On the Central Levels Problem

Petr Gregor

Department of Theoretical Computer Science and Mathematical Logic, Charles University, Prague, Czech Republic gregor@ktiml.mff.cuni.cz

Ondřej Mička

Department of Theoretical Computer Science and Mathematical Logic, Charles University, Prague, Czech Republic micka@ktiml.mff.cuni.cz

Torsten Mütze

Department of Computer Science, University of Warwick, Coventry, UK Department of Theoretical Computer Science and Mathematical Logic, Charles University, Prague, Czech Republic torsten.mutze@warwick.ac.uk

- Abstract

The central levels problem asserts that the subgraph of the (2m+1)-dimensional hypercube induced by all bitstrings with at least $m+1-\ell$ many 1s and at most $m+\ell$ many 1s, i.e., the vertices in the middle 2ℓ levels, has a Hamilton cycle for any $m\geq 1$ and $1\leq \ell\leq m+1$. This problem was raised independently by Savage, by Gregor and Škrekovski, and by Shen and Williams, and it is a common generalization of the well-known middle levels problem, namely the case $\ell=1$, and classical binary Gray codes, namely the case $\ell=m+1$. In this paper we present a general constructive solution of the central levels problem. Our results also imply the existence of optimal cycles through any sequence of ℓ consecutive levels in the n-dimensional hypercube for any $n\geq 1$ and $1\leq \ell\leq n+1$. Moreover, extending an earlier construction by Streib and Trotter, we construct a Hamilton cycle through the n-dimensional hypercube, $n\geq 2$, that contains the symmetric chain decomposition constructed by Greene and Kleitman in the 1970s, and we provide a loopless algorithm for computing the corresponding Gray code.

2012 ACM Subject Classification Mathematics of computing \rightarrow Combinatorial algorithms; Mathematics of computing \rightarrow Matchings and factors

Keywords and phrases Gray code, Hamilton cycle, hypercube, middle levels, symmetric chain decomposition

Digital Object Identifier 10.4230/LIPIcs.ICALP.2020.60

Category Track A: Algorithms, Complexity and Games

Related Version A full version of the paper is available at https://arxiv.org/abs/1912.01566.

Funding This work was supported by the Czech Science Foundation grant GA19-08554S. *Torsten Mütze*: was also supported by DFG grant 413902284.

Acknowledgements We thank Jiří Fink for several valuable discussions about symmetric chain decompositions, and for feedback on an earlier draft of this paper.

1 Introduction

The *n*-dimensional hypercube, or *n*-cube for short, is the graph Q_n formed by all $\{0,1\}$ -strings of length n, with an edge between any two bitstrings that differ in exactly one bit. This family of graphs has numerous applications in computer science and discrete mathematics, many of which are tied to famous problems and conjectures, such as the sensitivity conjecture of Nisan and Szegedy [29], recently proved by Huang [23]; Erdős and Guys' crossing number

problem [8] (see [9]); Füredi's conjecture [13] on equal-size chain partitions (see [41]); Shearer and Kleitman's conjecture [37] on orthogonal symmetric chain decompositions (see [39]); the Ruskey-Savage problem [32] on matching extendability (see [11, 12]), and the conjectures of Norine, and Feder and Subi on edge-antipodal colorings [10, 30], to name just a few.

The focus of this paper are Hamilton cycles in the *n*-cube and its subgraphs. A Hamilton cycle in a graph is a cycle that visits every vertex exactly once, and in the context of the *n*-cube, such a cycle is often referred to as a *Gray code*. Gray codes have found applications in signal processing, circuit testing, hashing, data compression, experimental design, and in solving puzzles like the Towers of Hanoi or the Chinese rings; see Savage's survey [35]. They are also fundamental for efficient exhaustive generation algorithms, a topic that is covered in depth in the most recent volume of Knuth's *'The Art of Computer Programming'* [25].

1.1 Our results

In this paper we consider the central levels problem, a broad generalization of the middle levels problem: Does the subgraph of the (2m+1)-cube induced by the middle 2ℓ levels, i.e., by levels $m+1-\ell,\ldots,m+\ell$, have a Hamilton cycle for any $m\geq 1$ and $1\leq \ell\leq m+1$? This problem was raised independently by Savage [34], Gregor and Škrekovski [20], and by Shen and Williams [38]. Clearly, the case $\ell=1$ of the central levels problem is the aforementioned middle levels problem (solved in [27]). Moreover, the case $\ell=2$ was solved affirmatively in a paper by Gregor, Jäger, Mütze, Sawada, and Wille [16] presented at ICALP 2018. Also, the case $\ell=m+1$ is established by the binary reflected Gray code Γ_{2m+1} . Furthermore, the case $\ell=m$ was solved by El-Hashash and Hassan [7], and in a more general setting by Locke and Stong [26], and the case $\ell=m-1$ was settled in [20].

The main contribution of this paper is to solve the central levels problem affirmatively in full generality; see Figure 1 (a)–(d).

▶ **Theorem 1.** For any $m \ge 1$ and $1 \le \ell \le m+1$, the subgraph of the (2m+1)-cube induced by the middle 2ℓ levels has a Hamilton cycle.

The most general question in this context is to ask for a Hamilton cycle in Q_n that visits all vertices in any sequence of ℓ consecutive levels, i.e., the levels need not be symmetric around the middle, and the dimension n needs not be odd. These graphs are all bipartite, and to circumvent the imbalances that prevent the existence of a Hamilton cycle for general n and ℓ , we have to slightly generalize the notion of Hamilton cycles: Firstly, a saturating cycle in a bipartite graph is a cycle that visits all vertices in the smaller partition class (if it has

size 1, then a single edge is considered to be a cycle). Secondly, a *tight enumeration* in a (bipartite) subgraph of the cube is a cyclic listing of all its vertices where the total number of bits flipped is exactly the number of vertices plus the difference in size between the two partition classes. Clearly, if both partition classes have the same size, these two notions are equal to a Hamilton cycle. In fact, all cases of this more general problem, except the central levels problem, were solved already in [18], some of them conditional on a "yes" answer to the central levels problem. Combining Theorem 1 with these previous results, we now also obtain an unconditional result for this more general question.

▶ Corollary 2. For any $n \ge 1$ and $1 \le \ell \le n+1$, the subgraph of the n-cube induced by any sequence of ℓ consecutive levels has both a saturating cycle and a tight enumeration.

