*Sixth International Workshop on Optimal Codes and Related Topics*
*June 16-22, 2009, Varna, Bulgaria* *pp. 173-178*

# Some notes on the binary Gilbert-Varshamov bound

Dejan Spasov dejan@ii.edu.mk
Marjan Gusev marjan@ii.edu.mk
Institute of Informatics, Faculty of Natural Sciences,
Ss. Cyril and Methodius University, Skopje, MACEDONIA

**Abstract.** Given a linear code $[n, k, d]$ with parity check matrix $H$, we provide inequality that supports existence of a code with parameters $[n + l + 1, k + l, d]$. We show that this inequality is stronger than the Gilbert-Varshamov (GV) bound even if the existence of the code $[n, k, d]$ is guaranteed by the GV bound itself.

## 1 Introduction

Let $H$ is the parity check matrix of some binary code $[n-1, k-1, d]$. Let $H(n-k, d-2)$ denotes the set of all unique $(n-k)$-tuples that are linear combination of $d-2$ columns of $H$. Then a code with parameters $[n, k, d]$ does exist provided

$$|H(n-k, d-2)| \leq 2^{n-k} - 2 \tag{1}$$

Let $B_k(n, d)$ denotes the size of the largest (optimal) linear code over $F_2^n$. Then the existence of $[n-1, k-1, d]$ lower-bounds $B_k(n, d)$

$$B_k(n, d) \geq n - \lceil \log |H(n-k, d-2)| \rceil \tag{2}$$

We will write $V(n, d-2)$ to denote the number of all combinations of $d-2$ or less elements from an $n$-element set, namely

$$V(n, d-2) = \sum_{i=0}^{d-2} \binom{n}{i} \tag{3}$$

In coding theory this quantity is known as *the volume of a Hamming sphere with radius* $d$ in $F_2^n$. Since $|H(n-k, d-2)|$ cannot be larger than $V(n, d-2)$ we get an estimate of (1)

$$V(n, d-2) \leq 2^{n-k} - 2 \tag{4}$$

known as the Varshamov's lower bound. If the triplet $n$, $k$, and $d$ satisfies (4), then the code $[n, k, d]$ does exist [3]. Motivated by recent improvements of (4) by linear factor with code length $n$, [1], [2], we will present the following result:

The code $[n, k, d]$ can be extended to a code with parameters $[n + l + 1, k + l, d]$ provided that the following holds true

$$\sum_{i=1,\ i\ \text{odd}}^{\min(l,d-2)} \binom{l}{i} V\left(n, d-2-i\right) \leq 2^{n-k} \tag{5}$$

where the summation is over the odd values of the index $i$. To support the existence of the code $[n, k, d]$ we can use the GV bound (4) or we can use (5) recursively as long as $n - k > d - 1$. The stopping criterion for the recursion is the repetition code $[d, 1, d]$. In both cases (5) guarantees existence of a better code than (4).

It has been known that a simple greedy algorithm will produce a linear code with parameters that are at least as good as the parameters that satisfy (4). The greedy algorithm searches over $F_2^{n-k}$ and outputs a parity check matrix $H$. We will show that code parameters obtained by running the greedy algorithm also satisfy (5).

In the following section we will put side by side (5) with some previously known results on the GV bound. In Section III we will present constructive proof of (5). Then we will show additional improvement of the term $V\left(n, d-2-i\right)$. The main idea behind these improvements is to count only once as many linear combinations from $H\left(n-k, d-2\right)$ as possible.

## 2 Comparison with prior results

We will start with the GV bound. The inequality

$$2 \cdot \sum_{i=1,\ i\ \text{odd}}^{\min(l,d-2)} \binom{l}{i} V\left(n, d-2-i\right) \leq V\left(n+l, d-2\right) \tag{6}$$

holds true for any $n$ and $d$. This can be proved by applying the Vandermonde's convolution formula on the right-hand side of (6). From (6) we conclude that (5) improves the Varshamov's inequality (4).

Elia [5] reported the following result: Let the code $[n-2, k-1, d]$ does exist. Then the code $[n, k, d]$ does exist too, provided

$$V\left(n-2, d-3\right) \leq 2^{n-k-1} \tag{7}$$

If we restrict $l$ to be at most 1, $l \leq 1$, then (5) is precisely Elia's result. Moreover, letting $l \leq 2$ we obtain improvement of (7); namely, assuming prior existence of the code $[n-3, k-2, d]$, the code $[n, k, d]$ does exist if the following holds true

$$V\left(n-3, d-3\right) \leq 2^{n-k-2} \tag{8}$$

Barg, Guritman, and Simonis [7] reported the following remark: The code $[n, k, d]$ with covering radius $\rho \leq d-2$ can be extended to $[n+d-\rho-1, k+1, d]$. In this context, if the covering radius of $[n, k, d]$ is strictly less than $d-2$, then (5) guarantees existence of the trivial lengthening $[n+1, k, d]$. However if we have prior knowledge of the covering radius, we can modify (5) so that we obtain at least the same result as in [7]. For example, similar to (8), we can extend *remark 13* from [7], i.e. if

$$V(n, \alpha) \leq 2^{n-k-1}$$

for some $\alpha \leq d-1$, then any $[n, k, d]$ code can be extended to an $[n+d-\rho, k+2, d]$ code. For $\alpha = d-3$ this reduces to (8).

