A new non-asymptotic upper bound for snake-in-the-box codes

A. Lukito and A. J. van Zanten

Department of Mediamatics
Delft University of Technology
P.O. Box 5301, 2600 GA
Delft, The Netherlands

ABSTRACT. A snake in a graph is a simple cycle without chords. A snake-in-the-box is a snake in the n-dimensional cube Q_n. Combining the methods of G. Zemor (Combinatorica 17 (1997), 287-298) and of F.I. Solov'jeva (Diskret Analiz. 45 (1987), 71-76) a new upper bound for the length of a snake-in-the-box is derived for $16 \leq n \leq 19081$.

1 Introduction

A snake in a graph is an induced cycle, that is a simple cycle with no chords. Let Q_n be an n-dimensional cube. This is the graph with all binary words of length n as vertices, and all pairs of vertices which differ in exactly one coordinate as edges. A snake in Q_n is called *a snake-in-the-box code* or, shortly, *a snake-in-the-box*. More precisely, suppose

$$S = X_0, X_1, \cdots, X_{m-1}$$

is a list of m words in Q_n. The distance $d(X_i, X_j)$ between two words of S is defined to be the Hamming distance, that is the number of coordinates in which X_i and X_j differ. The list distance $d_S(X_i, X_j)$ is defined as the minimum number of words from X_i until X_j in S, i.e.

$$d_S(X_i, X_j) = \min\{|i-j|, m-|i-j|\}.$$

Then, S is a snake-in-the-box if for every i, j with $0 \leq i < m$, $0 \leq j < m$,

$$d(X_i, X_{i+1}) = 1,$$

$$d(X_i, X_j) = 1 \Rightarrow d_S(X_i, X_j) = 1$$

where subscripts are reduced modulo m.

JCMCC 39 (2001), pp. 147-156

For applications of snakes we refer to [6, 7, 8]. A central problem in the study of snakes is to determine the maximal length of a snake. We denote this maximal length by $s(n)$. At present, the exact value of $s(n)$ has been determined only for six values of n, i.e. $s(2) = 4$, $s(3) = 6$, $s(4) = 8$, $s(5) = 14$, $s(6) = 26$, and $s(7) = 48$. For all remaining values of n one has to be satisfied with lower and upper bounds for $s(n)$ for the time being. The best lower bound at this moment is the one established by Abbot and Kachalski. In [1] they proved that $s(n) \geq \lambda 2^n$, with $\lambda = 0.300781....$
In the course of time several upper bounds have been derived. We mention here several upper bounds found since 1987. In [13] Solov'jeva derived

$$s(n) \leq \left(1 - \frac{2}{n^2 - n + 2}\right) 2^{n-1}, \qquad n \geq 7. \tag{1}$$

Her method, based on counting four-cycles intersecting S in i vertices, $0 \leq i \leq 4$, is discussed in Section 2.
In [12] Snevily improved this bound by deriving

$$s(n) \leq \left(1 - \frac{1}{20n - 41}\right) 2^{n-1}, \qquad n \geq 12. \tag{2}$$

His method is based on estimating the average number of vertices in S adjacent to a vertex not in S.

By refining some details of the proof in [12], Emelyanov [3] found

$$s(n) \leq \left(1 - \frac{1}{6n - 13}\right) 2^{n-1}, \qquad n \geq 19. \tag{3}$$

And by using some parts of Snevily's proof and Solov'jeva's method, Emelyanov and Lukito [4] showed that for $n \geq 7$

$$\begin{aligned} s(n) &\leq \left(1 - \frac{2\,(n^5 - 18n^4 + 68n^3 + 7n^2 - 298n + 270)}{5n^6 - 31n^5 + 47n^4 - 57n^3 + 362n^2 - 806n + 540}\right) 2^{n-1} \\ &\leq \left(1 - \frac{2}{5n + 59 + o(1)}\right) 2^{n-1}. \end{aligned} \tag{4}$$

Finally, Zemor [15] proved that for all n

$$\begin{aligned} s(n) &\leq 1 + \frac{6n}{6n + \frac{\sqrt{n}}{6\sqrt{6}} - 7} 2^{n-1} \\ &\leq \left(1 - \frac{1}{89\sqrt{n}} + O\left(\frac{1}{n}\right)\right) 2^{n-1}, \end{aligned} \tag{5}$$

applying a technique of counting vertices of "high degree" outside S. We shall discuss parts of his method in Section 2. We remark that Zemor appears to have introduced an essentially new element, since his upper bound is not of the type $r(n)2^n$, where $r(n)$ is some rational function of n, but rather shows a $\sqrt{n}$-dependence. We also remark that although, asymptotically, Zemor's bound is the best bound thusfar, bound (4) is better on the interval $7 \leq n \leq 19079$.

