Monotone Gray Codes and the Middle Levels Problem

Carla D. Savage *
Department of Computer Science
North Carolina State University
Raleigh, North Carolina 27695-8206
cds@adm.csc.ncsu.edu

Peter Winkler
Bellcore
445 South St.
Morristown, New Jersey 07960-6438
pw@bellcore.com

*Research supported by National Science Foundation Grant No. CCR8906500, National Security Agency Grant No. MDA904-H-1025, and DIMACS

Proposed Running Head:

Gray Codes and the Middle Levels

Direct correspondence, proofs, etc. to:

Carla D. Savage
Department of Computer Science
Box 8206
North Carolina State University
Raleigh, NC 27695-8206

Abstract

An n-bit binary Gray code is an enumeration of all n-bit binary strings so that successive elements differ in exactly one bit position; equivalently, a hamilton path in the Hasse diagram of $\mathcal{B}_n$ (the partially ordered set of subsets of an n-element set, ordered by inclusion.) We construct, for each n, a hamilton path in $\mathcal{B}_n$ with the following additional property: edges between levels $i-1$ and i of $\mathcal{B}_n$ must appear on the path before edges between levels i and $i+1$. Two consequences are an embedding of the hypercube into a linear array which simultaneously minimizes dilation in both directions, and a long path in the middle two levels of $\mathcal{B}_n$.

Using a second recursive construction, we are able to improve still further on this path, thus obtaining the best known results on the notorious "middle levels" problem (to show that the graph formed by the middle two levels of $\mathcal{B}_{2k+1}$ is hamiltonian for all k). We show in fact that for every $\epsilon > 0$, there is an $h \geq 1$ so that if a hamilton cycle exists in the middle two levels of $\mathcal{B}_{2k+1}$ for $1 \leq k \leq h$, then there is a cycle of length at least $(1-\epsilon)N(k)$ for *all* $k \geq 1$, where $N(k) = 2\binom{2k+1}{k}$. Using the fact that hamilton cycles are currently known to exist for $1 \leq k \leq 11$, the construction guarantees a cycle of length at least $.839N(k)$ in the middle two levels of $\mathcal{B}_{2k+1}$ for all k.

1. Introduction

An n-bit binary Gray code is an enumeration of all n-bit strings so that successive elements differ in one bit position. In 1953, Frank Gray patented a simple and efficient scheme to generate a binary Gray code for any value of n [7, 9]. His scheme, now called the "binary reflected Gray code," can be defined recursively as follows. If L_n denotes the listing of n-bit strings, then

$$L_1 = 0,\ 1$$
$$L_n = 0L_{n-1},\quad 1\overline{L_{n-1}},\qquad \text{for } n > 1.$$

Here, aL denotes the list formed from L by adding a to the front of every element, and $\overline{L}$ denotes the reverse of the list L. For example, $L_2 = 00, 01, 11, 10$; $L_3 =$ 000, 001, 011, 010, 110, 111, 101, 100. Since the first and last elements of L_n also differ in one bit position, the code is in fact a cycle. It can be implemented efficiently in the sense that successive elements can be generated in worst case constant time [2].

In practice, Gray codes with certain additional properties may be desirable (see [8] for a survey). For example, note that as the elements of L_n are scanned, the lowest order (rightmost) bit changes 2^{n-1} times, whereas the highest order bit changes only twice, counting the return to the first element. A *balanced* Gray code, in which the 2^n bit changes are distributed as equally as possible among the n bit positions, has long been sought and heuristics have been proposed [15, 19]. Only in the case where n is a power of two is a balanced Gray code known to exist [18]. In other applications, the requirement is to maximize the shortest maximal consecutive sequence of zeroes (or ones) among all bit positions [8].

Here we consider a new constraint. Define the *weight* of a binary string to be the number of 1's in the string. For some applications strings are treated differently according to their weights, in which case it may be desirable to generate first strings of weight 0, then weight 1, etc. Clearly this cannot be done with a Gray code since strings of the same weight are not adjacent. However, there might conceivably be a code which runs through the weight levels two at a time; in other words, a code in which consecutive pairs of strings of weights $i, i+1$ precede those of weights $j, j+1$ for all $i < j$.

One of the main results of this paper is that such a code, which we call a *monotone* Gray code, exists for all values of n. An example is illustrated in Figure 1 and a construction is given in Section 2.

Let $\mathcal{B}_n$ be the n-atom Boolean lattice, i.e. the partially ordered set of subsets of the

00000	11000	01010	11110
00001	10000	01011	11100
00011	10001	01001	11101
00010	10101	01101	11001
00110	10100	00101	11011
00100	10110	00111	10011
01100	10010	01111	10111
01000	11010	01110	11111

Figure 1: A monotone Gray code for $\mathcal{B}_5$.

set $[n] = \{1, 2, \ldots, n\}$, ordered by inclusion. Let $H(\mathcal{B}_n) = (V_n, E_n)$ be the Hasse diagram, or covering graph, of $\mathcal{B}_n$, so that the vertices V_n are the subsets of $[n]$ with two subsets adjacent when they differ in just one element. The correspondence

$$b_1 b_2 \ldots b_n \to \{i \mid b_i = 1\}$$

is a bijection from n-bit binary numbers to subsets of $[n]$, and under this bijection, a Gray code corresponds to a hamilton path in $H(\mathcal{B}_n)$. (See Figure 2.)

Let us partition V_n into levels $\{V_n(i)\}_{0 \le i \le n}$ and E_n into $\{E_n(i)\}_{0 \le i \le n-1}$, where $V_n(i)$ is the set of i-element subsets of $[n]$ and $E_n(i)$ is the set of edges in E_n joining vertices in $V_n(i)$ to vertices in $V_n(i+1)$. Then a monotone Gray code corresponds to a hamilton path p in $H(\mathcal{B}_n)$ with the following property: if edge $e \in E_n(i)$ precedes edge $e' \in E_n(j)$ on p, then $i \le j$.

In Section 4 we show that the monotone Gray code also has an application to the theory of interconnection networks, providing an embedding of the hypercube into a linear array which minimizes dilation in both directions.

Monotone Gray codes split naturally into paths between adjacent levels of $\mathcal{B}_n$. For $i = 0, \ldots, n-1$, let $G_n(i)$ denote the subgraph of $H(\mathcal{B}_n)$ induced by $V_n(i) \cup V_n(i+1)$. It is an open problem to determine whether $G_n(i)$ has a path which includes every vertex of the smaller of the two sets $V_n(i)$, $V_n(i+1)$. In particular, the notorious *middle levels problem* is to determine whether $G_{2k+1}(k)$ has a hamilton path or cycle [3, 5, 11, 12, 16]. Recently, Felsner and Trotter have shown that $G_{2k+1}(k)$ has a cycle of length at least one-fourth the length of a hamilton cycle, giving the first constant factor approximation [6]; we show in Section 3 that monotone Gray codes do even better, yielding paths of more than half the

$\emptyset$

1 2 3 4 5

1,2 2,3 1,3 3,4 1,4 2,4 4,5 2,5 3, 5 1, 5

1,2,3 1,3,4 1,2,4 2,3,4 2,4,5 2,3,5 3,4,5 1,3,5 1,4,5 1,2,5

1,2,3,4 2,3,4,5 1,3,4,5 1,2,4,5 1,2,3,5

1,2,3,4,5

Figure 2: The hamilton path in $H(\mathcal{B}_5)$ corresponding to the monotone Gray code in Figure 1.

length of a hamilton cycle.

