Complexity measures on the symmetric group

Yuval Filmus, Noam Lifshitz, Nathan Lindzey, Marc Vinyals

June 24, 2020

1 Introduction and motivation

The study of complexity measures on functions was initiated by Noam Nisan in his paper *CREW PRAMs and Decision Trees*. The goal of this paper was to understand the time required to compute a Boolean function $f\colon \{0,1\}^n \to \{0,1\}$ on a CREW PRAM. Improving on an earlier lower bound of $\Omega(\log s(f))$, Nisan proves a lower bound of $\Omega(\log \mathrm{bs}(f))$, and a matching upper bound of $O(\log D(f))$. To show that these bounds are indeed matching, Nisan shows that the following complexity measures are all polynomially related: block sensitivity, certificate complexity, and decision tree complexity (deterministic and randomized). This uses earlier results of Blum and Impagliazzo, *Generic oracles and oracle classes*.

Nisan and Szegedy continued this study in their paper *On the degree of Boolean functions as real polynomials*, where they add two new measures to the mix: degree and approximate degree.

Later on, other complexity measures were added to the mix. Perhaps the most prominent one is query complexity, which is lower bounded by degree or approximate degree (depending on whether errors are allowed). A matching upper bound was proven by Beals, Buhrman, Cleve, Mosca and de Wolf in their paper *Quantum lower bounds by polynomials*; see also the survey *Complexity measures and decision tree complexity* by Buhrman and de Wolf.

Conspicuously missing from this mix is *sensitivity*. This has captured the attention of researchers in the field, who wrote many papers and surveys on the topic, but the final resolution came from outside: combinatorialist Hao Huang came up with a simple argument showing that $s(f) \geq \sqrt{\deg(f)}$ in his paper *Induced subgraphs of hypercubes and a proof of the sensitivity conjecture*.

Huang's paper prompted stunned complexity theorists with its simplicity. Soon mathematicians offered several alternative views of the argument, and computer scientists refined his statement to make better sense of the square root relation. Our own response went in a different direction. We asked the following question:

> Does Huang's argument generalize to other settings, say functions on the symmetric group?

The first step in answering this question is defining appropriate complexity measures for functions on the symmetric group. Indeed, although Huang's sensitivity theorem relates sensitivity and degree, it is natural to try and adapt the entire theory developed by Nisan and others. This ties in with a different thread of research I have been involved with:

> What do almost degree d Boolean functions on the symmetric group look like?

The analogous question on the Boolean cube has been answered for $d = 1$ by the Friedgut–Kalai–Naor theorem: they are dictators. For larger d, the Kindler–Safra theorem shows that such functions are close to juntas, a robust version of the characterization of Boolean degree d functions due to Nisan and Szegedy.

On the symmetric group, Ellis–Filmus–Friedgut show that Boolean functions close to degree 1 are close to dictators, for an appropriate notion of dictator: a function depending only on $\pi(i)$ or only on $\pi^{-1}(j)$. What is the corresponding result for larger d?

The first step in answering this question is understanding the structure of Boolean functions of degree d (exactly). These are no longer juntas: the indicator of "1 belongs to a 2-cycle" has degree 2 but depends on all coordinates. However, this function can be expressed as a depth 2 "matching decision tree", a concept arising in proof complexity as part of proving lower bounds on the pigeonhole principle. This suggests that the connection between degree and decision tree complexity might extend to the symmetric group.

A final, post hoc motivation for studying complexity measures in this setting, comes from Erdős–Ko–Rado theory:

Find t-intersecting families of permutations of the maximum size $(n-t)!$.

A t-intersecting family of permutations is one in which every two permutations agree on the image of at least t points. Ellis, Friedgut and Pilpel showed that for large n, such families contain at most $(n-t)!$ permutations. This bound is attained by "t-double cosets", which consist of all permutations sending $i_1, \ldots, i_t$ to $j_1, \ldots, j_t$, respectively. Ellis, Friedgut and Pilpel showed that (for large n) these are the unique families attaining the bound, but their proof contains a mistake. The result also follows from subsequent work of Ellis, but with a much more complicated proof. Using complexity measures, we are able to give a simple and short proof of uniqueness.

2 Complexity measures on the symmetric group

Let us start by exploring various complexity measures of functions on the symmetric group. For each measure we introduce, we first explain how the measure is defined for functions on the Boolean cube, and then how the definition generalizes to functions on the symmetric group.

