Quantum Lower Bounds for Fanout

M. Fang* S. Fenner† F. Green‡ S. Homer* Y. Zhang†

May 25, 2005

Abstract: We consider the resource bounded quantum circuit model with circuits restricted by the number of qubits they act upon and by their depth. Focusing on natural universal sets of gates which are familiar from classical circuit theory, several new lower bounds for constant depth quantum circuits are proved. The main result is that parity (and hence fanout) requires log depth quantum circuits, when the circuits are composed of single qubit and arbitrary size Toffoli gates, and when they use only constantly many ancillæ. Under this constraint, this bound is close to optimal. In the case of a non-constant number a of ancillæ and n input qubits, we give a tradeoff between a and the required depth, that results in a non-constant lower bound for fanout when $a = n^{1-o(1)}$. We also show that, regardless of the number of ancillæ arbitrary arity Toffoli gates cannot be simulated exactly by a circuit family with gates of bounded arity.
Classification: Current challenges - quantum computing

1 Introduction

It is clear that for the foreseeable future, realizable quantum computations will have very limited duration, due to short coherence times. This has led to the investigation of parallelizing quantum computation in the form of small-depth quantum circuits. Such models also yield added insight into the nature of quantum computation, since they provide a setting in which quantum analogs of classical computational models are *provably* more powerful.

There has been significant recent progress in understanding the power of constant depth quantum circuits. Much of the progress in this area has been in showing that a variety of families of constant depth circuits are more powerful than their classical counterparts. These models make use of (reversible, unbounded) quantum *fanout* gates. A fanout gate takes an arbitrary number of bits and fans out one of them by taking its XOR with each of the others.

Fanout gates have proved to be unexpectedly powerful. Indeed, Høyer and Špalek [5] have shown that, in a sense, fanout gates are *universal* for all sets of commuting gates. By parallelizing commuting gates, fanout can be used to do simulations in constant depth. Thus, for example, together with single qubit gates, fanout gates can be used to efficiently simulate *generalized Toffoli* gates, which represent the simplest form of unbounded fan*in* gates.

Fanout gates might actually be feasible to build via ion trap or bulk NMR techniques. The same is true of Toffoli gates (see [12] for one such proposal). Thus it is potentially of practical

*Computer Science Department, Boston University, Boston, MA 02215, {heroes|homer}@bu.edu
†Dept. of CS and Eng., University of South Carolina, Columbia, SC 29208, {fenner|zhang29}@cse.sc.edu
‡Dept. of Math and CS, Clark University, Worcester, MA 01610, fgreen@black.clarku.edu

importance to compare the power of the two.

Here we seek to fill in a gap in our understanding of circuits with Toffoli and fanout gates. Roughly speaking, we know that given fanout, we can do fanin. Here we ask if, given fanin, we can do fanout. More precisely, we ask if generalized Toffoli gates and fanout gates are equivalent in power, up to polynomial size and constant depth.

In answering this question we find it necessary to grapple with another likely limitation of real quantum computers. It is evident that they will be limited not only in their run-time duration but also by the number of qubits used in the computation, due to the difficulty in controlling the interactions of multiple qubits. It will be necessary to restrict that number of work bits, or *ancillæ*, as much as possible. Thus we consider here the number of ancillæ that a circuit uses as an additional computational resource and investigate cases where this resource is limited as well.

To explain our results and motivations in greater detail, we review some of the history of this investigation. Moore [7] first observed that fanout gates and parity gates, in the presence of single qubit gates using 0 ancillæ, are equivalent up to depth 3. This was extended by Green et al. [4]: fanout is also equivalent to *any* MOD_q function (for $q \geq 2$), which determines if the number of 1's in the input is not divisible by q. Here the equivalence is again up to constant depth, but using $O(n)$ ancillæ. One may interpret this result by defining quantum circuit classes analogous to classical constant-depth circuit classes. For example, a reasonable analog of the classical unbounded fanin and fanout class AC^0 is QAC^0_{wf}, the class of constant depth quantum circuit families composed of single qubit, generalized Toffoli, and fanout gates. (Here the subscript "*wf*" denotes "with fanout.") Similarly one may define quantum analogs of $\mathsf{ACC}(q)$ (called $\mathsf{QACC}(q)$) and ACC (called QACC). Thus the equivalence of fanout with MOD_q implies that, for any $q > 2$, $\mathsf{QAC}^0_{wf} = \mathsf{QACC}(q) = \mathsf{QACC}$. Contrast this with the fact that $\mathsf{AC}^0 \neq \mathsf{ACC}$ (Furst, Saxe and Sipser [3]), and, for any distinct primes q, p, $\mathsf{ACC}(q) \neq \mathsf{ACC}(p)$ [9, 11]. Høyer and Špalek [5] have improved these results by proving these same QAC^0_{wf} circuits can compute threshold functions. Thus $\mathsf{QAC}^0_{wf} = \mathsf{QTC}^0_{wf}$, an even sharper contrast with the classical classes. Indeed, this result implies that we can approximate the quantum fast Fourier transform in constant depth using fanout. Thus the "quantum part" of Shor's renowned quantum factoring algorithm [10] can be carried out with a quite simple, constant depth quantum circuit that uses the fanout operator.

A theme running through these results is the following. In the classical setting, in the class AC^0, we take both fanout and fanin for granted. (This is not realistic from a practical point of view, since classical wires do have a thickness and can be fanned out only to a limited extent; but it is a useful abstraction.) But it is evident that, in the quantum setting, we cannot take either for granted. It is surprising indeed that the mere operation of fanning out a set of bits is universal, in the sense mentioned above, and can even be used to implement powerful fanin operations such as the threshold gate. Nothing remotely like this holds in the classical case. We are thus led to ask if the quantum world is even more different than one might think, and that if, given fanin, we can also do fanout. We may phrase the question in terms of complexity classes. In [4], the class QAC^0 is the class of circuit families containing single-qubit gates and unbounded fanin Toffoli gates. QAC^0_{wf} is the same class but with fanout gates allowed. In [4] it is observed that AC^0 and QAC^0 are different. Here we ask if QAC^0 and QAC^0_{wf} are different. We believe they are.

