in [5] without any success. In [5, Page 1551], it has been commented that:

> "The search space corresponding to this case is very large and exhaustive search is infeasible. It will be of interest to develop some heuristic methods to find solutions to this system of linear inequalities."

In this paper we revisit the problem with more disciplined effort and indeed obtained Patterson-Wiedemann type functions on 21 variables having nonlinearity strictly greater than the bent concatenation bound.

CAVEAT: The nonlinearity of the functions f_{21} that we achieve in this paper is $2^{21-1} - 2^{\frac{21-1}{2}} + 61$. Note that this nonlinearity is less than $2^{21-1} - 2^{\frac{21-1}{2}} + 20 \cdot 2^{\frac{21-15}{2}}$ as obtained by composing the Patterson-Wiedemann function f_{15} on 15 variables and a bent function g_6 on 6 variables, i.e., $f'_{21} = f_{15} \oplus b_6$, where the functions f_{15} and b_6 are on distinct variables. However, such functions f'_{21} are not of Patterson-Widemann type and it does not answer the challenge of obtaining such functions on 21 variables beating the bent concatenation bound as posed in [12]. We solve this problem for the first time. Further, we could not make exhaustive search as the integer programming problem in this case is on 115 binary variables. It may very well happen that with the dissemination of our results, the problem may be solved with better efficiency in obtaining solutions with higher nonlinearity. In case such a Patterson-Wiedemann type function on 21-variables with nonlinearity greater than $2^{21-1} - 2^{\frac{21-1}{2}} + 160$ could be obtained, that will provide highest known nonlinearity for all the odd variable Boolean functions for 21 variables and more. That we leave as an open problem in this direction.

Next we provide necessary background in this area. For this, we mostly follow to the explanations in [12], [5]. One may also note that several modifications of Patterson-Wiedemann type functions have been studied in literature as evident from [4], [10], [13] and the references therein.

A. Background

By $\mathcal{F}_n$, let us denote the set of Boolean functions from $GF(2^n)$ to $GF(2)$ and consider $f \in \mathcal{F}_n$. The support of f is defined as $Supp(f) = \{x \in GF(2^n) | f(x) = 1\}$, and its weight is $wt(f) = |Supp(f)|$. Let a and b be two distinct odd primes such that $n = ab$. Let $M = GF(2^{ab})$, $L = GF(2^a)$, $J = GF(2^b)$ and $K = GF(2)$. Now consider the tower of subfields $K \hookrightarrow L \hookrightarrow M$. The index of the multiplicative group L^* in M^* is $m = \frac{2^{ab}-1}{2^a-1}$. One may note that M^* can be written as $M^* = \cup_{i=1}^{m} L^* x_i$ where $\{x_1, x_2, \ldots, x_m\}$ (the complete set of coset representatives of L^* in M^*). It is well known that one can characterize any function from $M \to K$ by specifying its support. Let us consider functions in $\mathcal{F}_n$ whose supports are of the form $\cup_{i=1}^{l} L^* x_i$ for some positive integer l. Such functions have been considered by Dillon [3] towards the construction of bent functions and formal proofs could be devised using this idea to show that such functions provide the best known nonlinearity for even number of variables. Naturally, this idea was later exploited by Patterson and Wiedemann [12] to explore high nonlinearity for odd number of variables. Though there had been no clear proof of nonlinearity for odd number of variables as accepted in [12], such ideas along with some additional techniques produced functions with nonlinearity greater than bent concatenation bound.

We denote the set of all such functions by $I_{a,b}$. Any linear function in $\mathcal{F}_{ab}$ can be expressed as $l_\alpha(x) = Tr_1^{ab}(\alpha x)$ where $\alpha \in M$ and $Tr_1^n(x) = x + x^2 + x^{2^2} + \ldots + x^{2^{n-1}}$ for all $x \in GF(2^n)$. The support of l_α is $Supp(l_\alpha) = \{x \in$

$M|Tr_1^{ab}(\alpha x) = 1\}$, and the support of the affine function $h_\alpha(x) = l_\alpha(x) + 1$ is $Supp(h_\alpha) = \{x \in M|Tr_1^{ab}(\alpha x) = 0\}$. Let $H_\alpha = Supp(h_\alpha)$, which is a hyperplane in M when considered as a vector space over K.