Certificate complexity Let $f\colon \{0,1\}^n \to \{0,1\}$. A certificate for f at a point $x \in \{0,1\}^n$ is a subset S of the coordinates such that if $y|_S = x|_S$ then $f(y) = f(x)$; in other words, $x|_S$ certifies the value of the function. The certificate complexity of f at x, denoted $C(f,x)$, is the size on the minimal certificate, and the certificate complexity of f, denoted $C(f)$, is the maximum of $C(f,x)$ over all $x \in \{0,1\}^n$.

This definition readily extends to functions on the symmetric group. A certificate for a function $f\colon S_n \to \{0,1\}$ at a point $\pi \in S_n$ is again a set S of coordinates such that $\pi|_S$ certifies the value of f. The rest of the definition generalizes in exactly the same way.

For future reference, let us record that a function with certificate complexity 1 is a *dictator*: a function depending only on $\pi(i)$ or only on $\pi^{-1}(j)$.

Decision tree complexity Let $f\colon \{0,1\}^n \to \{0,1\}$. A decision tree is a binary tree whose nodes represent queries of the form "$x_i =$?", and whose leaves are marked by $0, 1$. A decision tree computes a function in the natural way. The decision tree complexity of f, denoted $D(f)$, is the minimum depth of a decision tree computing f.

In order to extend this definition to functions on the symmetric group, we need to decide on a set of allowed queries. The basic decision we have to make is whether we insist on Boolean queries, or allow more general ones. In the former case, the natural queries are "$\pi(i) = j$?". However, balanced dictators will then have decision tree complexity of order $\log n$. Since we want all of our measures to be polynomially related, this suggests allowing queries with more than two answers. The natural queries are "$\pi(i) =$?" and "$\pi^{-1}(j) =$?", and indeed, dictators have decision tree complexity 1 using this type of queries. This is the definition we adopt.

As mentioned in the introduction, such decision trees appear in proof complexity lower bounds for the pigeonhole and perfect matching principles.

The randomized decision tree complexity $R(f)$ is the minimum d such that there is a distribution on depth d decision trees which, for each input separately, computes the correct value with probability at least 2/3.

Degree and approximate degree Let $f\colon \{0,1\}^n \to \{0,1\}$. We can define the degree of f in many equivalent ways:

1. Degree of the Fourier expansion of f; equivalently, maximum d such that $f^{=d} \neq 0$.
2. Minimal degree of a polynomial representing f.
3. Minimum d such that f can be written as a sum of *d-juntas*, that is, functions depending on d coordinates.

These definitions extend to the symmetric group. All the following notions coincide for functions $f\colon S_n \to \{0,1\}$:

1. Minimal degree of a polynomial representing f. The variables of this polynomial are $x_{i,j}$, which are the indicators of "$\pi(i) = j$", or equivalently, the entries of the permutation matrix representing π.
2. Minimal d such that f can be written as a sum of d-juntas, that is, functions depending on d pieces of information of the form $\pi(i)$ or $\pi^{-1}(j)$.
3. Maximum d such that $f^{=\lambda} \neq 0$ for some partition λ with $\lambda_1 = n - d$; here $f^{=\lambda}$ is the component of f in the isotypical component indexed by λ.

The approximate degree of f is the minimum degree of a non-Boolean function g such that $\|f-g\|_\infty \leq 1/3$.

Block sensitivity The definition of these parameters is less intuitive, so we start by describing two additional measures, average certificate complexity and fractional block sensitivity, for functions on the Boolean cube.

For a function $f\colon \{0,1\}^n \to \{0,1\}$ and a point $x \in \{0,1\}^n$, we can write $C(f,x)$ as the solution of an integer program:

$$\begin{aligned} & \min z_1 + \cdots + z_n \\ \text{s.t.} \quad & \sum_{i\colon y_i \neq x_i} z_i \geq 1 \text{ for all } y \text{ such that } f(y) \neq f(x) \\ & z_1, \ldots, z_n \in \{0,1\} \end{aligned}$$

The corresponding linear programming relaxation is

$$\begin{aligned} & \min z_1 + \cdots + z_n \\ \text{s.t.} \quad & \sum_{i\colon y_i \neq x_i} z_i \geq 1 \text{ for all } y \text{ such that } f(y) \neq f(x) \\ & 0 \leq z_1, \ldots, z_n \leq 1 \end{aligned}$$