An essential tool in our proof of Theorem 1 are symmetric chain decompositions. This is a well-known concept from the theory of posets, which we now define specifically for the n-cube using graph-theoretic language. A symmetric chain in Q_n is a path $(x_k, x_{k+1}, \ldots, x_{n-k})$ in the n-cube where x_i is from level i for all $k \leq i \leq n-k$, and a symmetric chain decomposition, or SCD for short, is a partition of the vertices of Q_n into symmetric chains. It is well-known that the n-cube has an SCD for all $n \geq 1$, and the simplest explicit construction was given by Greene and Kleitman [15] (see Section 2.2 below). Streib and Trotter [40] first investigated the interplay between SCDs and Hamilton cycles in the n-cube, and they described an SCD in Q_n that can be extended to a Hamilton cycle; see Figure 1 (e). Their SCD, however, is different from the aforementioned Greene-Kleitman SCD. In this paper, we extend Streib and Trotter's result as follows; see Figure 1 (f).

▶ **Theorem 3.** For any $n \ge 2$, the Greene-Kleitman SCD can be extended to a Hamilton cycle in Q_n .

The Greene-Kleitman SCD has found a large number of applications in the literature, e.g., to construct symmetric Venn diagrams [21, 33], to solve the Littlewood-Offord problem [3, Chap. 4], or to learn monotone Boolean functions [25, Sec. 7.2.1.6] (see also [1, 6, 31, 37, 42]). Knowing that this SCD extends to a Hamilton cycle and that it is a crucial ingredient for solving the general central levels problem adds to this list of interesting properties and applications. Observe also that a Hamilton cycle that extends an SCD has the intriguing property that it minimizes the number of changes of direction from moving up to moving down, or vice versa, between consecutive levels in the cube. For comparison, the monotone paths constructed by Savage and Winkler [36] maximize these changes.

Motivated by these results and by the aforementioned conjecture of Ruskey and Savage [32] that every matching in Q_n extends to a Hamilton cycle, we raise the following conjecture:

▶ Conjecture 4. Every SCD can be extended to a Hamilton cycle in Q_n .

Although every SCD of Q_n is the union of two matchings, there are matchings in Q_n that do not extend to an SCD; take for example the two edges obtained by starting at the vertices 0^n and 1^n and flipping the same bit. Consequently, an affirmative answer to Conjecture 4 would cover only some cases of the Ruskey-Savage conjecture.

1.2 Efficient algorithms

Our proof of Theorem 1 is constructive and translates directly into an algorithm for computing the Hamilton cycle in time and space that are polynomial in the size of the graph (the middle 2ℓ levels of Q_n , n := 2m + 1), which is exponential in n. Often, it is desirable to have

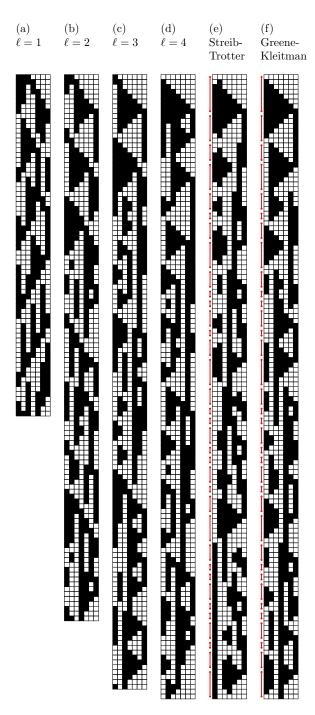


Figure 1 (a)−(d) The Hamilton cycles in $Q_{7,\ell}$ for $\ell=1,2,3,4$ constructed as in our proof of Theorem 1. (e) The Hamilton cycle in Q_7 containing an SCD obtained from the Streib-Trotter construction, with symmetric chains highlighted on the side. (f) The Hamilton cycle in Q_7 containing the Greene-Kleitman SCD obtained from our proof of Theorem 3. In this figure, 1-bits are drawn as black squares, 0-bits as white squares.

a "local" algorithm that uses only time and space that are polynomial in n. Ideally, one might hope for $\mathcal{O}(n)$ space to store the current bitstring and some additional data structures, and $\mathcal{O}(1)$ time to compute the next bitstring on the cycle. Such algorithms are known for the binary reflected Gray code Γ_n [2], and for the middle levels problem [28], i.e., for the extreme cases $\ell = m+1$ and $\ell = 1$ of the central levels problem. There are fundamental obstacles that prevent us to obtain such a local algorithm from our proof, and it remains a challenging open problem to find such an algorithm. Our Theorem 3 on the other hand, can be translated into a simple algorithm that uses only $\mathcal{O}(n)$ space and $\mathcal{O}(1)$ time in every iteration to compute the next bitstring along the Hamilton cycle. A pseudocode description of this algorithm is available in [17]. We also implemented it in C++, available for download and for demonstration on the Combinatorial Object Server [5].

1.3 Proof ideas

We first describe the ideas for proving Theorem 1. For any $m \geq 1$ we define n := 2m+1, and for $1 \leq \ell \leq m+1$ we let $Q_{n,\ell}$ denote the subgraph of Q_n induced by the middle 2ℓ levels. To prove that $Q_{n,\ell}$ has a Hamilton cycle for general m and ℓ , we combine and generalize the tools and techniques developed for the cases $\ell = 1$ and $\ell = 2$ in [19] and [16], respectively. Our proof proceeds in two steps: In a first step, we construct a cycle factor in $Q_{n,\ell}$, i.e., a collection of disjoint cycles which together visit all vertices of $Q_{n,\ell}$. In a second step, we use local modifications to join the cycles in the factor to a single Hamilton cycle. Essentially, this technique reduces the Hamiltonicity problem in $Q_{n,\ell}$ to proving that a suitably defined auxiliary graph is connected, which is much easier.

In fact, the predecessor paper [16] already proved the existence of a cycle factor in $Q_{n,\ell}$, but this construction does not seem to yield a factor that would be amenable to analysis. In this paper, we therefore construct another cycle factor in $Q_{n,\ell}$, based on modifying the aforementioned Greene-Kleitman SCD of Q_n by the lexical matchings introduced by Kierstead and Trotter [24]. The resulting cycle factor in $Q_{n,\ell}$ has a rich structure, in particular the number of cycles and their lengths can be described combinatorially.

The simplest way to join two cycles C and C' from this factor to a single cycle is to consider a 4-cycle F that shares exactly one edge with each of the cycles C and C' (the other two edges of F must then go between C and C'), and to take the symmetric difference of the edge sets of $C \cup C'$ and of F, yielding a single cycle $(C \cup C') \triangle F$ on the same vertex set as $C \cup C'$. We refer to such a cycle F as a flipping 4-cycle. For example, if we interpret the binary reflected Gray code Γ_n as a cycle in Q_n , we see that $\Gamma_{n+1} = (0\Gamma_n \cup 1\Gamma_n^R) \triangle F$ where F is the 4-cycle $F = 0^{n+1}, 010^{n-1}, 110^{n-1}, 10^n$. In addition to flipping 4-cycles, we also use flipping 6-cycles, which intersect with the two cycles to be joined in a slightly more complicated way, albeit with the same effect of joining them to a single cycle. The most technical aspect of this part of the proof is to ensure that all flipping cycles used are edge-disjoint, so that the joining operations do not interfere with each other.

