

An update on the middle levels problem

Ian Shields*

IBM, P.O. Box 12195, Research Triangle Park, North Carolina 27709, USA

Brendan J. Shields

Department of Physics, Massachusetts Institute of Technology, 77 Massachusetts Avenue Cambridge, MA 02139-4307, USA

Carla D. Savage¹

Department of Computer Science, North Carolina State University, Box 8206, Raleigh, North Carolina 27695-8206, USA

Abstract

The middle levels problem is to find a Hamilton cycle in the middle levels, M_{2k+1} , of the Hasse diagram of \mathcal{B}_{2k+1} (the partially ordered set of subsets of a $2k + 1$ -element set ordered by inclusion). Previously, the best known, from [1], was that M_{2k+1} is Hamiltonian for all positive k through $k = 15$. In this note we announce that M_{33} and M_{35} have Hamilton cycles. The result was achieved by an algorithmic improvement that made it possible to find a Hamilton path in a reduced graph (of complementary necklace pairs) having 129,644,790 vertices, using a 64-bit personal computer.

Key words: Hamilton cycles, middle levels, Boolean lattice, necklaces

1 Introduction

Let \mathcal{B}_n be the n -atom Boolean lattice, i.e., the partially ordered set of subsets of $[n] = \{1, 2, \dots, n\}$, ordered by inclusion. The Hasse diagram of \mathcal{B}_n is iso-

* Corresponding author.

Email addresses: `ishields@us.ibm.com` (Ian Shields),
`bshields@fas.harvard.edu` (Brendan J. Shields), `savage@cayley.csc.ncsu.edu`
(Carla D. Savage).

¹ Research supported in part by NSF grant DMS-0300034

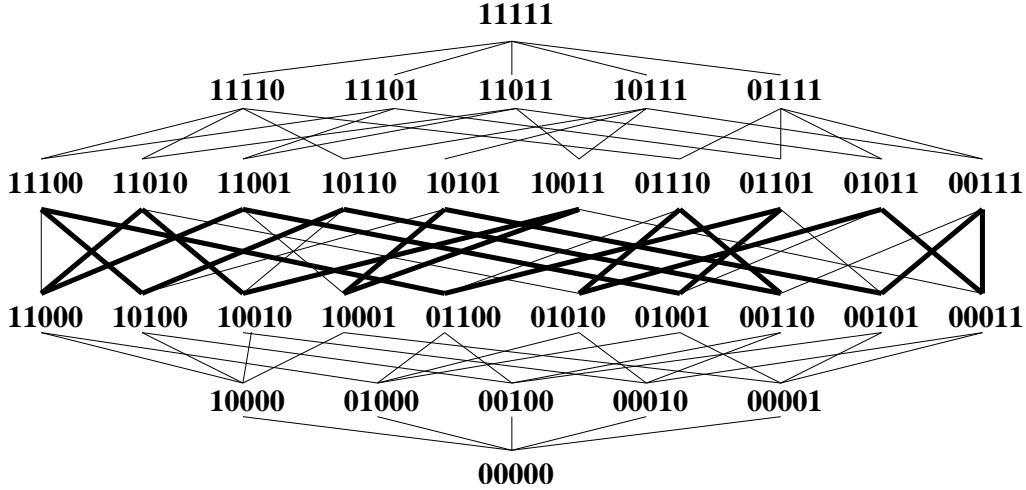


Fig. 1. The Hasse diagram of \mathcal{B}_5 showing a Hamilton cycle in M_5 .

morphic to the n -cube, whose vertices are the n -bit binary numbers, with two numbers adjacent if they differ in one bit position (Figure 1). The i th level of \mathcal{B}_n can thus be viewed as the set of n -bit binary numbers with i ones.

The *middle levels problem* is to determine if there is a Hamilton cycle in the subgraph M_{2k+1} of \mathcal{B}_{2k+1} induced by the middle levels k and $k+1$. The heavier lines in Figure 1 show a Hamilton cycle in M_5 . The graph M_{2k+1} gained notoriety as an example of a connected, vertex transitive graph, all of which were conjectured by Lovász [2] to have Hamilton paths.