This defines what Scott Aaronson called *randomized certificate complexity*. The dual linear program has a variable w_y for each input y such that $f(y) \neq f(x)$:

$$\begin{aligned} & \max \sum_{y\colon f(y) \neq f(x)} w_y \\ \text{s.t.} \quad & \sum_{i\colon y_i \neq x_i} w_y \leq 1 \text{ for all } 1 \leq i \leq n \\ & 0 \leq w_y \leq 1 \end{aligned}$$

The corresponding integer program is

$$\begin{aligned} \max \quad & \sum_{y\colon f(y)\neq f(x)} w_y \\ \text{s.t.} \quad & \sum_{i\colon y_i\neq x_i} w_y \leq 1 \text{ for all } 1\leq i\leq n \\ & w_y \in \{0,1\} \end{aligned}$$

We can think of the w_y as defining a subset $y_1,\ldots,y_b$ of inputs satisfying $f(y_j)\neq f(x)$. If we denote by B_j the set of coordinates on which y_j and x differ, then the constraint states that the B_j are disjoint. The integer program defines the so-called block sensitivity of f at x, denoted $\mathrm{bs}(f,x)$: it is the maximal number of disjoint "blocks" of coordinates, flipping each of which changes the value of the function. The block sensitivity $\mathrm{bs}(f)$ is the maximum of $\mathrm{bs}(f,x)$ over all x. The preceding fractional relaxation first appears in a paper of Tal, who called the parameter it defines *fractional block sensitivity*.

We can repeat the same steps for the symmetric group. The end result is as follows. The block sensitivity of a function $f\colon S_n\to\{0,1\}$ at a point $\pi\in S_n$ is the maximal number of permutations $\rho_1,\ldots,\rho_b$ such that $f(\rho_j)\neq f(\pi)$ and the sets $B_j=\{i:\rho_j(i)\neq\pi(i)\}$ are disjoint.

Sensitivity The definition of the sensitivity of a function $f\colon\{0,1\}^n\to\{0,1\}$ at a point $x\in\{0,1\}^n$ is very similar to the definition of block sensitivity. All we do is add a constraint on the blocks B_j, namely, $|B_j|=1$. In other words, $s(f,x)$ is the number of indices whose flipping changes the output of the function. We define $s(f)$, the sensitivity of f, by taking a maximum over x.

To generalize sensitivity to the symmetric group, we add a similar constraint on B_j. Since any two permutations which agree on $n-1$ coordinates are in fact identical, the minimum size of a block is 2, and accordingly, $s(f,\pi)$ is the maximum number of disjoint transposition whose application changes the output of the function; here two transpositions $(i\ j),(k\ \ell)$ are disjoint if the sets $\{i,j\},\{k,\ell\}$ are disjoint.

Another reasonable definition of $s(f,\pi)$ would have been a suitable normalization of the number of transpositions whose application changes the value of the function. For the dictatorship "$\pi(1)\leq m$", this number is $\max(m,n-m)=\Theta(n)$, suggesting that the correct normalization factor is n. However, this definition is problematic: for the function "the cycle decomposition of π contains at least one 2-cycle of the form $(2i-1\ 2i)$", the resulting measure would be $O(1)$, although the decision tree complexity is linear.

3 Relations between complexity measures

In order to polynomially relate all complexity measures other than sensitivity, we simply generalize classical arguments. Let us go over the various steps, and explain the technical ingredients necessary to carry this plan through.

Part 1: Bounding decision tree complexity in terms of certificate complexity This part, attributed by Nisan to Blum, repeatedly chooses a 1-certificate c not conflicting with inputs seen so far (if any), and determines the value of the input on these coordinates. After $C(f)$ iterations, one can prove that the values seen so far determine f, and so $D(f)\leq C(f)^2$.

Generalizing this to the symmetric group is straightforward. A certificate c is now a "partial permutation", that is, a list of pairs $(i_1,j_1),\ldots,(i_k,j_k)$ such that if $\pi(i_t)=j_t$ for $t\in[k]$ then $f(\pi)=1$. We need to ask all queries "touching" the certificate, that is, we need to determine $\pi(i_1),\ldots,\pi(i_k)$ as well as $\pi^{-1}(j_1),\ldots,\pi^{-1}(j_k)$. This increases the upper bound to $2C(f)^2$.