Our main result, proved in Section 4, is that one cannot compute parity (and hence fanout) with QAC^0 circuits using a constant number of ancillæ. This is the first hard evidence that QAC^0 and QAC^0_{wf} may be different, and that fanout may be necessary for all the upper bound results mentioned above (it certainly is if we limit our computations to only constantly many ancillæ).

The issue of the necessity of ancillæ in quantum computations is a murky one. It is generally accepted that a limited number (polynomially many relative to the number of inputs) are needed. This seems reasonable as it allows polynomially extra space in which to carry out a computation. However, it is possible to approximate any unitary operator with a small set of universal gates without ancillæ (although one apparently needs circuits of great depth and size in order to do so). Furthermore, to our knowledge, no systematic investigation into the absolute necessity of ancillæ for efficient quantum computation has been done. It is a natural question as the number of qubits needed to carry out a computation is likely to be a limiting factor in quantum computing. Ancillæ play a crucial role in the present result, in which we find the lower bound to be difficult to obtain when more than sublinearly many ancillæ are allowed.

To help clarify this problem, in Section 3 we provide a proof (implicitly stated in Cleve and Watrous [1]) that quantum circuits with gates of bounded size must be of log depth to compute parity (and hence fanout) exactly. In particular, we carefully address the problem of including ancillæ, and show that in this case the depth of the circuit must be $\log n$ to compute parity, no matter how many ancillæ are used. This proof serves as a revealing, though considerably simpler warm-up to our main theorem in Section 4. In this section we consider circuits which include Toffoli gates of unbounded size. It is easiest to see the log-depth lower bound in the case of zero ancillæ, so this result is given first, in Theorem 4.3. We then explain how the proof yields a depth/ancillæ trade-off, showing that with fewer ancillæ one needs greater depth to compute fanout.

We end with some open questions.

2 Preliminaries

In this section we set down most of our notational conventions and the circuit elements we use. Some acquaintance with quantum computational complexity as described in [8] or [6] is assumed.

The following notation and terminology will be convenient. Let $\mathcal{H}$ denote the 2-dimensional Hilbert space spanned by the computational basis states $|0\rangle, |1\rangle$. Let $\mathcal{H}_1, \ldots, \mathcal{H}_n$ be n copies of $\mathcal{H}$. By $\mathcal{B}_{\{1,\ldots,n\}}$ (or simply "$\mathcal{B}_n$" when the set notation is clearly understood) we denote the 2^n-dimensional Hilbert space $\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$ spanned by the usual set of computational basis states of the form $|x_1, \ldots, x_n\rangle$, where each $x_i \in \{0, 1\}$. We also consider "quotient spaces of $\mathcal{B}_{\{1,\ldots,n\}}$ over m bits," defined as $\mathcal{B}_{\{i_1,\ldots,i_m\}} = \mathcal{H}_{i_1} \otimes \cdots \otimes \mathcal{H}_{i_m}$, where $1 \le i_1 < \cdots < i_m \le n$, which obviously have dimension 2^m. A "state over a set of m bits" is a state in such a quotient space. A quantum gate G corresponds to a unitary operator (also denoted G) acting on some quotient space $\mathcal{B}_{\{i_1,\ldots,i_m\}}$ of $\mathcal{B}_n$. We will say that G *involves* the bits $i_1, \ldots, i_m$. We will freely identify G with any "extension by the identity" that acts on a bigger quotient space $\mathcal{B}_A$ for any set of bits $A \supseteq \{i_1, \ldots, i_m\}$, that is, G can be identified with the operator $G \otimes I$, where I is the identity on $\mathcal{B}_{A-\{i_1,\ldots,i_m\}}$. If we fix a state $|\Psi_m\rangle$ over m bits $\{i_1, \ldots, i_m\}$, we are effectively restricting $\mathcal{B}_{\{1,\ldots,n\}}$ to the 2^{n-m}-dimensional linear subspace $|\Psi_m\rangle \otimes \mathcal{B}_{\{1,\ldots,n\}-\{i_1,\ldots,i_m\}}$. The space $\mathcal{B}_{\{1,\ldots,n\}-\{i_1,\ldots,i_m\}}$ is referred to as the quotient space of $\mathcal{B}_{\{1,\ldots,n\}}$ *complementary* to $|\Psi_m\rangle$.

A *single-qubit gate* is a 2×2 unitary matrix (e.g., acting in $\mathcal{B}_{\{1\}}$). For example, the Hadamard gate H is the single-qubit gate,

$$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

A *generalized Toffoli gate*, which we refer to in this paper as simply a *Toffoli* gate T, transforms computational basis states as follows:

$$T|x_1, ..., x_n, b\rangle = |x_1, ..., x_n, b \oplus \wedge_{i=1}^n x_i\rangle$$

A *generalized Z-gate*, which we refer to as a *Z-gate* for brevity, has the following effect:

$$Z|x_1, ..., x_n, b\rangle = (-1)^{\bigwedge_{i=1}^n x_i \wedge b}|x_1, ..., x_n, b\rangle.$$

It is not hard to show that, $T = H_b Z H_b$ where the Hadamard gate H_b in this equation is applied to the target bit b of T, and the Z-gate is acting on $|x_1, ..., x_n, b\rangle$. Hence we may substitute Z-gates for T-gates in any circuit that allows Hadamards (which will be true throughout the paper). Z-gates are useful for our purposes since they are bosonic (that is, completely symmetric over their bits), and thus have no preferred target bit.