The middle levels problem remains open, in spite of the efforts of many [3–8]. In 1990, in unpublished work, Moews and Reid verified that M_{2k+1} is Hamiltonian for $1 \leq k \leq 11$. In 1999, we extended this for $12 \leq k \leq 15$ [1]. In this note we announce that M_{33} and M_{35} are Hamiltonian. The new results are due to an algorithmic improvement that made it possible to find a Hamilton path in a reduced graph of complementary necklace pairs having 129,644,790 vertices, using a 64-bit personal computer.

It is known that any connected vertex transitive graph G with $V(G)$ vertices has a cycle of length at least $\sqrt{3V(G)}$ [9]. For M_{2k+1} , the best absolute lower bound is given by the following theorem in [10].

Theorem 1 *The middle levels graph, M_{2k+1} , has a cycle of length at least*

$$\left(1 - \frac{\binom{2(j+1)}{j+1}}{2^{2(j+1)}}\right) V(M_{2k+1})$$

if for every $i \leq j$, M_{2i+1} has a Hamilton cycle.

Since our new results show M_{2k+1} Hamiltonian for $1 \leq k \leq 17$, it follows that M_{2k+1} has a cycle of length at least $0.867V(M_{2k+1})$.

Recently, it was shown that M_{2k+1} is “asymptotically Hamiltonian” in the following sense [11]: *There is a constant c such that for all k , M_{2k+1} has a cycle of length at least $(1 - c/\sqrt{k})V(M_{2k+1})$.* In the other direction, it has been shown that M_{2k+1} has a closed spanning walk in which no edge and no vertex occurs more than twice [12].

In Section 2 we describe a reduction of the problem which lessens the memory and computation requirements. In Section 3 we review the Hamilton cycle heuristic (SS) from [1]. In Section 4 we describe the improved SSS heuristic which made the present results possible.

2 Reducing the problem

We sketch here the reduction of the middle levels problem that was used in the earlier work of Dejter [3] and of Moews and Reid.

Given a string $x = x_1x_2\dots x_n$ of n symbols, define the cyclic shift σ by $\sigma(x_1x_2\dots x_n) = x_2x_3\dots x_nx_1$. Let $\sigma^1(x) = \sigma(x)$ and for $i \geq 0$, $\sigma^{i+1}(x) = \sigma(\sigma^i(x))$. Define a relation \sim on the set of n -bit binary numbers (regarded as n -bit strings) by $x \sim y$ if and only if $y = \sigma^i(x)$ for some integer i . The relation \sim is an equivalence relation and the equivalence classes are called *necklaces*. Denote the necklace of x by $\nu(x)$.

Fix $n = 2k + 1$. We use necklaces to define a quotient graph of the middle levels graph M_n . Let N_n be the graph whose vertices are the necklaces of the vertices of M_n (i.e., necklaces of $2k + 1$ -bit strings with k or $k + 1$ ones). The edges of N_n are those pairs $\nu(x)\nu(y)$ such that $xz \in E(M_n)$ for some $z \in \nu(y)$. Note that the necklace of a $2k + 1$ -bit binary number with k or $k + 1$ ones has exactly $2k + 1$ elements, so N_n is smaller than M_n by a factor of n .

The *complement* \bar{b} of a binary digit b is 1 if $b = 0$ and 0 if $b = 1$. Extend this to binary strings by bitwise complement and to necklaces by $\overline{\nu(x)} = \nu(\bar{x})$. Note that this is well-defined since $y = \sigma^i(x)$ if and only if $\bar{y} = \sigma^i(\bar{x})$.

We use these complementary necklace pairs to further reduce the problem size by a factor of 2 by observing that $\nu(x)\nu(y) \in E(N_n)$ if and only if $\overline{\nu(x)}\overline{\nu(y)} \in E(N_n)$. Now define an equivalence relation $\overset{\circ}{\sim}$ on the necklaces in $V(N_n)$ by $X \overset{\circ}{\sim} Y$ if either $X = Y$ or $X = \overline{Y}$ and denote the equivalence class of $\overset{\circ}{\sim}$ containing X by $\rho(X)$. Since every string in X has odd length, $X \neq \overline{X}$, and every equivalence class $\rho(X)$ has exactly 2 elements. Construct the reduced graph, R_n , whose vertices are the equivalence classes $\{\rho(X) \mid X \in V(N_n)\}$ with edges $\rho(X)\rho(Y) \in E(R_n)$ if $XY \in E(N_n)$ or $X\overline{Y} \in E(N_n)$. Observe that if $Z = \nu(0^k1^{k+1})$ then $\overline{Z} = \nu(0^k1^{k+1}) = \nu(1^k0^{k+1})$, and so $Z\overline{Z} \in E(N_n)$. Hence

