Searching for Patterns of Snakes in Hypercubes

1 Introduction

The problem of finding longest snakes in hypercubes was posed in 1958 [Kau58], and is still an open problem. Few characteristics of longest snakes are understood, but some recent observations of snakes have provided a possible route to predicting the length of longest snakes in higher-dimensional hypercubes.

In Section 2 we give an informal overview of snakes and hypercubes. In Section 3 we present some background information that lead to the present research. In Section 4 we present some relevant formal notation and definitions. In Sections 5 we present a proof of a property of snakes that we have observed. In Section 6 we summarize the significance and implications of our results.

2 Snakes in Hypercubes

Imagine a three-dimensional cube, which has eight corners and six faces, as an undirected graph, where the corners of the cube are the vertices of the graph and the lines where the faces meet are the edges of the graph. A *snake* is defined as a path, or a sequence of vertices in the graph, such that every pair of consecutive vertices is connected by an edge in the cube, while no two other vertices in the path are connected by an edge on the cube. Furthermore, this path must not double back on itself, that is, it must not form a cycle (loop). For example, in Figure 1, the path $c1, c2, c3$ is a snake of length 2 in the cube; the path $c1, c2, c3, c4$ is *not* a snake because the corners $c1$ and $c4$ are connected by an edge that is present in the cube; the path $c1, c2, c3, c7, c6, c2$ is *not* a snake because the path has a loop. The snake shown, $c1, c2, c3, c7, c8$ is a longest snake, of length 4, in the 3-dimensional cube.

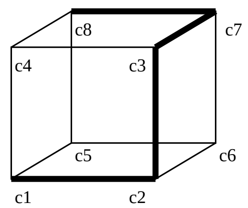

Figure 1: Longest Snake in a 3-dimensional Cube

This concept extends to hypercubes, which are "cubes" of dimensions greater than 3 (see Figure 2 for an example of a 4-dimensional hypercube). By referring to a snake in a hypercube, we are referring to the snake in the graphical representation of the hypercube. Alternatively, a snake in a hypercube is an *induced path* within the graph of the hypercube. (See [AK91] for a more formal definition.)

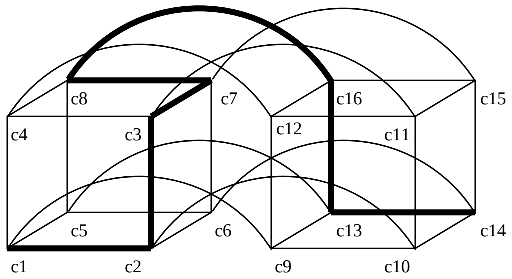

Figure 2: Longest Snake in a 4-dimensional Hypercube

The problem of finding a longest snake in a hypercube was proposed in the context of error-checking codes in 1958 [Kau58]. The vertices of an n-dimensional hypercube can be uniquely coded in n bits, and a snake in this hypercube is a sequence of vertices that differ by exactly one bit. Selecting a suitable subset of the vertices of a snake gives a 1-bit error detecting code. Selecting a more constrained subset gives a 1-bit error correcting code. This problem also has some importance in connection with other applied problems such as electronic locking schemes (see [Bla64, CFT64, Eve63, Jab74, Kau58, Kle70] for discussions of such applications), and is still an open problem.

Analytical approaches have produced broad-ranging results for the longest snake problem [JPP94]. Exhaustive search is the only known method of finding a longest snake in a hypercube, and hence experimental approaches are uncommon because this method is computationally intractable in any realistic time frame. For instance, an exhaustive search for the longest snake in dimension 7 took approximately one month on a supercomputer dedicated to the task [She]. Since exhaustive searches take an unacceptable amount of time, longest snakes are known only up through dimension 7. It is for this reason that we are continuing the analytical approach in an attempt to discover the properties of snakes.

