Quantum advantage of unitary Clifford circuits with magic state inputs

Mithuna Yoganathan, Richard Jozsa and Sergii Strelchuk

DAMTP, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, U.K.

Abstract

We study the computational power of unitary Clifford circuits with solely magic state inputs (CM circuits), supplemented by classical efficient computation. We show that CM circuits are hard to classically simulate up to multiplicative error (assuming PH non-collapse), and also up to additive error under plausible average-case hardness conjectures. Unlike other such known classes, a broad variety of possible conjectures apply. Along the way we give an extension of the Gottesman-Knill theorem that applies to universal computation, showing that for Clifford circuits with joint stabiliser and non-stabiliser inputs, the stabiliser part can be eliminated in favour of classical simulation, leaving a Clifford circuit on only the non-stabiliser part. Finally we discuss implementational advantages of CM circuits.

1 Introduction

A fundamental goal of quantum complexity theory is to prove that quantum computers cannot be efficiently simulated by classical computers. An approach to proving this was put forward by Bremner et al. [9], showing that if a particular class of quantum circuits, so-called IQP circuits, could be efficiently classically simulated up to multiplicative error then the polynomial hierarchy (PH) would collapse. However on physical grounds it is more natural to consider classical simulations with additive or $l_1$ error. In this vein, Aaronson and Arkhipov [1] showed that assuming the validity of two plausible complexity theoretic conjectures, the quantum process of boson sampling cannot be efficiently simulated up to additive error unless there is PH collapse. The conjectures are referred to as the anticoncentration conjecture and average-case hardness conjecture. Bremner, Montanaro and Shepherd [11] showed a similar result for IQP circuits, and furthermore they were able to prove the anticoncentration conjecture in their context. Since then, there have been further similar results for various classes [21, 6, 23, 20, 19].

In this paper we introduce a subclass of quantum computing that we call Clifford Magic (CM), inspired by the PBC (Pauli Based Computing) model of Bravyi, Smith and Smolin [8], and establish a variety of its properties. The class CM comprises quantum circuits of unitary Clifford gates with fixed input $|A\rangle^\otimes t$ (for $t$ qubit lines) where $|A\rangle = ...
\( \frac{1}{\sqrt{2}}(|0\rangle + e^{i\pi/4}|1\rangle) \) and with output given by final measurement of some number of qubits in the computational basis. For computational applications we will also allow classical polynomial time computation for assistance before and after the Clifford circuit is run, in particular to determine the structure of a CM process \( C_w \) for each computational input bit string \( w \). If the Clifford gates could adaptively depend on further intermediate measurements (not allowed here), the latter model would be universal for quantum computation, but our model appears to be a weaker than universal. Our main result is to show that nevertheless, this class is hard to classically simulate up to additive error, given any one of a broad variety of average-case hardness conjectures.

This result has been shown in the recent works [6] and [23] (and our results were developed independently concurrently) but only for a single particular hardness conjecture. Furthermore both papers prove the anticoncentration conjecture by using the fact that random Clifford circuits form a \( k \)-design for suitable \( k \). The idea of using \( k \)-designs to prove anticoncentration conjectures is explored in [16]. In this paper, we use a different approach. We show that this class, although unlikely to be universal, suffices to emulate the hardness of other classes of computations already known to have the desired properties, thereby establishing hardness of CM simulation up to additive error, given any one of a number of inherited hardness conjectures.

Along the way we also establish a generalised form of the Gottesman–Knill theorem viz. that any adaptive Clifford computation (now allowing intermediate measurements) with input \( \sigma \otimes \rho \), where \( \sigma \) is a stabiliser state, can be simulated by an adaptive Clifford circuit on just \( \rho \), with the help of polynomial time classical processing. This result amounts to a translation of the PBC model back into the circuit model, but has considerable conceptual interest in its own right, applying also to universal quantum computation. The standard Gottesman–Knill theorem [22] is obtained in the case that the whole input is a stabiliser state and then the simulation can be done entirely classically.

Finally we will consider the feasibility of experimentally implementing CM circuits. This has become an increasingly relevant topic with the expected imminent availability of small quantum computers that may allow physical implementation of quantum algorithms unlikely to be simulatable even by the best classical computers [17]. We show that CM circuits have several properties that may make them advantageous for prospective experimental realisation in the near term. We show that in the measurement based computing model (MQC), given the standard graph state, any CM circuit can be implemented without adaptions, and hence can be implemented in MQC depth one. We also show that CM has good properties when it is made fault tolerant in both the circuit and MQC models: while syndrome measurements must be performed, the associated correction operators need not be applied. Also, in MQC given an initial state that can be created offline with high fidelity, CM can be implemented fault tolerant with one further time step.
2 Preliminaries

$X$, $Y$ and $Z$ will denote the standard 1-qubit Pauli operations and $\mathcal{P}_n$ will denote the $n$-qubit Pauli group (generated by tensor products of the 1-qubit Pauli operations). $Z_i$ will denote the Pauli operation on $n$ qubits having $Z$ on the $i$th line and $I$ on all other lines. Pauli measurements for $P \in \mathcal{P}_n$ will have outcomes $\pm 1$. This applies to $Z_i$ measurements too, having outputs $\pm 1$ rather than bit values 0 and 1. We will state explicitly when the latter are used as output labels. A Pauli measurement $P$ is said to be dependent on Pauli measurements $Q_1, \ldots, Q_K$ if $P = \pm Q_{a_1}^{a_1} \cdots Q_{a_K}^{a_K}$ for some $a_1, \ldots, a_K \in \{0,1\}$. $\ket{A}$ will denote the 1-qubit magic state $\ket{A} = \frac{1}{\sqrt{2}} (\ket{0} + e^{i\pi/4} \ket{1})$.

Clifford circuits will always be assumed to be given as circuits of some chosen set of one and two qubit Clifford gates that suffice for any Clifford operation e.g. the Hadamard gate $H$, controlled NOT gate $CX$ and phase gate $S = \text{diag}(1, i)$. We will also use the non-Clifford $T$ gate defined by $T = \text{diag}(1, e^{i\pi/4})$. It is well known that the $T$ gate can be implemented by the so-called $T$-gadget [22], using an extra ancilla qubit line (labelled $a$) in state $\ket{A}$ and adaptive Clifford operations: to apply $T$ to a qubit line $k$ in a circuit, we first apply $CX_{ka}$ with the ancilla as target qubit, and then measure the ancilla qubit in the $Z$ basis giving outcome $(-1)^m$ for $m = 0, 1$ (always with equal probability). Finally $S^m$ is applied to the original qubit line. The ancilla qubit is never used again and may be discarded. The final result in every case is to apply $T$ to line $k$ up to overall phase.

It is well known that Clifford operations with $T$ gates are universal for quantum computation. Then using the $T$-gadget we see that any (universally general) circuit composed of Clifford gates and a number $t$ of $T$ gates can be rewritten as an adaptive circuit of only Clifford gates (and intermediate $Z$ basis measurements) with the addition of $t$ additional ancilla qubit lines initialised in state $\ket{A}^\otimes t$.

2.1 CM circuits

We begin by defining a class of quantum processes that we call “Clifford Magic”, written CM, and introduce an associated complexity class post-CM.

**Definition 2.1.** A CM circuit on $t$ qubits is a unitary Clifford circuit which has input state $\ket{A}^\otimes t$, and has output given by the result of measuring $r$ specified qubits (the output register $O$) in the $Z$ basis. A postselected CM circuit is a CM circuit with an additional register $P$ of $s$ qubits (called the postselection register) disjoint from $O$, which is also measured at the end. □

Let $B_n$ denote the set of all $n$-bit strings and $\mathcal{B} = \cup_n B_n$ denote the set of all finite-length bit strings. A family of circuits $C_w$ for $w \in \mathcal{B}$ is a uniform family if there is a classical poly($n$)-time computation which on input $w \in B_n$ (for any $n$) outputs a description of the circuit $C_w$.

We now introduce a complexity class that intuitively comprises languages decidable with bounded error by using postselected CM quantum processes assisted by classical poly-time computation.
Definition 2.2. A language $L$ is in the class $\text{Post-CM}(\epsilon)$ for error tolerance $0 < \epsilon < 1/2$ if there is a uniform family $C_w$ of postselected CM circuits with the following property: for each $w \in B_n$ let $a$ (an $r$-bit string) and $b$ (an $s$-bit string) denote the results of measuring the output and postselection registers respectively of $C_w$, where we relabel the +1 and −1 measurement outcomes as 0 and 1 respectively. Then there is a classical poly($n$)-time computable function $f : B_{n+r} \to B_1$ such that the bit $\xi = f(w, a) \in B_1$ has the following relation to $L$:

- If $w \in L$ then $\Pr[\xi = 0 | b = 00...0] \geq 1 - \epsilon$, and
- If $w \notin L$ then $\Pr[\xi = 1 | b = 00...0] \geq 1 - \epsilon$. □

We note that the class $\text{Post-CM}(\epsilon)$ is in fact independent of the parameter $\epsilon$, as is readily seen using standard amplification techniques: to improve the value of $1 - \epsilon$ we consider a number $K$ of parallel repetitions of the process (making a single larger process of the same kind) and take the majority vote of all the corresponding $\xi$ bits, with the majority vote computation being incorporated into the function $f$ (now on $n + Kr$ bits). Correspondingly henceforth we will drop the parameter $\epsilon$ and refer to the class as post-CM.