To prove Theorem 3, we proceed by induction from dimension n to n+2, treating the cases of even and odd n separately. We first specify a particular ordering of all chains of the Greene-Kleitman SCD, and then show that this ordering admits a matching that alternatingly joins the bottom or top vertices of any two consecutive chains in our ordering. In fact, there is a close relation between our proofs of Theorem 1 and 3: The aforementioned construction of a cycle factor in $Q_{n,\ell}$ is particularly nice for $\ell=m+1$, i.e., for the case where we consider the entire cube. Specifically, in this case our cycle factor contains all chains from the Greene-Kleitman SCD. These cycles can be joined to a single Hamilton cycle in such a way, so as to give exactly the aforementioned Hamilton cycle constructed for proving Theorem 3.

1.4 Outline of this paper

In Section 2 we discuss the Greene-Kleitman SCD and lexical matchings, and collect some other preliminaries. In Section 3 we describe our construction of a cycle factor in $Q_{n,\ell}$. Due to space constraints, in this extended abstract we are unable to provide the full details of the analysis of this cycle factor, and how to join its cycles to a Hamilton cycle. We rather give an informal high-level sketch of these steps in Section 4. In Section 5 we present our proof of Theorem 3. The omitted proof details, together with the pseudocode description of the corresponding loopless algorithm can be found in [17].

2 Preliminaries

For the reader's convenience, important notations that are introduced in the following and used repeatedly in the paper are summarized in Table 1 at the end of this paper.

2.1 Bitstrings and lattice paths

For any string x and any integer $k \ge 0$, we let x^k denote the concatenation of k copies of x. We often interpret a bitstring x as a path in the integer lattice \mathbb{Z}^2 starting at the origin (0,0), where every 0-bit is interpreted as a \setminus -step that changes the current coordinate by (+1,-1) and every 1-bit is interpreted as an \setminus -step that changes the current coordinate by (+1,+1); see Figure 2.

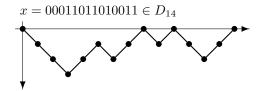


Figure 2 The correspondence between bitstrings (top) and lattice paths (bottom).

Let D_{2k} denote the set of bitstrings with exactly k many 1s and k many 0s, such that in every prefix, the number of 0s is at least as large as the number of 1s. We also define $D := \bigcup_{k \geq 0} D_{2k}$. Note that $D_0 = \{\varepsilon\}$, where ε denotes the empty bitstring. In terms of lattice paths, D corresponds to so-called Dyck paths that never move above the line y = 0 and end on this line. If a lattice path x contains a substring $x \in D$, then we refer to this substring $x \in D$ as a valley in x.

2.2 The Greene-Kleitman SCD

We now describe Greene and Kleitman's [15] construction of an SCD in the n-cube; see Figure 3. For any vertex x of the n-cube, we interpret the 0s in x as opening brackets and the 1s as closing brackets. By matching closest pairs of opening and closing brackets in the natural way, the chain containing x is obtained by flipping the leftmost unmatched 0 to ascend the chain, or the rightmost unmatched 1 to descend the chain, until no more unmatched bits can be flipped. It is easy to see that this indeed yields an SCD of the n-cube for any $n \ge 1$. We always work with this SCD due to Greene and Kleitman, and whenever referring to a chain, we mean a chain from this decomposition.

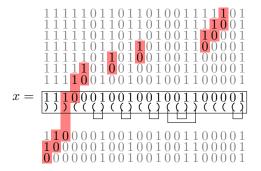


Figure 3 Construction of the Greene-Kleitman SCD containing a bitstring x via parenthesis matching. The highlighted bits are the leftmost unmatched 0 and the rightmost unmatched 1 in each bitstring.

Each chain C of length h in Q_n can be encoded compactly as a string of length n over the alphabet $\{0,1,*\}$ in the form

$$C = u_0 * u_1 * \dots * u_{h-1} * u_h, \tag{1}$$

where $u_0, \ldots, u_h \in D$. The symbols * represent unmatched positions, and the vertices along the chain are obtained by replacing the *s by 1s followed by 0s in all possible ways; see (2). For example, the chain shown in Figure 3 is C = *****01*01*010011***01, so we have $u_0 = u_1 = u_2 = u_3 = u_4 = u_8 = u_9 = \varepsilon$, $u_5 = u_6 = u_{10} = 01$, and $u_7 = 010011$.

We distinguish four types of chains depending on whether u_0 and u_h , i.e., the first and last valleys in (1), are empty or not. These chain types are denoted by [--], [+-], [++], and [-+], where the first symbol is - if $u_0 = \varepsilon$ and + otherwise, and the second symbol is - if $u_h = \varepsilon$ and + otherwise. For example, the chain in Figure 3 is a [-+]-chain. We also use the symbol? in these type specifications if we do not know whether a valley is empty or not. Note that there is no [--]-chain in Q_n of length h = 1 unless n = 1.

Given a chain C of length h as in (1), the ith vertex of C from the bottom is

$$x = u_0 1 \cdots u_{i-1} 1 u_i 0 u_{i+1} \cdots 0 u_h$$
 (2)

where $i=0,\ldots,h$, and this vertex belongs to level $k=\frac{n-h}{2}+i$. Note that every vertex x of Q_n can be written uniquely in the form (2), and we refer to this as the *chain factorization* of x. For the following arguments, it will be crucial to consider the lattice path representation of x, with the valleys u_0,\ldots,u_h that are separated by i many \nearrow -steps, followed by h-i many \searrow -steps, i.e., the valley u_i is the highest one on the lattice path.

We use $C_{h,i}$, $0 \le i \le h$, to denote the set of the *i*th vertices in all chains of length h. Moreover, we partition $C_{h,i}$ into two sets $C_{h,i}^-$ and $C_{h,i}^+$, depending on whether the valley u_i in (2) is empty or nonempty, respectively. Clearly, $C_{h,h}^+$ are exactly the top vertices of [?+]-chains of length h and $C_{h,0}^+$ are exactly the bottom vertices of [+?]-chains of length h, and similarly with - instead of +. Note that the sets $C_{h,i}$ are empty if n is odd and h is even, or vice versa.