The *fanout* gate F and the *parity* gate P are defined, respectively, by

$$\begin{aligned} F|x_1, ..., x_n, b\rangle &= |b \oplus x_1, ..., b \oplus x_n, b\rangle, \\ P|x_1, ..., x_n, b\rangle &= |x_1, ..., x_n, b \oplus \bigoplus_{i=1}^n x_i\rangle. \end{aligned}$$

There is no obvious *a priori* relation between these operators, but as was observed by Moore, F is conjugate to P via an $(n+1)$-fold tensor product of Hadamards applied to all the bits:

$$F = H^{\otimes(n+1)} P H^{\otimes(n+1)} \tag{1}$$

Recall that Hadamard, phase, CNOT (Toffoli gates for $n = 1$), and $\pi/8$ gates are a universal set of gates in that any unitary operator can be approximated to an arbitrary degree of precision with them. Our lower bound techniques work against *arbitrary* single-qubit gates combined with Z-gates, which together also form a universal set by the above discussion.

A quantum circuit is constructed out of layers. Each layer L is a tensor product of a certain fixed set of gates (in our main theorems, these will consist of single-qubit and Z-gates). A circuit is simply a (matrix) product of layers $L_1 L_2 \cdots L_d$. (Observe that the "last" layer L_d is actually the one that is applied directly to the inputs, and L_1 is the output layer.) The number of layers d is called the *depth* of C. A circuit C over n qubits is then a unitary operator in the 2^n-dimensional Hilbert space $\mathcal{B}_{\{1,...,n\}}$. Clearly, C computes a unitary operator U *exactly* if for all computational basis states, $C|x_1, ..., x_n\rangle = U|x_1, ..., x_n\rangle$. This is in general too restrictive, however. One must allow for the presence of "work bits," called *ancillæ*, that make extra space available in which to do a computation. In that case, in order to exactly compute the operator U we extend the Hilbert space in which C acts to the 2^{n+m}-dimensional space spanned by computational basis states $|x_1, ..., x_n, a_1, ..., a_m\rangle$, where again $x_i, a_i \in \{0, 1\}$, the a_i serving as ancillæ. Then we say that C *cleanly computes* U if, for any $x_1, ..., x_n$ and $y_1, ..., y_n$,

$$\langle y_1, ..., y_n, 0, ..., 0|C|x_1, ..., x_n, 0, ..., 0\rangle = \langle y_1, ..., y_n, 0, ..., 0|(U \otimes I)|x_1, ..., x_n, 0, ..., 0\rangle,$$

where I is the identity in the subspace that acts on the ancillæ, and the number of 0's in each state above is m. That is, C does a clean computation if the ancillæ begin and end all as 0's. We assume all of our circuits perform clean computations. This is a reasonable constraint, since only then is it easy to compose the circuits.

Lastly, all circuits should be understood to be elements of an infinite *family* of circuits $\{C_n | n \geq 0\}$, where C_n is a quantum circuit for n input qubits.

3 Fanout Requires Log Depth with Bounded Size Gates

It is easy to see that, by an obvious divide-and-conquer strategy, we can compute parity in depth $O(\log n)$ using just CNOT gates and 0 ancillæ. In this section we prove that $\Omega(\log n)$ bits are required for circuits with any bounded size multi-qubit gates, and furthermore that no number of ancillæ help to reduce the depth of the circuit. Note that to perform a clean computation, the target bit must be copied and the original circuit run backward. We can thus get an exact upper bound on the depth of a circuit that computes parity of $2 \log n + 1$. In the following section we obtain a lower bound of about $1.44 \log n$ in the case of a constant number of ancillæ; the techniques in this section do not seem sufficient to determine the exact multiplicative factor.

The intuition behind the proof of the next Lemma seems quite obvious. Namely, if a depth d circuit is composed of only one- and two- qubit gates, then any output qubit of the circuit can depend on at most 2^d input qubits. Furthermore, the Lemma and following Theorem provide formal verification of the intuitive fact that a quantum circuit must connect all the qbits on on which its output depends to the qubit we will measure for the output. However, as is often the case in this field, a formal proof of this fact is less obvious than first appears, and the techniques we use here form the basis for the proof of the lower bound theorem of the next section.

Let $C = L_1 \cdots L_d$ consist entirely of arbitrary two-qubit gates and single-qubit gates. (The extension to arbitrary, but fixed, size gates is straightforward.) Further suppose that M is an observable on a single qubit in the last layer. Let L_1' denote the gate whose output M is measuring. L_1' could be a two-qubit or a single-qubit gate. In either case, $L_1 = L_1' \otimes R_1$, where R_1 is the tensor product of all the other gates in that layer, if any. More generally, we decompose layer i similarly, writing $L_i = L_i' \otimes R_i$, where L_i' is a transformation that acts on some subset of the bits, and R_i acts on the rest.

Lemma 3.1. For each d, there are layers $L_1', ..., L_d'$ such that

$$L_d^\dagger L_{d-1}^\dagger \cdots L_1^\dagger M L_1 \cdots L_{d-1} L_d = L'^\dagger_d L'^\dagger_{d-1} \cdots L'^\dagger_1 M L'_1 \cdots L'_{d-1} L'_d$$

where, for each i, L_i' acts on at most 2^i bits. Furthermore, for each i, L'_i acts on bits with indices in some set S_i such that $S_d \supseteq S_{d-1} \supseteq ... \supseteq S_1$.

Figure 1 makes the notation a little clearer. Note that the *input* will, as usual, be on the *left*, but it doesn't enter the claim (or the following argument) at all.

Proof: The proof of Lemma 3.1 is by induction on d. First consider $d = 1$. Then consider the operator $L_1^\dagger M L_1$. By the observations above, we may write $L_1 = L_1' \otimes R_1$, where L_1' is either a single or two-qubit gate. So,

$$L_1^\dagger M L_1 = (L_1'^\dagger \otimes R_1^\dagger) M (L_1' \otimes R_1) = L_1'^\dagger M L_1',$$

by virtue of the fact that M and R_1 commute. Since L'_1 only depends on ≤ 2 qubits, this establishes the result for $d = 1$.