In Section 5 we improve the result still further, using another recursive construction to obtain a cycle in $G_{2k+1}(k)$ of length .839 times the number of vertices. This construction is capable of proving factors arbitrarily close to 1, given sufficiently strong base for the recursion.

2. Construction of Monotone Gray Codes

In this section, we give an inductive construction to show that a monotone Gray code exists for every $n \geq 1$. The construction varies slightly according to whether n is even or odd and it relies on the lemma below.

Let S_n denote the symmetric group of all permutations of n elements. For $\pi \in S_n$ and $A \subseteq [n]$, denote by $\pi(A)$ the set $\{\pi(a) \mid a \in A\}$.

Lemma 1 *Let $x_0, x_1, \ldots, x_n$ and $y_0, y_1, \ldots, y_n$ be sequences of subsets of $[n]$ satisfying:*

$$|x_i| = |y_i| = i \quad \text{for} \quad 0 \leq i \leq n$$

and

$$x_i \subset x_{i+1}, \quad y_i \subset y_{i+1} \quad \text{for} \quad 0 \leq i \leq n-1.$$

Then there is a permutation $\pi \in S_n$ such that $\pi(x_i) = y_i$ for $1 \leq i \leq n$.

Proof. For $1 \leq i \leq n$, let $\{a_i\} = x_i \setminus x_{i-1}$ and $\{b_i\} = y_i \setminus y_{i-1}$. Define π by $\pi(a_i) = b_i$ for $1 \leq i \leq n$. □

A path in a graph G will be specified by its sequence of vertices. If $p = x_1, \ldots, x_i$ and $q = y_1, \ldots, y_j$ are disjoint paths in G, and if x_i and y_1 are adjacent in G, then we use the notation $r = p, q$ to denote the path $r = p, q = x_1, \ldots, x_i, y_1, \ldots, y_j$ in G.

Recall that $G_n(i)$ is the subgraph induced by $V_n(i) \cup V_n(i+1)$, where the $V_n(i)$ are the i-element subsets of $[n]$. Note that any monotone Gray code p for $\mathcal{B}_n$ can be written uniquely as $p = p_0, p_1, \ldots, p_{n-1}$ where for $0 \leq i \leq n-1$, p_i is a path in $G_n(i)$. The monotone Gray code construction is given in the proof of Theorem 1 below.

Theorem 1 *For all $n \geq 1$, $\mathcal{B}_n$ has a monotone Gray code. In particular, $H(\mathcal{B}_n)$ has a hamilton path of the form $p_0, p_1, \ldots, p_{n-1}$, where for $0 \leq i \leq n-1$, p_i is a path in $G_n(i)$ of the form*

$$p_i = y_i, \ldots, x_{i+1} \quad \text{(if } i \text{ is even)},$$
$$p_i = x_{i+1}, \ldots, y_i \quad \text{(if } i \text{ is odd)},$$
where $x_i,\ y_i \in V_n(i)$ *for* $0 \le i \le n$ *and for* $0 \le i < n$, $x_i \subset x_{i+1}$ *and* $y_i \subset y_{i+1}$.

Proof. The construction proceeds by induction on n. If $n = 1$, then $x_0 = y_0 = \emptyset$ and $x_1 = y_1 = \{1\}$. The hamilton path is $p = p_0 = \emptyset, \{1\} = y_0, x_1$.

For $A \subseteq [n]$, let A^+ denote the set $A \cup \{n+1\}$. Let $\mathcal{B}_n^+$ denote the sublattice of $\mathcal{B}_{n+1}$ consisting of all subsets of $[n+1]$ which contain $n+1$. For a path, p, in $H(\mathcal{B}_n)$, let p^+ denote the corresponding path in $H(\mathcal{B}_n^+)$ in which each vertex x on p is replaced by x^+.

Assume inductively that $H(\mathcal{B}_n)$ has a hamilton path $p = p_0, \ldots, p_{n-1}$ with vertices $\{x_i\}_{0 \le i \le n}$ and $\{y_i\}_{0 \le i \le n}$ satisfying the properties described in the theorem (see Figure 3.) Note that V_{n+1} is the disjoint union of V_n and V_n^+ and both $H(\mathcal{B}_n)$ and $H(\mathcal{B}_n^+)$ are subgraphs of $H(\mathcal{B}_{n+1})$. We construct the desired hamilton path in $H(\mathcal{B}_{n+1})$ by inserting disjoint paths in $H(\mathcal{B}_n^+)$ which contain all elements of $\mathcal{B}_n^+$ between consecutive vertices on p.

Note that since $p_0, \ldots, p_{n-1}$ is a hamilton path in $H(\mathcal{B}_n)$, the paths $\{p_i^+\}_{0 \le i < n-1}$ are pairwise disjoint and contain all vertices of $H(\mathcal{B}_n^+)$. By Lemma 1, there is a permutation $\pi \in S_{n+1}$ satisfying

$$\pi(x_i) = y_i \quad \text{for} \quad 1 \le i \le n \quad \text{and} \quad \pi(n+1) = n+1.$$

Then the paths $\{\pi(p_i^+)\}_{0 \le i < n-1}$ are also pairwise disjoint and contain all vertices of $\mathcal{B}_n^+$.

Since $y_{i-1} \subset y_i$ and $|y_i| = |y_{i-1}| + 1$, y_i and y_{i-1} are adjacent in $H(\mathcal{B}_n)$ and therefore $\pi(y_{i-1}^+)$ and $\pi(y_i^+)$ are adjacent in $H(\mathcal{B}_n^+)$. If i is even, $\pi(p_i^+)$ starts at $\pi(y_i^+)$ and ends at $\pi(x_{i+1}^+) = y_{i+1}^+$. If i is odd, $\pi(p_i^+)$ starts at $\pi(x_{i+1}^+) = y_{i+1}^+$ and ends at $\pi(y_i^+)$. Then for odd $i < n$,

