On the Bit Extraction Problem

Joel Friedman*
Department of Computer Science
Princeton University
Princeton, NJ 08544

Abstract

Consider a coloring of the n-dimensional Boolean cube with $c = 2^s$ colors in such a way that every k-dimensional subcube is equicolored, i.e. each color occurs the same number of times. We show that for such a coloring we necessarily have $(k-1)/n \geq \theta_c = (c/2-1)/(c-1)$. This resolves the "bit extraction" or "t-resilient functions" problem (also a special case of the "privacy amplification" problem) in many cases, such as $c-1|n$, proving that XOR type colorings are optimal, and always resolves this question to within $c/4$ in determining the optimal value of k (for any fixed n and c). We also study the problem of finding almost equicolored colorings when $(k-1)/n < \theta_c$, and of classifying all optimal colorings.

1 Introduction

The bit extraction problem of [Vaz85], aka t-resilient functions problem (see [CFG+85]), aka a special case of the privacy amplification problem (see [BBR88],[Bra89]) is equivalent to the following coloring problem. The vertices of the Boolean cube, $\mathbf{B}^n = \{-1, 1\}^n$ are to be colored with $c = 2^s$ colors such that every k-*dimensional subcube* is *equicolored*. Given n and c, what is the smallest value of k for which this possible? Here, by a k-dimensional subcube we mean a subset of $\mathbf{B}^n$ determined by fixing the values of some $n-k$ coordinates on $\mathbf{B}^n$; we denote the set of all such subcubes by $\mathcal{H}_k$. By equicolored we mean that every color appears the same number of times in the subcube, i.e. $2^k/c$ times.

To fix ideas, we refer to such a coloring as a $(c; n, k)$-coloring and write $\kappa(c, n)$ for the smallest value of k achievable for a given n and c. In this paper we are primarily interested in viewing c as small or fixed, studying κ as a function of n. We will study this problem and the problem of constructing colorings which are approximately equicolorable.

To elaborate on the context in which the problem arises, consider n boolean variables, $X = \{x_1, \ldots, x_n\}$, whose values are set by the following process. An adversary fixes some subset of the variables, $T \subset X$, to specific values, neither the values nor T being knowing to us; then the remaining variables are randomly set to boolean values, independently and uniformly. Knowing only the size of T, $t = |T|$, we wish to extract s unbiased bits from X for s as large as possible; i.e. we wish to find a function $f : \mathbf{B}^n \to \mathbf{B}^s$ such that for any setting of any t of the variables of X in the above process, our f takes on each value in $\mathbf{B}^s$ with probability $1/2^s$. For example, to extract one unbiased bit for any T of size $\leq n-1$, we can take $f = x_1 \oplus \cdots \oplus x_n$, where $\oplus$ denotes the XOR (exclusive-OR); as long as there is one bit which is not set by the adversary, this f will be unbiased. While the obvious lower bound $s \leq n-t$ is achieved in extracting $s = 1$ bit, one cannot in general attain $s = n-t$. In [CFG+85] it proven that to extract $s = 2$ bits requires $t < 2n/3$ (we assume $3|n$ for simplicity); one can achieve this taking $f = (f_1, f_2)$, with f_1 the XOR of the first $2n/3$ of the variables, and f_2 that of the last $2n/3$ of them. The bit extraction problem is equivalent to our coloring problem since constructing an f as above gives a $(2^s; n, k)$ coloring with $k = n-t$, and vice versa (and thus the name "t-resilent function").