If $r$ and $s$ are each at most $O(\log n)$ then the output probabilities of the CM circuit can be classically efficiently computed [18] and the language $L$ will then be in classical poly-time $P$.

2.2 The Pauli based model of computation (PBC)

Our consideration of CM circuits is motivated by the PBC model of quantum computation [8], and in particular by a development of it (translated back into the circuit model), leading to an extended form of the Gottesman-Knill theorem. We now describe these ingredients, beginning with a review of the PBC model. We will be considering adaptive quantum computational processes and we start with a formal specification.

Definition 2.3. An adaptive quantum circuit $C$ on $n$ qubits, with input state $\alpha$ and output distribution $P_C$ comprises the following ingredients. We have a specified sequence of steps (on the $n$-qubit state $\alpha$) of length poly($n$), with the following properties:

(i) each step is either a unitary gate or a non-destructive $Z$ basis measurement. Post-measurement states from intermediate measurements may be used as inputs to the next step.

(ii) each step is specified in terms of a poly($n$) time classical computation, possibly involving previous measurement outcomes.

If no steps depend on previous measurement outcomes then the circuit is called non-adaptive, and if there are no intermediate measurements steps, then the circuit is called unitary.

The output distribution $P_C$ is the probability distribution of a specified set of measurements (called output measurements). Without loss of generality this may be taken to be the set of all measurements of the circuit $C$ and we often omit explicit mention of the output set. □
**Definition 2.4.** (PBC circuits and the Pauli based computing model PBC)

(i) A **PBC circuit** $C$ on $t$ qubits with any input state $\rho$, is a (generally adaptive) sequence $C$ of Pauli measurements $P_1, \ldots, P_s$ from $\mathcal{P}_t$, which pairwise commute.

(ii) For computational applications (**the PBC model of computing**) we will use a uniform family $\{C_w : w \in \mathcal{B}\}$ of PBC circuits on $t = \text{poly}(n)$ qubits where $n$ is the length of the bit string $w$, and furthermore, each $C_w$ is required to have the input state $\rho = |A\rangle^{\otimes t}$. The result of the computation is given by a specified $\text{poly}(n)$ time (randomised) classical computation on $w$ together with the measurement outcomes of the circuit $C_w$. □

**Definition 2.5.** We say that a circuit $C$ (on $n$ qubits, with input state $\alpha$, and output distribution $P_C$) can be **weakly simulated** by a circuit $\tilde{C}$ (on $m$ qubits, with input state $\beta$, and output distribution $P_{\tilde{C}}$) if

(i) a description of the circuit $\tilde{C}$ may be given by a classical $\text{poly}(n)$ time (possibly randomised) translation from a description of $C$, and

(ii) a sample of the distribution $P_C$ can be produced from a sample of $P_{\tilde{C}}$ together with $\text{poly}(n)$ time classical (randomised) computation. □

(More precisely, in the above definitions the $\text{poly}(n)$ bounds refer to a situation in which we are considering a uniform family of circuits depending on an associated parameter $n \in \mathbb{N}$, which will be clear from the context when needed.)

**Theorem 2.1.** (after [8]). Let $C$ be any (generally adaptive) quantum circuit on $n + t$ qubits with input state $\alpha = \sigma \otimes \rho$ where $\sigma$ is a stabiliser state of $n$ qubits and $\rho$ is any state of $t$ qubits. Suppose also that the unitary steps of $C$ are all Clifford gates. Then:

(i) $C$ may be weakly simulated by a (generally adaptive) PBC circuit $P_1, \ldots, P_s$ on $t$ qubits with input state $\rho$, and with $s \leq t$ steps that are independent Pauli measurements.

(ii) If $C$ was non-adaptive (with final $Z$ basis measurement outputs) then the PBC circuit $P_1, \ldots, P_s$ in (i) can also be chosen to be non-adaptive. □

Before giving the proof we remark that Theorem 2.1 shows that the PBC model of computation is universal for quantum computation (providing its main motivation). Indeed, starting with the circuit model of quantum computation we may assume that any circuit consists of Clifford gates and $T$ gates (known to be a universal set). The circuit input state has the form $\sigma_w = |w\rangle |0\rangle \ldots |0\rangle$ i.e. a computational basis state $|w\rangle$ for $w \in \mathcal{B}_n$ supplemented by extra $|0\rangle$ qubits, so a stabiliser state. The circuit size is $O(\text{poly}(n))$ so has at most $t = \text{poly}(n)$ $T$ gates. Replacing each $T$ gate by a corresponding $T$ gadget, we obtain an adaptive circuit of the kind in Theorem 2.1 with input state $\sigma_w \otimes |A\rangle^{\otimes t}$, which may then, according to the theorem, be weakly simulated by a PBC circuit on $|A\rangle^{\otimes t}$.

### 2.2.1 Proof of theorem 2.1

We give the proof in full (following the method of [8]) dividing it into labelled sections, for later reference back from a slightly more general context. We begin with a supporting lemma.
Lemma 2.2. [8] Let \( P, Q \in \mathcal{P}_n \) be anti-commuting Pauli operations and let \( |\psi\rangle \) be an eigenstate of \( P \) with \( P|\psi\rangle = \lambda_P|\psi\rangle, \lambda_P = \pm 1 \). Then:

(i) Measurement of \( Q \) on \( |\psi\rangle \) gives result \( \lambda_Q = \pm 1 \) with equal probabilities half.

(ii) The operator \( V(\lambda_P, \lambda_Q) = (\lambda_P P + \lambda_Q Q)/\sqrt{2} \) is always a unitary Clifford operation.

(iii) \( V(\lambda_P, \lambda_Q)|\psi\rangle \) is the normalised projection of \( |\psi\rangle \) onto the \( \lambda_Q \)-eigenspace of \( Q \).

Hence measurement of \( Q \) on \( |\psi\rangle \) is equivalent to classically choosing (offline) a uniformly random \( \lambda \in \{-1, +1\} \) and applying the Clifford unitary \( V(\lambda_P, \lambda) \) to \( |\psi\rangle \).

**Proof.** We have \( |\psi\rangle = \lambda_P P|\psi\rangle \).

For (i) we have \( \text{Prob}(Q \text{ measurement gives } \pm 1) = \left| \frac{1}{2}(I \pm Q)|\psi\rangle \right|^2 \). Replacing \( |\psi\rangle \) by \( \lambda_P P|\psi\rangle \), and using the fact that \( PQ = -QP \) and that \( P \) is unitary, we readily see that the two probabilities are equal.

For (ii), using \( P^2 = Q^2 = I \) and \( PQ = -QP \) we can check directly that \( V(\lambda_P, \lambda_Q) V(\lambda_P, \lambda_Q)^\dagger = I \). Similarly for any Pauli \( R \), for each of the four possible combinations of \( R \) commuting or anti-commuting with \( P \) and \( Q \), we can check directly that \( V(\lambda_P, \lambda_Q) R V(\lambda_P, \lambda_Q)^\dagger \) is a Pauli operation (being just a suitable product of \( P, Q \) and \( R \) in each case).

For (iii) the normalised post-measurement state after outcome \( \lambda \) is

\[
\frac{(I + \lambda Q)}{\sqrt{2}} |\psi\rangle = \frac{(\lambda P + \lambda Q)}{\sqrt{2}} |\psi\rangle = V(\lambda_P, \lambda) |\psi\rangle .
\]

**Proof of Theorem 2.1**

Let \( C \) be any adaptive circuit whose steps are either unitary Clifford gates or \( Z \) measurements. For clarity, we will give the proof for the case where \( \sigma \) is the pure state \( |0\rangle^\otimes n \). The general case of arbitrary (mixed) stabiliser state \( \sigma \) is proved similarly by just replacing each instance of \( Z_1, \ldots, Z_n \) below by a set of generators \( S_1, \ldots S_k \) \((k \leq n)\) of the stabiliser subgroup of \( \sigma \).

(a) Starting with the rightmost Clifford gate and working successively to the left, we commute each gate out to the end of the circuit beyond the last measurement. As a result each \( Z \) measurement will become conjugated into a Pauli measurement \( P_i \in \mathcal{P}_{n+t} \) which may be efficiently determined. Unitary gates applied after the measurements have no effect on the outcomes so we delete them, and we are left with a sequence \( P_1, P_2, \ldots, P_s \) of (generally adaptive) Pauli measurements (where \( s \) is the number of \( Z \) measurements in \( C \), acting on input state \( |0\rangle^\otimes n \otimes \rho \).