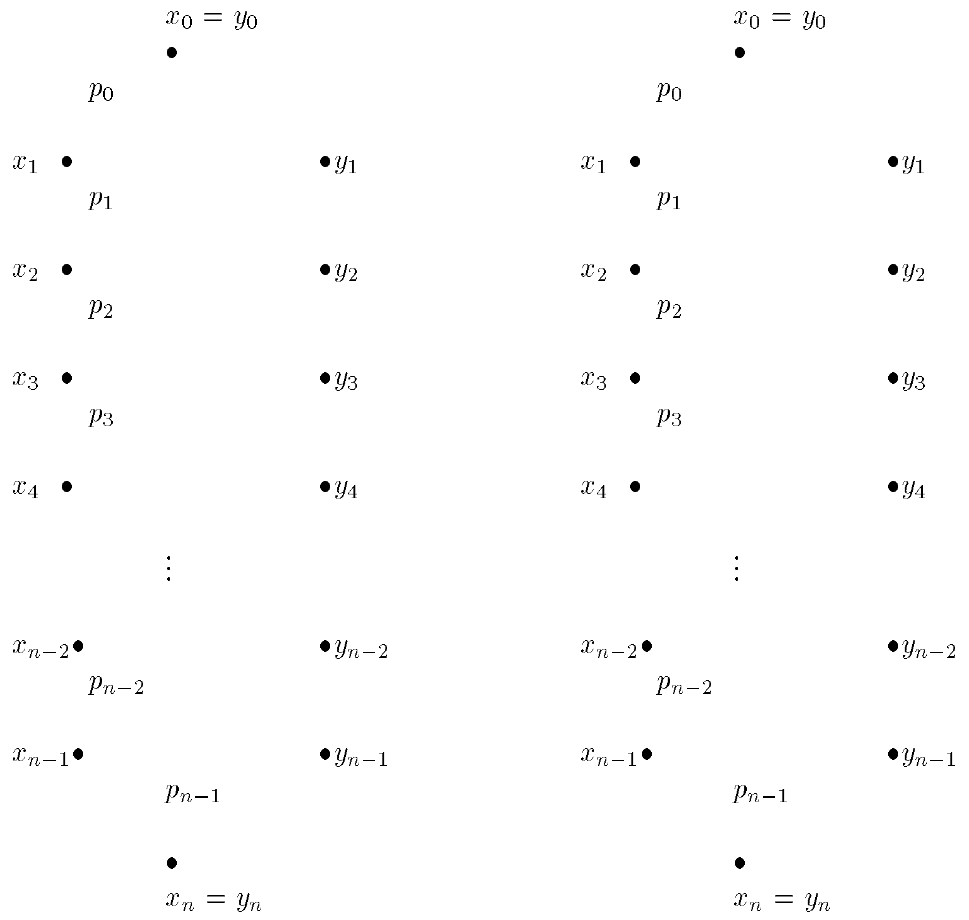
$$\pi(p_{i-1}^+)^{-1},\ \pi(p_i^+)^{-1} = y_i^+, \ldots, y_{i+1}^+$$

is a path in $H(\mathcal{B}_n^+)$ from y_i^+ to y_{i+1}^+.

For odd $i < n$, the last element of p_i is y_i and the first element of p_{i-1} is y_{i-1}, so

$$p_{i-1}, p_i = y_{i-1}, \ldots, y_i$$

is a path in $H(\mathcal{B}_n)$ from y_{i-1} to y_i.



(a) n odd

(b) n even

Figure 3: The monotone Gray code in $H(\mathcal{B}_n)$ guaranteed by the induction hypothesis in the proof of Theorem 1.

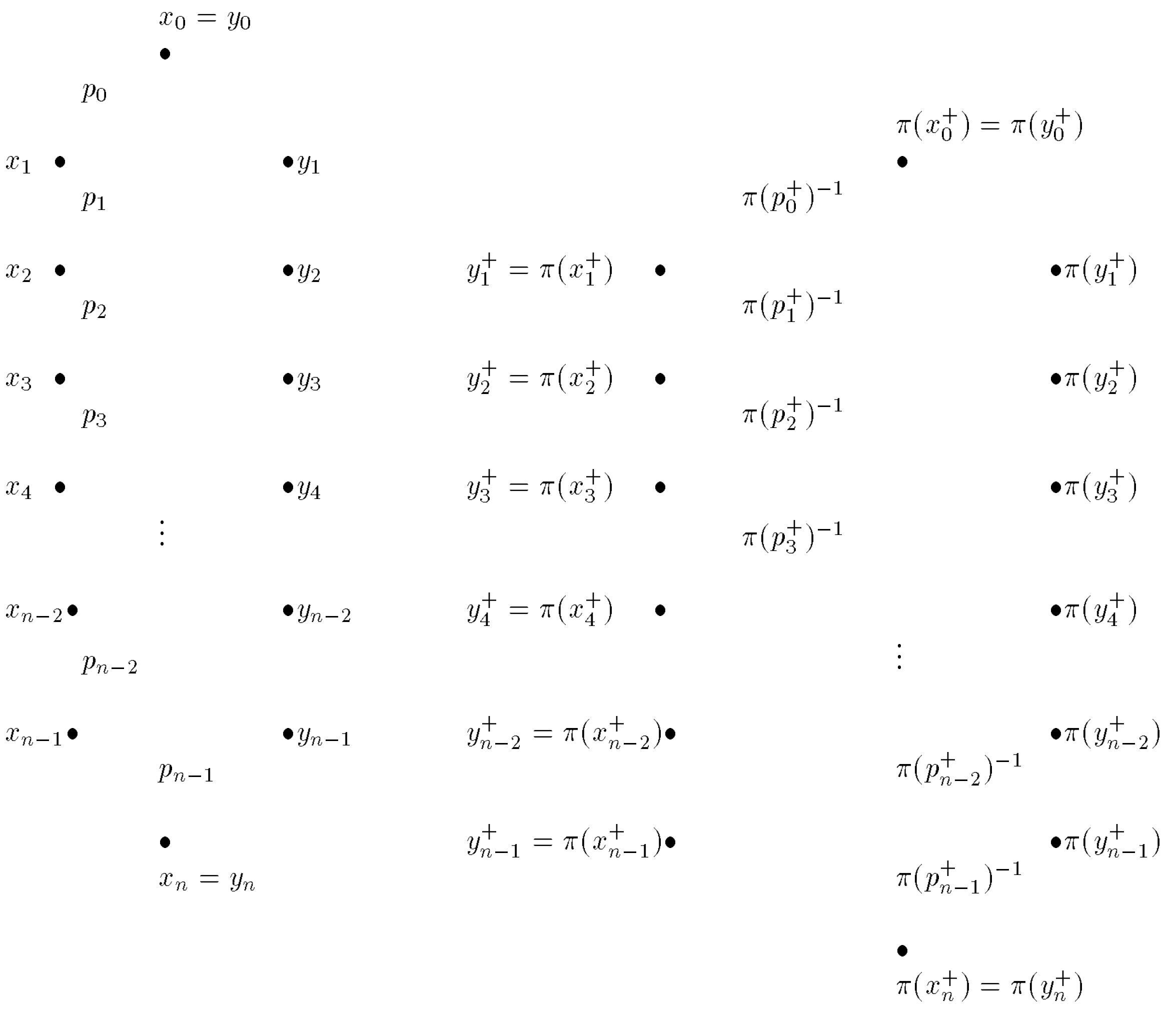


Figure 4: Construction of hamilton path in $H(\mathcal{B}_{n+1})$ from hamilton path in $H(\mathcal{B}_n)$ for n odd.

The required hamilton path, p^*, in $H(\mathcal{B}_{n+1})$ can be constructed as follows. For n odd (see Figure 4),

$$\begin{aligned} p^* \quad = \quad & p_0, p_1, \pi(p_0^+)^{-1}, \pi(p_1^+)^{-1}, \\ & p_2, p_3, \pi(p_2^+)^{-1}, \pi(p_3^+)^{-1}, \\ & p_{i-1}, p_i, \pi(p_{i-1}^+)^{-1}, \pi(p_i^+)^{-1}, \\ & \vdots \\ & p_{n-3}, p_{n-2}, \pi(p_{n-3}^+)^{-1}, \pi(p_{n-2}^+)^{-1}, \\ & p_{n-1}, \pi(p_{n-1}^+)^{-1}. \end{aligned}$$

For n even (see Figure 5),

$$\begin{aligned} p^* \quad = \quad & p_0, p_1, \pi(p_0^+)^{-1}, \pi(p_1^+)^{-1}, \\ & p_2, p_3, \pi(p_2^+)^{-1}, \pi(p_3^+)^{-1}, \\ & p_{i-1}, p_i, \pi(p_{i-1}^+)^{-1}, \pi(p_i^+)^{-1}, \\ & \vdots \\ & p_{n-2}, p_{n-1}, \pi(p_{n-2}^+)^{-1}, \pi(p_{n-1}^+)^{-1}. \end{aligned}$$

Clearly, p^* is a hamilton path in $H(\mathcal{B}_{n+1})$. Note that if n is even, p^* ends at $y_n^* = [n+1] \in V_{n+1}(n+1)$, but if n is odd, p^* ends at $\pi(y_{n-1}^+) \in V_{n+1}(n)$. To show that p^* preserves the properties of the theorem, first let

$$x_i^* = \begin{cases} x_i & \text{if } 0 \le i \le n \\ [n+1] & \text{if } i = n+1 \end{cases}$$

and

$$y_i^* = \begin{cases} y_0 & \text{if } i = 0 \\ \pi(y_{i-1}^+) & \text{if } 1 \le i \le n+1. \end{cases}$$

Then for $0 \le i \le n+1$, $x_i^*, y_i^* \in V_{n+1}(i)$ and for $0 \le i \le n$, $x_i^* \subset x_{i+1}^*$ and $y_i^* \subset y_{i+1}^*$.