3 Background

[RS99] conjectured that there exists in every dimension d a maximal snake of every length from $2(d-1)$ up to the length of the longest snake in d. However, through an exhaustive generation of maximal 7-snakes, it was found that there exists no maximal 7-snake of length 13, thus disproving that conjecture in general. It is known that there exists a maximal snake of length $2d-1$ for dimensions up through 6, so we speculated that there exists no maximal snake of length $2d-1$ in any dimension higher than 6. This is proven as Theorem 1.

It is our hope that continued examination of maximal snakes of length $2d$, $2d+1$, $2d+2$, and so on, will provide a method of determining for what dimensions a certain length maximal snake exists.

4 Notation and Definitions

Here we present some additional notation, as well as formal definitions of some of the previously discussed concepts. This will serve as a foundation for the proofs that follow.

For each $d > 0$, we will denote the set $\{0, 1, \ldots, (d-1)\}$ by $\mathbf{d}$. $\mathcal{P}^{\mathbf{d}}$ will denote the set of all subsets of $\mathbf{d}$, and $\mathbf{2^d}$ will denote the set of characteristic functions of $\mathcal{P}^{\mathbf{d}}$. A *d-dimensional hypercubical graph* (or simply *d-dimensional hypercube*), denoted by H^d, is, by definition, an undirected graph with 2^d vertices, so that

1. each vertex is labeled with the characteristic function for a unique subset of $\mathbf{d}$, and

2. for any two vertices, v and w, with labels f_v and f_w respectively, (v, w) is an edge in H^d (i.e., v and w are neighbors in H^d) iff f_v and f_w differ in exactly one place.

Henceforth, we will not make any distinction between vertices in H^d and the labels assigned to them. Figure 3 shows a labeled 4 dimensional cube.

Then, by definition, a d-snake, S_d, is a sequence, $(s_0, s_1, \ldots, s_k)$, of elements of $\mathbf{2^d}$, such that for all i, j, $0 \leq i < j < k$, s_i and s_j are neighbors in H^d iff $j = i + 1$. The length of this snake is k, the number of edges of H^d in the snake.

Without loss of generality, we can assume that every d-snake $S_d = (s_0, s_1, \ldots, s_k)$, is such that $s_0 = \emptyset$. Then, we will represent S_d by a string $c_0 c_1 \ldots c_{k-1}$ over $\mathbf{d}$ so that for each $i, 0 \leq i \leq (k-1), s_i(c_i) \neq s_{(i+1)}(c_i)$, that is, the labels s_i and $s_{(i+1)}$ differ in the c_i-th place. Thus,

$$S_d = (s_0 = \emptyset, s_1, \ldots, s_k) = c_0 c_1 \ldots c_{k-1}.$$

Let $C_d = (c_1 c_2 \ldots c_k)$ be the sequence of symbols(dimensions) that define the snake S_d.

A snake $S_d = (s_0, s_1, \ldots, s_k)$ is *maximal* in H^d iff, by definition, there is no vertex v in H^d such that $(v, s_0, \ldots, s_k)$ or $(s_0, \ldots, s_k, v)$ is a snake in H^d.

By *length* of a snake we will always mean the number of edges in that snake.

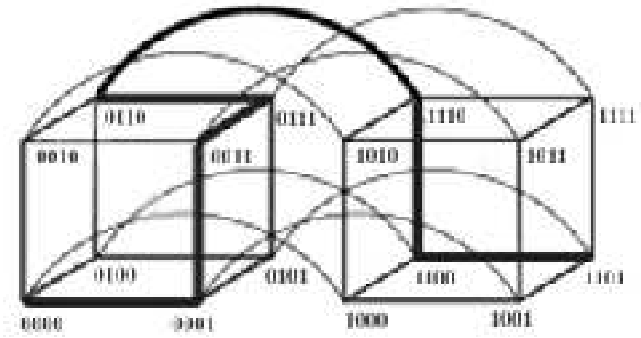

Figure 3: A Longest Snake in a 4-dimensional Hypercube

5 Maximal Snakes of Length $2d-1$

Here we present a proof that no maximal snake of length $2d-1$ exists for any dimension $d \geq 7$.