(b) Next we prefix the list with “dummy” \( Z \) measurements for each of the first \( n \) lines obtaining the list \( Z_1, Z_2, \ldots, Z_n, P_1, P_2, \ldots, P_s \). This has no effect as the input is \( |0\rangle \) on each of these lines (and the \( Z \) measurements all give result +1 with certainty).

(c) We now show how the extended list can be reduced to an equivalent (in the sense of weak simulation) shorter list comprising only pairwise commuting measurements that act non-trivially only on the last \( t \) lines. Let \( M \) be the first measurement in the list that anticommutes with some earlier measurement \( N \). \( M \) will be one of the \( P_k \)’s as \( Z_1, \ldots, Z_n \) all commute. Since all measurements prior to \( M \) commute the pre-measurement state for \( M \) will be a \( \lambda_N \)-eigenstate of \( N \) where \( \lambda_N = \pm 1 \) is the outcome of the measurement.
Thus we can delete $M$ from the list and replace it by the process of Lemma 2.2 viz. we classically randomly choose $\lambda \in \{+1, -1\}$ and replace $M$ by the (adaptively chosen) unitary gate $V(\lambda, \lambda_N) = \frac{1}{\sqrt{2}}(\lambda_N N + \lambda M)$. Finally (as in (a)) we commute $V(\lambda, \lambda_N)$ out to the right and delete it. This introduces possible further adaptations depending on the $N$ measurement outcome. Note that if $N$ was one of the first $r$ $Z$ measurements then its outcome is deterministically $\lambda_N = +1$ and no further adaptations are introduced.

As a result of eliminating $M$, the list has been shortened by one and all measurements to the left of where $M$ was, pairwise commute. We now repeat the process from the start of (c) until no anticommuting measurements remain in the list. This results in a list $Z_1, \ldots, Z_n, R_1, \ldots, R_v$ with $v \leq s$, of pairwise commuting Pauli measurements.

(d) Now, since each $R_k$ commutes with all of $Z_1, \ldots, Z_n$, the 1-qubit Pauli operation $A_j$, for $j = 1, \ldots, n$ in the first $n$ positions of $R_k = A_1 \otimes A_2 \otimes \ldots \otimes A_{n+t}$ must be either $I$ or $Z$. Consequently on input state $|0\rangle^\otimes n \otimes \rho$ each measurement projection (for either outcome) in the list preserves the $|0\rangle^\otimes n$ part of the state and the process amounts to just the state $\rho$ being subjected to a sequence of commuting Pauli measurements from $P_1$ given by the operators $A_{n+1} \otimes \ldots \otimes A_{n+t}$ for each $R_k$. Writing $R_k$ for $R_k$ restricted to its last $t$ positions, and now deleting the initial dummy $Z$ measurements too, we see that the given circuit $C$ is weakly simulated by the adaptive list of pairwise commuting Pauli measurements $\tilde{R}_1, \tilde{R}_2, \ldots, \tilde{R}_v$, on the input state $\rho$. The reduction from $C$ to this list is a randomised classical reduction (because of the random classical choices of the $\lambda$’s in (c)).

(e) The commuting (and generally adaptive) list $\tilde{R}_1, \tilde{R}_2, \ldots, \tilde{R}_v$ will not generally comprise independent operators. However any independent set of operators from $P_1$ has size at most $t$ and furthermore, given any independent set $S \subset P_1$ and operator $T \in P_t$, it is possible to efficiently determine whether $T$ is independent of $S$ or not [22]. Now starting from the left, we can check (as a function of all previous measurement outcomes, that define the possibly adaptive list up to that stage) whether each successive operator in the list $\tilde{R}_1, \tilde{R}_2, \ldots, \tilde{R}_v$ is independent or not, of the previous measurement operators. If it is dependent, its result is already classically fixed (from previous measurement outcomes) and we deleted it from the list. This finally produces a list of at most $t$ remaining measurements that are independent and commuting, completing the proof of (i).

(f) Finally we prove (ii). If $C$ is non-adaptive then we may assume without loss of generality that it is a unitary circuit $U$ followed by final measurements $Z_{i_1}, \ldots, Z_{i_s}$ on specified qubit lines $i_1, \ldots, i_s$ [18]. Then in (b) we will obtain the list

$$Z_1, Z_2, \ldots, Z_n, P_1, P_2, \ldots, P_s.$$ 

Here $P_k = UZ_{i_k}U^\dagger$ for $k = 1, \ldots, s$, which are commuting and independent. However some may anticommute with an initial dummy $Z$ measurement. Then following the process of (e) (with $M$ and $N$ as in (c) above) $M$ is a $P_1$ and $N$ must be one of the dummy $Z$’s, whose measurement outcome $\lambda_N = +1$ is deterministic. Thus the unitary gate $V(\lambda, \lambda_N)$ involves no adaptations, and the sequence remains non-adaptive after $V(\lambda, \lambda_N)$ is commuted out to the end (although it depends on the classical random
choice of $\lambda$). Continuing in this way, as in the second paragraph of (c), we note that if any subsequent operator $M$ anticommutes with any earlier operator $N$, then $M$ must always anticommute with one of the dummy $Z$’s too. This is because at any iteration stage, the operators after the dummy $Z$’s are given by initial $P_i$’s conjugated some number of times by operators $V$ that are always in the algebra generated by the $P_k$’s and dummy $Z$’s (i.e. the successive $V$’s that have been commuted out). Thus if $M$ commuted with all the dummy $Z$’s, it must also commute with all preceding operators $N$ (recalling that the $P_k$’s were all commuting).

Now by choosing an anticommuting $N$ to always be a dummy $Z$, $\lambda_N$ will always be $+1$ and no adaptation is ever introduced, so at the end of (c) above, we will obtain a non-adaptive commuting list $Z_1, \ldots, Z_n, R_1, \ldots, R_v$ with $v \leq s$. Continuing as in (d) we obtain a non-adaptive list $\tilde{R}_1, \tilde{R}_2, \ldots, \tilde{R}_v$ from which (as in (e)) an independent sublist of length at most $t$ can be efficiently selected. This proves (ii).

\section{2.3 Extending the Gottesman–Knill theorem}

A standard form of the Gottesman-Knill theorem asserts that any adaptive Clifford circuit (in which we require all intermediate measurements to be elementary $Z$ measurements) and stabiliser state input may be classically efficiently weakly simulated \cite{15, 22, 18}. As noted above, universal quantum computation can be performed using adaptive Clifford circuits which include additional (non-stabiliser) $|A\rangle$ state ancilla inputs, motivating the consideration of Clifford circuits on such more general inputs. In our extension of the Gottesman-Knill theorem we consider adaptive Clifford circuits but now allow the input to have a non-stabiliser part. We show that it may be weakly simulated by a hybrid classical-quantum process whose quantum part (obtained by an efficient classical reduction from the description of the original circuit) is an adaptive Clifford circuit acting now only on the non-stabiliser part of the original input, thereby relegating the “stabiliser-input part” of the original computation into efficient classical computation instead. In the special case where in the initial input is fully a stabiliser state, we recover the standard Gottesman–Knill theorem, as our hybrid process then has no residual quantum part.

A PBC circuit with general input state $\rho$ is similar to an adaptive Clifford circuit albeit with no unitary gate steps, except that the measurements are general Pauli measurements rather than just elementary $Z$ measurements. Correspondingly our extended Theorem 2.3 below is obtained as a translation of Theorem 2.1 into a standard circuit form. With the non-stabiliser part of the input being restricted to just copies of $|A\rangle$, it provides a simulation result for arbitrary quantum computation.

\textbf{Theorem 2.3.} \textit{(Extended Gottesman–Knill Theorem)} Let $C$ be any adaptive Clifford circuit with input state $\sigma \otimes \rho$, where $\sigma$ is a stabiliser state of $n$ qubits and $\rho$ is an arbitrary state of $t$ qubits, and the output is given by measurement of any specified qubit lines. (Usually we will also have $t = O(\text{poly}(n))$). Then

(i) $C$ can be weakly simulated by an adaptive Clifford circuit $C^*$ on $t$ qubits with input $\rho$, assisted by $\text{poly}(n + t)$-time classical computation;
(ii) if $C$ is non-adaptive then $C^*$ may be taken to be unitary (with $Z$ basis measurements only for outputs at the end).

Proof. According to Theorem 2.1(i), $C$ can be weakly simulated by a PBC circuit of commuting Pauli measurements $P_1, \ldots, P_s$ on input state $\rho$, and we just need to translate this back into an adaptive Clifford circuit with only $Z$ basis measurements. This follows immediately by applying lemma 2.4 below to each $P_i$ separately, expressing them as $P_i = U_i^\dagger Z_k U_i$ for unitary Clifford operations $U_i$ and any choice of line $k$ (which could even be independent of $i$), thus establishing (i).