Now to show that p^* has the required form, let

$$p_i^* = \begin{cases} p_0 & \text{if } i = 0 \\ p_i, \pi(p_{i-1}^+)^{-1} & \text{if } i \text{ is odd and } i < n-1. \\ \pi(p_{i-1}^+)^{-1}, \; p_i & \text{if } i \text{ is even and } i < n-1. \\ \pi(p_{n-1}^+)^{-1}, \; p_i & \text{if } i = n. \end{cases}$$

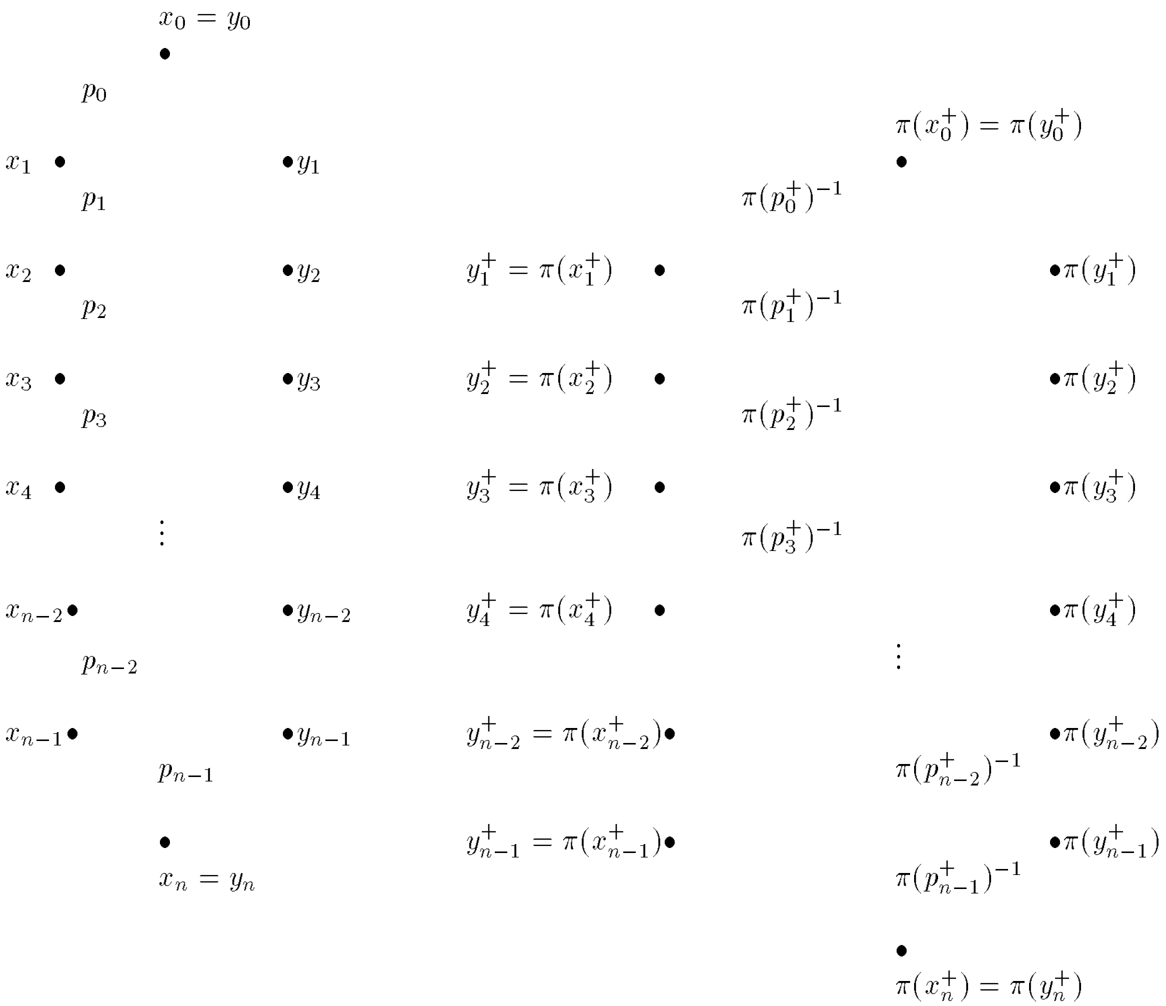


Figure 5: Construction of hamilton path in $H(\mathcal{B}_{n+1})$ from hamilton path in $H(\mathcal{B}_n)$ for n even.

Then,

$$p^* = p_0^*,\ p_1^*,\ p_2^*,\ \ldots, p_n^*$$

and for $0 \leq i \leq n$, p_i^* satisfies

$$p_i^* = \begin{cases} p_i^* = y_i^*, \ldots, x_{i+1}^* & \text{if } i \text{ is even} \\ p_i^* = x_{i+1}^*, \ldots, y_i^* & \text{if } i \text{ is odd} \end{cases}$$

□

3. Paths Between Adjacent Levels

One of the best known of combinatorial problems is the *middle levels* problem which has been variously attributed to Dejter, Erdős, Trotter, Hável, and Kelley (see [3, 5, 11, 12, 16]). The problem is to determine whether it is possible to list all of the k-element and $k+1$-element subsets of the set $\{1, \ldots, 2k+1\}$ in such a way that (i) each subset occurs exactly once, (ii) the k- and $k+1$-element subsets occur alternately, and (iii) consecutive sets on the list differ by exactly one element. Restated, the question is whether there is a hamilton cycle or path in the graph $G_{2k+1}(k)$ formed by the middle two levels in the Hasse diagram of $\mathcal{B}_{2k+1}$. The problem sounds simple, but has resisted the efforts of many researchers. The largest value of k for which a hamilton cycle is known to exist is $k = 11$ (attributed to David Moews and Mike Reid [14]).

Lovász [13] has asked whether *every* connected, vertex transitive graph has a hamilton path. One reason that the middle levels graph has attracted so much attention is that it belongs to this class; thus, either it has a hamilton path, or it provides a negative answer to the question of Lovász. Samples of approaches to the middle levels problem can be found in [4, 5, 11, 16, 17].