This coloring problem also arises in a cryptographic context (as in [BBR88],[Bra89]): player A communicates to player B via a channel of n bits, $x_1, \ldots, x_n$, but there is a spy who is able to see some t of the n bits, A and B not knowing which bits the spy can see. A and B would like the spy to obtain no information about the true message, say a message of $s < n$ bits. In other words, A must takes a message of s bits, encode this to a message of n bits, using some source of randomness, send this through the channel to B, who then applies a decoding function to obtain the origi-

*The author also wishes to acknowledge the National Science Foundation for supporting this research in part under a PYI grant, CCR–8858788.

nal s bit message; we require a scheme in which the spy obtains no information about the s bit message. For example, to communicate one bit, $y \in \mathbf{B}$, A can pick $x_1, \ldots, x_{n-1}$ at random and then set x_n to be the XOR of y and $x_1, \ldots, x_{n-1}$, which B will decode by computing the XOR of $x_1, \ldots, x_n$. Constructing such an encoding and decoding for general s turns out to be equivalent to the coloring problem, again with $c = 2^s$ and $k = n - t$.

Returning to the coloring problem, consider the colorings by $f\colon \mathbf{B}^n \to \mathbf{B}^s$, identifying $\mathbf{B}^s$ with the set of colors, for those f's formed by XOR's of the variables, i.e. $f = (f_1, \ldots, f_s)$ with each f_i being an XOR of a subset of the variables; we call such a coloring an *XOR coloring*. Let $\kappa_{\mathrm{XOR}}(c, n)$ denote the smallest value of k that an XOR coloring can achieve given c, n. It has been conjectured (e.g. [Vaz85]) that $\kappa_{\mathrm{XOR}}(c, n) = \kappa(c, n)$ for all c, n, i.e. that that XOR colorings can always achieve the optimal k; as mentioned before, this was proven for all n with $c = 4$ in [CFG$^+$85] (and is clear for $c = 2$).

It is easy to determine $\kappa_{\mathrm{XOR}}(c, n)$ to within $O(c)$. For example, for n divisible by $c - 1$, it is easy to see that $\kappa_{\mathrm{XOR}}(c, n) = n\theta_c + 1$, where $\theta_c = (c/2 - 1)/(c - 1)$. It follows that $\kappa_{\mathrm{XOR}}(c, n)$ is always within $(c - 2)/2$ of $n\theta_c + 1$.

Here we prove:

Theorem 1.1 *For any $(c; n, k)$ coloring we have $(k - 1)/n \geq \theta_c$, i.e. $\kappa(c, n) \geq n\theta_c + 1$.*

Corollary 1.2 *For $c - 1 | n$, $n > 0$, XOR colorings are optimal. For any n, c, XOR colorings are within $(c - 2)/2$ of optimal.*

It is a non-trivial problem to determine $\kappa_{\mathrm{XOR}}(c, n)$ exactly for general n, but one can do so in certain cases to obtain other bounds as a corollary. For example, one gets:

Proposition 1.3 *For $n \geq 2$ and $\equiv -2, -1, 0, 1, c/2 - 1, c/2, c/2 + 1 \pmod{c - 1}$, we have $\kappa_{\mathrm{XOR}}(c, n) = 1 + \lceil n\theta_c \rceil$ and hence $= \kappa(c, n)$. For $n > 2$ and $\equiv 2 \pmod{c - 1}$ with $c \geq 8$, $\kappa_{\mathrm{XOR}}(c, n) = 2 + \lceil n\theta_c \rceil$, and hence $\kappa_{\mathrm{XOR}}(c, n)$ is at worst within 1 of $\kappa(c, n)$. For $c = 8$ the former and latter congruences hold according to whether $n \not\equiv 2 \pmod 7$ or not. For any n, c, $\kappa_{\mathrm{XOR}}(c, n)$ is within $c/4$ of $\kappa(c, n)$.*

In particular, $n \equiv 2 \pmod 7$ is the simplest case in which we don't know if XOR colorings are optimal.

We are also interested in two related problems:

Problem 1.4 *For $k < \kappa(c, n)$, how close (in various metrics involving $\mathcal{H}_k$) to equicolorable can be achieved by a c-coloring of $\mathbf{B}^n$?*

Problem 1.5 *Classify all optimal colorings, at least for $c - 1 | n$. Are there optimal colorings not obtainable as XOR type colorings?*

The author knows of no optimal colorings other than XOR colorings. Of course, if one views n and k as fixed and asks for the optimal c, then there exist optimal colorings which are not XOR colorings (e.g. $k = (n/3) - 1$ with $3 | n$ and n large, so that $c = 1$ is the optimal c).