Note that the Lemma cannot be applied to all $P_i$ simultaneously (giving a single $U$) since although pairwise commuting, they are generally adaptively determined and not fixed a priori, and not necessarily independent. However if $C$ is non-adaptive then according to Theorem 2.1(ii), the sequence $P_1, \ldots, P_s$ can be chosen to be non-adaptive and independent, with $s \leq t$. Lemma 2.4 can then be applied to the whole list to give a single $U$ with $U_1^\dagger Z_k U_1 = P_k$ for $k = 1, \ldots, s$. The circuit $C^*$ is then just the unitary Clifford $U$ (as unitaries after the $Z$ measurements have no effect and can be deleted), thus establishing (ii).

Lemma 2.4. Let $\{P_1, \ldots, P_m\}$ be any set of independent and pairwise commuting Pauli operations on $n$ qubits (so $m \leq n$). Then there is a unitary Clifford operation $U$ such that $U_1^\dagger Z_k U_1 = P_k$ for $k = 1, \ldots, m$. Furthermore a circuit of basic Clifford gates of depth $O(n^2/\log(n))$ implementing $U$ may be determined in classical poly($n$) time.

Proof. We first extend the set $\{P_1, \ldots, P_m\}$ to a maximally sized set $\{P_1, \ldots, P_n\}$ of independent pairwise commuting Pauli operations. This extension is not unique, but see Section 7.9 of [24] for an efficient method of extension. Using similar techniques we also find generators of the ‘destabiliser group’ $\{D_1, \ldots, D_n\}$ (defined in [2, 25]). Then there is a unique (up to phase) Clifford $V$ such that $V Z_i V^\dagger = P_i$ and $V X_i V^\dagger = D_i$ for $i = 1, \ldots, n$. An $O(n^2/\log(n))$ circuit implementing $V$ may be determined in classical poly($n$) time by the construction of Theorem 8 in [2]. Finally take $U = V^\dagger$.

3 Hardness of classical simulation of CM circuits

We now establish lower bounds on the complexity of classical simulation of CM circuits, allowing either multiplicative or additive errors in the simulation. The scenario of additive error is generally regarded as a reasonable model of what is feasible to physically implement in practice.

A distribution $q(x)$ is an $\epsilon$-additive approximation of a distribution $p(x)$ if

$$\sum_x |p(x) - q(x)| \leq \epsilon. \quad (1)$$

A number $Y$ is an $\epsilon$-multiplicative approximation of a number $X$ if $|X - Y| \leq \epsilon X$. A distribution $q(x)$ is an $\epsilon$-multiplicative approximation of a distribution $p(x)$ if for each $x$, $q(x)$ is an $\epsilon$-multiplicative approximation of $p(x)$. Thus clearly $\epsilon$-multiplicative approximation of distributions implies $\epsilon$-additive approximation.
3.1 Classical simulation with multiplicative error

Although (uniform families of) CM circuits themselves are not likely to be universal for quantum computation, we first establish that postselected CM circuits suffice as a quantum resource for postselected universal quantum computation.

**Theorem 3.1.** Any postselected poly-sized unitary quantum circuit $C$ on $n$ qubits (with final $Z$ measurements) can be weakly simulated by a postselected poly-sized CM circuit on poly($n$) qubits.

**Proof.** We may suppose without loss of generality that $C$ has the following form: the input state is $|0\rangle^\otimes n$, followed by Clifford and $T$ gates, and finally some number of lines are measured in the $Z$ basis. Of these, some are postselected to outcome $k = 0$ (for $Z$ measurement outcome $(-1)^k$). To begin, we replace each $T$ gate with a $T$ gadget where the measurement is postselected to outcome $(-1)^m$ with $m = 0$ so the correction $S^m$ is not required. As no other part of the circuit acts on this ancilla line again this measurement can be performed at the end of the circuit. The resulting circuit $\tilde{C}$ then has input $|0\rangle^\otimes n|A\rangle^\otimes t$, which is acted on by a Clifford unitary $U$ followed by $Z$ measurements, some of which are postselected. Let $i_1, \ldots, i_s$ denote all the lines that are measured at the end of $\tilde{C}$.

Naively, one may now hope to prove the theorem by starting with a CM circuit, with the same gates as $\tilde{C}$ and input $|A\rangle^\otimes n|A\rangle^\otimes t$, measuring $n$ of the inputs in the $Z$ basis, and post selecting them to the outcome 0. Then the state of the computation is $|0\rangle^\otimes n|A\rangle^\otimes t$, and the rest of the computation can be done straightforwardly on this postselected state. However, our definition of CM circuit does not allow intermediate measurements (postselected or otherwise), and so this approach fails.

Instead we will adapt the constructions in the proof of Theorem 2.1, extending them to incorporate postselection, and then apply Lemma 2.2 (as it was used in Theorem 2.3(ii)). The labels (a) to (f) below will refer to the corresponding labelled parts of the Proof of Theorem 2.1.

We begin with the non-adaptive Clifford circuit given by $\tilde{C}$ but having none of its $Z$ measurements being postselected. Then following the process in (f) (after its use of (b)) we obtain the measurement list $Z_1, Z_2, \ldots, Z_n, P_1, P_2, \ldots, P_s$ with $P_k = U Z_{i_k} U^\dagger$ for $k = 1, \ldots, s$, and these $s$ measurements are commuting. Continuing with the process of (c) we eliminate anti-commuting measurements from the full list. If the anti-commuting measurement $M$ being treated arose from one that was postselected in $\tilde{C}$, then instead of choosing $\lambda \in \{+1, -1\}$ randomly, we simply set $\lambda$ to its postselected value $+1$ or $-1$, and replace $M$ by the unitary gate $V(\lambda, \lambda_N)$ with $\lambda_N = +1$ too. Here (as in (f)) $N$ is the earlier measurement with which $M$ anticommutes, so $N$ must be one of $Z_1, Z_2, \ldots, Z_n$.

Continuing in this way we complete the application of (c) and (d) to obtain a non-adaptive pairwise commuting list $R_1, R_2, \ldots, R_m$ of Pauli measurements from $P_t$ acting on input $|A\rangle^\otimes t$. Some of these are postselected and then the corresponding postselected process will weakly simulate $\tilde{C}$ (with its postselections). Next (as in (e)) we reduce the list to an independent list (of length at most $t$). To achieve this, first reorder the commuting list to have all postselected measurements at the start. Then for each successive
measurement in the reordered list we determine if it is dependent or not on those previous to it. If it is, then its outcome is already determined and we delete it from the list (while recording its classical relation to previous outcomes for weak simulation as in Definition 2.5). If its determined outcome is inconsistent with an imposed postselection at that stage, then this indicates that the postselection requirement of the original circuit has probability zero. Finally we apply Lemma 2.4 to the independent list (just as it was used in the proof of Theorem 2.3(ii)) to obtain a bonafide postselected CM circuit that weakly simulates the original given postselected circuit $C$.

**Corollary 3.2.** \( \text{post-CM} = \text{post-BQP} \). Thus if uniform families of CM circuits could be weakly classically simulated to within multiplicative error \( 1 \leq c < \sqrt{2} \), then the polynomial hierarchy would collapse to its third level.

**Proof.** The first claim follows immediately from Theorem 3.1, and then the second follows from [9].

### 3.2 Background for additive error case

Before considering simulation of CM circuits up to additive error, we first outline a general framework and argument (following [11, 11] but with some generalisation of context for our later purposes) that has been used in the literature [1, 11, 14, 21, 6, 5] to argue for hardness of classical simulation, up to additive error, of a variety of classes of quantum computational processes.

Consider a given class $C = \{ C_\theta : \theta \in \Theta \}$ of quantum circuits parameterised by $\theta \in \Theta$ (with each circuit also having its input state specified). The class will generally contain circuits with asymptotically large numbers of qubit lines and we generically denote the number of qubit lines of $C_\theta$ by $n$. For all our needs it will also be the case that the number of circuits in $C$ with any fixed $n$ is finite. Let the output be given by a measurement of all $n$ lines and let $p_\theta(x)$ with $x \in B_n$ denote the output probability distribution of $C_\theta$.

Introduce the following computational (sampling) task $T_C$ associated to the class $C$: for any given $\theta$, return $(\theta, y)$ where $y \in B_n$ has been sampled according to the output distribution $p_\theta$ of $C_\theta$. We will be interested in the complexity of simulating this task (and some approximate variants) as a function of $n$.

By an $\epsilon$-additive error simulation of the task $T_C$, we mean a process that given $\theta$, returns $(\theta, y')$ where $y'$ has been sampled according to a distribution $q_\theta$ on $B_n$ which is an $\epsilon$-additive approximation of the distribution $p_\theta$.

An alternative task (that neither a classical nor quantum computer is likely to be able to efficiently achieve) is to compute a value for $p_\theta(x)$ for given $\theta$ and $x$, up to a (suitably specified) multiplicative error. Indeed for relevant classes that are studied in the literature, it can be shown that computing such approximations is #P hard in the worst-case. This task is of computational significance since for suitably chosen classes $C$ the probability values can be used to represent quantities that are of independent physical or mathematical interest.