Now suppose that we can write,

$$L_d^\dagger L_{d-1}^\dagger \cdots L_1^\dagger M L_1 \cdots L_{d-1} L_d = L'^\dagger_d L'^\dagger_{d-1} \cdots L'^\dagger_1 M L'_1 \cdots L'_{d-1} L'_d$$

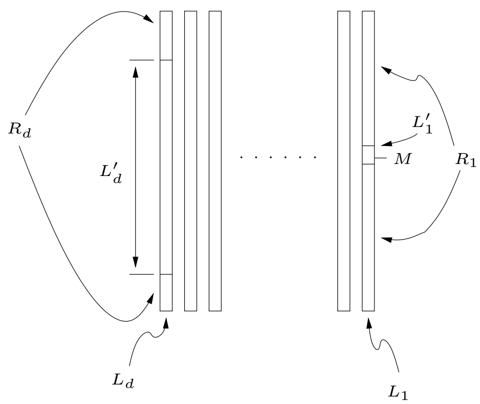

Figure 1: Decomposition of the layers of the circuit C.

where, for each i, L'_i acts on at most 2^i bits. In particular, note that L'_d acts on at most 2^d bits. Suppose that L'_d acts on indices in the set S_d (where S_d has size $\leq 2^d$). Now by the induction hypothesis,

$$L^\dagger_{d+1}L^\dagger_d \cdots L^\dagger_1 M L_1 \cdots L_d L_{d+1} = L^\dagger_{d+1}{L'}^\dagger_d \cdots {L'}^\dagger_1 M {L'}_1 \cdots {L'}_d L_{d+1},$$

and $S_d \supseteq S_{d-1} \supseteq \ldots \supseteq S_1$.

The gates in L'_d involve at most the bits in S_d. Since the circuit only contains at most two-qubit gates, all the gates in L_{d+1} involving bits in S_d can act on at most 2^{d+1} bits. Let the tensor product of these gates be denoted by L'_{d+1}, and let S_{d+1} denote the set of bits on which L'_{d+1} acts. Clearly $S_{d+1} \supseteq S_d$. Then for some tensor product of single and two-qubit gates R_{d+1} we may write $L_{d+1} = L'_{d+1} \otimes R_{d+1}$. Since R_{d+1} acts on bits not in S_{d+1}, it commutes with all the ${L'}_i$ and M, which only act on bits inside S_{d+1}. Hence R_{d+1} "cancels out" and we have the desired relation. $\square$

Theorem 3.2. Let C be a quantum circuit on n inputs of depth d, consisting of single-qubit and two-qubit gates, with any number of ancillæ that cleanly computes parity exactly. Then $d \geq \log n$. If C computes fanout in the same way, then $d \geq \log n - 2$.

Note that, the following proof shows that this theorem applies not only to parity but to any function which depends on all of its input bits.

Proof: Let $C = L_1 \cdots L_d$ as in Lemma 3.1. Suppose C uses m ancillæ, and that it cleanly computes the parity operator P in depth $d < \log n$. It follows that for any $x_1, ..., x_n, b$ and any measurement operator M on the target bit,

$$\langle x_1, ..., x_n, b, 0, ..., 0|C^\dagger M C|x_1, ..., x_n, b, 0, ..., 0\rangle = \langle x_1, ..., x_n, b|PMP|x_1, ..., x_n, b\rangle. \quad (2)$$

By Lemma 3.1,

$$C^\dagger M C = L^\dagger_d L^\dagger_{d-1} \cdots L^\dagger_1 M L_1 \cdots L_{d-1} L_d = {L'}^\dagger_d {L'}^\dagger_{d-1} \cdots {L'}^\dagger_1 M {L'}_1 \cdots {L'}_{d-1} {L'}_d,$$

where the operator ${L'}_1 \cdots {L'}_d$ acts on at most 2^d inputs. Since $2^d < n$, there is an input on which that operator does not act. Hence the value on the left hand side of eq. (2) remains unchanged if we can flip some x_i. However, the outcome of the measurement on the parity gate on the right hand side depends on every input, which is a contradiction.

The second assertion in the Theorem follows from eq. (1). $\square$

It is clear that if we have a family of circuits that use a fixed set of multi-qubit gates with arity independent of n, that a similar proof will work. Thus we have the following as a corollary of the proof of Theorem 3.2:

Corollary 3.3. Let C be a quantum circuit on n inputs of depth d, consisting of single-qubit and multi-qubit gates of size $O(1)$, with any number of ancillæ, that cleanly computes parity, or fanout, exactly. Then $d = \Omega(\log n)$.

This same proof technique can be used to establish that constant depth circuit families with gates of bounded arity are not capable of simulating Toffoli gates. Thus we have,

Corollary 3.4. Let C be a quantum circuit on n inputs of depth d, consisting of single-qubit and multi-qubit gates of size $O(1)$, with any number of ancillæ, that cleanly and exactly computes the Toffoli gate of arity n. Then $d = \Omega(\log n)$.

4 Parity Requires Log Depth with Few Ancillæ

In this section we treat circuits that contain Toffoli gates or, equivalently, Z-gates of arbitrary size (i.e., that can depend on n), and consider whether such circuits can compute parity. The technique of the preceding section does not work in this case. This is because the large gates in general do not cancel, since they may not commute with the measurement operator M.

To see how to proceed, it is useful to briefly consider classical circuits with similar constraints. Suppose we have a classical circuit with NOT gates and unbounded fan-in AND and OR gates, but that we do *not* allow any fanout. Once inputs (or outputs of other gates) are used in either an AND or an OR gate, they can not be used again. It is obvious that if such a circuit has constant depth, it cannot compute such functions as parity. The AND and OR gates can be killed off by restricting a small set of inputs, resulting in a constant function, while parity depends on all the inputs.