Next we define the Hadamard transform of $f \in \mathcal{F}_n$ as

$$\hat{f}(\lambda) = \sum_{x \in GF(2^n)} (-1)^{f(x)+Tr(\lambda x)}.$$

The nonlinearity of a Boolean function f is defined as

$$nl(f) = 2^{n-1} - \frac{1}{2} \max_{\lambda \in GF(2^n)} |\hat{f}(\lambda)|.$$

We can write the elements of $GF(2^n)$ in some order, say $\{\alpha_0, \alpha_1, \ldots, \alpha_{2^n-1}\}$. For $f, g \in \mathcal{F}_n$, we define the distance $d(f, g)$ between f, g as the Hamming distance between the 2^n-dimensional vectors $(f(\alpha_0), f(\alpha_1), \ldots, f(\alpha_{2^n-1}))$ and $(g(\alpha_0), g(\alpha_1), \ldots, g(\alpha_{2^n-1}))$. It is easy to see that if $f, g \in \mathcal{F}_n$ then $d(f, g) = |Supp(f)\Delta Supp(g)|$ where Δ is the symmetric difference between the sets $Supp(f)$ and $Supp(g)$. One may calculate that (see also [12]) if $Supp(f) = \cup_{i=1}^{l} L^* x_i$ then

$$\begin{aligned} d(f, \mathbf{0}) &= l(2^a - 1), \\ d(f, \mathbf{1}) &= 2^{ab} - l(2^a - 1), \\ d(f, h_\alpha) &= 2^{ab-1} - 2^a \cdot t(\alpha) + l, \\ d(f, l_\alpha) &= 2^{ab-1} + 2^a \cdot t(\alpha) - l, \end{aligned} \tag{1}$$

where $\mathbf{0}$ and $\mathbf{1}$ are constant functions with all 0 values and all 1 values respectively. Further, $t(\alpha)$ is the number of cosets of the form $L^* x_i$ that are totally contained in the hyperplane H_α, equivalently $t(\alpha)$ is the number of x_i for which $Tr_a^{ab}(x_i \alpha) = 0$. Thus, the nonlinearity of f is given by

$$nl(f) = \min_{\alpha \in M} \{l(2^a - 1), 2^{ab} - l(2^a - 1), 2^{ab-1} - 2^a \cdot t(\alpha) + l, 2^{ab-1} + 2^a \cdot t(\alpha) - l\}.$$

For an $f \in I_{a,b}$ with $nl(f) > 2^{ab-1} - 2^{(ab-1)/2}$, considering (1), we obtain that l and $t(\alpha)$ must satisfy:

$$\frac{2^{ab-1} - 2^{(ab-1)/2}}{2^a - 1} < l < \frac{2^{ab-1} + 2^{(ab-1)/2}}{2^a - 1}, \tag{2}$$

$$\frac{1}{2^a}\left\{\frac{2^{ab-1} - 2^{(ab-1)/2}}{2^a - 1} - 2^{(ab-1)/2}\right\} < t(\alpha) < \frac{1}{2^a}\left\{\frac{2^{ab-1} + 2^{(ab-1)/2}}{2^a - 1} + 2^{(ab-1)/2}\right\}. \tag{3}$$

Consider the two subgroups L^* and J^* in M^*. The intersection of L^* and J^* contains only the identity element and the group M^* is an abelian group. Hence the product $L^* \cdot J^*$ is direct. One can identify the group M^* to the group $\Phi(M^*)$ of left multiplications by the elements of M^* in $GL_K(M)$ and this correspondence is an isomorphism.

