Thrifty shadow estimation: re-using quantum circuits and bounding tails

Jonas Helsen[1] and Michael Walter[2]
[1]*QuSoft and CWI, The Netherlands*
[2]*Faculty of Computer Science, Ruhr University Bochum, Germany*

Shadow estimation is a recent protocol that allows estimating exponentially many expectation values of a quantum state from “classical shadows”, obtained by applying random quantum circuits and computational basis measurements. In this paper we study the statistical efficiency of this approach in light of near-term quantum computing. We propose a more practical variant of the protocol, *thrifty shadow estimation*, in which quantum circuits are reused many times instead of having to be freshly generated for each measurement. We show that reuse is maximally effective when sampling Haar random unitaries, and maximally ineffective when sampling from the Clifford group, i.e., one should not reuse circuits when performing shadow estimation with the Clifford group. We provide an efficiently simulable family of quantum circuits that interpolates between these extremes, which we believe should be used instead of the Clifford group. Finally, we consider tail bounds for shadow estimation and discuss when median-of-means estimation can be replaced with standard mean estimation.

A key aspect of the development of larger-scale quantum computers is the availability of protocols that can diagnose errors and noise in quantum computations. Over the years many such protocols have been proposed, optimizing either for informational completeness (various forms of tomography) or sampling efficiency (e.g., randomized benchmarking [1, 2] or direct fidelity estimation [2]), but not achieving both at the same time. See [3] for a general overview. Recently Huang, Küng and Preskill (HKP) went beyond this apparent dichotomy by proposing *shadow estimation* [4], a randomized protocol which extracts *exponentially* many expectation values $\mathrm{Tr}(O\rho)$ from *polynomially* many copies of the state ρ, with the only caveat being a restriction on the set of allowed observables O [5]. Shadow estimation has generated significant interest and led to several theoretical follow-up works [6, 7] and experimental applications [8, 9] (see also [10] for a comprehensive overview). At its face value, the protocol is extremely simple: upon receiving a state ρ one generates a random n-qubit circuit U from a circuit set $\mathbb{U}$, applies it to the state ρ, and then measures in the computational basis, obtaining a bit string x. The tuple (U, x) then forms a so-called *classical shadow* of the state ρ, from which expectation values $\mathrm{Tr}(O\rho)$ can be reconstructed by classical post-processing (see Fig. 1 (a) for details). The performance of the protocol depends on the circuit set $\mathbb{U}$, as well as the observables one considers. An important case is when the circuit set is the multi-qubit Clifford group $\mathbb{C}_n$, which is a 3-design. In this case, shadow tomography is efficient for observables O for which $\mathrm{Tr}(O^2)$ is bounded. The Clifford group furthermore has the advantage that if the observable is, e.g., a projection onto a stabilizer state, then the classical post-processing needed is also *computationally efficient* by the Gottesman-Knill theorem [11], which is very useful in practice.

A key component of the HKP proposal is that *every classical shadow requires an independent random circuit*. This is critical to the mathematical argument for its statistical efficiency, but can be undesirable in practice. Especially in near-term quantum computers, it is preferable in many systems to measure a fixed circuit multiple times to generate a large number of classical shadows. This can already be seen in experimental implementations of shadow estimation such as [9], which reports repeating each circuit $> 10^4$ times, and [8], which reports measuring each circuit 10^3 times. This is likely inspired by experience with randomized benchmarking, which similarly samples random circuits (and where the statistics of repeating circuits is well understood [12, 13]). In this work we systematically study the effect of circuit repetition for shadow tomography, which to the best of our knowledge has not been studied before, using tools from representation theory. We also apply those tools to the question of whether median-of-means estimators, another key component of the HKP proposal, are actually necessary for shadow tomography.

THRIFTY SHADOW ESTIMATION

We introduce *thrifty shadow estimation*, our variant of shadow estimation that re-uses quantum circuits and can be significantly more economic in practice. The standard and thrifty protocols protocols are summarized in Fig. 1.