In this paper we shall sharpen bound (4) even further, by combining Solov'jeva's approach and (part of) Zemor's approach. We shall do so by considering a configuration of four vertices such that one of these vertices is adjacent to the other three. Such a configuration usually is called a *claw*. In [15] it is shown that any subpath of S of length 7 contains four vertices adjacent to a claw outside S. Zemor exploits this property to derive a lower bound for the total degree of vertices not in S.
In Section 3 we discuss a relationship between such claws and four-cycles having only one vertex in common with S. This relationship enables us to refine Solov'jeva's technique of counting four-cycles, and hence, to improve her bound for $n \geq 17$. This improved bound is also better than Zemor's bound for $7 \leq n \leq 15603$. However it is not better than bound (4). Therefore, we extend Zemor's result on claws. In Section 4 we present four lemmas that play a major role in proving that any subpath of length 9 contains a set of vertices adjacent to two claws, not necessarily disjoint, outside S. Because of its technical arguments, with many subcases, the proof will not be presented in this paper. Instead we refer to [11]. This result is used in Section 5 to prove our main result, stated in Corollary 2, yielding an upper bound for snake-in-the-box codes better than all other bounds mentioned in this Introduction for $30 \leq n \leq 19081$. For sufficiently large n, Zemor's bound remains the best bound known thusfar.

As for our notation, edges in Q_n will be indicated by lowercase letters and vertices by uppercase letters. A snake subpath will be denoted by a sequence of edges and vertices

$$\stackrel{e_0}{-} V_0 \stackrel{e_1}{-} V_1 \stackrel{e_2}{-} V_2 \cdots \quad V_{i-1} \stackrel{e_i}{-} V_i \quad \cdots$$

where the edge $\{V_{i-1}, V_i\}$ is identified with the "unit vector" $e_i := V_{i-1} + V_i$. Remember that vertices in Q_n stand for binary n-tuples, and hence, that we can add vertices componentwise mod 2. Because of the snake properties, e_i is an n-tuple with zeros in all positions, except in the position where V_{i-1} and V_i differ, and we have

$$e_i \neq e_{i+1}, \qquad e_i \neq e_{i+2}, \tag{6}$$

for all relevant values of i.

2 Previous methods

Throughout the rest of this paper, let S be a fixed snake of length $s(n)$.

The method applied in [13] to derive an upper bound for $s(n)$ is by counting four-cycles (i.e. simple cycles of length 4) having i vertices in common with a snake S, for $i \in \{1, 2, 3, 4\}$.

Definition 1 *The set A_i is the set of four-cycles in Q_n having precisely i vertices in common with S.*

Any two vertices in Q_n at Hamming distance 2 determine precisely one four-cycle. Hence, the total number of four-cycles in Q_n is equal to $\frac{1}{4}\binom{n}{2}2^n$. Furthermore, assuming $n > 2$, $|A_4| = 0$. So, we can write

$$|A_0| + |A_1| + |A_2| + |A_3| = \frac{1}{4}\binom{n}{2}2^n. \tag{7}$$

Since any subpath of S of length 2 determines precisely one four-cycle having exactly the three vertices of this subpath in common with S, we have $|A_3| = s(n)$. In order to deal with the numbers $|A_1|$ and $|A_2|$ the following notion is introduced (cf. [13]).