Our aim is to argue for classical hardness of simulation of the sampling problem $T_C$ up to additive approximation. To do this we will need to conjecture that estimating the
value of $p_\theta(x)$ up to (suitable) multiplicative approximation remains $\#P$ hard not just in the worst-case, but in an average-case setting of the following kind.

To the class $C$ we associate two constants: a measure size $0 < f < 1$ and an error tolerance $\eta$. Also for each $m \in \mathbb{N}$ introduce the set

$$D = \{(\theta, x) : C_\theta \text{ has } m \text{ lines and } x \in B_m\}.$$

Finally (for each $m$) we have a given probability measure $\pi$ on the set of $\theta$’s that occur in $D$, and let $\nu$ denote the uniform probability measure on $B_m$. Then $\pi \times \nu$ is the product measure on $D$.

We assert the following conjecture that we will refer to as Hardness$(C, \pi)$.

**Average-case hardness conjecture for $C$ with $\pi$:** let $F \subseteq D$ be any chosen subset of $D$ having $\pi \times \nu$ probability measure $f$. Then it is $\#P$ hard to approximate the values $p_\theta(x)$ for all $(\theta, x) \in F$ up to multiplicative error $\eta$. □

Note that if $\pi$ is the uniform measure too, then the subsets $F$ (for each $m$) will also be of fractional size $f$. But for nonuniform $\pi$’s there will be subsets of measure $f$ that have smaller fractional size than $f$ and asserting their $\#P$ hardness is a stronger conjecture. The use of nonuniform distributions will also feature significantly in the anticoncentration property below.

As an example, in [11] classes of IQP circuits $C$ are considered but we are interested there only in the all-zero output probability $p(0^m)$ (for circuits $C$ on $m$ qubits). However it is noted that $p(x)$ for other outputs $x \in B_m$ are given by the all-zero output probability of a suitably modified IQP circuit $C_x$, and the classes of circuits considered are closed under this modification. Then conjectures 2 and 3 of [11] can be expressed as above, with $\pi$ being the uniform distribution, $f = 1/24$ and $\eta = 1/4 + o(1)$. In [10] on the other hand, a nonuniform $\pi$ is used that allows subsets admissible in the Conjecture to be dominated by so-called sparse IQP circuits.

The arguments below will use several complexity classes that we will loosely describe here in a way that suffices to express the hardness of simulation argument. For more complete descriptions see for example [3]. $\text{BPP}^{\text{NP}}$ is the class of decision problems that can be solved by randomised classical polynomial time computations armed with an oracle for any problem in NP. $\text{FBPP}^{\text{NP}}$ is the same except that the outputs can be bit strings rather than just a single bit. $\text{BPP}^{\text{NP}}$ is in the third level of the tower of complexity classes known as the polynomial hierarchy PH. $\text{P}^{\#P}$ is the class of decision problems solvable in classical polynomial time, given access to an oracle for any $\#P$ problem; and it is known (Toda’s theorem) that $\text{PH} \subseteq \text{P}^{\#P}$.

Now suppose that the sampling task $T_C$ can be solved up to additive error by a classical polynomial time algorithm $A$. The first step is to show this ability to sample implies the existence of an $\text{FBPP}^{\text{NP}}$ algorithm which, with use of $A$, can estimate $p_\theta(x)$ up to an additive error, for each $\theta$ and a constant fraction of choices of $x$. After that an anticoncentration result will be used to convert the additive error into a multiplicative one, at least for a good measure of instances of $(C_\theta, x)$. The final step is to then invoke the average-case hardness conjecture for $C$: if our multiplicative approximation determination (computable in $\text{FBPP}^{\text{NP}}$) is $\#P$ hard then $\text{P}^{\#P} \subseteq \text{D}^{\text{FBPP}^{\text{NP}}} = \text{BPP}^{\text{NP}}$.\text{12}
The latter class is in the third level of PH and then by Toda’s theorem, PH will collapse to its third level. However such a collapse is widely regarded as extremely implausible (similar to a collapse of NP to P), providing plausibility that the purported classical polynomial time algorithm $A$ for solving $T_C$ up to additive error, cannot exist (if the average hardness conjecture is accepted).

**Lemma 3.3.** (adapted from Lemma 4 of [11]) Suppose there is a classical polynomial time algorithm $A$ that simulates the sampling task $T_C$ up to additive error $\epsilon$. Then for any $0 < \delta < 1$ there is an FBBP$^{NP}$ algorithm that, for each $\theta$, approximates $p_\theta(x)$ up to additive error

$$\frac{p_\theta(x)}{\text{poly}(n)} + (1 + o(1)) \cdot \frac{\epsilon}{2^n \delta}$$

for at least a fraction $1 - \delta$ of all $x \in B_n$. Thus for any probability measure $\pi$, the subset of $D$ to which eq. (2) applies, has $\pi \times \nu$ measure at least $1 - \delta$ (since the measure of the full space of $\theta$’s is always unity).

This lemma is readily proved by following the argument of the proof of Lemma 4 in [11], with minor notational modifications.

To obtain a multiplicative error from this additive one, we require an anticoncentration property of the following form.

**Anticoncentration property for $C$ with $\pi$:** there are constants $\alpha > 0$ and $0 \leq \beta \leq 1$ such that $p_\theta(x) \geq \alpha/2^n$ holds on a subset of $D$ of $\pi \times \nu$ measure at least $\beta$. □

In the literature a property of this form is proved for some classes $C$ (e.g. in [11, 6, 21, 10]) and conjectured to hold for others (e.g. in [1]). Proofs of the property generally involve applying the Paley-Zygmund inequality to the probability measure $\pi \times \nu$.

Suppose now that the anticoncentration property holds for $C$. Then by choosing $\delta$ in Lemma 3.3 to be $\beta/2$ we guarantee an overlap $\Xi \subset D$ of probability measure at least $\beta/2$ on which the anticoncentration property $p_\theta(x)/\alpha \geq 1/2^n$ and the additive approximation bound of eq. (2) both hold.

Then substituting $p_\theta(x)/\alpha$ for $1/2^n$ in eq. (2) the approximation bound becomes

$$\frac{p_\theta(x)}{\text{poly}(n)} + (1 + o(1)) \cdot \frac{2\epsilon}{\alpha\beta p_\theta(x)}$$

giving a multiplicative approximation bound of size $\frac{2\epsilon}{\alpha\beta} + o(1)$ for $p_\theta(x)$, for a $\beta/2$ measure subset of $D$.

Finally collecting all the above, we arrive at the following conclusion.

**Theorem 3.4.** Let $C$ be any class of quantum circuits with associated measure $\pi$ for which the anticoncentration property holds (with constants $\alpha$ and $\beta$). Suppose that the sampling task $T_C$ can be efficiently classically simulated up to additive error $\epsilon$. Then if the average-case hardness conjecture holds with measure size $f = \beta/2$ and error tolerance $\eta = 2\epsilon/(\alpha\beta)$, the polynomial hierarchy will collapse to its third level.

For example in [11] we have $\epsilon = 1/192$, and the anticoncentration property is shown to hold with uniform $\pi$, $\alpha = 1/2$ and $\beta = 1/12$. So to obtain collapse of PH we need
the average-case hardness conjecture to be valid with error tolerance \( \eta = 2\epsilon/(\alpha\beta) = 1/4 \) and fraction \( f = \beta/2 = 1/24 \).

### 3.3 Classical simulation with additive error

We now show that CM circuits cannot be classically efficiently simulated with additive error unless PH collapses, given average-case hardness conjectures. While CM circuits have been shown before \([6, 23]\) to have this property for one particular average-case-conjecture, here we show that actually a broad variety of such conjectures apply, such that if any one of them is proven, it implies the hardness of CM circuit simulation. Furthermore, in previous work, this result was shown by invoking the fact that Clifford gates form a 2-design \([12]\) and that 2-designs anticoncentrate \([16, 19]\), to give the needed anticoncentration property. Here we follow a very different method, instead using the ability of CM circuits (via Theorem 2.3) to simulate any nonadaptive circuit. This allows CM circuits to simulate several other classes of circuits (not necessarily 2-designs) and inherit their average-case hardness conjecture as a basis for hardness of CM circuit simulation up to additive error.

Consider any class of unitary circuits \( C = \{C_\theta : \theta \in \Theta\} \) and associated measure \( \pi \) on \( \Theta \), for which a suitable anticoncentration property holds, and whose classical simulation up to additive error would imply collapse of PH if we assume \text{Hardness}(C, \pi). Suppose that these circuits have been expressed as circuits of gates from the universal set of basic Clifford gates with \( T \) and \( T^\dagger \). We can use any choice of such a representation. Now consider the expanded class \( C^T \) obtained by taking each circuit \( C_\theta \) and replacing each \( T \) and \( T^\dagger \) gate by either \( T \) or \( T^\dagger \) in all combinations. If \( C_\theta \) has \( t \) \( T \) and \( T^\dagger \) gates then it will give rise to \( 2^t \) circuits in \( C^T \), and these can be labelled by \( (\theta, \tau) \) where \( \tau \) is a \( t \)-bit string indicating the choices of \( T \) and \( T^\dagger \). Accordingly, we write \( C^T = \{C_{\theta, \tau} : \theta \in \Theta, \tau \in B_t\} \).