In the quantum case, it appears again that the only thing to do is to attempt to "kill off" the large Toffoli gates. However, the quantum case is much more subtle since we must face the fact that intermediate states are a superposition of computational basis states, and furthermore that the Z-gates, in combination with the single-qubit gates, may cause entanglement.

As before, write $C = L_1 L_2 \cdots L_d$. Thus the circuit C transforms the state $|\Psi\rangle$ to $L_1 \cdots L_d|\Psi\rangle$. We assume without loss of generality that each layer L_i is a tensor product of Z-gates and single-qubit gates. Further assume without loss of generality that a specific bit (say, the n^{th} bit) of C serves as the output or *target* bit (which eventually is supposed to agree with the output bit of a parity gate).

Our main technical lemma is easiest to see in the case that C has *no ancillæ*, which we assume until later in the section.

Lemma 4.1. Let C be a circuit with n inputs as described above, with no ancillæ. Then for each $1 \le k \le d$, there exists a state $|\Psi_k\rangle$ over at most 2^k bits such that for any state $|R\rangle$ in the quotient space of $\mathcal{B}_n$ complementary to $|\Psi_k\rangle$, the state $L_1 L_2 \cdots L_k(|R\rangle \otimes |\Psi_k\rangle)$ has a 0 in the target position of C.

Proof: The proof is by induction on k. First let $k = 1$. There are two cases:

1. In layer L_1, the target is the output of a single-qubit gate S. Then let the state $|\Psi_1\rangle = S^\dagger|0\rangle$ over the n^{th} bit. Now we may write $L_1 = L'_1 \otimes S$, where L'_1 acts on the quotient space $\mathcal{R}$ complementary to $|\Psi_1\rangle$. No matter what state $|R\rangle \in \mathcal{R}$ we choose over the bits $\{1, \ldots, n-1\}$, it follows that $L_1(|R\rangle \otimes |\Psi_1\rangle) = (L'_1|R\rangle) \otimes (S|\Psi_1\rangle) = (L'_1|R\rangle) \otimes |0\rangle$ is a state where the n^{th} qubit of the tensor product is unentangled with the rest of the circuit and is in state $|0\rangle$.

2. In layer L_1, the target is the output of a Z-gate. Write $L_1 = L'_1 \otimes G$, where G is this Z gate. In this case, we choose $|\Psi_1\rangle = |0\rangle$ over the n^{th} bit. Now G acts both on $|\Psi_1\rangle$ as well as the complementary quotient space $\mathcal{R}$ (via extension by the identity). But since G involves a bit that is 0 (i.e., the n^{th} bit), G is equivalent to the unit matrix in $\mathcal{R}$. Hence for any state $|R\rangle \in \mathcal{R}$, $L_1(|R\rangle \otimes |\Psi_1\rangle) = (L'_1 \otimes G)(|R\rangle \otimes |\Psi_1\rangle) = (L'_1|R\rangle) \otimes |0\rangle = (L'_1|R\rangle) \otimes |\Psi_1\rangle$ again is a state where the n^{th} qubit of the tensor product is unentangled with the rest of the circuit and is in state $|0\rangle$. (Note that $L'_1|R\rangle$ is well defined by extending L'_1 by the identity.)

Now suppose the assertion is true for $k-1$ where $k > 1$. We will show that it remains true for k. Suppose the $|\Psi_{k-1}\rangle$ in the assertion is a state over the (at most) 2^{k-1} bits in the set K_{k-1}. Let R_{k-1} denote the rest of the bits $\{1, \ldots, n\} - K_{k-1}$. Thus $|\Psi_{k-1}\rangle$ is a state in $\mathcal{B}_{K_{k-1}}$, and the quotient space complementary to $|\Psi_{k-1}\rangle$ is $\mathcal{B}_{R_{k-1}}$, which for convenience we denote by $\mathcal{R}_{k-1}$. We specify the state $|\Psi_k\rangle$ as follows: Start with $K_k := K_{k-1}$ and $R_k := R_{k-1}$. If a Z-gate G in L_k involves bits both in K_k and in R_k, we remove a *single* bit from R_k involved with G, add it to K_k, declare the gate G *killed*, and remove G from further consideration. Continue until all such Z-gates have been killed. Since each bit in K_{k-1} can be involved with at most one Z-gate in L_k, the number of bits added to K_k (and removed from R_k) in this process is at most 2^{k-1}. Let $L_k^{(K)}$ denote the gates in L_k that involve the bits in K_k, excluding the Z-gates that have been killed. Then finally, we define the state $|\Psi_k\rangle$ as $L_k^{(K)\dagger}$ applied to the tensor product of $|\Psi_{k-1}\rangle$ with the state in which all the bits in $K_k - K_{k-1}$ are 0.

Note that $|\Psi_k\rangle$ is a state over at most $2 \cdot 2^{k-1} = 2^k$ bits, as seen in Figure 2. Let $\mathcal{R}_k$ denote the quotient space complementary to $|\Psi_k\rangle$. Clearly, $\mathcal{R}_k = \mathcal{B}_{R_k}$. Now let $|R\rangle$ be any state in $\mathcal{R}_k$ (equivalently, over the bits in R_k), and apply L_k to $|R\rangle \otimes |\Psi_k\rangle$. Let $L_k^{(R)}$ denote the gates in L_k acting in $\mathcal{R}_k$, again excluding the Z-gates that have been killed. Note that any Z-gate in layer L_k that involves bits in K_k as well as R_k acts as the identity on $\mathcal{R}_k \otimes |\Psi_k\rangle$, by the construction of $|\Psi_k\rangle$. Thus we have eliminated these gates from L_k without any loss of generality. Thus,

$$L_k(|R\rangle \otimes |\Psi_k\rangle) = (L_k^{(R)} \otimes L_k^{(K)})(|R\rangle \otimes |\Psi_k\rangle) = (L_k^{(R)}|R\rangle) \otimes (L_k^{(K)}|\Psi_k\rangle).$$