2.3 Lexical matchings

Lexical matchings in Q_n were introduced by Kierstead and Trotter [24], and they are parametrized by some integer $p \in \{0, 1, ..., n-1\}$. These matchings are defined as follows; see Figure 4. We interpret a bitstring x as a lattice path, and we let x^{\setminus} denote the lattice

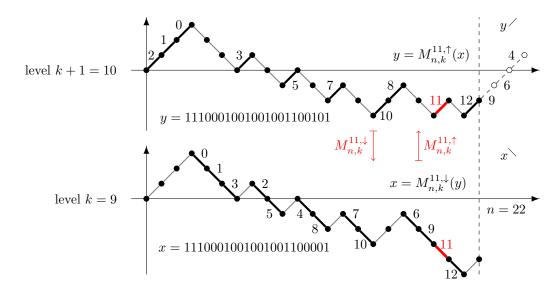


Figure 4 Definition of p-lexical matchings between levels 9 and 10 of Q_{22} , where steps flipped along the p-lexical edge are marked with p. Between those two levels, the vertex x is incident with p-lexical edges for each $p \in \{0, 1, \ldots, 12\}$, and the vertex y is incident with p-lexical edges for each $p \in \{0, 1, \ldots, 12\} \setminus \{4, 6, 9\}$.

path that is obtained by appending \setminus -steps to x until the resulting path ends at height -1. If x ends at a height less than -1, then x := x. Similarly, we let x' denote the lattice path obtained by appending \angle -steps to x until the resulting path ends at height +1. If x ends at a height more than +1, then $x^{\prime} := x$. We let $L_{n,k}$ denote the set of all vertices on level k of Q_n , and we define a matching by two partial mappings $M_{n,k}^{p,\uparrow}:L_{n,k}\to L_{n,k+1}$ and $M_{n,k}^{p,\downarrow}: L_{n,k+1} \to L_{n,k}$ defined as follows: For any $x \in L_{n,k}$ we consider the lattice path xand scan it row-wise from top to bottom, and from right to left in each row. The partial mapping $M_{n,k}^{p,\uparrow}(x)$ is obtained by flipping the pth \stackset step encountered in this fashion, where counting starts with $0, 1, \ldots$, if this \setminus -step is part of the subpath x of x^{\setminus} ; otherwise x is left unmatched. Similarly, for any $x \in L_{n,k+1}$ we consider the lattice path x^{\wedge} and scan it row-wise from top to bottom, and from left to right in each row. The partial mapping $M_{n,k}^{p,\downarrow}(x)$ is obtained by flipping the pth /-step encountered in this fashion if this /-step is part of the subpath x of x^{\prime} ; otherwise x is left unmatched. It is straightforward to verify that these two partial mappings are inverse to each other, so they indeed define a matching between levels k and k+1 of Q_n , called the *p-lexical matching*, which we denote by $M_{n,k}^p$. We also define $M_n^p := \bigcup_{0 \le k \le n} M_{n,k}^p$, where we omit the index n whenever it is clear from the context. In the following, we will only ever use p-lexical edges for p = 0, 1, 2. For instance, it is

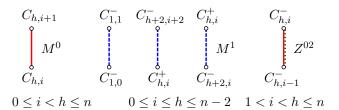


Figure 5 Perfect matchings described by Lemma 5. The $\{0, 1, 2\}$ -lexical edges are drawn solid, dashed, and dotted, respectively.

well-known that taking the union of all 0-lexical edges, i.e., the set M^0 , yields exactly the Greene-Kleitman SCD [24]. This property is captured by the following lemma, together with several other explicit perfect matchings, consisting of $\{0,1,2\}$ -lexical edges between certain sets of vertices from our SCD; see Figure 5.

To state the lemma, for a set M of edges of Q_n and disjoint sets X, Y of vertices, we let M[X,Y] denote the set of all edges of M between X and Y. Moreover, for any vertex $x \in C_{b,i}^-$, $1 < i < h \le n$, we consider the chain factorization $x = u_0 \, 1 \, \cdots \, u_{i-2} \, 1 \, u_{i-1} \, 1 \, 0 \, u_{i+1} \, \cdots \, 0 \, u_h$ with $u_0, \ldots, u_h \in D$, and we define a neighbor z(x) on the level below by

$$z(x) := \begin{cases} u_0 \, 1 \cdots u_{i-2} \, 1 \, 0 \, 0 \, u_{i+1} \cdots 0 \, u_h & \text{if } u_{i-1} = \varepsilon, \\ u_0 \, 1 \cdots u_{i-2} \, 1 \, 0 \, v \, 0 \, w \, 1 \, 0 \, u_{i+1} \cdots 0 \, u_h & \text{if } u_{i-1} = 0 \, v \, 1 \, w \text{ with } v, w \in D. \end{cases}$$
(3)

Note that (x, z(x)) is a 0-lexical or 2-lexical edge in the first or second case, respectively.

- ▶ **Lemma 5.** For every $n \ge 3$, the following sets of edges M[X,Y] are perfect matchings in Q_n between the vertex sets X and Y.
 - (i) $M^0[C_{h,i}, C_{h,i+1}]$ for every $0 \le i < h \le n$;
- $\begin{array}{ll} \text{(ii)} \ \ M^1[C_{1,0}^-,C_{1,1}^-], \ M^1[C_{h,i}^+,C_{h+2,i+2}^-], \ and \ M^1[C_{h,i}^+,C_{h+2,i}^-] \ for \ every \ 0 \leq i \leq h \leq n-2; \\ \text{(iii)} \ \ Z^{02}[C_{h,i-1}^-,C_{h,i}^-] \ for \ every \ 1 < i < h \leq n, \ where \ Z^{02} := \{(x,z(x)) \ | \ x \in C_{h,i}^-\}. \end{array}$

The proof of Lemma 5 can be found in [17].

3 Cycle factor construction

We now construct a cycle factor $C_{n,\ell}$ in the graph $Q_{n,\ell}$, n=2m+1, i.e., in the subgraph of the n-cube induced by the middle 2ℓ levels. Throughout this section we consider fixed $m \geq 1$ and $2 \le \ell \le m+1$. We construct the cycle factor incrementally, starting with chains from the Greene-Kleitman SCD and adding $\{0,1,2\}$ -lexical edges between certain sets of vertices, see Figure 6. In the following, when referring to a subgraph given by a set of edges, we mean the subgraph of $Q_{n,\ell}$ induced by those edges. Moreover, we say that a chain is *short* if its length is at most $2\ell-3$, i.e., if it does not span all levels of $Q_{n,\ell}$.