The correctness proof relies on the following property: if c_0,c_1 are two conflicting certificates (for example, a 0-certificate and a 1-certificate) then either $(i,j_0)\in c_0$ and $(i,j_1)\in c_1$ for some i and $j_0\neq j_1$, or $(i_0,j)\in c_0$ and $(i_1,j)\in c_1$ for some j and $i_0\neq i_1$.

Part 2: Bounding certificate complexity in terms of block sensitivity Let $f\colon \{0,1\}^n \to \{0,1\}$ be a Boolean function, and $x \in \{0,1\}^n$ a point. The idea here is to consider an optimal set of blocks $B_1, \dots, B_s$ for x, that is, a maximal number of disjoint blocks such that $f(x \oplus B_j) \neq f(x)$. The certificate consists of $x|_{B_1 \cup \cdots \cup B_s}$. If it is not a certificate, then this translates to an additional disjoint block, thus contradicting the choice of blocks.

In order for the certificate to be small, we require the blocks to be inclusion-minimal. As a result, the size of each block is at most $s(f)$. Indeed, a block B_j can be shortened by one element in $|B_j|$ ways. If $|B_j| > s(f)$ then by the definition of sensitivity, one of these shortenings B'_j must satisfy $f(x \oplus B'_j) = f(x \oplus B_j)$, since otherwise $s(f,x) = |B_j|$. In total, the size of the certificate is at most $s(f)\,\mathrm{bs}(f)$.

The argument generalizes, by and large, to the symmetric group. The only complicated bit is the shortening step. We need to show that a large "block" can be shortened in many different *disjoint* ways.

Let us assume that we are trying to bound $C(f,e)$ for some $f\colon S_n \to \{0,1\}$, where e is the identity permutation. We are given that $f(\pi) \neq f(e)$ for some permutation π with associated block $B = \{i : \pi(i) \neq i\}$. We want to show that if π is minimal with respect to the size of B, then $|B| = O(s(f))$.

Write π as a product of cycles, and consider one of those cycles $(i_1 \ \cdots \ i_\ell)$. Applying the transposition $(i_t \ i_{t+1})$, we shorten this cycle by one or two elements (the latter, when $\ell = 2$), thus decreasing the size of the corresponding block. This can be done in $\lfloor \ell/2 \rfloor \geq \ell/3$ many disjoint ways. Summing over all cycles, we get at least $|B|/3$ many disjoint ways to shorten the block, and so $|B| \leq 3s(f)$.

Part 3: Bounding block sensitivity in terms of degree The argument for functions on the Boolean cube uses the lower bound $\Omega(\sqrt{n})$ on the approximate degree of OR.

In this case there is no need to generalize the argument, since we can *reduce* to the case of the Boolean cube. Here's how. Consider some function $f\colon S_n \to \{0,1\}$, and suppose that the block sensitivity of f is attained at the identity permutation. Thus there are permutations $\rho_1, \dots, \rho_{\mathrm{bs}(f)}$ whose support is disjoint such that $f(\rho_j) \neq f(e)$.

We construct a new function $g\colon \{0,1\}^{\mathrm{bs}(f)} \to \{0,1\}$ given by:

$$g(y_1, \dots, y_{\mathrm{bs}(f)}) = \prod_j \rho_j^{y_j}.$$

This function has block sensitivity $\mathrm{bs}(f)$, and so degree $O(\deg(g)^2)$. On the other hand, by translating a polynomial representation of f monomial by monomial, it is not hard to check that $\deg(g) \leq \deg(f)$. Therefore $\mathrm{bs}(f) = O(\deg(f)^2)$.

Part 4: Bounding degree in terms of decision tree complexity This part is trivial: $\deg(f) \leq D(f)$ since a decision tree readily translates into a polynomial whose degree is at most the depth of the decision tree.

4 Sensitivity theorem

Let us briefly recall Huang's proof of the sensitivity theorem for functions on the Boolean cube. The starting point is a function $f\colon \{0,1\}^n \to \{\pm 1\}$ of degree d. Since f has degree d, it has non-zero correlation with some degree d Fourier character χ_S, where $|S| = d$. Substituting arbitrary values in the variables outside S, we obtain a degree d function $g\colon \{0,1\}^d \to \{\pm 1\}$. Multiplying g by the parity character, we obtain a function $h\colon \{0,1\}^d \to \{\pm 1\}$ such that $\mathbb{E}[h] \neq 0$, say $\mathbb{E}[h] > 0$. Huang's ingenious argument then shows that $h^{-1}(1)$ has a point with $\sqrt{d}$ neighbors, which translates to a point of sensitivity $\sqrt{d}$ in g, and so in f.