In this paper we study problem 1.4 for the metric $L^2(\mathcal{H}_k)$ (defined in section 3). We give a lower bound on the distance to "equicolor" one can achieve for $k < \kappa(c, n)$. One remarkable fact is:

Theorem 1.6 *For $c - 1 | n$, the optimal XOR coloring is closest to equicolorable in $L^2(\mathcal{H}_k)$ for any k.*

Both this theorem and theorem 1.1 are proven using the eigenvalues and eigenspaces of the adjacency matrix of the $\mathbf{B}^n$. Such methods can yield other inequalities expressable in terms of the distance distribution (see section 3) of a subset of $\mathbf{B}^n$. For example, the fact that the average influence of a variable on a boolean function which is 1 on half of the values is at least $1/n$ (see [KKL88]) follows immediately from the fact that the sum of the influences is just $\big((n - A)\chi_c, \chi_c\big) = 2^n \sum_{r>0} 2r\mu_r \geq 2^n \sum_{r>0} 2\mu_r = 2^n$, using the notation of section 2, where C is the subset of $\mathbf{B}^n$ where the function takes the value 1.

In general, XOR colorings are not closest to equicolor for many important metrics, such as the $\mathrm{RP}(\mathcal{H}_k)$ metric defined in section 3 (already for the case $n = c = 4$, $k = 2$). Roughly speaking, the problem is that for k one less than what can be achieved by the best XOR coloring for fixed c, n, the best XOR colorings are equicolored on almost all $H\mathcal{H}_k$, but on the other H's they avoid half the colors! It would be interesting to know about the closest to equicolor colorings for various sup-type norms, such as $\mathrm{RP}(\mathcal{H}_k)$ or $L^\infty(\mathcal{H}_k)$, and to give an explicit constructions.

In the proof that XOR colorings are optimal for $c = 4$ of [CFG$^+$85], one produces from an optimal coloring f_1, f_2 two subsets of variables X_1, X_2 whose XOR's yield an optimal coloring (any X_1, X_2 with the Fourier coefficient of $f_1, f_2, f_1 + f_2$ non-vanishing at, respectively, $X_1, X_2, X_1 + X_2$ will do). This proof does not directly generalize because of the possibility of cancellation of Fourier coefficients in computing a convolution, and theorem 1.1 somehow precludes very bad cancellation. It would be nice to find a generalization of the method in [CFG$^+$85], understanding precisely how much cancellation can occur in such convolutions. In our paper there is no explicit reference to Fourier coefficients of the f_i's. They occur only implicitly, in that the eigenspace of the adjacency matrix of $\mathbf{B}^n$ corresponding to the eigenvalue $n-2r$ is precisely the ($\mathbf{R}$-linear) span of the collection of all XOR's of r variables.

In section 2 we give a short proof of theorem 1.1 and prove some facts about the optimal XOR colorings, proving proposition 1.3. In section 3 study the approximate equicoloring problem for L^2. To do so we study, for a subset $C \subset \mathbf{B}^n$, its distance distribution. It seems that studying "higher-order" distance distributions may shed light on approximate equicolorings in other norms; for this and other intriguing relations between the distance distribution and the bit extraction problem, the reader is referred to the full version of the paper, [Fri91]. In section 4 we make some remarks about problem 1.5, giving a simple geometric characterization of all XOR colorings.

The author wishes to thank Kai Li, Bernard Chazelle, and Avi Wigderson for useful comments and discussions.

2 The Bit Extraction Problem

We begin by proving theorem 1.1. We do this via a somewhat stronger statement. We say that a subset $C \subset \mathbf{B}^n$ is $1/c$ dense in $\mathcal{H}_k$ if

$$\frac{|C \cap H|}{|H|} = \frac{1}{c} \qquad \forall H \in \mathcal{H}_k.$$

Theorem 2.1 *If there exists a $1/c$ dense in $\mathcal{H}_k$ subset of $\mathbf{B}^n$, then $(k-1)/n \geq \theta_c$.*

Theorem 1.1 follows by taking C to be the set of vertices of any fixed color of a $(c;n,k)$ coloring.