\( C^T \) is exactly the class of circuits we obtain if we implement the circuits \( C_\theta \) using \( T \) gadgets for each \( T \) and \( T^\dagger \) gate, but omit all the adaptive \( S \) gate corrections that are normally specified by the \( T \)-gadget measurement outcomes. Denote that non-adaptive circuit by \( U_\theta \) with outputs \( (x, \tau) \) where \( \tau \in B_t \) is the string of gadget measurement outcomes and \( x \) arises from the output lines from \( C_\theta \). Each of the \( 2^t \) possibilities for \( \tau \) will occur with equal probability. Note that the circuits \( U_\theta \) are unitary Clifford circuits (having only final \( Z \) measurements). Indeed the measurement within any (generally intermediate) \( T \)-gadget can now be moved to the end of the circuit as that line is not acted on again, and the measurement outcome is not used in any adaptations. Because these circuits are unitary Clifford circuits, they can be simulated by CM circuits using Theorem 2.3 (ii). Denote the associated CM circuit (with input state \( |A_0\rangle^{\otimes t} \)) by \( V_\theta \). Finally let \( p_{\theta}(x) \), \( p_{\theta, \tau}(x) \) and \( u_{\theta}(x, \tau) \) (with \( x \in B_n, \tau \in B_t \)) denote the output probabilities for the circuits \( C_\theta, C_{\theta, \tau} \) and \( U_\theta \) respectively.

Note that for each \( \theta \) there is a \( \tau_0 = \tau_0(\theta) \) for which \( p_{\theta, \tau_0}(x) = p_{\theta}(x) \), viz. \( \tau_0 \) just specifies the \( T \) and \( T^\dagger \) choices that actually occur in \( C_\theta \). Furthermore, since each \( \tau \) arises in the output of \( U_\theta \) with equal probability \( 1/2^t \), the relationship between \( C_{\theta, \tau} \) and \( U_\theta \)
gives (via conditional probabilities):

$$p_{\theta, \tau}(x) = u_{\theta}(x, \tau) 2^t.$$  

(3)

Finally in addition to distribution $\pi$ on the $\theta$’s, let $\nu$ and $\nu'$ denote the uniform distribution on the $x$’s and $\tau$’s respectively. Let $\text{prob}_{\pi \times \nu \times \nu'}(\theta, x, \tau)$ denote the probability of $(\theta, x, \tau)$ in the product distribution $\pi \times \nu \times \nu'$, and similarly for $\text{prob}_{\pi \times \nu'}(\theta, \tau)$, $\text{prob}_{\pi}(\theta)$ etc.

We will show that, for some classes $\mathcal{C}$ of circuits already proved to have the additive simulation hardness property of Theorem 3.4 (subject to an associated Hardness($\mathcal{C}, \pi$) conjecture), that $\mathcal{C}^T$ contains no new circuits that were not already present in $\mathcal{C}$. Thus the labels $(\theta, \tau)$ will label the circuits of $\mathcal{C}$ with generally high redundancy, and with an abuse of notation we write $\mathcal{C}^T = \mathcal{C}$ in this situation. Since such circuits can be simulated by CM circuits, classical simulation of CM circuits up to additive error can then imply collapse of PH, subject to the conjecture Hardness($\mathcal{C}, \pi$) of the class $\mathcal{C}$, as will be formalised in the Theorem below.

Suppose now that $\mathcal{C} = \mathcal{C}^T$. Then for each $(\theta, \tau)$ there is $\tilde{\theta} = \tilde{\theta}(\theta, \tau)$ with $C_{\theta, \tau}$ being $C_{\tilde{\theta}}$ so

$$p_{\theta, \tau}(x) = p_{\tilde{\theta}}(x).$$

We will also require the following $\theta$-sampling relation: the $C_{\theta}$ circuits occurring multiply in $\mathcal{C}^T$, occur with the same probability in $\mathcal{C}^T$ (wrt distribution $\pi \times \nu'$) as they did in $\mathcal{C}$ (wrt distribution $\pi$):

$$\sum_{(\theta, \tau): \tilde{\theta}(\theta, \tau) = \theta_0} \text{prob}_{\pi \times \nu'}(\theta, \tau) = \text{prob}_{\pi}(\theta_0).$$  

(4)

**Theorem 3.5.** Consider any class of circuits $\mathcal{C}$ with associated distribution $\pi$ for which the following hold:

(i) the anticoncentration property (with parameters $\alpha$ and $\beta$);

(ii) $\mathcal{C} = \mathcal{C}^T$ and the $\theta$-sampling relation eq. (4).

Then if every CM circuit can be efficiently classically simulated to additive error $\epsilon$, the average-case hardness conjecture for $(\mathcal{C}, \pi)$ with parameters $f = \beta/2$ and $\eta = 2\epsilon/(\alpha\beta)$ will imply that PH collapses.

**Proof.** We use the notations and definitions introduced above. Since $U_{\theta}$ can be simulated by a CM circuit, if every CM circuit can be efficiently classically simulated to additive error $\epsilon$, then so can the distribution $u_{\theta}(x, \tau)$. So by Lemma 3.3 applied in $(\theta, \tau, x)$ space, there is a $(1 - \beta/2)$ sized subset in $\pi \otimes \nu' \otimes \nu$ measure where an FBBP$^{\text{NP}}$ algorithm can calculate an additive approximation to $u_{\theta}(x, \tau)$ with additive error bound of

$$\frac{u_{\theta}(x, \tau)}{\text{poly}(n + t)} + (1 + o(1)) \cdot \frac{2\epsilon}{2^{n+t}\beta}$$  

(5)

(since we have $n + t$ lines now).
Next we will want a measure \( \beta \) subset of \((\theta, \tau, x)'s\) on which the anticoncentration property \( u_\theta(\tau, x) \geq \alpha/2^{n+t} \) holds. By \((\mathcal{C}, \pi)\) anticoncentration, there is a measure \( \beta \) subset of \((\theta, x)'s\) with \( p_\theta(x) \geq \alpha/2^n \). So by the \( \theta \)-sampling relation eq. \( [4] \) and eq. \( [5] \) there is a measure \( \beta \) subset of \((\theta, \tau, x)'s\) with

\[
\frac{u_\theta(x, \tau)}{2^t} \geq \frac{\alpha}{2^{n+t}} \tag{6}
\]

(notating that for any \( x \), \( \text{prob}_{\pi_{XV'}}(\theta, x) = \text{prob}_\pi(\theta)/2^n \)). Combining eqs. \( [6] \) and \( [5] \) we get a measure \( \beta/2 \) subset of \((\theta, \tau, x)'s\) on which \( u_\theta(x, \tau) \) can be calculated by an FBBP\(_{NP}\) algorithm to multiplicative approximation \( 2\epsilon/(\alpha \beta) + o(1) \), and this also applies to \( p_{\theta,\tau}(x) = u_\theta(x, \tau)2^t \) there (as multiplicative approximations are invariant under scale changes).

Finally we want to map this back to \((\theta, x)\) space. Note that for any \( x \)

\[
\text{prob}_{\pi_{XV'}}(\theta, \tau, x) = \frac{1}{2^n} \text{prob}_{\pi_{XV'}}(\theta, \tau) \leq \frac{1}{2^n} \text{prob}_\pi(\hat{\theta}(\theta, \tau)) = \text{prob}_{\pi_{XV'}}(\tilde{\theta}, x)
\]

(where the inequality follows from eq. \( [4] \)). Hence the map \((\theta, \tau, x) \mapsto (\tilde{\theta}(\theta, \tau), x)\) gives a subset of \((\theta, x)'s\) of measure \( \geq \beta/2 \) on which \( p_\theta(x) \) can be calculated to multiplicative approximation \( 2\epsilon/(\alpha \beta) + o(1) \) by an FBBP\(_{NP}\) algorithm. Hence the average-case hardness conjecture for \((\mathcal{C}, \pi)\) implies that \( \text{PH} \) collapses to its third level.

Examples of circuit classes in the literature for which a suitable anticoncentration property holds, \( \mathcal{C} = \mathcal{C}^T \) and the \( \theta \)-sampling relation eq. \( [4] \) holds, include the following.