Now $L_k^{(K)}|\Psi_k\rangle$ is the tensor product of $|\Psi_{k-1}\rangle$ with a number of $|0\rangle$ states. So we conclude that $L_k(|R\rangle \otimes |\Psi_k\rangle)$ is of the form $|R'\rangle \otimes |\Psi_{k-1}\rangle$ for some state $|R'\rangle \in \mathcal{R}_{k-1}$. Then,

$$L_1 L_2 \cdots L_{k-1} L_k(|R\rangle \otimes |\Psi_k\rangle) = L_1 L_2 \cdots L_{k-1}(|R'\rangle \otimes |\Psi_{k-1}\rangle).$$

By the induction hypothesis, the right hand side of the above equation has a 0 target bit, which proves the lemma. $\square$

With a bit more careful analysis, Lemma 4.1 can be improved to the following:

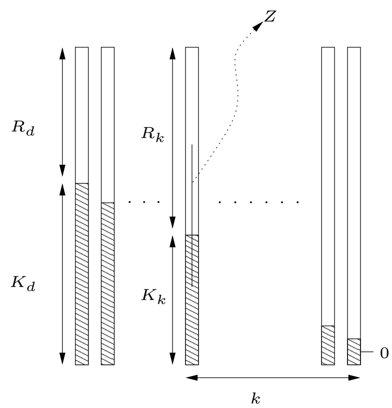

Figure 2: The sets K_k and R_k. A Z gate that involves bits in both sets is shown.

Lemma 4.2. Let C be a circuit as described above and $\phi = (1+\sqrt{5})/2$ be the golden ratio. Then for each $1 \le k \le d$, there exists a state $|\Psi_k\rangle$ over at most ϕ^{k+1} bits such that for any state $|R\rangle$ in the quotient space of $\mathcal{B}_n$ complementary to $|\Psi_k\rangle$, the state $L_1 L_2 \cdots L_k(|R\rangle \otimes |\Psi_k\rangle)$ has a 0 in the target position of C.

Remark. The difference is that now $|\Psi_k\rangle$ is over only ϕ^{k+1} bits instead of 2^k (recall $\phi \cong 1.62$). This results in a tighter lower bound on the depth.

Proof: We indicate how the proof of Lemma 4.1 is modified to obtain this result.

We claim that the size of the state $|\Psi_k\rangle$ that we create in step k of the induction is at most f_{k+1} where $\{f_k\}$ is the Fibonacci sequence.

To see this, observe that when some bit (the i^{th} bit, say) is moved from R_{k-1} to K_{k-1}, it is set to the $|0\rangle$ state in step $k-1$ of the induction. Hence that bit becomes a bit in K_{k-1} that is set to $|0\rangle$. Consider the gate G (if any) in L_k that involves this bit. If G is a single-qubit gate, then no Z-gate in L_k is killed involving the i^{th} bit, so no additional bit needs to be added to K_k for the sake of the i^{th} bit. If G is a Z-gate, then the i^{th} bit alone is enough to kill G, since this bit is already 0. So again, no additional bit must be added to K_k to kill G. This shows that it takes more than one step of the induction to double the number of bits in K_k.

Let r_{k-1} denote the number of bits that were moved from R_{k-1} to K_{k-1} (and hence set to the $|0\rangle$ state). Similarly, let ℓ_{k-1} denote the number of bits in K_{k-1} that were already in K_{k-1} (by virtue of their being in K_{k-2}, if $k > 2$). Thus the size of the state $|\Psi_{k-1}\rangle$ is $r_{k-1} + \ell_{k-1}$. For the base case, we start counting at $k = 1$. Note that after the first layer, $r_1 = 0$ (since the target bit was set to 0 and sent through either a Z-gate or a single qubit gate, so no bits need to be set to 0), and $\ell_1 = 1$. For the induction step, from the preceding argument, the only bits that might require that new bits be set to 0 are the ℓ_{k-1} bits that were not set to 0 in step $k-1$. We thus arrive at the following inequalities:

$$\begin{aligned} r_k &\le \ell_{k-1} \\ \ell_k &\le r_{k-1} + \ell_{k-1} \end{aligned}$$

It is easy to see by induction that $r_k \le f_{k-1}$ and $\ell_k \le f_k$. Thus $r_k + \ell_k \le f_{k-1} + f_k = f_{k+1}$, which proves the above claim.

Since $f_{k+1} \leq \phi^{k+1}$, the lemma follows.

$\square$

Theorem 4.3. Let C be a depth d circuit of n inputs, consisting of single-qubit gates and Z-gates, and using 0 ancillæ. If $d < \frac{\log n}{\log \phi} - 1 \cong 1.44 \log n - 1$, then C cannot compute P, the parity gate with $n-1$ inputs and one target.

Proof: Suppose $C = P$. Then for any input state, the target bit of C is 0 iff the target bit of P is 0. By Lemma 4.2, there exists a state $|\Psi\rangle$ on at most $\phi^{d+1} < n$ bits such that, for any state $|R\rangle$ on the remaining $n - \phi^{d+1}$ bits, $C(|R\rangle \otimes |\Psi\rangle)$ has a 0 value for the target. First let $|R\rangle$ be the state with 0's in all $n - \phi^{d+1}$ positions (since $n - \phi^{d+1} > 0$, such positions exist). Then $P(|R\rangle \otimes |\Psi\rangle)$ has a 0 target. This is only possible if the state $|\Psi\rangle$ is in a quotient space of $\mathcal{B}_n$ spanned by computational basis states in which an even number of the variables are 1. Now change one of the bits of $|R\rangle$ from 0 to 1. The target of $C(|R\rangle \otimes |\Psi\rangle)$ still has the value 0, but the target of $P(|R\rangle \otimes |\Psi\rangle)$ must change to 1, which contradicts the assumption that $C = P$. $\square$

This Theorem applies more broadly than stated here. It gives a circuit depth lower bound of at least $1.44 \log m - 1$ for any function with the property that, for any input string x, there is a set of m bits such that flipping any one of them changes $f(x)$. Since fanout and parity are equivalent up to depth 3 (with 0 ancillæ), we immediately have the following.