Definition 2 *Let X be a vertex of the snake S. Then the symbol $\alpha(X)$ denotes the total number of four-cycles containing X, but no other vertices of S.*

It is clear immediately that

$$|A_1| = \sum_{X \in S} \alpha(X). \tag{8}$$

Using simple counting arguments (cf. [13, 14]) one can infer that

$$|A_2| = (n-3)s(n) + \frac{1}{2}\sum_{X \in S}\left(\binom{n-2}{2} - \alpha(X)\right) \tag{9}$$

$$= (n-3)s(n) + \frac{1}{2}\binom{n-2}{2}s(n) - \frac{1}{2}\sum_{X \in S}\alpha(X). \tag{10}$$

Substituting expressions (8) and (9) in (7), and leaving out $|A_0|$ provide us with the inequality

$$\left(\binom{n}{2} - 1\right)s(n) \leq \binom{n}{2}2^{n-1} - \sum_{X \in S}\alpha(X). \tag{11}$$

The result in [13] uses the following:

Lemma 1 (Glagolev) *For $n \geq 7$ one has that $\alpha(X) \geq 2$.*

For the proof we refer to [13].

In [15] an upper bound for $s(n)$ is determined by deriving a lower bound for the number of edges in $\mathcal{Y} := Q_n - \mathcal{S}$. A basic tool in Zemor's method is counting vertices of high degree.

Definition 3 *Let A, B, C and D be distinct vertices of Q_n. We say that $\{A, B, C, D\}$ is a claw with center A, if A is adjacent to the other three vertices.*

Another important ingredient in [15] is the $\mathcal{Y}$-degree of some vertex $X \in Q_n$.

Definition 4 *Let X be some vertex of Q_n. The symbol $\delta(X)$ denotes the degree of X in $\mathcal{Y}$, or $\delta(X) := |N(X) \cap \mathcal{Y}|$, where $N(X)$ is the neighborhood of X in Q_n.*

The following property is essential in [15].

Proposition 1 (Zemor) *If $\{A, B, C, D\}$ is a claw, then $\delta(A) + \delta(B) + \delta(C) + \delta(D) \geq n$.*

Proof: From the definition of $\mathcal{S}$ it follows that not all vertices of a claw $\{A, B, C, D\}$ can be vertices of $\mathcal{S}$. The same holds for the claw $\{A+e_i, B+e_i, C+e_i, D+e_i\}$, which is a translation by an edge e_i of the first claw. Applying this observation for all i with $1 \leq i \leq n$ gives the above inequality. ■

It is shown in [15] that any subpath of $\mathcal{S}$ of length 7 contains four vertices being neighbors of four vertices in $\mathcal{Y}$ which form a claw. We call such a configuration an *adjacent claw*. Hence, applying Proposition 1, it follows that any such subpath contains a vertex V such that some neighbor Y of V has degree $\delta(Y) \geq \frac{n}{4}$ in $\mathcal{Y}$. Such a vertex Y is called a "vertex of high degree", and it contributes at least $\frac{n}{4}$ edges in $\mathcal{Y}$. However, disjoint subpaths of $\mathcal{S}$ of length 7 may produce the same vertex of high degree. Zemor continues by proving that whenever this happens, it in turn produces more vertices of high degree. A careful analysis, combining both arguments, yields a lower bound $\frac{1}{36}(s(n) - 10)(\sqrt{\frac{n}{6}} - 6)$ for the number of edges in $\mathcal{Y}$. This in turn provides us with upper bound (5) for the length of a snake $\mathcal{S}$ (cf. Section 1).

3 A combined approach

We are now going to combine the method of [13] and part of the techniques used in [15], which will result in a new upper bound, which is better than bound (1) for $n \geq 17$ and better than bound (5) for $n \leq 15603$.

Basic to our approach is the following property, relating the functions α and δ.

Proposition 2 *Let $X \in \mathcal{S}$ and $Y \in N(X) \cap \mathcal{Y}$. If $\delta(Y) = k$, then $\alpha(X) \geq k-2$.*

Proof: There exist k vertices in $\mathcal{Y}$ at distance 2 from X. Since a pair of vertices at distance 2 from each other determines a unique four-cycle, these k vertices determine k four-cycles through X. At most two of these four-cycles contain a second vertex of $\mathcal{S}$. ∎

From Propositions 1 and 2 it is clear that a claw the vertices of which are adjacent to four different vertices of $\mathcal{S}$ gives rise to a contribution of at least $n-4(2) = n-8$ to the sum $\sum_{X \in \mathcal{S}} \alpha(X)$. The more claws of this type that can be proved to exist, the better a lower bound for $\sum_{X \in \mathcal{S}} \alpha(X)$ we obtain, and the better an upper bound for $s(n)$ by applying (11).