First we introduce some additional concepts and definitions used in the proof:

First we introduce the concept of "covered" dimensions. A vertex "covers" the dimensions that are included in the vertex. For example, the $5-d$ vertex labeled 10110 "covers" dimensions 0, 2, and 3.

Let $p_0, p_1, \ldots, p_m$ be all the vertices, in order of appearance in S, with hamming distance 2 such that for all $i, 1 \leq i \leq m$, there are no vertices of hamming distance 2 between p_i and $p_{(i-1)}$.

Define the length of the snake S up to p_m as L_m and the length of the snake S after p_m as L'_m. Therefore, $L_m + L'_m = 2d - 1$.

Define for all $i, 1 \leq i \leq m, e(i)$ to be the number of edges between p_i and $p_{(i-1)}$. Note that $L_m = 2 + \sum_1^k e(i)$, since two edges are required to move from $s_0 = 0$ to p_0.

Recall that vertices are labeled by subsets of d. Define, for all $i, 1 \leq i \leq m, f(i) = |p_i \backslash p_{(i-1)}|$. Essentially, $f(i)$ is the number of dimensions "covered" by p_i that are not covered by $p_{(i-1)}$.

Since each p_i is at hamming distance 2, it contains two elements. Therefore, $0 \leq f(i) \leq 2$.

Note that when $f(i) = 1, e(i) \geq 2$, and when $f(i) = 2, e(i) \geq 4$. Therefore, for all $i, 1 \leq i \leq m$,

$$e(i) \geq 2 \cdot f(i) \tag{1}$$

[RS99] presented the following conjecture,

Conjecture 1 *For each d, for each k, $2(d-1) <= k <= s(d)$, there is a maximal d-snake of length k.*

They noted that although this conjecture has been verified for $d \leq 6$, it has been found that there is no maximal 7-snake of length 13, thus disproving the conjecture in general. The following is a proof that there exist no snakes of length $2d - 1$ for dimensions $d \geq 7$.

Let $S = (s_0, s_1, \ldots, s_k)$ be a maximal snake, where $k = 2d - 1$, and let $C = (c_1 c_2 \ldots c_k)$ be the corresponding sequence of symbols.

Note that since S is a maximal snake, each dimension must be "covered" by some point in S at hamming distance 2 from the first vertex, 0, of the snake. Were any dimension not covered by points at hamming distance 2, the snake could be extended from s_0 to a point at hamming distance 1 that included that dimension.

Lemma 1 *Suppose S is a maximal d-snake of length $2d-1$. The last vertex of S, s_k, is at hamming distance 3, and the vertex $s_{(k-1)}$ is at hamming distance 2.*

Proof. Note that since $2d - 1$ is odd, the snake must end at a vertex with an odd hamming distance. Also note that, since $\emptyset$ is on the snake, it cannot end at a vertex at hamming distance 1.

Since all dimensions $0 \ldots (d-1)$ must be "covered" by vertices at hamming distance 2 (and p_0 covers 2 dimensions), $2 + \sum_1^k f(i) \geq d$. Therefore, $\sum_1^k f(i) \geq d - 2$.

Using Equation (1), $\sum_1^k e(i) \geq 2 \cdot \sum_1^k f(i) \geq 2(d-2) = 2d - 4$.

Therefore, $L_m = 2 + \sum_1^k e(i) \geq 2d - 2$.

Therefore, $L'_m = 2d - 1 - L_m \leq (2d-1) - (2d-2) = 1$.

Since the snake S must end at a point of odd hamming distance, and the snake cannot be extended from p_m to a point at hamming distance 1, the snake must be extended to a point at hamming distance 3. Note then that $s_{(k-1)} = p_m$ and therefore is at hamming distance 2. □

Define the sub-snake $S_m = (s_2, s_3, \ldots, s_{(k-1)})$. In other words, S_m is the consisting of all of the vertices between and including the first vertex in S that is at hamming distance 2 and the last vertex in S that is at hamming distance 2.