Jiang and Vardy have developed a graph-theoretic approach to asymptotically improve the GV bound for nonlinear codes [1], [2]. They were able to show that the code $(n, M, d)$ does exist provided

$$c\frac{V(n, d-1)}{n} \leq 2^{n-\lceil \log_2 M \rceil} \tag{9}$$

where the constant $c$ is at least $1/2 + o(1)$, as reported in [2]. How does (5) compares with (9)? In our case, we were unable to prove that the left-hand of (6) is smaller than the right-hand by a factor $n$. Hence, one may assume that (9) guarantees existence of a code with better parameters than (5). However, in general inequality (9) guarantees existence of a non-linear code, while (5) pertains to the linear codes. Gaborit and Zemor [6] proved that some linear double circulant codes follow (9), but only for code rates of $1/2$ . If a linear code is proved to comply with (9), then (6) and (9) will complement each other. Namely, Jiang and Vardy reported [1] that (9) improves the GV bound when the relative distance $d/n$ is constant. On the other hand, (6) improves the GV bound even when the relative distance $d/n$ approaches to zero.

## 3 The proof

We will consider infinite family of linear codes $[n_m, n_m - m, d]$ with parity check matrices $H_m$, where $m \to \infty$. The parity check matrix for each $m$ is obtained by greedy adding vectors from $F_2^m$ to $H_m$ as long as (1) holds true. Hence, each $[n_m, n_m - m, d]$ code satisfies

$$|H(m, d-2)| = 2^m - 1 \tag{10}$$

The parity check matrix $H_m$ is in systematic form with $H_{m-1}$ embedded in it

$$H_m = \begin{bmatrix} 0 \ldots 0 & 1 & 1 \ldots 1 \\ H_{m-1} & 0 & L_m \end{bmatrix} \tag{11}$$

To prove this, simply pick the first row from $F_2^m$ and keep its value 0, while running the greedy algorithm. This will generate $H_{m-1}$. Add the vector $\begin{bmatrix}1 & 0\end{bmatrix}^T$, then flip the value of the coordinate to 1 and continue as long as (1) is true. The systematic nature of the parity check matrix $H_m$ was first noticed in [4].

**Proposition 1.** *Given a code $[n_m, n_m - m, d]$ with parity check matrix $H_m$. The code $[n_m + l + 1, n_m + l - m, d]$ does exist provided*

$$\sum_{i=1,\ i\ odd}^{\min(l,d-2)} \binom{l}{i} V\left(n_m, d-2-i\right) \leq 2^m - 1 \tag{12}$$

*Proof.* The set $H\left(m+1, d-2\right)$ of $d-2$ linear combinations from $H_{m+1}$ can be divided into three subsets $A$, $B$, and $C$. Set $A$ is made of all $d-2$ linear combinations from $\begin{bmatrix}0\\ H_m\end{bmatrix}$. Set $B$ consists of all $d-2$ linear combinations that include odd number of vectors from $\begin{bmatrix}1 & 1\\ 0 & L_{m+1}\end{bmatrix}$. Similar to $B$, set $C$ consists of all $d-2$ linear combinations that include even number of vectors from $\begin{bmatrix}1 & 1\\ 0 & L_{m+1}\end{bmatrix}$. Every $(d-2)$-linear combination from $H_{m+1}$ belongs to at least one of these sets. However, the first row of the vectors from set $C$ is 0, thus every vector from set $C$ also belongs to set $A$. Hence, we conclude that $H\left(m+1, d-2\right)$ is union of two disjoint sets $A$ and $B$. Since $H_m$ satisfies (10), set $A$ has at most $2^m$ elements. Let $l$ is the number of columns of $\begin{bmatrix}1 & 1\\ 0 & L_{m+1}\end{bmatrix}$. Then the right-hand side of (12) is estimate of $|B|$ □

If we compare results obtained from running the greedy algorithm for small $d$ [8] with the corresponding solutions of (12), we will observe a considerable gap. For example, for $m = 32$ the greedy algorithm will produce the code $[8752, 8720, 5]$, while (12) predicts the existence of $[3186, 3154, 5]$. Can we find a better estimate for the code parameters of the greedy algorithm? Following the idea of counting only once as many linear combinations from $H_m$ as possible we have obtained proposition 2, which improves the term $V\left(n_m, d-2-i\right)$ in (12).