If we apply Zemor's result that any subpath of $\mathcal{S}$ of length 7 has an adjacent claw, we obtain for $n \geq 7$

$$\sum_{X \in \mathcal{S}} \alpha(X) \geq n \left\lfloor \frac{s(n)}{8} \right\rfloor \geq \frac{n\,(s(n)-6)}{8}. \tag{12}$$

Here, we used the property that $s(n)$ is even.

Substituting this lower bound for $\sum_{X \in \mathcal{S}} \alpha(X)$ in (11) yields the following upper bound for the length of a snake

$$s(n) \leq \left(1 - \frac{n-8}{4n^2-3n-8}\right) 2^{n-1} + \frac{6n}{4n^2-3n-8}, \qquad n \geq 7 \tag{13}$$

(cf. also [14]).

This bound is better than bounds (1) and (2) for $n \geq 17$. It is also better than bound (3). It is an improvement of Zemor's bound (5) for $7 \leq n \leq 15603$. However, bound (13) does not improve (4) which was derived by different means. For this reason we shall extend our method. More precisely, we shall prove that any subpath $\mathcal{I}$ of $\mathcal{S}$ of length 9 has two adjacent claws in $\mathcal{Y}$. The vertices of these two claws are not necessarily different and neither are their neighbors in $\mathcal{I}$, as long as the edges connecting the vertices of the claws to the vertices in $\mathcal{I}$ can be chosen such that these all are distinct. In order to deal with the situation that two different vertices of adjacent claws (or of the same claw) are adjacent to the same vertex of $\mathcal{I}$, we prove the following property.

Proposition 3 *Let $X \in \mathcal{S}$ and let $Y, Z \in N(X) \cap \mathcal{Y}$. Then $\alpha(X) \geq \delta(Y) + \delta(Z) - 5$.*

Proof The statement follows immediately from Proposition 2 since four-cycles through X and Y differ from four-cycles through X and Z, except the uniquely determined four-cycle through X, Y, and Z. ■

4 Four lemmas

In this section we shall present four lemmas (cf. Section 1) dealing with claws which exist under typical, rather frequently occurring conditions.

Lemma 2 *Let $\stackrel{e_0}{-} V_0 \stackrel{e_1}{-} V_1 \stackrel{e_2}{-} V_2 \stackrel{e_3}{-} V_3 \stackrel{e_4}{-} V_4 \stackrel{e_5}{-} V_5$ be a subpath of $\mathcal{S}$. If $e_0, e_1 \neq e_4$ and $e_2 \neq e_5$, then $\mathcal{W} := \{V_0 + e_4, V_1 + e_4, V_2 + e_4, V_4 + e_2\}$ is a claw in $\mathcal{Y}$.*

Proof: From the conditions of the Lemma and from inequalities (6) it follows that all vertices of $\mathcal{W}$ are in $\mathcal{Y}$. It is clear that $V_0 + e_4$, $V_1 + e_4$ and $V_2 + e_4$ are all different vertices, and that $V_0 + e_4$ and $V_2 + e_4$ are neighbors of $V_1 + e_4$. Now, we can write $V_4 + e_2 = V_0 + e_1 + e_2 + e_3 + e_4 + e_2 = (V_0 + e_4) + e_1 + e_3$. Hence, $V_4 + e_2 \neq V_0 + e_4$. Similarly, we have $V_4 + e_2 = V_2 + e_3 + e_4 + e_2 = (V_2 + e_4) + e_2 + e_3 \neq V_2 + e_4$. Finally, $V_4 + e_2 = (V_1 + e_4) + e_3$, which shows that $V_4 + e_2$ and $V_1 + e_4$ are adjacent.