Proof Consider the adjacency matrix, A of the Boolean cube, and χ_C, the characteristic function of C in $\mathbf{B}^n$. Clearly $(A\chi_C, \chi_C) \geq 0$. On the other hand, the eigenvalues of A are $n-2r$ with $r = 0, 1, \ldots n$, and the corresponding eigenspaces, E_r, are just the spans of all XOR's of r variables, viewing $\mathbf{B}$ as $\{-1,+1\}$. It follows that if v_r is the projection of χ_C onto E_r, then the assumption that C is $1/c$ dense in $\mathcal{H}_k$ implies that $v_1 = \cdots = v_{n-k} = 0$ (and conversely[1]). Setting $\mu_r = |v_r|^2/|\chi_c|^2$, we have

$$\sum_{r=0}^{n} \mu_r = \frac{\sum |v_r|^2}{|\chi_C|^2} = 1$$

[1] "Conversely" follows from the invertibility of the standard $2^n \times 2^n$ Hadamard matrix. In modern lingo, all the weight 1 thru $n-k$ Fourier coefficients vanish for the function taking the value 1 on C and 0 elsewhere, as in [CFG+85], section 5.1.

and

$$(A\chi_C, \chi_C) = \sum_{r=0}^{n} (n-2r)|v_r|^2 = |\chi_C|^2 \sum_{r=0}^{n} (n-2r)\mu_r.$$

Also, E_0 corresponds to the trivial eigenvector, $(1,\ldots,1)$, and so $\mu_0 = |C|/n = 1/c$; this and $\mu_1 = \cdots = \mu_{n-k} = 0$ give

$$0 \leq n\mu_0 + \sum_{r=n-k+1}^{n} (n-2r)\mu_r \leq$$

$$n\mu_0 + (2k-2-n) \sum_{r=n-k+1}^{n} \mu_r = n\frac{1}{c} + (2k-2-n)\frac{c-1}{c},$$

and so

$$2(k-1)\frac{c-1}{c} \geq n\frac{c-2}{c},$$

which is the desired result.

□

We will now evaluate $\kappa_{\mathrm{XOR}}(c,n)$ for some special values of n to deduce corollary 1.2 and proposition 1.3. First recall that in general $f_1, \ldots, f_s$ yield an $\mathcal{H}_k$ equicolored coloring iff all XOR's of a subsets of $\{f_1, \ldots, f_s\}$ yields a function which is half 1, half -1 on every $H \in \mathcal{H}_k$ (see, for example, [CFG+85]; this is just to say that the standard $2^s \times 2^s$ Hadamard matrix is invertible). So for a subset $T \subset S = \{1, \ldots, s\}$, consider the XOR of the f_i with $i \in T$, which we denote f_T. If the f_i are XOR's of a subsets the variables $X = \{x_1, \ldots, x_n\}$, then so is each f_T. Furthermore, an XOR of the variables is half 1, half -1 on $\mathcal{H}_k$ iff it is the XOR of at least $n-k+1$ variables. This reduces the analysis of optimal XOR colorings to a question about the possible Venn diagrams of s subsets, $X_1, \ldots, X_s$, of X, X_i being the subset of variables of which f_i is an XOR (which is equivalent to a question about error correcting codes, as in [CFG+85]).

Namely, for a non-empty $T \subset S$, consider the size of the corresponding component of the Venn diagram on the X_i's,

$$I_T = |\left(\cap_{i\in T} X_i\right) \cap \left(\cap_{i\notin T} \overline{X}_i\right)|,$$

where $\overline{X}_i$ is the complement of X_i in X. The X_i's correspond to an $\mathcal{H}_k$ equicolored coloring iff for all $U \subset S$,

$$\sum_{|T\cap U| \equiv 1 \pmod 2} I_T \geq n - (k-1). \tag{2.1}$$

Furthermore if there exist non-negative, integral I_T satisfying the above equation with $n \geq k$, then clearly there exists a $(c;n,k)$ XOR coloring.