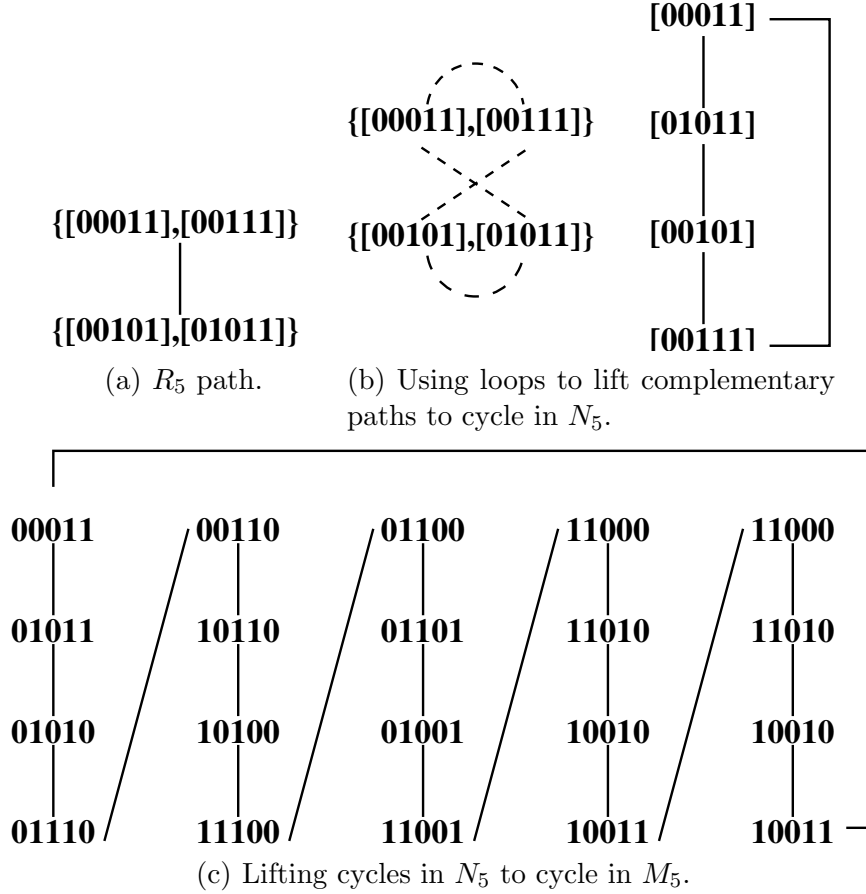


Fig. 2. Lifting a path in R_5 to a cycle in N_5 and then a cycle in M_5 .

$\rho(Z)\rho(\bar{Z}) \in E(R_n)$ and R_n has loops, so it is not a simple graph.

We exploit the fact that R_n has loops to show that a Hamilton path from the distinguished vertex $r_1 = \rho(\nu(0^k 1^{k+1}))$ to the distinguished vertex $r_l = \rho(\nu(0(01)^k))$ in R_n can be used to construct a cycle in N_n which can be lifted to a cycle in M_n . The path in R_n gives rise naturally to a pair of paths in N_n where corresponding pairs of vertices from each path are complements of each other. Because the distinguished vertices in R_n each have incident loops, these paths link to form a cycle in N_n . Finally, since k and $2k + 1$ are relatively prime, we can use suitably chosen necklace representatives to extend the cycle in N_n to a cycle in M_n . The case for M_5 is illustrated in Figure 2, where for $x \in V(M_5)$, $[x]$ denotes the necklace $\nu(x) \in V(N_5)$ and for necklace $X \in V(N_5)$, $\{X, \bar{X}\}$ denotes $\rho(X)$ in $V(R_5)$.