We write N for the total number of measurements, R for the number of times that a random circuit is re-used (including the first time) and K for the number of batches in the median-of-means estimator. We will assume that N is a multiple of KR throughout this section. Thus, the protocol uses N/R random quantum circuits. Note that thrifty shadow estimation reduces to ordinary shadow estimation for $R = 1$. To analyze its statistical performance, note that the thrifty estimator $\hat{o}_R$ is the median-of-means estimator for N/R many i.i.d. copies of the random variable

$$\mathbf{X}_R = \frac{1}{R}\sum_{r=1}^{R} \mathbf{X}^{(r)}, \tag{1}$$

where $\mathbf{X}^{(r)} = \mathrm{Tr}\left(O\mathcal{F}^{-1}(\mathbf{U}\,|\mathbf{x}^{(r)}\rangle\!\langle\mathbf{x}^{(r)}|\,\mathbf{U}^\dagger)\right)$, with $\mathbf{U}$ drawn uniformly at random from the circuit set $\mathbb{U}$, and $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(R)}$ drawn i.i.d. from the conditional distribution

$$p(x|U) = \langle x|\, U\rho U^\dagger\, |x\rangle\,.$$

(a) Original shadow estimation protocol [4]:

Data acquisition: For $t = 1, \dots, N$:

1. Draw a random circuit $\hat{U}_t \in \mathbb{U}$.
2. Prepare ρ, apply the unitary $\hat{U}_t$, and measure in the computational basis. Record the outcome $\hat{x}_t \in \{0,1\}^n$.

Prediction of expectation values:

1. For $k = 1, \dots, K$, compute the batch mean

$$\hat{\rho}_{(k)} = \frac{K}{N} \sum_{t=(k-1)\frac{N}{K}+1}^{k\frac{N}{K}} \mathcal{F}^{-1}\big(\hat{U}_t^\dagger \left|\hat{x}_t\middle\rangle\!\middle\langle\hat{x}_t\right| \hat{U}_t\big).$$

2. For $O \in \mathbb{O}$, output $\hat{o} = \mathrm{median}\,\{\mathrm{Tr}(O\hat{\rho}_{(k)})\}_{k=1}^{K}$.

(b) Thrifty shadow estimation (this work):

Data acquisition: For $t = 1, \dots, N/R$:

1. Draw a random circuit $\hat{U}_t \in \mathbb{U}$.
2. For $r = 1 \dots, R$: Prepare ρ, apply $\hat{U}_t$, and measure in the computational basis. Record the outcome $\hat{x}_{t,r} \in \{0,1\}^n$.

Prediction of expectation values:

1. For $k = 1, \dots, K$, compute the batch mean

$$\hat{\rho}_{(k)} = \frac{RK}{N} \sum_{t=(k-1)\frac{N}{RK}+1}^{k\frac{N}{RK}} \frac{1}{R} \sum_{r=1}^{R} \mathcal{F}^{-1}\big(\hat{U}_t^\dagger \left|\hat{x}_{t,r}\middle\rangle\!\middle\langle\hat{x}_{t,r}\right| \hat{U}_t\big).$$

2. For $O \in \mathbb{O}$, output $\hat{o}_R = \mathrm{median}\,\{\mathrm{Tr}(O\hat{\rho}_{(k)})\}_{k=1}^{K}$.

FIG. 1. (a) *The shadow estimation protocol of [4]:* A total of N measurements is performed, and each random circuit is used to obtain a single quantum measurement outcome. The parameter K corresponds to the number of batches in the median-of-means estimator. We assume that N is a multiple of K. The quantum channel $\mathcal{F}$ depends on the circuit set, and is given explicitly in Eq. (2). (b) *Thrifty shadow estimation as introduced in this work:* Each random circuit is re-used R times. We again use N to indicate the total number of measurements, and thus N/R random circuits are generated. The parameter K again corresponds to the number of batches of the median-of-means estimator. We assume here that N is a multiple of RK. Note that we sample at least one random circuit per batch, as required for the median-of-means estimator.

Finally, $\mathcal{F}^{-1}$ is the inverse of the quantum channel $\mathcal{F}$,

$$\mathcal{F}(A) := \sum_{x\in\{0,1\}^n} \mathbb{E}_{U\in\mathbb{U}} U^\dagger \left|x\middle\rangle\!\middle\langle x\right| U \left\langle a\right| UAU^\dagger \left|x\right\rangle, \qquad (2)$$

associated with the circuit set $\mathbb{U}$ and the computational basis POVM [14]. It can be shown that $\mathbb{E}(\mathbf{X}_R) = \mathrm{Tr}(O\rho)$.

This directly suggests the following protocol for estimating expectation values $\mathrm{Tr}(O\rho)$: sample R times from the distribution $p(x|U)$, compute the corresponding states $\mathcal{F}^{-1}(U^\dagger \left|x\middle\rangle\!\middle\langle x\right| U)$ and construct an estimator for the mean. Concretely, it was shown (for standard shadow tomography) in [4] that if one obtains N random samples $\{(\hat{U}_t, \hat{x}_t)\}_{t=1}^N$, corresponding to N independent random circuits, groups those into K equal-size batches, and computes the *median-of-means* estimator $\hat{o}$ as in Fig. 1 (a), then one can obtain with high probability an accurate estimate of the desired expectation value. This directly generalizes to thrifty shadow estimation.