Our construction starts by taking all those short chains, formally

$$X^{0} := \bigcup_{0 \le i \le h \le 2\ell - 3} M^{0}[C_{h,i}, C_{h,i+1}]; \tag{4a}$$

recall Lemma 5 (i). From Lemma 5 (ii) we know that 1-lexical edges perfectly match all bottom vertices of [-+]-chains of length 1 with all top vertices of [+-]-chains of length 1 along the edges

$$X_{\mathbf{m}}^{1} := M^{1}[C_{1,0}^{-}, C_{1,1}^{-}]. \tag{4b}$$

Furthermore, for $1 \le h \le 2\ell - 5$, 1-lexical edges perfectly match all top vertices of [?+]-chains of length h with all top vertices of [?-]-chains of length h+2, and all bottom vertices of [+?]-chains of length h with all bottom vertices of [-?]-chains of length h+2 along the edges

$$X_{\mathsf{t}}^{1} := \bigcup_{1 \le h \le 2\ell - 5} M^{1}[C_{h,h}^{+}, C_{h+2,h+2}^{-}], \quad X_{\mathsf{b}}^{1} := \bigcup_{1 \le h \le 2\ell - 5} M^{1}[C_{h,0}^{+}, C_{h+2,0}^{-}], \tag{4c}$$

respectively. Note that the only vertices of short chains that have degree 1 in the set

$$X := X^0 \cup X_{\mathtt{m}}^1 \cup X_{\mathtt{t}}^1 \cup X_{\mathtt{b}}^1 \tag{4d}$$

are exactly the vertices of $C^+_{2\ell-3,2\ell-3}$ and $C^+_{2\ell-3,0}$; that is, the top vertices of [?+]-chains of length $2\ell-3$ and the bottom vertices of [+?]-chains of length $2\ell-3$.

Next, between every pair of consecutive levels of $Q_{n,\ell}$ we take all 0-lexical and 1-lexical edges that are not incident to a degree-2 vertex in X. Specifically, between these pairs of levels we take all 0-lexical edges from chains that are not short and all 1-lexical edges between chains that are not short. In addition, between the top two levels we take all 1-lexical edges between top vertices of [?+]-chains of length $2\ell-3$ and top vertices of [?-]-chains of length $2\ell-1$, and symmetrically, between the bottom two levels we take all 1-lexical edges between bottom vertices of [+?]-chains of length $2\ell-3$ and bottom vertices of [-?]-chains of length $2\ell-1$. Formally, these sets of edges are

$$Y_1 := Y_1' \cup M^1[C_{2\ell-3,0}^+, C_{2\ell-1,0}^-], \quad Y_\ell := Y_\ell' \cup M^1[C_{2\ell-3,2\ell-3}^+, C_{2\ell-1,2\ell-1}^-], \quad Y_k := Y_k' \quad (5a)$$

for $1 < k < \ell$ where

$$Y_{k}' := \bigcup_{\substack{h \ge 2\ell - 1 \\ i := (h - (2\ell - 1))/2 + 2(k - 1)}} M^{0}[C_{h,i}, C_{h,i+1}] \cup M^{1}[C_{h,i}^{+}, C_{h+2,i+2}^{-}] \cup M^{1}[C_{h,i+1}^{+}, C_{h+2,i+1}^{-}]$$

$$(5b)$$

for $1 \le k \le \ell$. Note that Y_1 and Y_ℓ contain all $\{0,1\}$ -lexical edges between the bottom two levels or the top two levels of $Q_{n,\ell}$, respectively. We also define

$$Y := \bigcup_{1 \le k \le \ell} Y_k. \tag{5c}$$

As a consequence of these definitions and Lemma 5 (i) and (ii), the only vertices of $Q_{n,\ell}$ that have degree 1 in the set $X \cup Y$ are exactly the vertices of $C_{2\ell-1,i}^-$ for $1 \le i \le 2\ell-2$. We thus add the edges

$$Z := \bigcup_{i=1,3,5,\dots,2\ell-3} Z^{02}[C_{2\ell-1,i}^{-}, C_{2\ell-1,i+1}^{-}]$$

$$\tag{6}$$

defined in part (iii) of Lemma 5, which makes

$$C_{n,\ell} := X \cup Y \cup Z \tag{7}$$

a cycle factor in the graph $Q_{n,\ell}$.

3.1 Comparison with previous constructions

Our cycle factor construction generalizes the construction for $\ell=1$ presented in [19, 27], which simply consisted in taking the union of all 0-lexical and 1-lexical edges between the middle two levels. It also generalizes the construction for $\ell=2$ presented in [16], which also only used $\{0,1,2\}$ -lexical matchings. In fact, all these earlier papers actually used $\{m,m-1,m-2\}$ -lexical matching edges, but these are isomorphic to $\{0,1,2\}$ -lexical edges by reversing bitstrings. The earlier construction for $\ell=2$ seemed rather arbitrary at the time, but now nicely fits into the general picture shown in Figure 6^1 .

¹ As the picture of this construction resembles a rocket, with the tip on the left and the boosters on the right, one might be tempted to consider this rocket science.

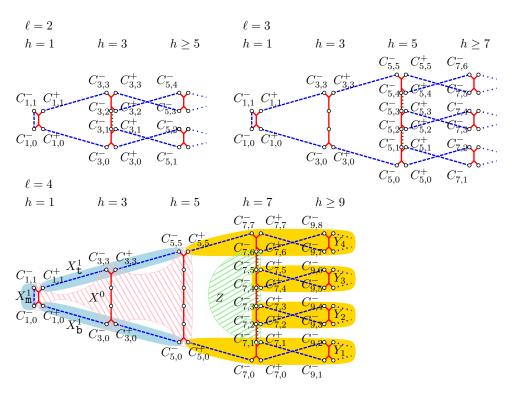


Figure 6 Illustration of the cycle factor $C_{n,\ell}$ for $\ell = 2, 3, 4$. Each bullet represents an entire set of vertices, as specified in the figure, lines between them specify perfect matchings between these sets. The $\{0, 1, 2\}$ -lexical edges are drawn with solid, dashed, and dotted lines, respectively. In the bottom part, various sets of matching edges are highlighted.

4 Sketch of the remaining proof steps

It turns out that each cycle from the factor $C_{n,\ell}$ defined in (7) visits vertices from an interval of 2r levels, where $2 \le r \le \ell$, around the middle. We refer to the number 2r as the range of the cycle, and we say the cycle is short if $2 \le r < \ell$, and long if $r = \ell$. One can show that any short cycle with range r has length 8(r-1), contains exactly one [--]-chain of length 2r-1, one [-+]- and one [+-]-chain of length 2r-3 each, and one [++]-chain of length 2r-5 (the latter only if $r \ge 3$), i.e., short cycles are in bijection with short [--]-chains. For long cycles, on the other hand, we are lacking such a detailed understanding of their structure. However, we are able to identify certain vertices on them, and to describe the operation of moving along one cycle from one such special vertex to the next one in terms of certain rotation operations on ordered rooted trees. Consequently, long cycles are obtained as equivalence classes of ordered rooted trees under such rotations.