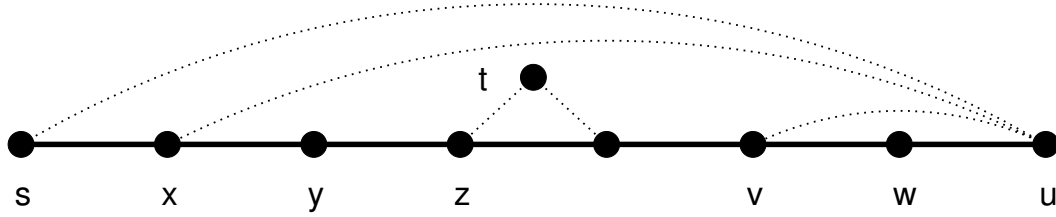


Fig. 3. Backtrack search from s has reached u and cannot continue.

3 The Hamilton cycle heuristic

Given a graph G and vertices $s, t \in V(G)$ we would like to find, if possible, a Hamilton path starting at s and ending at t .

A standard backtrack search attempts to construct such a path by starting at s and extending the path to a new vertex as long as possible. Whenever it is impossible to further extend the path from a vertex x , the search “backs up” to the predecessor, y , of x on the path and the path is extended (if possible) from y to one of its other neighbors. Figure 3 illustrates a path P from a starting vertex s to a vertex u whose only neighbors in G are s, x and v , which are already on P . In this case, backtrack search would back up to w , then v and eventually back to z , at which time t would be added to the path.

Pósa [13] observed that reaching a dead end in backtrack search could be used as an opportunity to modify the current path by a *rotation* and thus possibly to continue. If $P = (u_1, u_2, \dots, u_k)$ is a path in G and there is an edge $u_k u_j$ for some $0 < j < k$, then the *rotation of p at u_j* is the path $P' = (u_1, u_2, \dots, u_j, u_k, u_{k-1}, \dots, u_{j+1})$ obtained by deleting edge $u_j u_{j+1}$ from P and inserting $u_j u_k$. Path P' has the same length as P , but a different endpoint. Figure 4(a) shows the rotation at x of the path in Figure 3.

Define *PosaSearch* to be the Hamilton path heuristic which constructs a path P by starting at a vertex s and iterating the following: extend P , avoiding t until the end, until no longer possible; When P cannot be extended from an endpoint u , select a neighbor v of u and perform a rotation of P at v . *PosaSearch* may not succeed in finding a Hamilton path even if one exists. Furthermore, it may run indefinitely.

For the path P in Figure 3, *PosaSearch* can only transform P into one of the paths shown in Figure 4. None of these transformations will ever allow the vertex, t , to be added to the path.

Rotations transform a path and alter the order of vertices on it. For a path P , let $P(i)$ be the i th vertex of P , let $pos(P, v)$ be the position of vertex v on P , and let $|P|$ be the number of vertices on P . If P is a path with endpoint u and if $P(j)$ is adjacent to u , we can rotate P at $P(j)$ to arrive at a new path

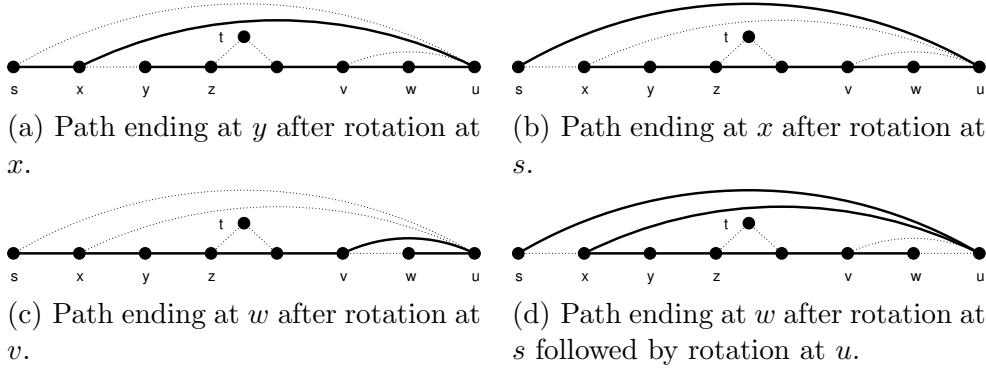


Fig. 4. Possible endpoints after one or more rotations

P' . The position of a vertex v on P' is then given by

$$pos(P', v) = \begin{cases} pos(P, v) & \text{if } pos(P, v) \leq j \\ |P| - pos(P, v) + j + 1 & \text{otherwise.} \end{cases} \quad (1)$$