**IQP circuits associated with the Ising model [11]**

This is the class of circuits \( \mathcal{C} \) having input \( |0\rangle^{\otimes n} \) acted on by \( H^{\otimes n} U H^{\otimes n} \), where \( U \) is unitary and chosen in the following way: apply \( T^{v_i} \) to each qubit line \( i \), and \( CS^{w_{ij}} \) to each pair of qubits \( i, j \), where \( v_i \) and \( w_{ij} \) (all collectively comprising the label \( \theta \)) are chosen in all possible combinations from \( \{0, ..., 7\} \) and \( \{0, ..., 3\} \) respectively, and \( CS \) is the controlled-\( S \) gate. Furthermore the \( CS \) gate is implemented in terms of Clifford+\( T + T^\dagger \) gates using the gadget of Figure 1. The distribution \( \pi \) is the uniform distribution.

To see that \( \mathcal{C} = \mathcal{C}^T \) note first that if any initial \( T \) or \( T^\dagger \) gates are changed (to the other choice), the resulting circuit is clearly still a circuit in the original set. However, there are also \( T \) and \( T^\dagger \) gates within the \( CS \) gadget of Figure 1 to consider. If the \( T \) or \( T^\dagger \) gates at either end are changed, this can be corrected by applying further \( T \) gates. If the middle \( T^\dagger \) gate is swapped, the result is \( CS(T \otimes T) \). So in each of these cases, the resulting circuit is still from the original set. The \( \theta \)-sampling relation eq. \( [4] \) holds because for each \( \theta \) there is a \( \tau_0 = \tau_0(\theta) \) with \( \hat{\theta}(\theta, \tau_0) = \theta \) and the fact that for any fixed \( \tau' \) (and varying \( \theta \)) the mapping \((\theta, \tau_0(\theta)) \mapsto (\theta, \tau_0 \oplus \tau')\) is one-to-one on the underlying \( \tilde{\theta} \)'s (with \( \oplus \) being addition of \( t \)-bits strings at each entry).

**Sparse IQP circuits [10]**

This class is the same as the above (so \( \mathcal{C} = \mathcal{C}^T \)) but with a different distribution \( \pi \). Specifically, having chosen each \( v_i \) and \( w_{ij} \) uniformly, each \( CS^{w_{ij}} \) gate is applied only with some probability \( p \), while each \( T^{v_i} \) is applied as in the above case. This amounts
to $w_{ij} = 0$ being chosen with probability $\frac{1}{4} + \frac{3}{4}(1 - p)$ and other $w_{ij}$’s with probability $p/4$ (and $v_i$’s chosen uniformly as before). Also as before when a $T$ gate inside of $CS$ is swapped, it always becomes $CS$ with some extra $T$ gates. The $\theta$-sampling relation eq. (4) holds since reassigning $T$ and $T^\dagger$ gates always preserves the number of two qubit gates in the circuit. □

**Random Circuit Sampling [4]**

Another class of circuits was put forward by the Google/UCSB team, and called random circuit sampling. The gates used in these circuits are from $\{CZ, X^{1/2}, Y^{1/2}, T\}$. In [10] it is shown that circuits from this set anticoncentrate if they are chosen as follows: Let $G = \{CZ, X^{1/2}, X^{-1/2}, Y^{1/2}, Y^{-1/2}, T, T^\dagger\}$ (i.e. the previous set closed under inverses). In each time step either $U_{1,2} \otimes U_{3,4} \otimes \ldots \otimes U_{n-1,n}$ or $U_{2,3} \otimes U_{4,5} \otimes \ldots \otimes U_{n-2,n-1}$ is applied, for all possible choices of $U_{j,j+1}$ from $G$ (with 1-qubit gates $U$ appearing as $I \otimes U$ or $U \otimes I$). Finally all $n$ lines are measured in the computational basis. the distribution $\pi$ over $C$ is the uniform distribution. All gates in $G$ besides $T$ and $T^\dagger$ are Clifford, so reassigning $T$ and $T^\dagger$ gates clearly results in circuits from the same class i.e. $C = C^T$, and a uniform distribution for $\pi$ satisfies eq. (4). □

By Theorem 3.5, if the relevant average-case hardness conjecture is proven for any one of the above three classes then it shows that CM circuits cannot be classically efficiently simulated to additive error unless the polynomial hierarchy collapses.

For other classes of circuits we generally have $C \neq C^T$ i.e. $C^T$ contains circuits that were not already present in $C$. However, if $C^T$ also has a suitable anticoncentration property, then up to an average-case hardness conjecture, PH will collapses if $C^T$ circuits can be classically simulated to additive error. Note that if $C$ has a worst-case hardness result (as is generally the case for classes considered), then so does $C^T$ since its circuits always form a superset of $C$. This provides evidence for a suitably analogous average-case conjecture for $C^T$. Hence, in the case that $C^T$ also anticoncentrates, it is also likely to be hard to classically simulate. For any $C$, the circuits in $C^T$ can always be simulated by CM circuits (in the sense above, used in Theorem 3.5 taking the uniform distribution over the $\tau$’s as above) and we obtain the following result.
Theorem 3.6. Suppose that $\mathcal{C}^T$ (arising from $(\mathcal{C}, \pi)$ as described above) satisfies an anticoncentration property with constants $\alpha$ and $\beta$. Then if every CM circuit can be efficiently classically simulated to additive error $\epsilon$, PH will collapse to the third level if we assume an average hardness conjecture for $\mathcal{C}^T$ with parameters $f = \beta/2$ and $\eta = 2\epsilon/(\alpha\beta)$, extending the corresponding conjecture for $\mathcal{C}$. Furthermore, if $\mathcal{C}$ had the worst-case hardness property, then so does $\mathcal{C}^T$.

One example of circuits for which $\mathcal{C} \subsetneq \mathcal{C}^T$ and $\mathcal{C}^T$ also anticoncentrates, is the class of Conjugated Clifford circuits introduced in [6]. Here we have circuits of the form $V^{\otimes n}U V^{\otimes n\dagger}$, where $V$ is any fixed 1-qubit gate and $U$ is any Clifford circuit (so we get a class for each choice of $V$), and $\pi$ is the uniform distribution. The representation of $V$ in terms of Clifford+T+T$^\dagger$ gates generally contains $T$ and $T^\dagger$ gates, and when these are reassigned in all combinations in $V^{\otimes n}$, the result is no longer necessarily a gate of the form $W^{\otimes n}$ i.e. the gates applied on different lines will generally be different, and the $n$-qubit gate on one end will also not necessarily be the inverse of the one on the other end. Hence $\mathcal{C} \subsetneq \mathcal{C}^T$. However, this new class of circuits does anticoncentrate. This follows from the original anticoncentration proof in Ref [6] (Lemma 4.3 there) which still applies for arbitrary $n$-qubit gates replacing $V^{\otimes n}$ and $V^{\otimes n\dagger}$ on the ends.

We expect there to be many other classes to which Theorem 3.6 can be applied, providing further corresponding average hardness conjectures which suffice to make CM circuits hard to classically simulate up to additive error. That is because a common strategy for proving that a class of circuits anticoncentrates is to show that it is an $\epsilon$-approximate 2 design and then use the result [16, 19], that such 2-designs have the anticoncentration property. We conjecture the following:

Conjecture 3.7. Suppose $\mathcal{C}$ with $\pi$ is an $\epsilon$-approximate 2 design. Then $\mathcal{C}^T$ with $\pi \times \nu$ is also an approximate 2 design.

The circuits in $\mathcal{C}^T$ depend on the choice of representation of those in $\mathcal{C}$, in terms Clifford+T+T$^\dagger$ gates, and we require this conjecture to hold only for any one choice of such representation, but could we further conjecture that then it will be true for any representation.

If Conjecture 3.7 holds, then for any $(\mathcal{C}, \pi)$ that is an $\epsilon$-approximate 2 design (and that satisfies the conditions of Theorem 3.4), Theorem 3.6 will apply.

4 Experimental feasibility of CM circuits

CM circuits offer several advantages for fault tolerant implementation and for implementation in the measurement based quantum computing model (MQC), inherited in part from such benefits for Clifford circuits.

4.1 Fault tolerance for CM circuits

In the circuit model, fault tolerance is often achieved by replacing $T$ gates by $T$ gadgets, with magic state distillation being used to create high fidelity $|A\rangle$ states offline.
A stabiliser code can be used to encode all states and Clifford operations. Syndrome measurements must be performed after every step, and corrected before the next step. In this framework CM circuits offer a number of advantages. The circuits comprise only Clifford gates, so no intermediate adaptive measurements for $T$ gadgets are introduced. Furthermore, although syndrome measurements still need to be performed, the associated adaptive correction operations (which are all Pauli operations) can be commuted past all the Clifford gates to the end of the circuit and accounted for there by simple reinterpretation of final measurement outcomes (viz. just flipping the bit for $X$ corrections and no action for $Z$ corrections). Thus the entire fault tolerant circuit is non-adaptive and there is no need to implement syndrome corrections as quantum gates, all providing potential savings on coherence time for the computation. A further benefit of CM circuits being Clifford circuits is that any such circuit on $t$ qubit lines can be expressed as a circuit of depth bounded by $O(t^2/\log(t))$ [2], again providing potential benefits for shorter coherence times in implementation, depending on the depth of the original circuit being simulated.