Summing the above over all U shows that

$$\kappa_{\mathrm{XOR}}(c,n) \geq 1 + n\theta_c, \tag{2.2}$$

and if equality holds then each of the inequalities of equation 2.1 holds with equality; the invertibiliy of the standard Hadamard matrix implies that equality holding in all the above inequalities necessitates $I_T = n/(c-1)$. So for $c-1|n$, any choice of X_i with $I_T = n/(c-1)$ for all T yields an optimal coloring, and any optimal coloring occurs in this way. For $c-1|n$ we use the term *balanced* XOR coloring for any optimal XOR coloring, to emphasise the fact that all I_T's are equal. Furthermore we have

Proposition 2.2 *For $n \geq 2$ and $\equiv -2,-1,0,1,c/2-1,c/2,c/2+1 \pmod{c-1}$, $\kappa_{\mathrm{XOR}}(c,n) = 1 + \lceil n\theta_c \rceil$, and thus optimal colorings can be achieved by XOR colorings in these cases.*

Proof All cases follow easily from the case $n = c/2$ and the trivial case $n = c-1$. For $n = c/2$ we take the "odd coloring," namely I_T is $= 1, 0$ according to whether or not $|T|$ is odd. It is easy to see than any U has $|T \cap U|$ odd for at least half the T with $|T|$ odd, and so $\kappa_{\mathrm{XOR}}(c, c/2) \leq c/4+1$. Rest of details omitted.

□

For general n the problem of determining $\kappa_{\mathrm{XOR}}(c,n)$ is more difficult. However, for fixed c it suffices to check the cases $n = 1, 2, \cdots, O(c^2)$ to determine $\kappa_{\mathrm{XOR}}(c,n)$ for all n. Indeed, for a fixed $r \in [0, c-2]$ let $K = K(r)$ be the smallest integer such that for all m sufficiently large there exists a $(c; (c-1)m + r, (c/2-1)m + K + 1)$ XOR coloring.

Lemma 2.3 *An XOR coloring with $I_T = 0$ for some T has $(k-1)/n \geq 1/2$. For any r there exists a unique $m_0 = m_0(r) \leq 2K(r) - r$ such that there exist $(c; (c-1)m + r, (c/2-1)m + K + 1)$ for all $m \geq m_0$.*

Proof The first statement follows from summing over all U with $|T \cap U| \equiv 1 \pmod 2$. For the second part, m_0 obviously exists, and the coloring at $n = (c-1)m_0 + r$ must have at least one I_T equal zero, or else we could subtract 1 from all the I_T's to get a coloring as above with $m = m_0 - 1$. So the first statement applies to yield $m_0 \leq 2K(r) - r$.

□

In particular, checking $\kappa_{\mathrm{XOR}}(c,n)$ for $n = 1, 2, \ldots, O(c^2)$, we can determine all $K(r)$'s, and therefore all $\kappa_{\mathrm{XOR}}(c,n)$ with $n \geq m_0(r)(c-1) + r = O(c^2)$ by the above. Another consequence of the above is:

Proposition 2.4 *$K(2) = 2$ for $c \geq 8$, and for $n > 2$ and $\equiv 2 \pmod{c-1}$ we have $\kappa_{\mathrm{XOR}}(c,n) = 2 + \lceil n\theta_c \rceil$.*

Proof $K(1) = 1$ implies that either $K(2)$ is 1 or 2. The lemma implies that to rule out $K(2)$ being 1 it suffices to check the case $m_0 = 0$, i.e. $n = 2$, where there is nothing to check (since $2^n < c$). Hence $K(2) = 2$, and then clearly $m_0(2) = 1$, i.e. for $n = c+1$ we can achieve $\kappa_{\mathrm{XOR}}(c,n) = 2 + \lceil n\theta_c \rceil$.