Similarly, the vertex now in position i on P' is

$$P'(i) = \begin{cases} P(i) & \text{if } i \leq j \\ P(|P| - i + j + 1) & \text{otherwise.} \end{cases} \quad (2)$$

The SS heuristic from [1] uses a variation of PosaSearch that looks ahead before performing any rotation. SS extends the path as is done in PosaSearch until all neighbors of the endpoint are already on the path. If the path is not already a Hamilton path, it then uses breadth-first search (BFS), and repeated application of eqs. (1) and (2) to search for a sequence of one or more rotations guaranteed to result in a path which can be further extended. If such a sequence is found, the sequence of rotations is performed and the path is extended. If no such sequence exists, the SS heuristic terminates without having found a Hamilton path. Figure 4 shows four paths obtainable by one or more rotations from the path shown in Figure 3. In this example, there are only four possible endpoints, u , v , w , and x . The path cannot be extended from any of these, so SS would terminate after evaluating all four.

4 The new SSS heuristic

The original work on the SS heuristic [1] used a 400MHz Intel Pentium-II system with 192MB RAM. Results were later run on a 2.4GHz Intel Pentium 4 system with 512MB of RAM (See [14]). We recently acquired an AMD Athlon 3500+ system with 2GB of RAM and converted the program to use 64-bit

values for most internal computations. With this system and the SS heuristic we were able to find a Hamilton path in R_{33} in about 3.5 weeks. The program ran entirely in memory, using approximately 1GB of RAM. However, it was unlikely that R_{35} would be feasible.

Therefore, using performance profiling tools, we analyzed the code performance and discovered that, contrary to expectations, a large part of the time was spent on performing the rotation operations, rather than on the BFS to find promising sequences of rotations. As a result, we made the following changes to the heuristic that resulted in dramatic speed improvements.

First, when BFS finds a sequence of rotations that will enable an extension of the path, instead of actually performing each rotation of the path the SSS heuristic encodes the sequence of rotations as a list of ordered pairs, each representing the number of vertices on the path at time of rotation and the rotation point. This list, together with repeated application of eqs. (1) and (2), suffices to calculate the actual vertex in a given position or the current position of a given vertex, when the sequence of rotations represented by the list of ordered pairs has not been explicitly performed.

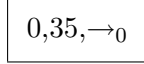
Secondly, the SSS heuristic periodically (but rarely) goes ahead and performs all the rotations encoded by the list of ordered pairs. Given a path v_0, \dots, v_l , if we perform a rotation at v_i to make v_{i+1} the new endpoint, then there is a section of $i + 1$ vertices up to v_i that will not be moved and a section of $l - i$ vertices after v_i that will be reversed. We may describe each of these sections (or *blocks*) of the path by a triple, $\{s, n, d\}$, where s is the index of the first vertex in the block, n is the number of vertices in the block and d is a direction indicating whether the block is now in original (forward) order or reversed. Our single rotation at v_i is represented by the triples $\{0, i + 1, \rightarrow\}$ and $\{i + 1, l - i, \leftarrow\}$. As we process a rotation from the list of ordered pairs we first add a block in the forward direction, representing any vertices added by extension since the last rotation. Next we determine the block containing the new endpoint and (possibly) split it into two blocks. All blocks before the new endpoint remain unchanged, while the sequence of blocks after the new endpoint, as well as the direction indicator of each such block, is reversed. After the list of blocks has been created representing all the stored rotations, the path can be copied block by block to effect the series of rotations in a single copy operation.

These modifications allow a sequence of rotations to be accumulated, without being performed, at the relatively smaller cost of increasing time for calculation of vertex position or vertex in a given position. The accumulated sequence of rotations is performed on a schedule discussed in the next subsection. This significantly reduced the average work per rotation.