Similarly, let C_m be the sequence of symbols(dimensions) that define the sub-snake S_m.

Lemma 2 *For each new symbol added to C_m at hamming distance 2, an average of 2 edges are added to S.*

PROOF. Let α be the average number of edges added to S per new symbol added to C_m at hamming distance 2. As noted earlier, in order for S to be a maximal snake, all dimensions must be covered by points at hamming distance 2. The hamming distance 2 points of C_m are responsible for covering $(d-2)$ of the d dimensions, since at the first point of C_m 2 dimensions are already covered. Therefore, there are $\alpha(d-2)$ edges in S_m. Since the length of S_m is $2d-1-3 = 2d-4 = 2(d-2)$, $\alpha \leq 2$. By Equation 1, $a \geq 2$. Therefore, $\alpha = 2$. $\square$

Lemma 3 *All vertices in S are at hamming distance less than or equal to 4.*

PROOF. Suppose, by way of contradiction, there exists a vertex v in S at hamming distance 5. To get from a vertex $p_{(i-1)}$ of hamming distance 2 to v and back to a vertex p_i of hamming distance 2 would require at least 6 edges. Furthermore, at most 2 new dimensions are covered by p_i. Therefore, each new covered dimension requires at least 3 edges to be added to S, thus violating Lemma 2. $\square$

Recall that vertices are subsets of d. It is now useful to label the vertices of S by the set of dimensions included in the vertices. Since by Lemma 1 s_k is at hamming distance 3, let $s_k = \{\alpha\beta\gamma\}$. Since $s_{(k-1)}$ is at hamming distance 2, one of these dimensions was added to $s_{(k-1)}$ to get to s_k. Without loss of generality, let that dimension be α. Therefore, $s_{(k-1)} = \{\beta\gamma\}$.

Lemma 4 *In the string $C_q = (c_1, c_2, \ldots, c_{(k-1)})$, exactly $d-2$ symbols appear twice, and all others appear once.*

PROOF. In order for the symbols β and γ to appear in the vertex $s_{(k-1)}$, they must have appeared an odd number of times in the string C_q. Also, because all dimensions must be "covered" by vertices at hamming distance 2 and because none of the remaining $d-2$ dimensions appear in $s_{(k-1)}$, the remaining $d-2$ symbols must appear an even number of times in C_q. Note that there are $2d-2$ symbols in C_q. Discounting the edges where the symbols β and γ were last added, there are $2d-4 = 2(d-2)$ edges in which symbols appear more than once. Therefore, only $d-2$ symbols could have appeared more than once, and they must have appeared exactly twice. $\square$

Corollary 1 *Suppose $s_{(k-1)} = \{\beta\gamma\}$. Then $\beta\gamma$ appear only once in C.*

Lemma 5 *Between any two vertices in S at hamming distance 4, there are at least 2 vertices.*

PROOF. Suppose, by way of contradiction, there are two vertices v_1 and v_2 in S at hamming distance 4 that are separated by one vertex. Let i be such that p_i is the last vertex of S at hamming distance 2 just before v_1 and $p_{(i+1)}$ is the first vertex at hamming distance 2 after v_2. Then the substring $p_i, \ldots, v_1, \ldots, v_2, \ldots, p_{(i+1)}$ contains at least 6 edges. Therefore $f(i+1) = 2$. Therefore, an average of at least 3 edges was required per dimension covered by $p_{(i+1)}$, thus violating Lemma 2. $\square$

Lemma 6 *There are $d-1$ vertices in S at hamming distance 3.*

PROOF. Recall also that d dimensions must be covered by vertices at hamming distance 2 and that 2 dimensions are covered by the vertex p_0. Since for each i, such that $e(i) = 4$, p_i covers 2 additional dimensions, $d - 2 - 2t$ dimensions must be covered by vertices p_i such that $e(i) = 2$. Note that when $e(i) = 4$ two vertices are added at hamming distance 3 and when $e(i) = 2$ one vertex is added at hamming distance 3. Therefore, the number of vertices at hamming distance 3 from s_0 to $p_m = s_{(k-1)}$ is $2t + (d - 2 - 2t) = d - 2$. Since by Lemma 1 the vertex s_k is at hamming distance 3, the total number of vertices at hamming distance 3 in S is $d - 2 + 1 = d - 1$. □

Lemma 7 $s_{(k-2)} = \{\beta\gamma x\}$, $0 \leq x \leq d, x \neq \alpha, \beta, \gamma$, *and no other vertex at hamming distance 3 contains* $\beta\gamma$.