In an attempt to make some progress on the problem, it is natural to look for long cycles in order to obtain at least a good lower bound on the length of the longest cycle in the middle two levels graph. As a starting point, a result of Babai [1] on vertex transitive graphs establishes that there is a cycle of length at least $(3N)^{1/2}$, where N is the total number of vertices in the middle two levels. In [16] a technique was presented for constructing a cycle of length at least $N^{.836}$, and then in a recent breakthrough, Felsner and Trotter [6] showed how to construct cycles of length at least $.25N$, giving the first constant factor approximation.

In this section, we show that our monotone Gray code already provides a path of length $.5N$; in Section 5 we use another construction to make a further improvement.

Let p^* be a monotone Gray code in $H(\mathcal{B}_n)$. Let $E(p^*)$ denote the set of edges on p^* and for $0 \leq i \leq n-1$, let

$$e_n(i) = |E_n(i) \cap E(p^*)|.$$

For example, in Figure 2, $e_5(0) = 1,\ e_5(1) = 9,\ e_5(2) = 11,\ e_5(3) = 9,\ e_5(4) = 1$. Then for each i, $0 \leq i \leq n-1$, p^* contains, as a subpath, a path of length $e_n(i)$ in $G_n(i)$, the graph on levels i and $i+1$ of $\mathcal{B}_n$. In particular, $e_{2k+1}(k)$ is the length of the subpath of p^* which passes through the middle levels graph $G_{2k+1}(k)$ of $\mathcal{B}_{2k+1}$.

To obtain a recurrence for e_{2k+1}, note that for $1 \leq i \leq n-1$, if $x \in V_n(i)$ then x is incident with exactly two edges in $S_i = (E_n(i) \cup E_n(i-1)) \cap E(p^*)$. Conversely, every edge in S_i is incident with exactly one vertex of $V_n(i)$. So,

$$2|V_n(i)| = e_n(i) + e_n(i-1) \quad \text{for} \quad 1 \leq i \leq n-1.$$

Since $|V_n(i)| = \binom{n}{i}$, this gives the recurrence

$$\begin{aligned} e_n(0) &= 1 \\ e_n(i) &= 2\binom{n}{i} - e_n(i-1) \quad \text{for} \quad 1 \leq i \leq n-1 \end{aligned}$$

which has solution, for $0 \leq i \leq n-1$,

$$e_n(i) = 2\left(\sum_{j=0}^{i} (-1)^{i-j} \binom{n}{i}\right) - (-1)^i = 2\binom{n-1}{i} - (-1)^i.$$

Thus, for each i, $0 \leq i \leq n-1$, a monotone Gray code p^* in $H(\mathcal{B}_n)$ contains a subpath q_i^* through the graph $G_n(i)$, of length given by $e_n(i)$ above. Assume, without loss of generality, that $i < n/2$, so that $|V_n(i)| \leq |V_n(i+1)|$. Then no path in $G_n(i)$ can have length greater than $2|V_n(i)| = 2\binom{n}{i}$. So, the ratio of the length of q_i^* to the longest path in $G_n(i)$ is

$$\frac{e_n(i)}{2\binom{n}{i}} = \frac{2\binom{n-1}{i} - (-1)^i}{2\binom{n}{i}} \geq \frac{n-i}{n} - \frac{1}{2\binom{n}{i}}.$$

For the middle two levels graph, $G_{2k+1}(k)$, this ratio becomes

$$\frac{2k+1-k}{2k+1} - \frac{1}{2\binom{2k+1}{k}} = \frac{1}{2} + \left(\frac{1}{2(2k+1)} - \frac{1}{2\binom{2k+1}{k}} \right) \geq \frac{1}{2}$$

for $k \geq 1$.

4. Bijections between Hypercubes and Linear Arrays

An *embedding* of a graph $G = (V(G), E(G))$ into a graph $H = (V(H), E(H))$ is a one-to-one mapping $\phi : V(G) \to V(H)$. The *dilation* of the embedding ϕ is

$$\max\{d_H(\phi(u), \phi(v)) \mid u, v \in V(G),\ (u, v) \in E(G)\},$$

where for $x, y \in V(H)$, $d_H(x, y)$ denotes the length of the shortest path from x to y in H.

The n-dimensional hypercube is just our graph $H(\mathcal{B}_n)$. A linear array is a graph A_m with vertex set $\{1, \ldots, \mathrm{m}\}$ in which vertices i and $i+1$ are joined by an edge for $1 \leq i < m$. Any binary Gray code gives an embedding of the array A_{2^n} into the hypercube $H(\mathcal{B}_n)$ which achieves the minimum possible dilation of 1. In the opposite direction, dilation of any embedding of the hypercube into the array is $\Omega(2^n/\sqrt{n})$; in fact, Harper [10] shows that the precise minimum dilation (called the "bandwidth" of the hypercube) is

$$\sum_{j=0}^{n-1} \binom{j}{\lfloor j/2 \rfloor}.$$

This dilation is achievable by embedding level-by-level, ordering lexicographically within each level; but like any level-by-level embedding its inverse has dilation $\Omega(n)$.

Let p^* be a monotone Gray code for $\mathcal{B}_n$ and define $\theta : A_{2^n} \to H(\mathcal{B}_n)$ by $\theta(i) =$ ith vertex of p^*. Then θ has dilation 1 since p^* is a Gray code. We show now that θ^{-1} has dilation $O(2^n/\sqrt{n})$.

For $0 \leq i \leq n$, let w_i be the first, and z_i the last, vertex of $V_n(i)$ on p^*. Then, for $1 \leq i \leq n-1$, the distance between w_{i-1} and z_i on p^* is no more than $e_n(i-2)+e_n(i-1)+e_n(i)-2$. Therefore,

$$\max_{2 \leq i \leq n-1}\{e_n(i-2) + e_n(i-1) + e_n(i) - 2\}$$

gives an upper bound on the dilation of the monotone Gray code embedding. So, using the recurrence for e_n in Section 3, the monotone Gray code embedding has dilation no more than

$$e_n(i-2)+e_n(i-1)+e_n(i) \;=\; 2\binom{n}{i-1}+e_n(i)$$

$$\leq\; 2\binom{n}{i-1}+\binom{n}{i} \;\leq\; 3\binom{n}{\lfloor n/2\rfloor} \in O(2^n/\sqrt{n}).$$

5. Longer Cycles in the Middle Levels

It turns out that with another recursive construction, specialized to paths between adjacent levels of $\mathcal{B}_n$, we can "in principle" achieve constant factors arbitrarily close to 1. In particular, we show that for every $\epsilon > 0$, there is an $h \geq 1$ so that if a hamilton cycle exists in the middle two levels of $\mathcal{B}_{2k+1}$ for $1 \leq k \leq h$, then there is a cycle of length at least $(1-\epsilon)N$ in the middle two levels of $\mathcal{B}_{2k+1}$ for all $k \geq 1$. Using the fact that hamilton cycles are currently known to exist for $1 \leq k \leq 11$, the construction guarantees a cycle of length at least $.839N$ in $\mathcal{B}_{2k+1}$ for all $k \geq 1$.

Let $H_n(k)$ denote the edge-labeled graph whose vertex set is the set of all k-element subsets of $\{1,\ldots,n\}$ and in which vertices S and T are joined by an edge, labeled $S \cup T$, if and only if $|S \cap T| = k-1$. Note, then, that the edge labels are all of the $k+1$-element subsets of $\{1,\ldots,n\}$, with each edge label occurring many times. In this context, the middle levels problem is to find a hamilton cycle in $H_{2k+1}(k)$ in which no edge label occurs more than once.