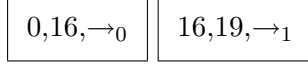
Saved list of three rotations: $\{35, 15\}$, $\{36, 9\}$, $\{40, 32\}$

1. $\{35, 15\}$ - 35 vertices on path, rotation at v_{15}

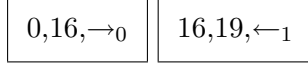
Create initial block 0



Split block 0, into blocks 0 and 1

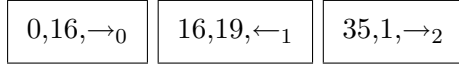


Rotate block 1

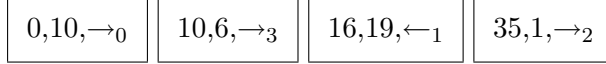


2. $\{36, 9\}$ - 36 vertices on path, rotation at v_9

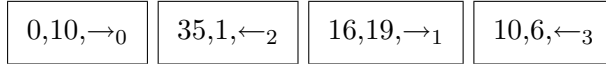
Add block 2



Split block 0, into blocks 0 and 3

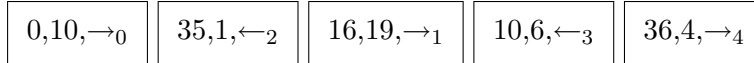


Rotate blocks 3 through 2

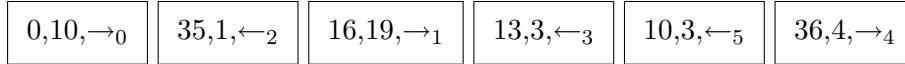


3. $\{40, 32\}$ - 40 vertices on path, rotation at v_{32}

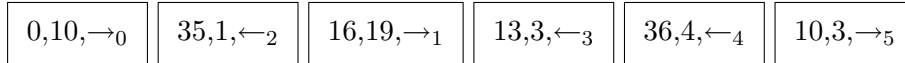
Add block 4



Split block 3, into blocks 3 and 5



Rotate blocks 5 and 4



Final path:

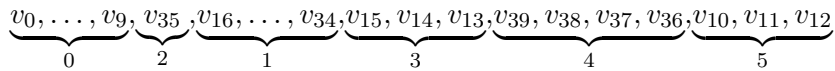


Fig. 5. Performing a sequence of rotations from a list.

Figure 5 illustrates the process for a list of three rotations, as seen in one run of the program. The subscripts in each block indicate the order of creation of the block. For computational convenience, a rotation point, v , is actually represented in the list of rotations by its position on the path.

4.1 Graph representation and storage

In the middle levels of \mathcal{B}_{35} , the reduced graph, R_{35} , has over 129 million vertices. An array of 32-bit integers with one entry per vertex takes approximately half a gigabyte of memory. Our system had only 2GB of RAM, so storage for basic information on vertices and paths was limited. In this environment, it is infeasible to store adjacency lists, so we recalculate adjacency lists as needed rather than storing them (see [14]). We reduced all arrays to the smallest native size (char, short, int, or long) that was sufficient for the data and still experienced unacceptable levels of memory swapping.

The vertices of R_{35} are represented by the lexicographically least elements (as sets) of each 35-bit necklace with 17 ones. Thus the two higher order bits will always be 0 and we need only use 33 bits for the internal representation. However, we store the lower 32 bits in 32-bit integer arrays and add the high-order bit according to the position of a vertex in the array. This saves half a gigabyte of RAM as compared to using 64-bit long integers. We still use 64-bit long integers for internal calculations, but none of our major storage arrays needs larger than 32-bit entries.

When copying the path to execute a series of rotations, we copy it to the position array (Pos) and then rebuild the position array after the copy as this results in less swapping than using memory that has not recently been referenced, such as the parent array used in building a BFS tree.

Fine tuning in this way resulted in the program using approximately 85% to 90% of the CPU for most of the run while finding a path in R_{35} . Somewhat higher CPU usage was observed toward the end of the run, even though BFS trees, with their storage requirement, were being computed much more frequently.

4.2 Performance of the heuristic on the middle levels graph

The SSS heuristic was applied to search the reduced middle levels graphs R_{2k+1} , $k = 1, 2, \dots, 17$, to try to find a Hamilton path from vertex $\rho(\nu(0^{k+1}1^k))$ to vertex $\rho(\nu(0(01)^k))$, and therefore a Hamilton cycle in M_{2k+1} . A Hamilton path meeting these requirements was found for each $k \leq 17$.