4.2 CM circuits in the MQC model

In our discussion below we will assume the following standard form of MQC (cf for example [13]). The starting resource state is the standard cluster state. $CZ$ operations in circuits are implemented by exploiting $CZ$’s that were used in the construction of the cluster state. 1-qubit measurements applied to the cluster state are either $Z$ measurements or else $M(\alpha)$ measurements in the basis $\{|\pm\alpha\rangle\}$, where $|\pm\alpha\rangle = 1/\sqrt{2}(|0\rangle \pm e^{-i\alpha}|1\rangle)$. The latter provide implementation of 1-qubit gates $J(\alpha) = H(|0\rangle\langle 0| + e^{i\alpha}|1\rangle\langle 1|)$, appearing as $X^s J(\alpha)$ where $s = 0, 1$ is the measurement outcome and $X^s$ is the associated byproduct operator. The $J(\alpha)$ gates together with $CZ$ provide a universal set.

**Theorem 4.1.** A CM circuit $C$ including preparation of its input $|A\rangle^\otimes t$, can be implemented in the MQC model in depth 1.

**Proof.** Note first that $|A\rangle = H J(\pi/4)|+\rangle$. Thus $C$ may be viewed as having input $|+\rangle$ on all lines, followed by a round of $J(\pi/4)$ gates, followed by Clifford gates (comprising a round of $H$ gates followed by the gates of $C$). Hence for MQC implementation the measurement pattern comprises a line of $M(\pi/4)$ measurements laid out next to implementations of Clifford gates. The $X^s$ byproducts of the $M(\pi/4)$ measurements can be commuted over the Clifford gates to the end, without incurring any adaptations. Similarly it is well known that Clifford circuits can be implemented without adaptation to the byproduct operators that arise. Hence the entire measurement pattern is non-adaptive and can be implemented in depth 1. 

Miller et al. [20] also propose a scheme for quantum supremacy without error correction that is depth 1 in MQC. Their scheme requires a nonstandard resource state that may not be simple to prepare, whereas our proposal uses the standard cluster state as the input. Furthermore our scheme can be made fault tolerant as follows.
**Theorem 4.2.** A CM circuit $C$ can be implemented fault tolerantly in the MQC model in depth 1, given a particular initial resource state that can be created offline with high fidelity.

**Proof.** For simplicity, we will consider a fault tolerance scheme using the 7-qubit Steane code. The initial resource state can be created as follows. Create an encoded magic state $\tilde{|A\rangle}^\otimes t$. Create the other parts of the encoded graph state by making the encoded states $|\tilde{+}\rangle$ and using the encoded version of $CZ$. The usual syndrome measurements and corrections are required during this process. Inclusion of $\tilde{|A\rangle}^\otimes t$ into the resource state allows us to avoid a later need for implementing encoded $M(\pi/4)$ measurements fault tolerantly, and our CM circuit is a circuit of only Clifford gates. Now we have $H = J(0)$ and $S = H J(\pi/2)$, with $M(0)$ and $M(\pi/2)$ being $X$ and $Y$ measurements respectively. Thus in MQC, Clifford gates are implemented using only Pauli measurements, and in our encoded setup we need to apply their corresponding fault tolerant encoded versions. These are transversal. Furthermore, syndrome measurements can be carried out using the usual fault tolerant construction in terms of Clifford operations and ancillas. These Clifford gates themselves can be implemented using MQC using ancillas. All these ancillas are included in the initial state. Hence every physical operation applied to the initial state is a 1 qubit Pauli measurement. Then, as before, Pauli errors can be corrected via classical post processing, and so the circuit is depth 1.

Finally we make a few further remarks on the classical simulatability of CM circuits. In [2] a classical algorithm is given for simulating stabiliser circuits with non stabiliser inputs. This algorithm requires time $O(t2^t)$ to simulate a CM circuit with $t$ lines. In [7] an improved classical algorithm is given for the case we are interested in, with run time $O(\text{poly}(t)2^{\beta t})$, where $\beta \approx 0.228$.

However, for a CM circuit on $t$ lines, it is possible to classically simulate $\log t$ lines in time $O(\text{poly}(t))$, using a method similar to that in the proof of Theorem 1 in [18]. In fact, this case can also be strongly simulated since the polynomially many outcome probabilities can even all be calculated in poly time. This is a useful fact for experimentally verifying that a CM circuit is outputting the correct outcomes. Interestingly, it seems that many schemes for quantum supremacy have this property, that a logarithmic number of lines are easy to classically simulate despite the fact that linearly many lines are provably hard to classically simulate given the relevant conjectures.

**Acknowledgements**

We thank M. Bremner and A. Montanaro for helpful discussions and clarifications. MY is supported by a Cambridge Australia Bragg Scholarship. RJ is supported in part by the QuantERA ERA-NET Cofund project QuantAlgo. SS is supported by a Leverhulme Early Career Fellowship. We acknowledge Mem Fox for suggesting terminology.
References

[1] S. Aaronson and A. Arkhipov. The computational complexity of linear optics. In Proceedings of the 43rd annual ACM symposium on Theory of computing - STOC ’11, page 333, New York, New York, USA, 2011. ACM Press.

[2] S. Aaronson and D. Gottesman. Improved simulation of stabilizer circuits. Physical Review A, 70(5):052328, 2004.

[3] S. Arora and B. Barak. Computational complexity: a modern approach. Cambridge University Press, 2009.

[4] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven. Characterizing Quantum Supremacy in Near-Term Devices. arXiv:1608.00263, 2016.

[5] A. Bouland, B. Fefferman, C. Nirkhe, and U. Vazirani. Quantum Supremacy and the Complexity of Random Circuit Sampling. arXiv: 1803.04402, mar 2018.

[6] A. Bouland, J. F. Fitzsimons, and D. E. Koh. Quantum Advantage from Conjugated Clifford Circuits. arXiv: 1709.01805, 2017.

[7] S. Bravyi and D. Gosset. Improved Classical Simulation of Quantum Circuits Dominated by Clifford Gates. Physical Review Letters, 116(25):250501, 2016.

[8] S. Bravyi, G. Smith, and J. A. Smolin. Trading Classical and Quantum Computational Resources. Physical Review X, 6(2):021043, 2016.

[9] M. J. Bremner, R. Jozsa, and D. J. Shepherd. Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 467(2126):459–472, 2011.

[10] M. J. Bremner, A. Montanaro, and D. J. Shepherd. Achieving quantum supremacy with sparse and noisy commuting quantum computations. arXiv:1610.01808, 2016.

[11] M. J. Bremner, A. Montanaro, and D. J. Shepherd. Average-Case Complexity Versus Approximate Simulation of Commuting Quantum Computations. Physical Review Letters, 117(8):080501, 2016.

[12] C. Dankert, R. Cleve, J. Emerson, and E. Livine. Exact and approximate unitary 2-designs and their application to fidelity estimation. Physical Review A, 80(1):012304, 2009.

[13] V. Danos, E. Kashefi, and P. Panangaden. The measurement calculus. Journal of the ACM, 54(2):8–es, 2007.

[14] B. Fefferman and C. Umans. On the Power of Quantum Fourier Sampling. arXiv:1507.05592, 2015.
[15] D. Gottesman. The Heisenberg representation of quantum computers. *arXiv: quant-ph/9807006*, 2008.

[16] D. Hangleiter, J. Bermejo-Vega, M. Schwarz, and J. Eisert. Anti-concentration theorems for schemes showing a quantum speedup. *arXiv: 1706.03786*, 2017.

[17] A. W. Harrow and A. Montanaro. Quantum computational supremacy. *Nature*, 549(7671):203–209, 2017.

[18] R. Jozsa and M. Van den Nest. Classical simulation complexity of extended Clifford circuits. *Quantum Information and Computation*, 14, 2013.

[19] R. L. Mann and M. J. Bremner. On the Complexity of Random Quantum Computations and the Jones Polynomial. *arXiv: 1711.00686*, 2017.

[20] J. Miller, S. Sanders, and A. Miyake. Quantum supremacy in constant-time measurement-based computation: A unified architecture for sampling and verification. *Physical Review A*, 96(6):062320, 2017.

[21] T. Morimae. Hardness of classically sampling the one-clean-qubit model with constant total variation distance error. *Physical Review A*, 96(4):040302, 2017.

[22] M. Nielsen and I. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 10th anniv edition, 2010.

[23] H. Pashayan, S. D. Bartlett, and D. Gross. From estimation of quantum probabilities to simulation of quantum circuits. *arXiv: 1712.02806*, 2017.

[24] J. Preskill. Lecture notes for Physics 219/Computer Science 219 Quantum Computation.

[25] T. J. Yoder. A generalization of the stabilizer formalism for simulating arbitrary quantum circuits. *www.scottaaronson.com/showcase2/report/ted-yoder.pdf*, 2012.