In particular, if we set the batch size $K = \lceil 8\log(1/\delta)\rceil$ for $\delta \in (0,1)$, then [15, Theorem 2] implies that for any fixed observable O,

$$\left|\hat{o}_R - \mathrm{Tr}(O\rho)\right| \leq \sqrt{\frac{32\,\mathbb{V}_R(O,\rho)\log(1/\delta)}{N/R}},$$

with probability at least $1-\delta$, where $\mathbb{V}_R(O,\rho)$ is the variance of the random variable $\mathbf{X}_R$. Our first result characterizes this variance:

Lemma 1. *The variance of the random variable $\mathbf{X}_R$ is given by*

$$\mathbb{V}_R(O,\rho) = \frac{1}{R}\mathbb{V}(O,\rho) + \frac{R-1}{R}\mathbb{V}_*(O,\rho), \qquad (3)$$

where $\mathbb{V}(O,\rho)$ is the variance of the random variable $\mathbb{X}_1$, as in ordinary shadow estimation, while

$$\begin{aligned}\mathbb{V}_*(O,\rho) &:= \mathbb{V}(\mathbb{E}(\mathbf{X}_1|\mathbf{U})) \\ &= \mathbb{V}_U\big(\mathbb{E}_x \mathrm{Tr}\left(O\mathcal{F}^{-1}(U^\dagger \left|x\middle\rangle\!\middle\langle x\right| U)\right)\big).\end{aligned} \qquad (4)$$

For the sake of brevity we postpone the proof of this lemma and all following results to the Supplemental Material [16]. For $R = 1$ we recover the performance guarantee of ordinary shadow estimation [4]. To analyze the thrifty case, $R > 1$, we need to estimate the term $\mathbb{V}_*(O,\rho)$, which depends on the fourth moment of the random circuits. A straightforward corollary of Lemma 1 (using the law of total variance) is

$$\mathbb{V}_R(O,\rho) \leq \mathbb{V}(O,\rho).$$

However, this does *not* imply that thrifty shadow estimation is always better than ordinary shadow estimation. In fact the above argument allows for a range of possibilities, going from $\mathbb{V}_R(O,\rho) \approx \mathbb{V}(O,\rho)/R$, in which case thrifty shadow estimation recovers the guarantees of ordinary shadow estimation for the same number of measurements (but might be preferable due to the lower cost of circuit reuse, as discussed in the introduction), to $\mathbb{V}_R(O,\rho) \approx \mathbb{V}(O,\rho)$, in which case setting $R > 1$ would be useless. We will see that both scenarios arise naturally when one performs thrifty shadow estimation with a unitary 4 design or the multiqubit Clifford group respectively. We also give a parametrized family of circuit models that elegantly interpolates between these extremes.

Unitary 4-designs. We begin by analyzing the variance of thrifty shadow estimation for any circuit set that is a unitary 4-design. Our objective is to calculate Eq. (3). We are interested in the limit of many qubits, meaning we will be happy with estimates that include $\mathcal{O}(2^{-n})$ terms in all expressions. We obtain the following theorem.

Theorem 2. *The variance of thrifty shadow estimation with any 4-design circuit set satisfies*

$$\mathbb{V}_R(O,\rho) = \frac{1}{R}\mathbb{V}(O,\rho) + \frac{R-1}{R}\mathcal{O}(2^{-n}\operatorname{Tr}(O^2))$$

for any traceless observable O, with $\mathbb{V}(O,\rho)$ the variance associated with standard shadow estimation.

The proof of this theorem follows from Schur-Weyl duality for the unitary group, which is matched by any unitary 4 design up to fourth order expressions (such as the variance). A primer on Schur-Weyl duality (in this context also referred to as Weingarten calculus [17]) can be found in the Supplemental Material [16]. We know that $\mathbb{V}(O,\rho) \approx \operatorname{Tr}(O^2)$, which means shadow estimation is scalable precisely when $\operatorname{Tr}(O^2)$ is polynomially bounded. In this case Theorem 2 tells us that $\mathbb{V}_R(O,\rho)$ and $\mathbb{V}(O,\rho)/R$ are exponentially close in the number of qubits – in other words, circuit reuse essentially does not impact the statistical accuracy of shadow estimation. This means that thrifty shadow tomography, with access to a unitary 4-design will in practice be preferable to standard shadow tomography. However, demanding access to an exact unitary 4-design is a strong requirement, which we address in more detail later.