There is an obvious bijection between cycles of length t in $H_n(k)$, in which no edge label occurs more than once, and cycles of length $2t$ in $G_n(k)$. We will show that if $h \geq 1$ is such that $G_{2i+1}(i)$ has a hamilton cycle for $1 \leq i \leq h$, then for all $n-3 \geq k \geq 2$ and $n \geq 6$, $H_n(k)$ has a cycle of length $f_n(k,h)$, defined below, in which no edge label is repeated. For non-negative integers n, k and h let

$$f_n(k,h) = \sum_{i=0}^{h}\binom{n-1-2i}{k-i}\left(\binom{2i}{i}-\binom{2i}{i-1}\right). \tag{1}$$

(The binomial coefficient $\binom{a}{b}$ is defined to be zero unless $a \geq b \geq 0$.) First we establish the behavior of f.

Lemma 2

(a) f satisfies the recursion $f_n(k,h) = f_{n-1}(k,h) + f_{n-1}(k-1,h)$ except when $n = 2k+1 \le 2h+1$;

(b) when $n \le 2h+1$, $f_n(k,h) = \binom{n}{k}$ for $k < n/2$ and $f_n(k,h) = \binom{n}{k+1}$ for $k \ge n/2$;

(c) $f_n(k,h) = f_n(n-k-1,h)$ for all n, k and h;

(d) $f_n(k,h) = \binom{n}{k}$ when $k \le h$ and $k < n/2$.

Proof. Note that $\binom{s}{t} = \binom{s-1}{t} + \binom{s-1}{t-1}$ for all values of s and t *except* $s = t = 0$. Thus,

$$f_n(k,h) = \sum_{i=0}^{h} \left(\binom{n-2-2i}{k-i} + \binom{n-2-2i}{k-i-1} \right) \left(\binom{2i}{i} - \binom{2i}{i-1} \right)$$

$$= f_{n-1}(k,h) + f_{n-1}(k-1,h)$$

unless there is a summand

$$\binom{0}{0} \left(\binom{2i}{i} - \binom{2i}{i-1} \right)$$

in the expression for $f_n(k,h)$. This happens precisely when $n = 2k+1$ and $0 \le k \le h$, in which case

$$f_{2k+1}(k,h) = f_{2k}(k,h) + f_{2k}(k-1,h) + \binom{2k}{k} - \binom{2k}{k-1}.$$

This proves part (a) of the statement of the lemma; (b) is proved by induction on n as follows.

Note that $f_n(k,h) = 0$ for $n < 1$, and that $f_1(k,h) = 0$ except when $k = 0$ and $h \ge 0$ where $f_1(k,h) = 1$ as predicted by the exception to the recursion.

Now assuming that the second statement of the Lemma holds for $n-1$, and that $n \le 2h+1$, we have for $k < (n-1)/2$ (since (a) applies:)

$$f_n(k,h) = f_{n-1}(k,h) + f_{n-1}(k-1,h) = \binom{n-1}{k} + \binom{n-1}{k-1} = \binom{n}{k};$$

for $k = (n-1)/2$:

$$f_n(k,h) = f_{2k+1}(k,h) = f_{2k}(k,h) + f_{2k}(k-1,h) + \binom{2k}{k} - \binom{2k}{k-1}$$

$$= \binom{2k}{k+1} + \binom{2k}{k-1} + \binom{2k}{k} - \binom{2k}{k-1} = \binom{n}{k};$$

for $k = n/2$:

$$f_n(k,h) = f_{2k}(k,h) = f_{2k-1}(k,h) + f_{2k-1}(k-1,h)$$
$$= \binom{2k-1}{k+1} + \binom{2k-1}{k-1} = \binom{2k-1}{k+1} + \binom{2k-1}{k} = \binom{n}{k+1};$$

and finally, for $k > n/2$:

$$f_n(k,h) = f_{n-1}(k,h) + f_{n-1}(k-1,h) = \binom{n-1}{k+1} + \binom{n-1}{k} = \binom{n}{k+1}.$$

Parts (c) and (d) are now routine consequences of (a) and (b), with the aid of an induction on n starting at $n = 2h+1$. $\square$

Say that a cycle C in $H_n(k)$ satisfies property $\mathcal{P}$ if all edge labels on C are distinct.

Lemma 3 *For $k < n/2$, if $C : A_1, A_2, \ldots, A_t$ is a cycle in $H_n(k)$ satisfying $\mathcal{P}$, then the following sequence*

$$C' : [n] \setminus (A_1 \cup A_2), \quad [n] \setminus (A_2 \cup A_3), \quad \ldots, \quad [n] \setminus (A_t \cup A_1)$$

is a cycle in $H_n(n-k-1)$ satisfying $\mathcal{P}$.

Proof. The elements of C' are distinct since C satisfies $\mathcal{P}$. To show that the edge labels on C' are distinct, first note that since $A_i \cup A_{i+1}$ and $A_{i+1} \cup A_{i+2}$ are distinct,

$$k \geq |(A_i \cup A_{i+1}) \cap (A_{i+1} \cup A_{i+2})| = |A_{i+1} \cup (A_i \cap A_{i+2})|$$

and therefore, $A_i \cap A_{i+2} \subseteq A_{i+1}$. So,

$$[n] \setminus (A_i \cup A_{i+1}) \quad \cup \quad [n] \setminus (A_{i+1} \cup A_{i+2})$$
$$= [n] \setminus (A_{i+1} \cup (A_i \cap A_{i+2})) \;=\; [n] \setminus A_{i+1}.$$

Since the A_i are distinct, the unions of successive pairs of vertices on C' must be distinct. $\square$

Lemma 4 *If $C : A_1, A_2, \ldots, A_t$ is a cycle of $H_n(k)$ satisfying $\mathcal{P}$, then there is another cycle in $H_n(k)$, $C' : B_1, B_2, \ldots, B_t$, satisfying $\mathcal{P}$, in which*

$$B_1 = \{n, n-1, \ldots, n-k+1\} \quad \text{and} \quad B_t = \{n-1, n-2, \ldots, n-k\}.$$

Proof. A_1 and A_t must have the form

$$A_1 = Z \cup \{a\}, \quad A_t = Z \cup \{b\},$$

for some $a, b \in [n]$ and a $k-1$-subset, $Z \subseteq [n] \setminus \{a, b\}$. Let π be a permutation of $1 \dots n$ with

$$\pi(Z) = \{n-1, n-2, \dots, n-k+1\}, \quad \pi(a) = n, \quad \pi(b) = n-k.$$

Then the required cycle is $C' : \pi(A_1), \pi(A_2), \dots, \pi(A_t)$. □

Say that a cycle C in $H_n(k)$ satisfies property $\mathcal{X}_n(k)$ if no edge label of C is the set $\{n-1, n-2, \dots, n-k-1\}$. If J is a graph (path) whose vertices and edges are labeled by subsets of S, then for $y \notin S$, let $y \circ J$ be the graph (path) obtained from H by adding element y to the label of each vertex and each edge of J. If P is the path $P = A_1, A_2, \dots, A_t$ then P^{-1} denotes the path $P^{-1} = A_t, A_{t-1}, \dots, A_1$.