□

Proposition 1.3 is a consequence of the above.

3 Almost Equicolored Colorings and Profiles

For a c-coloring, $\gamma\colon \mathbf{B}^n \to \mathbf{B}^s$, we define its $L^p(\mathcal{H}_k)$ distance from equicolor via

$$\|\gamma\|^p_{L^p(\mathcal{H}_k)} = \sum_{v \in \mathbf{B}_s} \|\gamma^{-1}(v)\|^p_{L^p(\mathcal{H}_k)},$$

where for a $C \subset \mathbf{B}^n$ (and a fixed $c = |C|/n$ in mind) we define the summand via

$$\|C\|^p_{L^p(\mathcal{H}_k)} = \sum_{H \in \mathcal{H}_k} \Big| |C \cap H| - |H|/c \Big|^p.$$

This is one sense in which we can measure how close to being equicolored a coloring is. In this section we study the case $p = 2$. There are other important metrics suggested by the applications, and we mention

$$\|\gamma\|_{\mathrm{RP}(\mathcal{H}_k)} = \max_{H \in \mathcal{H}_k, G \subset \mathbf{B}^s, |G| = 2^{s-1}} |\gamma^{-1}(G) - 2^{k-1}| \ ;$$

this measures how well the bits γ extracts work as a random source to an RP algorithm, G being interpreted as the set of witnesses.

We study $\|C\|_{L^2(\mathcal{H}_k)}$ for a $C \subset \mathbf{B}^n$ via C's *distance distribution* in the following sense:

Definition 3.1 *The* distance distribution *of a $C \subset \mathbf{B}^n$ is the collection of numbers, N_i, defined to be the number of pairs $(c_1, c_2) \in C \times C$ of points in C of distance i (i.e. in the $\mathbf{B}^n$ or Hamming distance), for $i = 0, \ldots, n$.*

The N_i is essentially the *weight enumerator* of coding theory (see [CS88] or [vL82]).

Consider, for a $C \subset \mathbf{B}^n$ with $|C| = n/c$,

$$\mathcal{E}_j(C) \equiv \sum_{H \in \mathcal{H}_j} |C \cap H|^2.$$

Since

$$\|C\|^2_{L^2(\mathcal{H}_j)} = \sum_{H\in\mathcal{H}_j} (|C\cap H| - |H|/c)^2 = \mathcal{E}_j(C) - \frac{2^{n+j}}{c^2}\binom{n}{j},$$

$\mathcal{E}_j(C)$ can be used to measure how close C is to being equicolored with respect to $\mathcal{H}_j$. Routine calculations demonstrate that one can calculate $\mathcal{E}_j$ knowing the μ_i's of section 2 (see [Fri91] for details):

Lemma 3.2

$$N_i = (q_i(A)\chi_C, \chi_C) = \sum_{r=0}^{n} q_i(n-2r)|C|\mu_r$$

where q_i is a polynomial of degree i given by $q_0 = 1$, $q_1(x) = x$, and

$$q_{i+1}(x) = \left(\frac{x}{i+1}\right) q_i(x) - \left(\frac{n-i+1}{i+1}\right) q_{i-1}(x) \quad (3.1)$$

for $i \geq 1$. Furthermore,

$$\mathcal{E}_j = \sum_{l=0}^{j} \binom{n-l}{j-l} N_l = (s_j(A)\chi_C, \chi_C) =$$

$$\sum_{r=0}^{n} s_i(n-2r)|C|\mu_r,$$

where s_j is a polynomial of degree j with leading coefficient $\big(n(n-1)\cdots(n-j+1)\big)^{-1}$, given generally as

$$s_j(x) = \sum_{l=0}^{j} \binom{n-l}{j-l} q_i(A). \quad (3.2)$$

The key point in the proof of theorem 1.6 is to understand the polynomials s_j; they are determined by the following proposition. The author admits to having discovered this proposition by example with computer-aided calculations, which once known is easy to prove (yet the author knows no direct combinatorial proof using only equations 3.2 and 3.1).