Multi-qubit Cliffords. Next we will *lower bound* the variance of thrifty shadow estimation for the multi-qubit Clifford group. In particular we will show that there are states ρ and observables O such that the variance $\mathbb{V}_R(O,\rho)$ in Eq. (3) is *independent* of R (in the limit of many qubits). Concretely:

Theorem 3. *Consider thrifty shadow estimation with the n-qubit Clifford group $\mathbb{C}_n$. For any pure stabilizer state $\rho = |S\rangle\langle S|$ and the traceless observable $O = |S\rangle\langle S| - 2^{-n}I$, we have*

$$\mathbb{V}_R(O,\rho) = 2 + \mathcal{O}(2^{-n}).$$

The proof of this theorem hinges on the recently developed Schur-Weyl duality theory for the Clifford group [18] (see also[19–21]) and can be found in the Supplemental Material [16]. There is thus a striking divergence in behavior between the Clifford group and and 4-design when it comes to re-using circuits. This result formalizes an observation already made in experiment [9], and serves as a warning for future experiments using this circuit set.

An interpolating family. From the preceding results one would prefer to perform thrifty shadow estimation with 4-design circuits where circuit reuse is maximally useful. Unfortunately no exact constructions of unitary 4-designs are known (for an arbitrary number of qubits n). Moreover the Clifford group is not only useful due to its statistical properties, but also because it allows for the estimator $\hat{o}_R$ to be computed efficiently in classical post-processing whenever the associated observable is a stabilizer state or a Pauli operator (or a well-behaved combination of these). This is a property we would like to preserve as much as possible. With this in mind, and inspired by [22], we consider a family of circuit sets that interpolates between the extreme cases discussed above. Recall that the T-gate is the non-Clifford unitary

$$\mathrm{T} = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix}.$$

For a system of n qubits, we denote by T this gate but acting on the first of n qubits. Then we can consider the following finite set of quantum circuits for any natural number k:

$$\mathbb{U}_k = \{C_k \mathrm{T} C_{k-1} \cdots \mathrm{T} C_0 \quad | \quad C_0, \dots, C_k \in \mathbb{C}_n\}, \qquad (5)$$

This set is at least a 3-design for any k. With increasing k, it is an approximate t-design of any order [22]. Moreover, computing classically the overlap $\operatorname{Tr}(OU|x\rangle\langle x|U^\dagger)$ for stabilizer O and $U \in \mathbb{U}_k$, which is required for computing the estimator $\hat{o}$, can be achieved in time $O(2^{0.396k})$ [23].

We now show that in the limit of large system sizes n, the variance of thrifty shadow estimation with the circuit set $\mathbb{U}_k$ approaches the result for 4-designs, which is $\mathbb{V}_R(O,\rho) \approx \mathbb{V}(O,\rho)/R$ for observables of bounded Hilbert-Schmidt norm (Theorem 2), up to an error that decreases exponentially with k, leading to a classical simulation cost that is inverse polynomial in the desired error.

Theorem 4. *The variance of thrifty shadow estimation with the circuit set $\mathbb{U}_k$ defined in Eq. (5) satisfies*

$$\begin{aligned}\mathbb{V}_R(O,\rho) - \frac{1}{R}\mathbb{V}(O,\rho) \leq & \frac{R-1}{R}\mathcal{O}(2^{-n}\operatorname{Tr}(O^2)) \\ & + \frac{R-1}{R} 30\operatorname{Tr}(O^2)\big(1+\mathcal{O}(2^{-n})\big) \\ & \times \left(\frac{3}{4} + \mathcal{O}(2^{-n})\right)^k\end{aligned}$$

for any traceless observable O, with $\mathbb{V}(O,\rho)$ the variance associated with standard shadow estimation.

While our result is inspired by [22], we do not know how to deduce it directly from their approximate 4-design result. The reason for this is again that the support of the shadow estimation probability distribution $p(U,x)$ grows as $\mathcal{O}(2^n)$, and hence any additive error term will blow up correspondingly.