Theorem 2 *Let $h \geq 1$ be such that $H_{2i+1}(i)$ has a hamilton cycle for all $1 \leq i \leq h$. Then for all $n \geq 6$ and $2 \leq k \leq n-3$, $H_n(k)$ has a cycle $C_n(k)$ which*

(1) has length $f_n(k, h)$,

(2) has the form $C_n(k) = \{n, n-1, \dots, n-k+1\}, \dots, \{n-1, n-2, \dots, n-k\}$,

(3) satisfies property $\mathcal{P}$, and

(4) satisfies $\mathcal{X}_n(k)$ if $k \leq \lfloor n/2 \rfloor - 1$.

Proof. Consider first the case $k = 2$. We show by induction on n the existence of a cycle $C_n(2)$ in $H_n(2)$ which satisfies the requirements of the theorem and has the form

$$C_n(2) = \{n, n-1\}, \{n, 1\}, \{n, n-2\}, \{n, 2\}, \dots, \{n-1, n-2\}.$$

For $n = 6$, the cycle

$$C_n(6) = \{6,5\}, \{6,1\}, \{6,4\}, \{6,2\}, \{6,3\}, \{3,4\}, \{2,4\},$$

$$\{2,5\}, \{1,2\}, \{1,4\}, \{1,3\}, \{2,3\}, \{3,5\}, \{1,5\}, \{4,5\}$$

satisfies $\mathcal{P}$ and $\mathcal{X}_6(2)$ and has length $\binom{6}{2}$, which is equal to $f_6(2)$ by Lemma 2(b), if $6 \leq 2h+1$, and otherwise, by Lemma 2(d). For $n > 6$, assume the desired cycle $C_{n'}(2)$ exists for $6 \leq n' < n$. Then, by induction, the cycle

$$C_n(2) = \{n, n-1\}, \{n, 1\}, \{n, n-2\}, \{n, 2\}, \{n, 3\} \dots, \{n, n-3\}, C_{n-1}^{-1}(2))$$

(1) has length $n + \binom{n-1}{2} = \binom{n}{2} = f_n(2,h)$ by Lemma 2(b), if $n < 2h+1$, and otherwise by Lemma 2(d),

(2) has the form $C_n(2) = \{n, n-1\}, \{n, 1\}, \{n, n-2\}, \{n, 2\}, \ldots, \{n-1, n-2\}$, since the first vertex of $C_{n-1}(2)$ is $\{n-1, n-2\}$,

(3) satisfies $\mathcal{P}$ since both subpaths do. Further, since the last vertex of $C_{n-1}(2)$ is $\{n-2, n-3\}$,

$$\{n, n-3\} \cup \mathrm{last}(C_{n-1}(2)) = \{n, n-2, n-3\}$$

which is not the union of any other consecutive sets on $C_n(2)$.

(4) Finally, $C_n(k)$ satisfies $\mathcal{X}_n(k)$ since the first and last sets of $C_{n-1}(2)$ are $\{n-1, n-2\}$ and $\{n-2, n-3\}$ which have the forbidden union $\{n-1, n-2, n-3\}$, but these sets are no longer consecutive on $C_n(2)$.

We proceed with the proof of the theorem by induction on (n,k), with basis $k = 2$, just covered. Let (n,k) satisfy $n \geq 6$ and $2 < k \leq n-3$ and assume inductively that the theorem is true for all (n', k') with $n' \geq 6$, $2 \leq k' \leq n'-3$ and either $k' < k$ or ($k' = k$ and $n' < n$). If $k > \lfloor (n-1)/2 \rfloor$, then $n-k-1 \leq \lfloor (n-1)/2 \rfloor < k$, so by induction, there is a cycle $C_n(n-k-1)$ in $H_n(n-k-1)$ satisfying the properties of the theorem, in particular, it satisfies $\mathcal{P}$ and has length $f_n(n-k-1,h) = f_n(k,h)$ by Lemma 2(c). By Lemmas 3 and 4, $C_n(n-k-1)$ can be transformed into a cycle $C_n(k)$ in $H_n(k)$ which has the same length and also satisfies $\mathcal{P}$.

Otherwise, we have $2 < k \leq \lfloor (n-1)/2 \rfloor$. If $n = 2k+1$ and $k \leq h$, then let $C_n(k)$ be a known hamilton cycle in $H_{2k+1}(k)$. By Lemma 4, we may assume $C_n(k)$ satisfies properties (2) and (3) of the theorem. The length of $C_n(k)$ is $\binom{n}{k} = f_n(k,h)$ by Lemma 2(b). Note in this case that $k > \lfloor n/2 \rfloor - 1$ so $\mathcal{X}_n(k)$ need not be satisfied.

Finally, if $2 < k \leq \lfloor (n-1)/2 \rfloor$ but ($n \neq 2k+1$ or $k > h$), we proceed as follows. Since

$$2 \leq k-1 \leq \lfloor (n-1/2) \rfloor - 1 \leq \lfloor n/2 \rfloor - 1,$$

by induction there is a cycle $C_{n-1}(k-1)$ in $H_{n-1}(k-1)$ of length $f_{n-1}(k-1,h)$ satisfying $\mathcal{P}$ and $\mathcal{X}_{n-1}(k-1)$ and having the form

$$C_{n-1}(k-1) = \{n-1, \ldots, n-k+1\}, \ldots, \{n-2, \ldots, n-k\}.$$

Also, since $n \geq 7$, then $k \leq (n-1)-3$, so by induction, there is a cycle $C_{n-1}(k)$ of length

$f_{n-1}(k,h)$ in $H_{n-1}(k)$ satisfying $\mathcal{P}$ (but not necessarily $\mathcal{X}_{n-1}(k)$), and having the form

$$C_{n-1}(k) = \{n-1,\ldots,n-k\},\ldots,\{n-2,\ldots,n-k-1\}.$$

We now show that

$$C_n(k) = n \circ C_{n-1}(k-1),\ C_{n-1}^{-1}(k)$$

is a cycle in $H_n(k)$ which satisfies all the requirements of the theorem.

Clearly the length of $C_n(k)$ is

$$f_{n-1}(k-1,h) + f_{n-1}(k,h) = f_n(k,h),$$

by Lemma 2(a), and it is easy to check that it has the form (2) of the theorem. To see that it has property $\mathcal{X}_n(k)$, note that since $C_{n-1}(k)$ satisfies $\mathcal{P}$, the forbidden set $\{n-1,\ldots,n-k-1\}$ appears only as the union of the first and last sets on $C_{n-1}(k)$. But these sets are no longer consecutive on $C_n(k)$. To check property $\mathcal{P}$, it is only necessary to check the the new links between the two cycles, since both cycles satisfy $\mathcal{P}$. Except for the element n, the union of the first and last sets of $C_n(k)$ is the same as the union of the first and last sets of $C_{n-1}(k-1)$, which are no longer consecutive in $C_n(k)$. As for the middle link, the last element of $n \circ C_{n-1}(k-1)$ and the first element of $C_{n-1}^{-1}(k)$ have union

$$\{n\} \cup \{n-2,\ldots,n-k-1\}$$

which cannot be the union of consecutive sets in $n \circ C_{n-1}(k-1)$, since $C_{n-1}(k-1)$ satisfies property $\mathcal{X}_{n-1}(k-1)$. □

Assume that h is such that $H_{2i+1}(i)$ has a hamilton cycle for $1 \le i \le h$. Since a cycle of length t in $H_n(k)$ which satisfies $\mathcal{P}$, corresponds to a cycle of length $2t$ in the graph, $G_n(k)$, formed by levels k and $k+1$ of $\mathcal{B}_n$, Theorem 2 shows that $G_n(k)$ has a cycle of length $2f_n(k,h)$. So, there is a cycle in the middle levels, $G_{2k+1}(k)$ of length

$$2f_{2k+1}(k,h) \;=\; 2\sum_{i=0}^{h} \binom{2(k-i)}{k-i} \left(\binom{2i}{i} - \binom{2i}{i-1} \right).$$