**Proposition 2.** *Given some code $[n_m, n_m - m, d]$ with parity check matrix $H_m$. The code $[n_m + l + 1, n_m + l - m, d]$ does exist provided*

$$\sum_{i=1,\ i\ odd}^{\min(d-2,l)} \binom{l}{i} C_{d-2-i} + V\left(n, d-3\right) - C_{d-3} \leq 2^m \tag{13}$$

*where $C_{d-2-i}$ is the largest term in*

$$C_{d-2-i} = \max_{\max(d-1-i,i)\leq p\leq d-2} \{C_{d-2-i}(p)\} \tag{14}$$

*and*

$$C_{d-2-i}(p) = \sum_{z=d-2-p}^{d-2-i} \sum_{j=0}^{\lfloor \frac{z+p-d+2}{2} \rfloor} \binom{p}{j} \binom{n-p}{z-j} \tag{15}$$

*Proof.* Let the greedy algorithm has generated the matrix $H_m$. Then the algorithm adds the vector $\begin{bmatrix}1 & 0\end{bmatrix}^T$ and continues with the search over the quotient space $\begin{bmatrix}1 & F_2^m\end{bmatrix}^T$. In order to improve (12) we will develop a mechanism that counts only once each vector that is linear combination of $\begin{bmatrix}1 & 0\end{bmatrix}^T$ and $d-3$ columns from $H_m$.

Let a vector $y$ is linear combination of $i$ columns of the matrix $\begin{bmatrix}1 & 1\\ 0 & L_{m+1}\end{bmatrix}$. The same vector $y$ can be represented as linear combination of $p$ vectors from $H_m$,

$$y = a_1 + a_2 + \cdots + a_p \tag{16}$$

Clearly, $p$ cannot be smaller than $d-1-i$, because this will violate the criteria of independence of each $(d-1)$-linear combination of the columns of $H_{m+1}$. On the other hand $p$ cannot be larger than $d-2$ because this would mean that the code $[n_m, n_m - m, d]$ did not satisfy (10). Hence, we can write $d-1-i \leq p \leq d-2$.

Let some vector $x$ is linear combination of at most $d-2-i$ vectors from $H_m$, $x \in H(m, d-2-i)$. We want to find the number $C_{d-2-i}$ of linear combinations $y+x$ that could possibly lead to linear combination of exactly $d-2$ vectors from $H_m$. For example, let $x$ is linear combination of $d-3-p$ columns from $H_m$; then the vector $y+x$ cannot result in a linear combination of exactly $d-2$ vectors from $H_m$. For $i \geq (d-2)/2$ , we have that $p$ must be greater than or equal to $i$ in order to be able to obtain linear combination of exactly $d-2$ columns from $H_m$.

We will denote the columns of $H_m$ as follows

$$H_m = \begin{bmatrix}a_1 & a_2 & \cdots & a_p & h_{p+1} & h_{p+2} & \cdots & h_n\end{bmatrix}$$

where the columns $a_i$ belong to the vector $y$ (16). Let assume that the vector $x$ is linear combination of $z$ vectors from $H_m$, where $d-2-p \leq z \leq d-2-i$. We can say that $x$ is made of $j$ vectors from $a_1, a_2, \ldots, a_p$ and $z-j$ vectors from $h_{p+1}, h_{p+2}, \ldots, h_n$. The resulting vector $y+x$ can possibly be a linear

combination of exactly $d-2$ vectors if $j \leq \left\lfloor \frac{p+z-d+2}{2} \right\rfloor$. Assuming that we have prior knowledge of $p$, we obtain that (15) can be inserted in (12) instead of $V(n_m, d-2-i)$. Since $p$ is not known in advance we must find the largest term $C_{d-2-i}(p)$, namely we must solve (14). Finally, we must add the term $V(n, d-3) - C_{d-3}$ so that each linear combination of $\begin{bmatrix} 1 & 0 \end{bmatrix}^T$ and $d-3$ columns from $H_m$ is counted only once. □

## References


[1] T. Jiang, A. Vardy, Asymptotic improvement of the Gilbert-Varshamov bound on the size of binary codes, *IEEE Trans. Inform. Theory* 50, 2004, 1655-1664.

[2] V. Vu, L. Wu, Improving the Gilbert-Varshamov bound for $q$-ary codes, *IEEE Trans. Inform. Theory* 51, 2005, 3200-3208.

[3] F. J. MacWilliams, N. J. A. Sloane, *The Theory of Error-Correcting Codes*, Amsterdam: North-Holland/Elsevier, 1977, 33-34.

[4] V. I. Levenshtein, A class of systematic codes, *Sov. Math. Dokl.* 1, 1960, 368-371.

[5] M. Elia, Some results on the existence of binary linear codes, *IEEE Trans. Inform. Theory* 29, 1983, 933-934.

[6] P. Gaborit, G. Zemor, Asymptotic improvement of the Gilbert-Varshamov bound for binary linear codes, *IEEE Trans. Inform. Theory* 54, 2008, 3865-3872.

[7] A. Barg, S. Guritman, J. Simonis, Strengthening the Gilbert-Varshamov bound, *Lin. Alg. Appl.* 307, 2000, 119-129.

[8] D. Spasov, Implementing the lexicographic construction, online: http://nislab.bu.edu/nislab/projects/lexicode/index.html