■

Lemma 3 *Let $V_0 \stackrel{e_1}{-} V_1 \stackrel{e_2}{-} V_2 \stackrel{e_3}{-} V_3 \stackrel{e_4}{-} V_4 \stackrel{e_5}{-} V_5 \stackrel{e_6}{-} V_6 \stackrel{e_7}{-} V_7$ be a subpath of $\mathcal{S}$. If $e_1 = e_5$ and $e_2 = e_6$, then $\mathcal{V} := \{V_4 + e_3, V_5 + e_3, V_6 + e_3, V_2 + e_5\}$ is a claw in $\mathcal{Y}$.*

Proof: Since $e_2 = e_6$ it follows that $e_5 \neq e_2$ and $e_6 \neq e_3$, because of the first inequality of (6). Furthermore, $e_7 \neq e_3$, since otherwise we would have $V_7 = V_0 + e_4$, and hence $d(V_0, V_7) = 1$, contradicting the second condition for a snake. So, we can apply Lemma 2, yielding the statement of the Lemma. ■

Lemma 4 *Let $V_0 \stackrel{e_1}{-} V_1 \stackrel{e_2}{-} V_2 \stackrel{e_3}{-} V_3 \stackrel{e_4}{-} V_4 \stackrel{e_5}{-} V_5 \stackrel{e_6}{-} V_6 \stackrel{e_7}{-} V_7 \stackrel{e_8}{-}$ be a subpath of $\mathcal{S}$. If $e_1 = e_4$ and $e_3 = e_6$, then $\mathcal{U} := \{V_1 + e_7, V_2 + e_7, V_3 + e_7, V_7 + r\}$ is a claw in $\mathcal{Y}$ for some $r \in \{e_4, e_5\}$. Moreover, if $e_8 \neq e_2$, then $\mathcal{T} := \{V_7 + e_2, V_6 + e_2, V_5 + e_2, V_1 + e_5\}$ is another claw in $\mathcal{Y}$.*

Proof: Since $e_1 = e_4$ and $e_3 = e_6$ it follows that $e_7 \neq e_2$, since otherwise we would have $V_7 = V_0 + e_5$, and hence $d(V_7, V_0) = 1$, contradicting the snake conditions. For similar reasons we have $e_7 \neq e_1, e_3, e_4$. Since $e_7 \neq e_4, e_5$, either $V_7 + e_4 \in \mathcal{Y}$ or $V_7 + e_5 \in \mathcal{Y}$. Assume $V_7 + e_4 \in \mathcal{Y}$. Then $V_7 + e_4 = (V_1 + e_7) + e_2 + e_5$, $V_7 + e_4 = (V_2 + e_7) + e_5$ and $V_7 + e_4 = (V_3 + e_7) + e_5 + e_6$. Hence, $\mathcal{U}$ is a claw in $\mathcal{Y}$ with center $V_2 + e_7$. In case that $V_7 + e_5 \in \mathcal{Y}$ there is a similar proof.

Now, assume $e_8 \neq e_2$. Since we also have $e_5, e_6, e_7 \neq e_2$ and $e_1 \neq e_5$, the vertices $V_7 + e_2, V_6 + e_2, V_5 + e_2$ and $V_1 + e_5$ are all in $\mathcal{Y}$. From the relations $V_7 + e_2 = (V_1 + e_5) + e_4 + e_7$, $V_6 + e_2 = (V_1 + e_5) + e_4$ and $V_5 + e_2 = V_1 + e_2 + e_3 + e_4 + e_5 + e_2 = (V_1 + e_5) + e_2 + e_3$ it follows that $\mathcal{T}$ is a claw in $\mathcal{Y}$ with center $V_6 + e_2$. ■

Lemma 5 *Let* $V_0 \xrightarrow{e_1} V_1 \xrightarrow{e_2} V_2 \xrightarrow{e_3} V_3 \xrightarrow{e_4} V_4 \xrightarrow{e_5} V_5 \xrightarrow{e_6} V_6 \xrightarrow{e_7} V_7 \xrightarrow{e_8}$ *be a subpath of* S *with* $e_1 = e_4 = e_7$. *If* $e_2 \neq e_6$ *and* $e_3 \neq e_8$, *then* $\mathcal{Z} := \{V_1 + e_6, V_2 + e_6, V_3 + e_6, V_7 + e_3\}$ *is a claw in* $\mathcal{Y}$.