Proposition 3.3 *s_j has $-n, -n+2, \ldots, -n+2j-2$ as its j roots; i.e.*

$$s_j(x) = \frac{(x+n)(x+n-2)\cdots(x+n-2j+2)}{n(n-1)\cdots(n-j+1)} =$$

$$2^j \binom{(x+n)/2}{j}.$$

Proof It is clear from the definition of the s_j's that they are polynomials of degree j satisfying

$$(s_j(A)f, f) = \sum_{H\in\mathcal{H}_j} \left(\sum_{x\in H} f(x)\right)^2,$$

for any $f \in L^2(\mathbf{B}^n)$. The right-hand-side of the above clearly vanishes if f is the XOR of r variables with $r \geq n-j+1$. Since such an f has eigenvalue $n-2r$, we conclude that s_j has a root $n-2r$ for any $r \geq n-j+1$.

□

From this proposition we see that $s_j(x)$ restricted to $x = -n, -n+2, \ldots, n$ is a convex function, as is its restriction to $[-n+2j-2, +\infty)$, and we easily obtain the following stronger version of theorem 1.6:

Theorem 3.4 *For $c-1|n$ and $j \leq 1 + \theta_c n$, the balanced XOR coloring is closest in $L^2(\mathcal{H}_j)$ to equicolor, and any coloring as close must have the same distance distribution as that of the balanced XOR coloring. In other words,*

$$\frac{\Big(s_j(A)\chi_C, \chi_C\Big)}{(\chi_C, \chi_C)} \geq \frac{1}{c} s_j(n) + \frac{c-1}{c} s_j\Big(-n/(c-1)\Big),$$

equality holding iff $c-1|n$ and all μ_i's vanish except for $\mu_0 = 1/c$ and $\mu_{n(1-\theta_c)} = (c-1)/c$. For $c-1 \not| n$, setting $-n/(c-1) = t + \alpha$ with t an integer and $\alpha \in (0,1)$, we have the sharper bound

$$\frac{\Big(s_j(A)\chi_C, \chi_C\Big)}{(\chi_C, \chi_C)} \geq \frac{1}{c} s_j(n) + \frac{c-1}{c}\Big((1-\alpha)s_j(t) + \alpha s_j(t+1)\Big),$$

with equality iff $\mu_t = (1-\alpha)(c-1)/c$, $\mu_{t+1} = \alpha(c-1)/c$.

4 Locally Symmetric Colorings and Concluding Remarks

We consider the problem of classifying for $c-1|n$ all optimal colorings, i.e. with $k = 1 + n\theta_c$. We call a coloring, $\gamma\colon \mathbf{B}^n \to \mathbf{B}^s$ *locally symmetric* if there is an η such that for all $v \in \mathbf{B}_n$, exactly $n\eta$ of v's neighbors are colored $\gamma(v)$, and every other color occurs among v's neighbors exactly $n(1-\eta)/(c-1)$ times. In addition we say that the coloring is *sparse* if $\eta = 0$; the balanced coloring is an example.

It is easy to see that locally symmetric colorings are also symmetric in their "distance i neighborhoods" for all i, and that each color has the same distance distribution which is uniquely determined by η. In particular, a locally symmetric coloring which is optimal for $c-1|n$ must be sparse.

These considerations, and theorems such as theorem 3.4 suggest dividing problem 1.5 into two parts:

Problem 4.1 *Are all optimal colorings, i.e. with $k = \kappa(c,n)$, necessarily locally symmetric, at least if $c-1|n$?*

Problem 4.2 *Are all locally symmetric sparse colorings necessarily XOR colorings?*

Regarding the latter question, and XOR colorings in general, one can make the following observation. Consider a general cycle of length 4 in $\mathbf{B}_n$, $(v_0, v_1, v_2, v_3, v_0)$. XOR colorings satisfy the following two conditions:

1. $\gamma(v_3)$ depends only on $\gamma(v_0), \gamma(v_1), \gamma(v_2)$, not on the particular cycle,
2. $\gamma(v_0) = \gamma(v_2)$ implies $\gamma(v_1) = \gamma(v_3)$.