Let us try to replicate this argument in the symmetric group. Looking ahead, the analog of the parity character is the sign character, of degree $d-1$. Our goal is thus to reduce the input domain from S_n from S_{d+1}. The idea is that if f is a degree d function on S_n for $n > d+1$, then we can restrict to a copy of S_{n-1} on which f also has degree d. Surprisingly, this is not always possible! The reason is the *branching rule* of the symmetric group, which describes what happens to the λ'th homogeneous parts of f under restriction.

They find themselves in homogeneous parts corresponding to all different ways of removing a box from λ. If $\lambda = (n/2)^2$ then all possible ways result in reducing the degree from $n/2$ to $n/2 - 1$!

The original reduction from $\{0,1\}^n$ to $\{0,1\}^d$ actually used a different idea: the observation that if $\deg f = d$ then f correlates with some Fourier character χ_S. An analogous statement holds for the symmetric group as well. Let V^λ denote the space of λ-homogeneous functions. This space is spanned by functions $\chi_{A,B}$, where A, B are two Young tableaux of shape λ.

To describe $\chi_{A,B}$, let us first explain a simpler function, $e_{A,B}$. This is the indicator of the following event: π sends the i'th row of A to the i'th row of B. The function $\chi_{A,B}$ is given by

$$\chi_{A,B} = \sum_\sigma (-1)^\sigma e_{A,B^\sigma},$$

where σ goes over all possible ways of permuting the columns of B (resulting in B^σ), and $(-1)^\sigma$ is the sign of the permutation σ.

If $\deg f = d$ then $f^{=\lambda} \neq 0$ for some partition λ with $\lambda_1 = n - d$, and so f has non-zero correlation with some $\chi_{A,B}$, where A, B have shape λ.

For each column c of A, choose $\lfloor |c|/2 \rfloor \geq (|c| - 1)/2$ disjoint transpositions. In total, this gives a set of at least $\sum_c (|c| - 1)/2 = (n - (n - d))/2 = d/2$ many transpositions T. The Cayley graph of S_n with respect to T consists of a disjoint union of many copies of the $|T|$-dimensional Boolean cube. On each such cube, either $\chi_{A,B}$ is identically zero, or its restriction to the cube behaves like the parity character. It follows that f has non-zero correlation with the parity character on one of the cubes C. By Huang's sensitivity theorem, $f|_C$ has sensitivity at least $\sqrt{|T|} = \Omega(\sqrt{d})$, as a function on the cube. This implies that f itself has sensitivity $\Omega(\sqrt{d})$ as a function on S_n.

5 Application to EKR theory

In a breakthrough paper, Ellis, Friedgut and Pilpel showed that a t-intersecting family of permutations (that is, a subset of S_n in which any two permutations agree on at least t points) consists of at most $(n - t)!$ permutations, for large enough n (as a function of t); previously only the easy case $t = 1$ was known. They also show that the unique families attaining this bound are "t-stars", that is, families of the form "$\pi(i_1) = j_1, \ldots, \pi(i_t) = j_t$". Unfortunately, the proof of this part is wrong, though the result does follow from subsequent work of Ellis.

The proof of the upper bound uses spectral technique. In particular, they show that if F is a t-intersecting family of size $(n - t)!$ then $\deg(f) = t$, where f is the characteristic function of F. Since $\deg(f) = t$, we know that $C(f) \leq C_t$, where C_t is some constant depending polynomially on t. We will show that for large enough n, this implies that F must be a subset of some t-star, and so (given its size) a t-star.

Suppose that F is not contained in any t-star; in particular, it is non-empty. Let σ be some 1-certificate of F. A short argument shows that $|\sigma| \geq t$ (as long as $n > t + 1$). For any subset τ of σ of size t, we know that F contains some element π_τ not containing τ, and so covered by some 1-certificate ρ_τ, which necessarily doesn't contain τ.

Now consider an arbitrary element $\pi \in \tau$. Since π intersects all extensions of σ, a short argument shows that π must contain some subset τ of σ of size t (again, as long as $n > t + 1$). Similarly, π must contain some subset of ρ_τ of size t, and in particular, some element (i, j) in $\rho_\tau \setminus \tau$. In total, π lies in the $(t + 1)$-star given by τ and (i, j). There are $\binom{C_t}{t}$ choices for τ and C_t choices for (i, j), and so

$$|F| \leq \binom{C_t}{t} C_t (n - (t + 1))!,$$

which for large enough n (polynomial in t suffices) is less than $(n - t)!$.