The results are summarized in Table 1. We measured elapsed time using a timer with a resolution of 1 second. Runs that start and complete in the same second show a time of 0 seconds. Table 1 also shows earlier results obtained with the SS heuristic on a 2.4GHz Intel Pentium 4 system with 512MB of RAM. Many factors affect the difference in performance between

the two systems, including operating system and compiler differences as well as processor, RAM, and disk speed. The newer system ran the SS heuristic approximately twice as fast as the older system. Thus for $k = 15$, the largest value for which we have both results, we see that hardware doubled the speed and the switch to the SSS algorithm achieved a further 16-fold increase.

Table 1
Running time to find a Hamilton cycle in the middle levels graph

k	$n = 2k + 1$	# vertices in R_n	# vertices in M_n	Time (Secs)	
				SS (32-bit)	SSS (64-bit)
8	17	1,430	48,620	0	0
9	19	4,862	184,756	0	0
10	21	16,796	705,432	1	0
11	23	58,786	2704,156	5	2
12	25	208,012	10,400,600	105	10
13	27	742,900	40,116,600	1,732	99
14	29	2,674,440	155,117,520	24,138	799
15	31	9,694,845	601,080,390	307,976 (3.6 days)	9,446
16	33	35,357,670	2,333,606,220	-	106,118 (1.2 days)
17	35	129,644,790	9,075,135,300	-	1,765,497 (20.4 days)

As noted earlier, the deferral of rotations also involves a cost. We experimented with the number of deferred rotations using values ranging from $\log V(G)$ to $\sqrt{V(G)}$. Our results were obtained using a value of $\sqrt{V(G)}$, or 11,387 for the reduced graph, R_{35} . There is some evidence to suggest that a small performance improvement could be obtained by varying this number during the running of the program, particularly in very large graphs.

References

- [1] I. Shields, C. D. Savage, A Hamilton path heuristic with applications to the middle two levels problem, in: Proceedings of the Thirtieth Southeastern

International Conference on Combinatorics, Graph Theory, and Computing (Boca Raton, FL, 1999), Vol. 140, 1999, pp. 161–178.

- [2] L. Lovász, Problem 11, in: *Combinatorial Structures and their Applications*, Gordon and Breach, 1970.
- [3] I. J. Dejter, Hamilton cycles and quotients of bipartite graphs, in: *Graph theory with applications to algorithms and computer science* (Kalamazoo, Mich., 1984), Wiley, New York, 1985, pp. 189–199.
- [4] D. Duffus, B. Sands, R. Woodrow, Lexicographic matchings cannot form Hamiltonian cycles, *Order* 5 (2) (1988) 149–161.
- [5] G. Hurlbert, The antipodal layers problem, *Discrete Math.* 128 (1-3) (1994) 237–245.
- [6] H. A. Kierstead, W. T. Trotter, Explicit matchings in the middle levels of the Boolean lattice, *Order* 5 (2) (1988) 163–171.
- [7] C. D. Savage, Long cycles in the middle two levels of the Boolean lattice, *Ars Combin.* 35 (A) (1993) 97–108.
- [8] S. Felsner, W. T. Trotter, Colorings of diagrams of interval orders and α -sequences of sets, *Discrete Math.* 144 (1-3) (1995) 23–31, *combinatorics of ordered sets* (Oberwolfach, 1991).
- [9] L. Babai, Long cycles in vertex-transitive graphs, *J. Graph Theory* 3 (3) (1979) 301–304.
- [10] C. D. Savage, P. Winkler, Monotone Gray codes and the middle levels problem, *J. Combin. Theory Ser. A* 70 (2) (1995) 230–248.
- [11] J. R. Johnson, Long cycles in the middle two layers of the discrete cube, *J. Combin. Theory Ser. A* 105 (2) (2004) 255–271.
- [12] P. Horák, T. Kaiser, M. Rosenfeld, Z. Ryjáček, The prism over the middle-levels graph is Hamiltonian, *Order* 22 (1) (2005) 73–81.
- [13] L. Pósa, Hamiltonian circuits in random graphs, *Discrete Math.* 14 (4) (1976) 359–364.
- [14] I. B. Shields, Hamilton cycle heuristics in hard graphs, Ph.D. thesis, North Carolina State University, Raleigh, www.lib.ncsu.edu/theses/available/etd-03142004-013420/ (2004).