Proposition 4.3 *Any locally symmetric coloring satisfying the above two conditions is an XOR coloring.*

Proof Fix any coloring, γ, of $\mathbf{B}^n$, and a vertex, $v \in \mathbf{B}_n$. We define a group law on the colors as follows. We can assume $\eta < 1$, or else there is nothing to prove. So for any $\gamma_1 \neq \gamma_2$ neither equal $\gamma(v)$, there exist neighbors v_1, v_2 of v with $\gamma(v_i) = \gamma_i$. There exists a unique v_3 making v, v_1, v_3, v_2 a simple cycle of length 4; define $\gamma_1 + \gamma_2$ to be $\gamma(v_3)$. If $\gamma_1 = \gamma_2$ define their sum to be $\gamma(v)$, and if one of γ_1, γ_2 is $\gamma(v)$ define their sum to be the other γ_i. That this defines defines a commuatative, associative group law, with identity $\gamma(v)$, and every other element of order 2, is an easy consequence of the above conditions; for example, commutativity follows from the fact that if v_0, v_1, v_2, v_3 is a simple cycle then so is v_0, v_3, v_2, v_1; associativity follows from the fact that the sum of three colors is the antipodal point to v_0 in a subcube of $\mathbf{B}^n$ isomorphic to $\mathbf{B}^3$. This sets up an isomorphism between the colors and $\mathbf{B}^s$. One can similarly show that for every cycle $(v_0, v_1, v_2, v_3, v_0)$ we have the sum of the colors vanishing (i.e. $= \gamma(v)$). Setting coordinates on $\mathbf{B}^n$ so that v is the origin, the coloring of the neighbors of v determine XOR's, $f_1, \ldots, f_s$ such that the f_i's induce the given coloring on v and its neighbors. The conditions on the coloring imply, by induction on k, that for any $w \in \mathbf{B}^n$ of distance k to v, the coloring determined by the $f_1, \ldots, f_s$ agrees with the original coloring.

$\square$

Of course, in the above proposition, one can replace the local symmetry condition by the condition that, say, each $v \in \mathbf{B}^n$ has all colors appearing among its neighbors except for, possibly, its own color. Call such a coloring *connected*. We can restate the problems as:

Problem 4.4 *Is any optimal coloring necessarily connected and does it necessarily satisfy the above two conditions?*

References

[BBR88] C.H. Bennett, G. Brassard, and J.-M. Robert. Privacy amplification by public discussion. *SIAM J. on Computing*, 17(2):210–229, 1988.

[Bra89] G. Brassard. Cryptology— Column 1. *SIGACT News*, 20(3):15–19, Summer 1989.

[CFG+85] B. Chor, J. Friedman, O. Goldreich, J. Håstad, S. Rudich, and R. Smolensky. The bit extraction problem or t-Resilient functions. In *26th Annual Symposium on Foundations of Computer Science*, pages 396–407, 1985.

[CS88] J.H. Conway and N.J.A. Sloan. *Sphere Packings, Lattices, and Groups.* Springer-Verlag, 1988.

[Fri91] J. Friedman. On the bit extraction problem. Technical report, Princeton University, December 1991.

[KKL88] J. Kahn, G. Kalai, and N. Linial. The influence of variables on boolean functions. In *29th Annual Symposium on Foundations of Computer Science*, pages 68–80, 1988.

[Vaz85] U. Vazirani. Towards a strong communication complexity theory, or generating quasi-random sequences from two communcating slightly-random sources. In *Proc. of 17th ACM Symposium on Theory of Computing*, pages 366–378, 1985.

[vL82] J.H. van Lint. *Intorduction to Coding Theory.* Springer-Verlag, 1982.