Proof: We have $e_6 \neq e_7$, because of (6), and hence $e_6 \neq e_1$. We also have $e_6 \neq e_3$, since otherwise we would have $d(V_2, V_7) = 1$. Therefore, it follows that all vertices of $\mathcal{Z}$ are in $\mathcal{Y}$. Furthermore, we have $e_2 \neq e_5$, since otherwise $d(V_0, V_5) = 1$. The Lemma now follows from the relations $V_2 + e_6 = (V_7 + e_3) + e_5$, $V_1 + e_6 = (V_7 + e_3) + e_2 + e_5$ and $V_3 + e_6 = (V_7 + e_3) + e_3 + e_5$. ■

5 A new upper bound for snakes

The following result can be proved, applying the lemmas of Section 4. Since the proof is a matter of routine case analysis, with many subcases, we refer to [11] for the technical details.

Theorem 1 *Any subpath of* S *of length 9 contributes at least* $2n - 10$ *four-cycles each of which has precisely one vertex in common with* S.

Now the following corollary is immediate.

Corollary 1

$$\alpha(S) = \sum_{X \in S} \alpha(X) \geq \frac{(2n-10)\,(s(n)-8)}{10}, \qquad n \geq 7.$$

Proof:

$$\alpha(S) \geq (2n-10) \left\lfloor \frac{s(n)}{10} \right\rfloor \geq \frac{(2n-10)\,(s(n)-8)}{10}$$

since $s(n)$ is even. ■

Next, using inequality (11), we obtain our main result.

Corollary 2. *For $n \geq 7$, one has*

$$s(n) \leq \left(1 - \frac{2n-20}{5n^2-3n-20}\right) 2^{n-1} + \frac{16n-80}{5n^2-3n-20}$$
$$= \left(1 - \frac{2}{5n+47+\frac{450}{n-10}}\right) 2^{n-1} + \frac{16n-80}{5n^2-3n-20}.$$

We remark that this bound is better than bound (4) for $n \geq 7$ and better than bound (1) for $n \geq 16$. It is an improvement of Zemor's bound (5) for $7 \leq n \leq 19081$.

References

[1] H.L. Abbott and M. Katchalski, On the construction of snake in the box codes, *Utilitas Math.* 40 (1991), 97-116.

[2] K. Deimer, A new upper bound for the length of snakes, *Combinatorica* 5 (1985), 109-120.

[3] P.G. Emelyanov, An upper bound for the length of a snake in the n-dimensional unit cube, *Operation Research and Discrete Analysis* (ed. A. Korshunov), (1997), 23-30.

[4] P.G. Emelyanov and A. Lukito, On the maximal length of a snake in hypercubes of small dimension, to be published in *Discrete Mathematics.*

[5] V.V. Glagolev, An upper bound for the length of a cycle in an unit n-dimensional cube, *Diskret. Analiz.* 6 (1966), 3-7 (in Russian).

[6] W.H. Kautz, Unit distance error checking codes, *IEEE Trans. Electron. Comput.* EC-7 (1958), 179-180.

[7] V. Klee, The use of circuit codes in analog-to-digital conversion, *Graph Theory and its Applications* (ed. B. Harris), Academic Press, New York, 1970, 121-132.

[8] V. Klee, What is the maximum length of a d-dimensional snake?, *Amer. Math. Monthly* 77 (1970), 63-65.

[9] A. Lukito, Upper and lower bounds for the length of snake-in-the-box codes, *Rept.* **98-07**, Faculty of Information Technology and Systems, Delft University of Technology, 1998.

[10] A.Lukito, An upper bound for the length of snake-in-the-box codes, *Proc. of the Sixth Int. Workshop on Algebraic and Combinatorial Coding Theory*, Pskov, Russia, September 6-12, 1998.

[11] A. Lukito and A.J. van Zanten, Stars and snake-in-the-box codes, *Rept.* **98-43**, Faculty of Information Technology and Systems, Delft University of Technology, 1998.

[12] H.S. Snevily, The snake-in-the-box problem: A new upper bound, *Discrete Math.* **133** (1994), 307-314.

[13] F.I. Solov'jeva, An upper bound for the length of a cycle in an n-dimensional unit cube, *Diskret. Analiz.* **45** (1987), 71-76.

[14] A.J. van Zanten and A. Lukito, Snakes and bounds, *Proc. of the Sixth Int. Workshop on Algebraic and Combinatorial Coding Theory*, Pskov, Russia, September 6-12, 1998.

[15] G. Zemor, An upper bound on the size of the snake-in-the-box, *Combinatorica* **17** (1997), 287-298.