Comparing this with the length of a hamilton cycle in $G_{2k+1}(k)$ gives the ratio

$$\frac{2f_{2k+1}(k,h)}{2\binom{2k+1}{k}} \ge \sum_{i=0}^{h} \frac{1}{2^{2i+1}} \left(\binom{2i}{i} - \binom{2i}{i-1} \right) \tag{2}$$

since $\binom{2(k-i)}{k-i}/\binom{2k+1}{k} \geq 1/2^{2i+1}$. A routine induction yields

$$\sum_{i=0}^{h} \frac{1}{2^{2i+1}} \left(\binom{2i}{i} - \binom{2i}{i-1} \right) = 1 - \frac{\binom{2(h+1)}{h+1}}{2^{2(h+1)}}. \qquad (3)$$

Theorem 3 *The middle levels graph, $G_{2k+1}(k)$, has a cycle of length at least $.839N(k)$ where $N(k)$ is the number of vertices of $G_{2k+1}(k)$.*

Proof. Since hamilton cycles have been shown to exist in $G_{2i+1}(i)$ for all $1 \leq i \leq 11$, $G_{2k+1}(k)$ has a cycle of length at least $2f_{2k+1}(k, 11)$. Since $N(k) = 2\binom{2k+1}{k}$, from (2) and (3) we have

$$\frac{2f_{2k+1}(k,11)}{N(k)} \;\geq\; 1 - \frac{\binom{24}{12}}{2^{24}} \;\geq\; .839.$$

□

Using Stirling's approximation on the right-hand side of (3) gives, for some positive constant, c,

$$\frac{2f_{2k+1}(k,h)}{N(k)} \;\geq\; 1 - \frac{c}{\sqrt{h+1}}$$

which converges to 1 as h approaches infinity. This establishes the following result.

Theorem 4 *For any $\epsilon > 0$, there is an $h \geq 1$ so that if $G_{2i+1}(i)$ has a hamilton cycle for $1 \leq i \leq h$, then the middle levels graph, $G_{2k+1}(k)$, has a cycle of length at least $(1-\epsilon)N(k)$ for all $k \geq 1$ where $N(k)$ is the number of vertices in $G_{2k+1}(k)$.*

6. Concluding Remarks

We mention an open problem related to monotone Gray codes, due to Felsner and Trotter. As in [6], define an $\alpha-$sequence for $\mathcal{B}_n$ to be a sequence $C_1, \ldots, C_h$ of subsets of $[n]$ satisfying $C_j \not\subseteq C_{i-1} \cup C_i$ for $1 < i < j \leq h$. It is shown in [6] that an $\alpha-$sequence for $\mathcal{B}_n$ can have length at most $2^{n-1} + \lfloor (n+1)/2 \rfloor$. Furthermore, an $\alpha-$sequence C achieving this length must be a subsequence of a monotone Gray code p^* for $\mathcal{B}_n$ in which every edge of p^* is incident with some element of C.

Can $\alpha-$sequences of length $2^{n-1} + \lfloor (n+1)/2 \rfloor$ be constructed for all n? The answer is known to be yes only for $n \leq 8$ and for general n, Felsner and Trotter show how to construct $\alpha-$sequences of length at least $2^{n-2} + \lfloor (n+1)/2 \rfloor$. This result is used in [6] to establish a lower bound of $1 + \lceil \log_2(\mathrm{ht}(I)) \rceil$ on the chromatic number of the diagram of any interval order, I. The existence of $\alpha-$sequences of length at least 2^{n-1} would improve the lower bound to $2 + \lceil \log_2(\mathrm{ht}(I)) \rceil$, thereby matching the upper bound.

The middle levels problem remains open, as does the following generalization: Does the bipartite graph formed by any two adjacent levels of $\mathcal{B}_n$ have a cycle containing every vertex of the smaller partite set? This is equivalent to asking whether, for $k < n/2$, $G_n(k)$ has a hamilton path satisfying property $\mathcal{P}$. The construction of this paper also gives long cycles in $G_n(k)$.

Another variation is the *antipodal layers problem:* Is it true for all values $n \geq k \geq 0$ that there is a hamilton path among the k-subsets and $(n-k)$ - subsets of $\{1, \ldots, n\}$, where two sets are joined by an edge if and only if one is a subset of the other? Results for limited values of k and n are given in [11] and [17]. It would be interesting in investigate long cycles in these graphs for general n and k.

Acknowledgement: Special thanks to Tom Trotter who inspired this work by telling us about $\alpha-$sequences.

References

[1] L. Babai, Long cycles in vertex-transitive graphs, *Journal of Graph Theory* **3** (1979), 301-304.

[2] J. R. Bitner, G. Ehrlich, and E. M. Reingold, Efficient generation of the binary reflected Gray code and its applications, *Communications of the ACM* **19** No. 9 (1976), 517-521.

[3] I. J. Dejter, Hamilton Cycles and quotients of bipartite graphs, *in* "Graph Theory and Applications to Algorithms and Computer Science," (eds. Y. Alavi *et al.*) Wiley, New York, 1985, 189-199.

[4] I. J. Dejter and J. Quintana, Long cycles in revolving door graphs, *Congressus Numerantium* **60** (1987), 163-168.

[5] D. Duffus, B. Sands, and R. Woodrow, Lexicographic matchings cannot form hamiltonian cycles, *Order* **5** (1988), 149-161.

[6] S. Felsner and W. T. Trotter, Colorings of diagrams of interval orders and α-sequences, preprint (1992).

[7] E. N. Gilbert, Gray codes and paths on the n-cube, *Bell Systems Technical Journal* (1958), 815-826.

[8] L. Goddyn, G. M. Lawrence, and E. Nemeth, Gray codes with optimized run lengths, *Utilitas Mathematica* **34** (1988), 179-192.

[9] F. Gray, Pulse code communication, U. S. Patent 2632058 (1953).

[10] L. H. Harper, Optimal Numberings and Isoperimetric Problems on Graphs, *J. Comb. Theory* **1** (1966), 385-393.

[11] G. Hurlbert, The antipodal layers problem, to appear in *Discrete Math.*.

[12] H. A. Kierstead and W. T. Trotter, Explicit matchings in the middle levels of the Boolean lattice, *Order* **5** (1988), 163-171.

[13] L. Lovász, Problem 11 **in** "Combinatorial Structures and their Applications", Gorden and Breach, London, 1970.

[14] D. Moews and M. Reid, electronic mail communication (via J. Gallian.)

[15] J. E. Ludman and J. L. Sampson, A technique for generating Gray codes, *J. Stat. Plan. Infer.* **5** (1981), 171-180.

[16] C. D. Savage, Long cycles in the middle two levels of the Boolean lattice, *Ars Combinatoria*, to appear.

[17] J. E. Simpson, Hamiltonian bipartite graphs, preprint.

[18] D. G. Wagner and J. West, Construction of uniform Gray codes, *Congressus Numerantium* 80.

[19] V. E. Vickers and J. Silverman, A technique for generating specialized Gray codes, *IEEE Transactions on Computers* **C-29** (1980), 329-331.