Corollary 4.4. Let C be a circuit of depth d consisting of single-qubit gates and Z-gates, and using 0 ancillæ. Then, if $d < \frac{\log n}{\log \phi} - 3$, C cannot compute the fanout operation.

The proof of Theorem 4.3 actually yields a stronger result: the circuit C cannot even approximate P within distance $1/\sqrt{2}$ in the operator norm.

Theorem 4.5. Let C be a circuit of depth d consisting of single-qubit gates and Z-gates, using 0 ancillæ. If $d < \frac{\log n}{\log \phi} - 1$, then $\|P - C\| \geq \frac{1}{\sqrt{2}}$, where $\| \cdot \| = \| \cdot \|_\infty$ is the operator norm.

Proof: Let C be such a circuit, and let $|\Psi\rangle$ be as in the proof of Theorem 4.3. Let $|R_0\rangle$ be the all-0 state on the remaining $n - \phi^{d+1}$ qubits, which do not include the target qubit n, and let $|R_1\rangle = X_i|R_0\rangle$, where X_i is the bit-flip gate applied to one of the qubits of $|R_0\rangle$. Thus $|R_1\rangle$ is the same as $|R_0\rangle$ but with some qubit i flipped to 1. Set $|R\rangle = (|R_0\rangle + |R_1\rangle)/\sqrt{2} = (I + X_i)|R_0\rangle/\sqrt{2}$. Then

$$C(|R\rangle \otimes |\Psi\rangle) = |S\rangle \otimes |0\rangle \tag{3}$$

by the definition of $|\Psi\rangle$, where $|S\rangle$ is some state of the $n-1$ nontarget qubits. Also, note that $PX_i = X_n X_i P$, where X_n is the bit-flip gate on the target qubit n. Letting P' be the parity gate applied to the qubits of $|\Psi\rangle$, and setting $P'|\Psi\rangle = |T_0\rangle \otimes |0\rangle + |T_1\rangle \otimes |1\rangle$ for some (nonnormalized) states $|T_0\rangle$ and $|T_1\rangle$ of the nontarget qubits of $|\Psi\rangle$, we have

$$\begin{aligned} P(|R\rangle \otimes |\Psi\rangle) &= \frac{P(I + X_i)(|R_0\rangle|\Psi\rangle)}{\sqrt{2}} && (4) \\ &= \frac{(I + X_n X_i)P(|R_0\rangle|\Psi\rangle)}{\sqrt{2}} && (5) \end{aligned}$$

$$= \frac{(I + X_n X_i)(|R_0\rangle P'|\Psi\rangle)}{\sqrt{2}} \tag{6}$$

$$= \frac{(I + X_n X_i)(|R_0\rangle(|T_0\rangle|0\rangle + |T_1\rangle|1\rangle))}{\sqrt{2}} \tag{7}$$

$$= \frac{(|R_0\rangle|T_0\rangle + |R_1\rangle|T_1\rangle)|0\rangle + (|R_0\rangle|T_1\rangle + |R_1\rangle|T_0\rangle)|1\rangle}{\sqrt{2}}. \tag{8}$$

Combining (3) with (4–8), we get

$$(P - C)(|R\rangle \otimes |\Psi\rangle) = |U\rangle|0\rangle + \frac{(|R_0\rangle|T_1\rangle + |R_1\rangle|T_0\rangle)|1\rangle}{\sqrt{2}}$$

for some unnormalized state $|U\rangle$. Since the two terms on the right-hand side are orthogonal, and also $\langle R_0 \mid R_1\rangle = 0$, we have

$$\begin{aligned}
\|P - C\| &\geq |(P - C)(|R\rangle \otimes |\Psi\rangle)| \\
&\geq \left|\frac{|R_0\rangle|T_1\rangle + |R_1\rangle|T_0\rangle}{\sqrt{2}}\right| \\
&= \sqrt{\frac{\langle T_0 \mid T_0\rangle^2 + \langle T_1 \mid T_1\rangle^2}{2}} \\
&= \frac{|P'|\Psi\rangle|}{\sqrt{2}} \\
&= \frac{1}{\sqrt{2}}.
\end{aligned}$$

□

We now consider the case in which our circuit has a non-zero number of ancillæ. Firstly, it is clear that Lemmas 4.1 and 4.2 work if we set a target and all ancillæ to 0 at the same time. If there are a many ancillæ, then we are setting $a + 1$ "outputs." The conclusion of the analogous Lemma for a ancillæ would then be that the state $|\Psi\rangle$ is over $(a+1)\phi^{d+1}$ bits (since the number of "committed" bits increases like the Fibonacci sequence with each layer, as in Lemma 4.2). These bits may include ancillæ, and assuming that C does a clean computation, $|\Psi\rangle$ will be 0 on the ancillæ (since they must all start out as 0 in order to return to their final value of 0). Therefore, if $n > (a+1)\phi^{d+1}$, the state $|R\rangle$ is over at least one bit but no ancillæ and is thus free to take on any value. Thus if $n > (a+1)\phi^{d+1}$, the output of C is insensitive to changes in at least one of the inputs, and hence the circuit is defeated as before. Note we have a depth/ancillæ trade-off as a result. We thus have the following corollary of the proof of Theorem 4.3:

Corollary 4.6. Let C be a circuit of depth d consisting of single-qubit gates and Z-gates. Then, if C cleanly computes the parity function with a ancillæ, then $d \geq (\log \phi)^{-1} \log(n/(a+1)) - 1 \cong 1.44 \log(n/(a+1)) - 1$.