6 Degree one functions

Ellis, Friedgut and Pilpel describe all Boolean degree 1 functions. It is an almost trivial fact that all Boolean degree 1 functions on the Boolean cube are dictators. In contrast, the corresponding result in the symmetric group is much less obvious. We have a slightly different take on their proof, which allows us to extend it to the perfect matching scheme (more on this, below).

Let us recall that we identify each permutation with the corresponding permutation matrix. The convex hull of all permutation matrices forms the so-called Birkhoff polytope, given by the following defining system:

$$
\begin{aligned}
&\sum_i x_{i,j} = 1 && \text{for all } j \\
&\sum_j x_{i,j} = 1 && \text{for all } i \\
&x_{i,j} \geq 0 && \text{for all } i, j
\end{aligned}
$$

Let f be a non-constant degree 1 Boolean function. The function f is non-negative on the entire Birkhoff polytope, since the minimum of a linear function on a polytope is attained at a vertex. The set $f^{-1}(0)$ is a face of the polytope, and in particular, consists of the set of solutions to $x_{i_1,j_1} = \cdots = x_{i_t,j_t} = 0$. In particular, we can find i, j such that $f > 0$ if $x_{i,j} = 1$. Since f is Boolean, $f > 1$ when $x_{i,j} = 1$, and so $f - x_{i,j}$ is another Boolean degree 1 function, with larger zero-set.

Continuing in this way, we write f as a sum of $x_{i,j}$'s. Since f is Boolean, the summands have to be mutually exclusive, and so all of them lie on the same row or all of them lie on the same column.

7 Other domains

So far we have considered functions on the symmetric group. However, the entire framework generalizes to many other domains:

1. Arbitrary product domains, such as $[m]^n$.
2. The perfect matching scheme, which consists of all perfect matchings in K_{2n}, and hypergraphical analogs of the symmetric group and perfect matching scheme.
3. Slices and multislices, that is, subsets of $[m]^n$ with fixed histogram.

All of these are examples of domains given by a set of linear equations. They can also all be viewed as "generalized permutations".

We have identified several combinatorial parameters whose values figure in the arguments above:

1. Maximum degree: maximum number of queries mentioning any given element.
2. Conflict bound: a bound on the size of certificates that guarantees that if two certificates conflict (have no joint extension) then they are separated by some query.
3. Sensitivity ratio: the worst ratio between the size of a block and the number of disjoint shortenings.

In most cases, all of these parameters do not depend on n. However, in the case of unbalanced slices, the conflict bound depends on the weight, and this has the effect that our results only hold for low complexity functions. For example, every function on the k'th slice has certificate complexity at most k, but its decision tree complexity could be linear.

In order for a sensitivity theorem to hold, we need analogs of the functions $\chi_{A,B}$. These exist for arbitrary product domains, the perfect matching scheme, and slices and multislices. Proving a sensitivity theorem for hypergraphical analogs remains open.

The application to Erdős–Ko–Rado theory extends with few changes to the perfect matching scheme. The characterization of Boolean degree 1 functions also extends to the scheme, but the argument is more complicated, and at one point, we need to use the fact that degree 1 functions have sensitivity 1. The end result is that a Boolean degree 1 function is either a dictator or it depends on whether the input perfect matching intersects some fixed triangle.

8 Open question

What is the bounded depth circuit complexity of computing the sign of a permutation? The input is given as the n^2 entries of the corresponding permutation matrix.

On the one hand, the sign is the parity of the number of inversions, that is, of the $\Theta(n^4)$ expressions $x_{i,j} \wedge x_{k,\ell}$, where $i < k$ and $j > \ell$. Thus, a depth-$(d-1)$ circuit for parity on $\Theta(n^4)$ elements gives rise to a depth-d circuit for computing the sign.

On the other hand, we can reduce parity on $n/2$ bits to the sign of a permutation in S_n (assuming n is even), using the mapping

$$(y_1, \ldots, y_{n/2}) \mapsto (1\ 2)^{y_1}(3\ 4)^{y_2} \cdots (n-1\ n)^{y_{n/2}}.$$

This shows that a depth-d circuit for computing the sign gives rise to a depth-d circuit for parity on $n/2$ elements.

Can we close the gap between the two bounds?