As outlined in Section 1.3, to join the cycles in our factor to a Hamilton cycle, we explicitly construct flipping 4-cycles and 6-cycles. The 4-cycles are used to join short cycles among each other and to long cycles, in such a way that every short cycle is joined to some long cycle, possibly via other short cycles. For this we exploit the fact that certain pairs of short chains from the Greene-Kleitman SCD are connected by many 4-cycles. Specifically, consider any short chain C of length $h \geq 3$, and any chain C' of length h = 2 obtained from C by replacing two consecutive *s at positions a and b by 0 and 1, respectively. Using the definition of Greene-Kleitman chains, it is easy to check that C and C' are connected by h = 2 many 4-cycles, each using a distinct edge of C and C', except the two consecutive edges of C that

flip the coordinates a and b. The Greene-Kleitman SCD has an abundance of such pairs of heavily connected pairs of chains, and as our cycle factor contains all these short chains, we can exploit this to join short cycles to each other and to some long cycle in a tree-like fashion, by considering the short cycles by increasing range. Particular care must be taken to ensure that all selected flipping 4-cycles are edge-disjoint from each other, so that they do not interfere with each other in the joining process.

The remaining task is to join long cycles to each other, and for this we use flipping 6-cycles between the topmost two levels of $Q_{n,\ell}$, ensuring that they are edge-disjoint from any flipping 4-cycles, which all live in the levels below. Such a flipping 6-cycle can be used to connect two long cycles with each other, and this operation can again be interpreted in terms of an operation on ordered rooted trees, which we call a pull operation. These 6-cycles have been described and used heavily already in the predecessor papers [16, 19], where it was shown that they are all edge-disjoint. To complete the proof of Theorem 1, we show that all long cycles can be joined to each other by flipping 6-cycles, by showing that all equivalence classes of ordered rooted trees under the aforementioned rotations (which correspond to long cycles) can be transformed into each other by pull operations (which correspond to flipping 6-cycles). This step of the proof reduces the Hamiltonicity problem in $Q_{n,\ell}$ to proving that a suitably defined auxiliary graph is connected, which turns out to be much easier.

5 Proof of Theorem 3

In this section, we prove Theorem 3. All lemmas stated below follow from straightforward calculations; see [17] for details. For any chain C, we let |C| denotes its length, i.e., the number of *s in C. For any chain C with $|C| \ge 2$, we let f(C) and $\ell(C)$, respectively, denote the chains obtained by replacing the first two *s or the last two *s in C by 0 and 1. Note that if $|C| \ge 2$, then we have $f(\ell(*C*)) = \ell(f(*C*))$.

Our goal is to order the chains of the Greene-Kleitman SCD in Q_n , $n \geq 2$, so that any consecutive pair of chains is joined at their top end vertices or bottom end vertices alternatingly, with the exception of any two consecutive chains of length 1 that are connected from the bottom end of one of them to the top end of the other, so as to form a Hamilton cycle. We call such an ordering of chains a cycle ordering. The following simple but powerful lemma, valid for arbitrary SCDs, shows that the direction in which each chain is traversed along the Hamilton cycle (upwards or downwards) is determined only by the chain length.

▶ **Lemma 6.** Let Λ_n be a cycle ordering of chains of an SCD in Q_n , $n \geq 2$. In this Hamilton cycle, any two chains C and C' with $|C| \equiv |C'| \pmod{4}$ are traversed in the same direction.

We now define a cycle ordering Λ_n , $n \geq 2$, for the Greene-Kleitman SCD. The corresponding Hamilton cycle is oriented so that it traverses the longest chain $*^n$, which will be the first in the ordering Λ_n , from bottom to top. Our construction works inductively, and the induction step goes from n to n+2, with separate rules for even and odd n. The base cases are n=0 and n=1, for which the entire cube consists only of a single vertex and a single edge, respectively, so for these cases the notion of a cycle ordering is not defined.

For even n, we define $\Lambda_0 := \varepsilon$, and for $n \geq 0$ and given $\Lambda_n =: C_1, \ldots, C_N$ we define $\Lambda_{n+2} := \rho(\Lambda_n) = \rho(C_1), \ldots, \rho(C_N)$ with

$$\rho(C) := \begin{cases} \lambda(C) & \text{if } |C| \equiv n \pmod{4}, \\ \lambda(C)^R & \text{if } |C| \not\equiv n \pmod{4}, \end{cases}$$
(8)

and

$$\lambda(C) := \begin{cases} *C*, f(*C*), f(\ell(*C*)), \ell(*C*) & \text{if } |C| \ge 2, \\ *C*, 0C1 & \text{if } |C| = 0. \end{cases}$$
(9)

We call the chains of $\lambda(C)$ arising from C the descendants of C. This rule replaces each chain C in Λ_n by its descendants $\lambda(C)$, where the order of descendants can be reversed, indicated by the superscript R, depending on the length of C modulo 4.

For odd n, we define $\Lambda_1 := *$, and for $n \ge 1$ and given Λ_n we define $\Lambda_{n+2} := \rho(\Lambda_n)$, where ρ is as before and

$$\lambda(C) := \begin{cases} *C *, \ell(*C *), \ell(f(*C *)), f(*C *) & \text{if } |C| \ge 3, \\ *C *, \ell(*C *), f(*C *) & \text{if } |C| = 1. \end{cases}$$
(10)

▶ **Lemma 7.** Λ_n contains every chain of the Greene-Kleitman SCD exactly once.

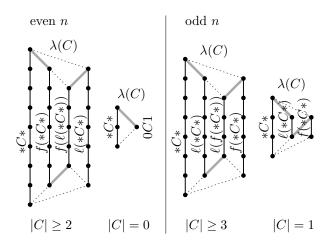


Figure 7 Connections between top and bottom ends of the descendants $\lambda(C)$ of a chain C, as guaranteed by Lemma 8. Bold gray edges are used along the Hamilton cycle. Dotted edges are present but not used.

To complete the proof of Theorem 3, it remains to show that any two consecutive chains in Λ_n can be joined by an edge between their top ends or bottom ends alternatingly. For this we need the following simple lemmas that guarantee these connecting edges.

▶ **Lemma 8.** For any $n \ge 2$ and any chain C with $|C| \ge 2$, the chains C and f(C), and the chains C and $\ell(C)$ are connected both at their top and bottom ends in Q_n .

All connecting edges between top and bottom ends among the descendants of a chain guaranteed by Lemma 8 are shown in Figure 7. The next two lemmas are illustrated in Figure 8.