PROOF. To get from $s_{(k-2)}$ to $s_{(k-1)}$, some $x, 0 \leq x \leq d, x \neq \alpha, \beta, \gamma$ must have been dropped. Therefore, $s_{(k-2)} = \{\beta\gamma x\}$. Since $s_{(k-1)} = \{\beta\gamma\}$, any other vertex at hamming distance 3 that contains $\beta\gamma$ is 1 away from $s_{(k-1)}$ but not its neighbor on S, thus violating the properties of snakes. □

Now, since the previous vertex at hamming distance 3 cannot be of the form $\{\beta\gamma x\}$, either β or γ must have been added to get from that point to $s_{(k-2)}$. Without loss of generality, let γ be the symbol added. In other words, $s_{(k-3)} = \{\beta x\}$.

Then,

Lemma 8 *No vertex in* S*, other than* s_k*, at hamming distance 3 contains* $\alpha\gamma$.

PROOF. By Corollary 1, γ can only appear once in the string C, and the only 2 vertices at hamming distance 3 after γ was added are $\{\beta\gamma x\}$ and $\{\alpha\beta\gamma\} = s_k$. □

Lemma 9 *There are at least* $d-4$ *vertices in* S_m *that have the form* $\{\alpha\beta x\}$, $0 \leq x \leq d, x \neq \alpha, \beta, \gamma$.

PROOF. Because S is a maximal snake, the vertex $s_k = \{\alpha\beta\gamma\}$ must be "blocked" from being extended to vertices at hamming distance 2 or 4 by vertices of the form $\{\alpha\beta x\}$, $\{\beta\gamma x\}$, or $\{\alpha\gamma x\}$, $0 \leq x \leq d, x \neq \alpha, \beta, \gamma$. $d - 3$ such vertices are required. The point $s_{(k-2)} = \{\beta\gamma x\}$ provides one such block, leaving $d - 4$ required blocking vertices. Lemmas 7 and 8, assert that none of these vertices can be in the form $\{\beta\gamma x\}$ or $\{\alpha\gamma x\}$. Therefore, the remaining $d - 4$ required blocking vertices must be in the form $\{\alpha\beta x\}$. □

Lemma 10 *There can exist in* S_m *no more than 2 vertices of the form* $\{\alpha\beta x\}$, $0 \leq x \leq d, x \neq \alpha, \beta, \gamma$, *and they must appear in sequence.*

PROOF. First we will show that any vertices of the form $\{\alpha\beta x\}$ must appear in sequence. Suppose, by way of contradiction, there exist two vertices v_1 and v_2 in S_m of the form $\{\alpha\beta x\}$ that are separated by a vertex not of that form. Let $c_i, \ldots, c_j$ be the sequence of symbols that define the path from v_1 to v_2. Either the symbol α or the symbol β was dropped and added again, implying that either α or β appears twice in $c_i, \ldots, c_j$. Moreover, α and β must have each appeared once in $c_0 \ldots, c_i$. Therefore, α or β appears 3 times in C, thus violating Lemma 4.

Now suppose, by way of contradiction, there exist more than two vertices of the form $\{\alpha\beta x\}$. Since all such vertices must appear in sequence, there are 4 possible arrangements of these vertices, as shown in Figure 4. According to Lemma 5, arrangement D is not possible. The remaining 3 arrangements include the substrings $xyyz$, $xyzy$, $yxyz$ respectively, each of which violate the properties of snakes.