As mentioned at the beginning of Section 3.1, one can compute parity in depth $2 \log n + 1$. We conjecture that this is optimal no matter what a is, so Corollary 4.6 leaves considerable room for improvement.

We offer an alternative interpretation of our result that arose out of conversations with Luc Longpré. Let us say that a quantum circuit C *robustly computes* a unitary operator U if C computes

U cleanly and, in addition, if its output is insensitive to the initial state of the ancillæ. Thus the ancillæ of C can start out in any state whatsoever; the circuit C is guaranteed to return the ancillæ to that state in the end, and always gives the same answer. This of course puts a much stronger constraint on the circuit (since in the usual model we only insist on a clean computation when the ancillæ are initialized to 0), but such circuits can be useful (e.g., see exercise 8.5 in Kitaev et al. [6]). It is not hard to see that in this case, if C consists only of single-qubit and Z-gates, then it must have depth about $1.44 \log n$ to compute parity, regardless of the number of ancillæ.

5 Conclusions and Open Problems

Following the line of earlier work of Green et al., Høyer and Špalek, and Cleve and Watrous [4, 5, 1], our main result gives an optimal, $O(\log n)$ lower bound on the depth of QAC-type circuits computing fanout, in the presence of limited (slightly sublinear) numbers of ancillæ. It would clearly be desirable to extend our result to obtain the same conclusion when polynomially many (or an unlimited number of) ancillæ are allowed, and thus to prove that $\mathsf{QAC}^0 \neq \mathsf{QAC}^0_{wf}$.

The role of ancillæ in quantum computation has not received much detailed attention. Prompted by our considerations here, there are several interesting questions that arise. One issue is the necessity of ancillæ for specific quantum computations or classes of quantum computations. Is there a problem that can be done in constant depth with ancillæ but which requires $\log n$ depth without ancillæ? Similarly, are there computational problems for which $\log n$ depth is possible with ancillæ but without ancillæ, polynomial depth is needed? In general, how many ancillæ are needed for specific problems? Is there a general tradeoff that can be proved between numbers of ancillæ and circuit depth?

While much has recently been learned concerning constant depth circuit classes, interesting questions remain. It would be worthwhile to be able to distinguish between the power of quantum gates of unbounded arity. We have seen that Toffoli and Z gates (which are equivalent up to constant depth) are weaker than parity and fanout (which are equivalent not only to each other but also, for all intents and purposes, to other mod gates, threshold gates and the quantum Fourier transform). Are there other natural types of gates that lie between these two classes, or is every gate either equivalent, up to constant depth, to either single qubit and CNOT gates, or to Toffoli gates, or to parity? It would also be of interest to characterize exactly what can be computed in constant depth using only single qubit and CNOT gates as, even from an optimistic point of view, this is the kind of circuit that might be built in the not too distant future. A further study of these limited quantum circuits can be found in Fenner et al [2]. Finally, there is a locality to quantum interactions which makes it difficult to implement gates which act on qubits outside of a small local neighborhood of each other. It would be interesting to consider non-locality as a resource and investigate which computations can be carried out by circuits with gates acting only on a neighborhood of “nearby” qubits.

6 Acknowledgements

We thank Luc Longpré for helpful discussions and comments on this paper. This work was supported in part by the National Security Agency (NSA) and Advanced Research and Development Agency (ARDA) under Army Research Office (ARO) contract numbers DAAD 19-02-1-0058 (for M. Fang, S. Homer, and F. Green) and DAAD 19-02-1-0048 (for S. Fenner and Y. Zhang).

References

[1] R. Cleve and J. Watrous, "Fast parallel circuits for the quantum Fourier transform," *Proceedings of the 41st Annual Symposium on Foundations of Computer Science* (2000), 526–536.

[2] S. Fenner, F. Green, S. Homer and Y. Zhang, "Bounds on the Power of Constant Depth Quantum Circuits," In Los Alamos Preprint Archives (2003), quant-ph/0312209.

[3] M. Furst, J.B. Saxe, and M. Sipser, "Parity, circuits, and the polynomial-time hierarchy." *Math. Syst. Theory* **17** (1984) 13–27.

[4] F. Green, S. Homer, C. Moore and C. Pollett, "Counting, Fanout and the Complexity of Quantum ACC," *Quantum Information and Computation* **2** (2002) 35–65.

[5] P. Høyer and R. Špalek, "Quantum circuits with unbounded fan-out," *20th STACS Conference, 2003*, LNCS 2607, 234–246.

[6] A. Yu. Kitaev, A. H. Shen, and M. N. Vyalyi, Classical and Quantum Computation, American Mathematical Society, 2002.

[7] Cristopher Moore. Quantum Circuits: Fanout, Parity, and Counting. In Los Alamos Preprint archives (1999), quant-ph/9903046.

[8] M. A. Nielsen and I. L. Chuang, Quantum Computation and Quantum Information, Cambridge University Press, 2001.

[9] A. A. Razborov, Lower bounds on the size of bounded depth networks over a complete basis with logical addition, *Matematicheskie Zametki* **41** (1987) 598-607. English translation in *Mathematical Notes of the Academy of Sciences of the USSR* **41** (1987) 333-338.

[10] P. W. Shor, "Polynomial-time algorithms for prime number factorization and discrete logarithms on a quantum computer." *SIAM J. Computing* **26** (1997) 1484–1509.

[11] R. Smolensky, Algebraic methods in the theory of lower bounds for Boolean circuit complexity, in *Proceedings of the 19th Annual ACM Symposium on Theory of Computing* (1987) 77-82.

[12] X. Wang, A. Sorenson and K. Molmer, "Multi-bit gates for quantum computing", *Phys. Rev. Lett.* 86 (17) (2001)3907-3910.