Let $\phi_2 \in GL_K(M)$ be the Frobenius automorphism of M. This is defined by $\phi_2(x) = x^2$ for all $x \in M$. The group $\langle \phi_2 \rangle$ generated by ϕ_2 is a cyclic group of order ab, which is contained in $GL_K(M)$. Patterson and Wiedemann [12] considered the action of the group $G = [\Phi(L^*) \cdot \Phi(J^*)]\langle \phi_2 \rangle / \Phi(L^*)$, where $[\Phi(L^*) \cdot \Phi(J^*)]\langle \phi_2 \rangle$ is the semidirect product of $\Phi(L^*) \cdot \Phi(J^*)$ by $\langle \phi_2 \rangle$.

The result of [12] was for $n = 15$. Here $L^* = GF(2^5)^*$ and $J^* = GF(2^3)^*$. The support of the function f is invariant under the action of L^* and J^*. That is, the support of f is invariant under the action of the product $L^* \cdot J^*$ which is also a cyclic subgroup of $M^* = GF(2^{15})^*$ of order $(2^5 - 1)(2^3 - 1) = (31)(7) = 217$. All the elements of $L^* \cdot J^*$ should have the same value and it is also true for the elements in each of its cosets in M^*. Note that $\frac{2^{15}-1}{217} = 151$. In the form of interleaved sequence [6], [5], this can be seen as 217 rows of length 151 each. The value in each column is the same. Thus one row of 151 elements will define the complete Boolean function at $2^{15} - 1$ inputs. Generally we consider the output as zero for the all zero input point. Next comes the constraint that the function f is invariant under Frobenius automorphism, i.e., $f(x) = f(x^2)$ for all $x \in M^*$. Due to this, the 151 elements are divided into 10 groups of size 15 each and one group of size 1. One can initially assign the output zero corresponding to the inputs of the group of size 1. Due to the weight conditions, it is evident that inputs corresponding to the 5 groups should have the output 0 and rest should have the output 1. Thus, we need to search $\binom{10}{5} = 252$ many different Boolean functions on 15 variables. As described in [12], two distinct functions with nonlinearity 16276 could be obtained in this class up to affine equivalence. The problem could also be seen as solutions to certain specific inequalities as explained in [5, Pages 1549-1550].

II. Patterson-Wiedemann type construction on 21 variables

In this section, we consider the case for $n = 21 = 7 \cdot 3$. As explained in [5], we have $L^* = GF(2^7)^*$ and $J^* = GF(2^3)^*$ here. Now $L^* \cdot J^*$ is a cyclic subgroup of $M^* = GF(2^{21})^*$ of order $(2^7 - 1)(2^3 - 1) = (127)(7) = 889$. Note that $\frac{2^{21}-1}{889} = 2359$. That is, we can consider the interleaved sequence, where we have 889 similar rows and each row has 2359 elements. Further, the function f is invariant under Frobenius automorphism, i.e., $f(x) = f(x^2)$ for all $x \in M^*$. Thus, 2359 elements are divided as 112 groups of size 21, 2 of size 3 and 1 of size 1. Thus, we have total 115 binary variables here. We consider the following preliminary things towards obtaining a solution.

- We consider that the outputs corresponding to the group of size 1 as zero.
- For satisfying the weight conditions, we need the following.
  - The inputs corresponding to the 56 groups of size 21 should have the output 0 and the rest 56 should have the output 1.
  - The inputs corresponding to one group of size 3 should have the output 0 and the other one should have the output 1.