- ▶ **Lemma 9.** For any $n \ge 2$ and any two chains C, C' connected at their bottom ends in Q_n , we have that *C* and *C'* are connected at their bottom ends in Q_{n+2} .
- ▶ Lemma 10. For any $n \ge 2$ and any chain C with $|C| \ge 2$ in Q_n , we have that $\ell(*C*)$ and $\ell(*f(C)*)$ are connected at their bottom ends in Q_{n+2} .

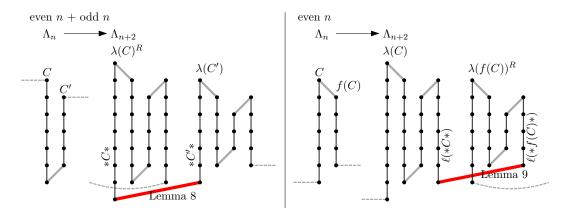


Figure 8 Joining of the descendants of two consecutive chain from Λ_n in the induction step $n \to n+2$, via the thick edges guaranteed by Lemmas 9 and 10. The dashed edges are connections to preceding and subsequent chains on the Hamilton cycle.

Proof of Theorem 3 (even n). We show that Λ_n , $n \geq 2$ even, defined in (9) is a cycle ordering of the Greene-Kleitman chains, by proving that any consecutive pair of chains is connected at their top or bottom ends alternatingly, starting with the first chain $*^n$ of length n that is traversed from bottom to top. We will also establish the following additional property P: For any two consecutive chains C and C' connected at their top ends, we either have C = f(C') or f(C) = C'. These invariants can easily be checked for the induction base case n = 2, which is given by $\Lambda_2 = **, 01$.

For the induction step consider $n \geq 2$ to be even, and assume that Λ_n is a cycle ordering satisfying property P. By Lemma 8, the descendants $\lambda(C)$ for any chain C from Λ_n can be joined as shown on the left hand side of Figure 7, so we only need to check the connections between the first and last chains among consecutive groups of descendants. Indeed, if Cand C' are consecutive in Λ_n and joined at their bottom ends, then C is traversed from top to bottom and C' from bottom to top in the Hamilton cycle; see the left part of Figure 8. Consequently, by Lemma 6, we have $|C| \not\equiv |*^n| = n \pmod{4}$ and $|C'| \equiv n \pmod{4}$, i.e., by (8) the sequence Λ_{n+2} contains $\lambda(C)^R$ and $\lambda(C')$, and indeed, the bottom vertex of the last chain of $\lambda(C)^R$, namely *C*, is connected to the bottom vertex of the first chain of $\lambda(C')$, namely *C'*, by Lemma 9. Similarly, if C and C' are consecutive in Λ_n and joined at their top ends, then C is traversed from bottom to top and C' from top to bottom in the Hamilton cycle; see the right part of Figure 8. Consequently, by Lemma 6, we have $|C| \equiv n \pmod{4}$ and $|C'| \not\equiv n \pmod{4}$, i.e., by (8) the sequence Λ_{n+2} contains $\lambda(C)$ and $\lambda(C')^R$, and indeed, the bottom vertex of the last chain of $\lambda(C)$, namely $\ell(*C*)$, is connected to the bottom vertex of the first chain of $\lambda(C')^R$, namely $\ell(*C'*)$, using that by property P we have either C = f(C') or f(C) = C', so we can invoke Lemma 10. Moreover, property P still holds for Λ_{n+2} by the definition (9) (note that if |C|=0, then we have 0C1=f(*C*)).

The proof of Theorem 3 for odd n is very similar. In [17] we provide all details and a loopless algorithm for computing this Gray code. An implementation of this algorithm in C++ is available for download and for demonstration [5].

Table 1 A glossary for notation used in the paper.

Q_n	$n \ge 1$	the n -dimensional hypercube
$Q_{n,\ell}$	$1 \le \ell \le m+1$	the subgraph of Q_n induced by the middle 2ℓ levels
	$n=2m+1,m\geq 1$	
D_{2k}	$k \ge 0$	the set of all Dyck paths (bitstrings) of length $2k$
D		the set of all Dyck paths
C		a chain $C = u_0 * u_1 * \cdots * u_{h-1} * u_h$ of length $h \ge 0$ in
		the Greene-Kleitman decomposition, $u_i \in D$ for every i
$C_{h,i}$	$0 \le i \le h$	the set of the i th vertices in all chains of length h
$C_{h,i}^-$	$0 \le i \le h$	as above but only in chains with $u_i = \varepsilon$
$C_{h,i}^-$ $C_{h,i}^+$	$0 \le i \le h$	as above but only in chains with $u_i \neq \varepsilon$
$L_{n,k}$	$0 \le k \le n$	the set of vertices on level k in Q_n
$M_{n,k}^p$	$0 \le k < n, 0 \le p < n$	the p-lexical matching between $L_{n,k}$ and $L_{n,k+1}$
M_n^p, M^p	$0 \le p < n$	the set of all p -lexical edges in Q_n
C		the length of a chain C , i.e., $ C = h$ for C as above
f(C)	$ C \ge 2$	the chain $f(C) = u_0 0 u_1 1 u_2 * \cdots * u_h$ for C as above
l(C)	$ C \ge 2$	the chain $l(C) = u_0 * \cdots * u_{h-2} 0 u_{h-1} 1 u_h$ for C as above
$\lambda(C)$		a sequence of descendant chains for a chain C , see (9) , (10)

References

- M. Aigner. Lexicographic matching in Boolean algebras. J. Combin. Theory Ser. B, 14:187–194, 1973.
- 2 J. Bitner, G. Ehrlich, and E. Reingold. Efficient generation of the binary reflected Gray code and its applications. *Comm. ACM*, 19(9):517–521, 1976.
- 3 B. Bollobás. Combinatorics: set systems, hypergraphs, families of vectors and combinatorial probability. Cambridge University Press, Cambridge, 1986.
- 4 M. Buck and D. Wiedemann. Gray codes with restricted density. *Discrete Math.*, 48(2-3):163–171, 1984. doi:10.1016/0012-365X(84)90179-1.
- 5 The Combinatorial Object Server: http://www.combos.org/chains.
- 6 N. de Bruijn, C. van Ebbenhorst Tengbergen, and D. Kruyswijk. On the set of divisors of a number. *Nieuw Arch. Wiskunde* (2), 23:191–193, 1951.
- M. El-Hashash and A. Hassan. On the Hamiltonicity of two subgraphs of the hypercube. In Proceedings of the Thirty-second Southeastern International Conference on Combinatorics, Graph Theory and Computing (Baton Rouge, LA, 2001), volume 148, pages 7–32, 2001.
- 8 P. Erdős and R. K. Guy. Crossing number problems. *Amer. Math. Monthly*, 80:52–58, 1973. doi:10.2307/2319261.
- 9 L. Faria, C. M. H. de Figueiredo, O. Sýkora, and I. Vrt'o. An improved upper bound on the crossing number of the hypercube. J. Graph Theory, 59(2):145–161, 2008. doi: 10.1002/jgt.20330.
- T. Feder and C. Subi. On hypercube labellings and antipodal monochromatic paths. *Discrete Appl. Math.*, 161(10-11):1421–1426, 2013. doi:10.1016/j.dam.2012.12.025.
- J. Fink. Perfect matchings extend to Hamilton cycles in hypercubes. J. Combin. Theory Ser. B, 97(6):1074–1076, 2007. doi:10.1016/j.jctb.2007.02.007.
- 12 J. Fink. Matchings extend into 2-factors in hypercubes. Combinatorica, 39(1):77-84, 2019. doi:10.1007/s00493-017-3731-8.
- 13 Z. Füredi. Problem session. In Kombinatorik geordneter Mengen, Oberwolfach, BRD, 1985.
- 14~ F. Gray. Pulse code communication, 1953. March 17, 1953 (filed Nov. 1947). U.S. Patent 2,632,058.