Therefore, there can be no more than 2 vertices of the form $\{\alpha\beta x\}$, and they must be in sequence. $\square$

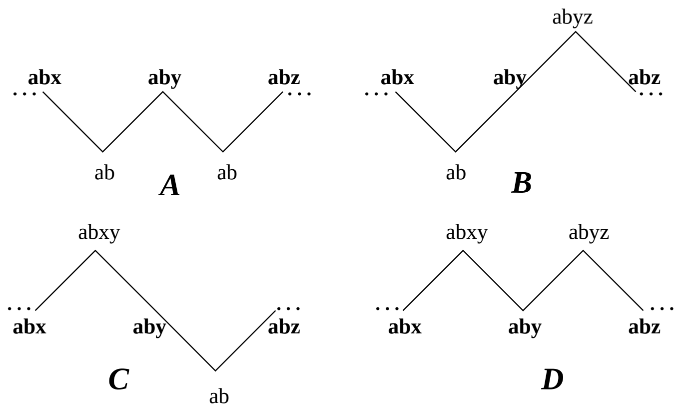

Figure 4:

Theorem 1 *Suppose that $S = (s_0 = \emptyset, s_1, \ldots, s_k)$ is a maximal snake where $k = 2d - 1$. $d \leq 6$.*

Proof. Let g be the number of vertices in S of the form $\{\alpha\beta x\}$. By Lemma 9, $g \geq d - 4$. By Lemma 10, $g \leq 2$. Therefore, $(d - 4) \leq 2$, i.e., $d \leq 6$. $\square$

6 Implications

To solve the problem of longest snakes in hypercubes, we must be able to find the longest snakes, or at least determine the length of the longest snakes, in any given dimensional hypercube. Thus, the property of length $2d - 1$ maximal snakes that we have presented here has little significance by itself. However, because of the existing proof that there exist no maximal snakes, in any dimension, of length $2d - 2$, and a forthcoming conclusion about the existence of length $2d$ maximal snakes, we how have knowledge about the existence of length $2d - 2$, $2d - 1$, and $2d$ maximal snakes in all dimensions. It is our hope that, by examining the methods of proof that were used to reach these conclusions, we may discover a general proof technique that can be applied to maximal snakes of any length. The ability to determine the existence of any length maximal snake in any dimension would be a significant step in the search for knowledge of longest snakes in hypercubes.

References

[AK91] H. L. Abbott and M. Katchalski. On the construction of snake in the box codes. *Utilitas Mathematica*, 40:97–116, 1991.

[Bla64] W. L. Black. Electronic combination locks. Technical Report 73, Research Laboratory of Electronics, M. I. T., 1964.

[CFT64] R. T. Chien, C. V. Freiman, and D. T. Tang. Error correction and circuits on the n-cube. In *Proceedings of the 2nd Annual Allerton Conference on Circuit and System Theory*, pages 899–912, 1964.

[Eve63] S. Even. Snake in the box codes. *IEEE Trans. Electronic Computers*, page 18, 1963.

[Jab74] S. V. Jablonskiĭ. Discrete mathematics and mathematical problems of cybernetics. *Nauka*, 1, 1974.

[JPP94] M. Juric, W. D. Potter, and M. Plaksin. Using the parallel virtual machine for hunting snake-in-the-box codes. In *Proceedings of the 7th Conference of the North American Transputer Users Group*, October 1994.

[Kau58] W. H. Kautz. Unit distance error checking codes. *IRE Transactions on Electronic Computers*, 7:179–180, 1958.

[Kle70] V. Klee. The use of circuit codes in analog to digital conversion. In B. Harris, editor, *Graph Theory and its Applications*. Academic Press, New York, 1970.

[RS99] Dayanand S. Rajan and Anil M. Shende. Maximal and reversible snakes in hypercubes. In *24th Annual Australasian Conference on Combinatorial Mathematics and Combinatorial Computing*, 1999.

[She] Anil Shende. Private Communications.