With these constraints, we have $2\binom{112}{56}$ many options to search which is computationally infeasible. For each option one may get back to the Boolean function of 21 variables and check the nonlinearity. However, this checking is time consuming and following [5, Remark 3] a much better strategy is to consider the concept of inequalities as explained in [5, Section 2.1] for $n = 15$. We implement the strategy for $n = 21$ and generate the inequalities as completely described in Appendix A. We consider the degree 21 primitive polynomial $z^{21} + z^2 + 1$ over GF(2) for realization of the field. Following (3) and putting the values $a = 7, b = 3$, we obtain $57 \leq t(\alpha) \leq 72$. Thus, the overall inequality is of the form $57 \leq \sum_{i=0}^{114} A_{i,j}x_i \leq 72$, where each i, j varies from 0 to 114 (total 115 elements) and there are 115 binary variables x_0 to x_{114}. The coefficient matrix $[A_{i,j}]$ is described in Appendix A. We like to present an observation here that if one leaves the rows and columns indexed by 0, 93 and 114, then the resulting 112×112 matrix becomes symmetric.

A. The functions that we obtained

As described above, we consider binary strings of length 115 such that the 0-th location is 0, one location (out of two locations corresponding to the groups of size 3) is 0 and the other is 1 and finally 56 locations (out of 112 locations corresponding to the groups of size 21) are 0 and the rest 56 are 1. This is clearly an integer programming problem, and seems to be a hard one given its dimension. Thus, we have attempted several heuristics and with one such heuristic, described in Section II-B, we obtain 4 solutions with an effort of more than a month. The solutions are as below. All the solutions have nonlinearity $2^{21-1} - 2^{\frac{21-1}{2}} + 61 = 1047613$ and the minimum absolute values in the autocorrelation spectra (see [5], [9] for details of autocorrelation spectra) of these four functions are 2948, 3436, 3940, 5116. These different values show that the functions are not affine equivalent.

```
01110001101111100111111010100100101110011101000010011000101001001011100010000010001001110100101110111011001111100000
01100101001110010101100001000011001111100111001110000001000101000001111110011100001111110010010111111001010011111000
01110100101000011111100100110010101111110001110011010010110011001100100001111101110110100000000000000101111001
00110001111001011110001000110110101000011010101010110111111110100010000010101110101001010001011001100101011100101000
```

B. Details of our heuristic

We utilize the steepest-descent-like iterative search algorithm [1] to attain the solutions. In this search algorithm, the current solution is replaced by another solution from a predefined set of candidate solutions, called the neighborhood. Using a cost function, the entire neighborhood is evaluated in each step of the search algorithm. The best one, i.e., the one with the best cost value, is then delivered to the next step, even if it is worse than the previously selected best solutions. To prevent the algorithm from looping, these best solutions are stored in memory.

Here, we constitute the neighborhood by swapping two dissimilar bits (of the current solution) at the locations corresponding to the groups of the same size. This is required to satisfy the necessary condition on the weight of the Boolean function. Hence, the neighborhood consists of 6272 ($= 2 \times 56^2$) candidate solutions and for each of them we compute the cost function that we define as the sum of squared errors, which is a measure of the distance of a solution from the inequality bounds. Let $x = (x_0, x_1, \ldots, x_{114})$ be a solution, $A_i = (A_{i,0}, A_{i,1}, \ldots, A_{i,114})$

be the i-th row of the coefficient matrix, and $(x, A_i) = \sum_{j=0}^{114} A_{i,j}x_j$ be the inner product of x and A_i, where $i = 0, 1, \ldots, 114$. The cost function is then given by

$$\begin{aligned} Cost(x) &= \sum_{i=0}^{114} C_i^2, \text{ where} \\ C_i &= \begin{cases} |(x, A_i) - 72|^2 & \text{if } (x, A_i) > 72, \\ 0 & \text{if } 57 \leq (x, A_i) \leq 72, \\ |(x, A_i) - 57|^2 & \text{if } (x, A_i) < 57. \end{cases} \end{aligned}$$

Thus, any solution x with zero cost value provides a 21-variable Patterson-Wiedemann type Boolean function having nonlinearity greater than the bent concatenation nonlinearity. The search algorithm stops after a fixed number of iterations, which we set to 40000 in our experiments. The search is performed on a workstation with Intel Xeon CPU E5-1650v3 (15M Cache, 3.50 GHz, 6 cores) and 16 GB RAM under Windows 7 Professional 64-bit operating system. Exploiting all the cores, it took more than a month to extract the aforementioned 4 solutions.