- 15 C. Greene and D. J. Kleitman. Strong versions of Sperner's theorem. J. Combin. Theory Ser. A, 20(1):80–88, 1976.
- P. Gregor, S. Jäger, T. Mütze, J. Sawada, and K. Wille. Gray codes and symmetric chains. In 45th International Colloquium on Automata, Languages, and Programming, ICALP 2018, July 9-13, 2018, Prague, Czech Republic, pages 66:1-66:14, 2018. doi:10.4230/LIPIcs.ICALP. 2018.66.
- P. Gregor, O. Mička, and T. Mütze. On the central levels problem. arXiv:1912.01566. Full preprint version of the present article, 2019.
- 18 P. Gregor and T. Mütze. Trimming and gluing Gray codes. *Theoret. Comput. Sci.*, 714:74–95, 2018. doi:10.1016/j.tcs.2017.12.003.
- 19 P. Gregor, T. Mütze, and J. Nummenpalo. A short proof of the middle levels theorem. Discrete Analysis, 2018:8:12 pp., 2018.
- 20 P. Gregor and R. Škrekovski. On generalized middle-level problem. *Inform. Sci.*, 180(12):2448–2457, 2010. doi:10.1016/j.ins.2010.02.009.
- J. Griggs, C. E. Killian, and C. D. Savage. Venn diagrams and symmetric chain decompositions in the Boolean lattice. *Electron. J. Combin.*, 11(1):Paper 2, 30 pp., 2004. URL: http://www.combinatorics.org/Volume_11/Abstracts/v11i1r2.html.
- 22 I. Havel. Semipaths in directed cubes. In Graphs and other combinatorial topics (Prague, 1982), volume 59 of Teubner-Texte Math., pages 101–108. Teubner, Leipzig, 1983.
- H. Huang. Induced subgraphs of hypercubes and a proof of the sensitivity conjecture. Ann. of Math. (2), 190(3):949–955, 2019. doi:10.4007/annals.2019.190.3.6.
- 24 H. A. Kierstead and W. T. Trotter. Explicit matchings in the middle levels of the Boolean lattice. *Order*, 5(2):163–171, 1988. doi:10.1007/BF00337621.
- 25 D. E. Knuth. The Art of Computer Programming. Vol. 4A. Combinatorial Algorithms. Part 1. Addison-Wesley, Upper Saddle River, NJ, 2011.
- S. Locke and R. Stong. Problem 10892: Spanning cycles in hypercubes. Amer. Math. Monthly, 110:440–441, 2003.
- 27 T. Mütze. Proof of the middle levels conjecture. Proc. Lond. Math. Soc., 112(4):677-713, 2016. doi:10.1112/plms/pdw004.
- T. Mütze and J. Nummenpalo. A constant-time algorithm for middle levels Gray codes. In Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, pages 2238–2253. SIAM, Philadelphia, PA, 2017. doi:10.1137/1.9781611974782.147.
- N. Nisan and M. Szegedy. On the degree of Boolean functions as real polynomials. Comput. Complexity, 4(4):301–313, 1994. Special issue on circuit complexity (Barbados, 1992). doi: 10.1007/BF01263419.
- 30 S. Norine. Edge-antipodal colorings of cubes. The Open Problem Garden. Available at http://www.openproblemgarden.org/op/edge_antipodal_colorings_of_cubes, 2008.
- 31 O. Pikhurko. On edge decompositions of posets. Order, 16(3):231-244 (2000), 1999. doi: 10.1023/A:1006419611661.
- F. Ruskey and C. Savage. Hamilton cycles that extend transposition matchings in Cayley graphs of S_n . SIAM J. Discrete Math., 6(1):152–166, 1993. doi:10.1137/0406012.
- F. Ruskey, C. D. Savage, and S. Wagon. The search for simple symmetric Venn diagrams. Notices Amer. Math. Soc., 53(11):1304–1312, 2006.
- 34 C. D. Savage. Long cycles in the middle two levels of the Boolean lattice. *Ars Combin.*, 35(A):97–108, 1993.
- 35 C. D. Savage. A survey of combinatorial Gray codes. SIAM Rev., 39(4):605–629, 1997. doi:10.1137/S0036144595295272.
- 36 C. D. Savage and P. Winkler. Monotone Gray codes and the middle levels problem. J. Combin. Theory Ser. A, 70(2):230-248, 1995. doi:10.1016/0097-3165(95)90091-8.
- J. Shearer and D. J. Kleitman. Probabilities of independent choices being ordered. Stud. Appl. Math., 60(3):271-276, 1979. doi:10.1002/sapm1979603271.

- 38 X. S. Shen and A. Williams. A middle levels conjecture for multiset permutations with uniform-frequency. Williams College Technical Report CSTR-201901. Available at http://tmuetze.de/papers/multiset.pdf, 2019.
- 39 H. Spink. Orthogonal symmetric chain decompositions of hypercubes. SIAM J. Discrete Math., 33(2):910–932, 2019. doi:10.1137/18M1187179.
- 40 N. Streib and W. T. Trotter. Hamiltonian cycles and symmetric chains in Boolean lattices. Graphs Combin., 30(6):1565–1586, 2014. doi:10.1007/s00373-013-1350-8.
- 41 I. Tomon. On a conjecture of Füredi. *European J. Combin.*, 49:1-12, 2015. doi:10.1016/j.ejc.2015.02.026.
- 42 D. E. White and S. G. Williamson. Recursive matching algorithms and linear orders on the subset lattice. *J. Combin. Theory Ser. A*, 23(2):117–127, 1977.