III. Conclusion

In this paper, for the first time, we could demonstrate Patterson-Wiedemann type Boolean functions on 21 variables that exceed the bent concatenation nonlinearity. This problem remained open for more than three decades even after significant effort as evident from the literature. We deploy heuristics to obtain such functions that can be seen as solutions to an integer programming problem with 115 binary variables. Indeed the problem is quite hard and exhaustive search seems to be elusive given the present computational power. The functions we obtain are of nonlinearity $2^{21-1} - 2^{\frac{21-1}{2}} + 61 = 1047613$ and we believe that further search effort may discover solutions with better nonlinearity. In this direction, we provide every details of the inequalities that need to be satisfied to obtain the solutions.

References

[1] M. Bartholomew-Biggs. Chapter 5: The Steepest Descent Method, pp. 51–64. Nonlinear Optimization with Financial Applications. Springer, 2005.

[2] E. R. Berlekamp and L. R. Welch. Weight distributions of the cosets of the $(32, 6)$ Reed-Muller code. IEEE Transactions on Information Theory, IT-18(1):203–207, January 1972.

[3] J. F. Dillon. Elementary Hadamard difference sets. In Proceedings of 6th S. E. Conference of Combinatorics, Graph Theory, and Computing. Utility Mathematics, Winnipeg, pp. 237-249, 1975.

[4] C. Fontaine. On some cosets of the First-Order Reed-Muller code with high minimum weight. IEEE Transactions on Information Theory, 45(4), pp. 1237-1243, (1999)

[5] S. Gangopadhyay, P. H. Keskar and S. Maitra. Patterson-Wiedemann construction revisited. Discrete Mathematics, Volume 306, Issue 14, pp. 1540-1556 (2006)

[6] G. Gong. Theory and applications of q-ary interleaved sequences. IEEE Transactions on Information Theory, 41(2), pp. 400-411, (1995)

[7] S. Kavut, S. Maitra and M. D. Yücel. Search for Boolean functions with excellent profiles in the rotation symmetric class. IEEE Transactions on Information Theory, Volume IT-53(5), pp. 1743-1751 (2007)

[8] S. Kavut and M. D. Yücel. 9-variable Boolean functions with nonlinearity 242 in the generalized rotation symmetric class. Information and Computation, Volume 208, No. 4, pp. 341-350 (2010)

[9] S. Kavut. Correction to the paper: Patterson-Wiedemann construction revisited. Discrete Applied Mathematics. Available online 1 September 2015.
http://www.sciencedirect.com/science/article/pii/S0166218X15004126

[10] S. Maitra and P. Sarkar. Modifications of Patterson-Wiedemann functions for cryptographic applications. IEEE Transactions on Information Theory, Volume 48(1), pp. 278-284, (2002)

[11] J. J. Mykkeltveit. The covering radius of the $(128, 8)$ Reed-Muller code is 56. IEEE Transactions on Information Theory, IT-26(3):358–362, (1983).

[12] N. J. Patterson and D. H. Wiedemann. The covering radius of the (2^{15}, 16) Reed-Muller code is at least 16276. IEEE Transactions on Information Theory, Volume IT-29(3), pp. 354-356 (1983) (See also correction in Volume IT-36(2), p. 443 (1990))

[13] S. Sarkar and S. Maitra. Idempotents in the neighbourhood of Patterson-Wiedemann functions having Walsh spectra zeros. Designs, Codes and Cryptography, Volume 49, Issue 1-3, pp. 95-103 (2008)

APPENDIX A: THE COEFFICIENT MATRIX A

We have the 115×115 coefficient matrix, where each location contains an integer. The single digit integers 0 to 9 are written as they are. For brevity, the two digit values 14 and 21 are written as u and v respectively.