The advantage of thrifty shadow estimation with the interpolating family is best seen by considering a simple cost model. Set $\operatorname{Tr}(O^2) = 1$ for simplicity, and assume that generating a new random circuit has cost $\alpha \geq 1$ and re-using it has unit cost. Then we can express the cost C for a total of N samples as $\mathrm{C} = (N/R)(\alpha + R - 1)$. When using a median of means estimator, the accuracy of thrifty (or standard for $R = 1$) shadow estimation with N samples is proportional to $\mathbb{V}_R/(N/R)$ (provided $N \geq K\lceil 8\log(1/\delta)\rceil$, so that the estimator is well-defined). We can express this in terms of the total cost C as

$$\frac{\mathbb{V}_R}{N/R} = \frac{\alpha + R - 1}{\mathrm{C}}\mathbb{V}_R. \qquad (6)$$

This can be minimized to obtain the optimal number of repetitions R for a fixed random circuit generation cost α and a fixed total cost budget C. For the homeopathic circuit family with k T-gates, using Theorem 4 and taking the maximal possible value of $\mathbb{V}_R$, Eq. (6) reads (suppressing small factors)

$$\frac{\mathbb{V}_R}{N/R} \approx \frac{1}{\mathrm{C}}\left[\left(\frac{\mathbb{V}_1}{R} + 30\,\frac{R-1}{R}\left(\frac{3}{4}\right)^k\right)(\alpha + R - 1)\right],$$

leading to an optimal choice of R (every value of R corresponds to a value of N at fixed cost C) given by

$$R \approx \sqrt{\frac{(1-\alpha)|(\mathbb{V}_1 - 30\,(3/4)^k)|}{30\,(3/4)^k}}.$$

This implies (if $\mathbb{V}_1 \neq 0$ and $\alpha > 1$) that for any value of α and C, there is a value of k such that the optimal choice of R is $R = N/K$. (accounting for the batching requirement in the median-of-means estimator), corresponding to a protocol where one samples a single circuit per batch and repeats it N/K times. The computational cost of strongly simulating a quantum circuit with k many T-gates currently [24] scales as $\mathcal{O}(2^{0.396k})$ [23]. Hence the optimal value of R scales roughly with an inverse square $(2\log(3/4)/0.396 \approx -2.08)$ with the cost of simulation. This means that thrifty shadow tomography with maximal reuse can be implemented using a circuit set that requires only polynomially more classical computational resources as compared (traditional) shadow tomography with the multi-qubit Clifford group. We emphasize that this is a heuristic calculation, since we are ignoring some small terms in the expression for the homeopathic variance. In particular it is only accurate in the regime of many qubits (when the $O(2^{-n})$ terms are small). However it shows that thrifty shadow estimation using the homeopathic interpolation circuit set can be a powerful alternative to standard shadow estimation when the cost of generating new circuits is high.

Finally, we note that sampling at least one circuit per batch is a requirement for the median-of-means estimator to function. To see this, consider the extreme scenario where only one random circuit is sampled, and repeated many times, with the measurement outcomes grouped into batches as above. As the number of repetitions increases, the median-of-means estimator will converge to the average value for the single random circuit, with the only remaining randomness due to circuit choice. However, this randomness can still be ill-behaved (in the sense that the distribution is heavy-tailed), precluding exponential concentration of the estimator, which is required for shadow tomography.

TAIL BOUNDS FOR SHADOW ESTIMATION

In this section we revisit the use of median-of-means estimation in shadow estimation with circuit sets that are (at least) 3-designs. It has been noted before that for the circuit sets of single-qubit Clifford gates and matchgates, the median-of-means estimator can be replaced by the standard mean estimator without a loss in performance [6]. In this section we show that this is also true if the circuit set is the entire unitary group, but is not true if the circuit set is the multi-qubit Clifford group. Hence, just like for thrifty shadow estimation, the Clifford group fails to fully emulate the statistical behavior of Haar random unitaries.

For simplicity we consider shadow estimation but both results also hold for thrifty shadow estimation with $R > 1$. Throughout this section we will write $\mathbf{X}_n = \mathbf{X}$ to explicitly indicate the number of qubits n in a subscript.

Unitary group We first consider shadow estimation with the full unitary group. Somewhat surprisingly, we can prove sub-exponential behavior for shadow tomography with the standard mean estimator.

Theorem 5. *Consider shadow estimation with the n-qubit unitary group as circuit set, state ρ, and traceless observable O. For N i.i.d. copies $\mathbf{X}_n^{(1)}, \dots \mathbf{X}_n^{(N)}$ of $\mathbf{X}_n$, we have a Bernstein-like tail bound:*

$$\mathbb{P}\left(\left|\frac{1}{N}\sum_{i=1}^{N}\mathbf{X}_n^{(i)} - \mathbb{E}(\mathbf{X}_n)\right| \geq \varepsilon\right) \leq \begin{cases} 2\exp\!\left(-\frac{N\varepsilon^2}{48\|O\|_{HS}^2}\right) \\ \qquad \textit{if } \varepsilon \leq 12\,\|O\|_{HS}, \\ 2\exp\!\left(-\frac{N\varepsilon}{4\|O\|_{HS}}\right) \\ \qquad \textit{if } \varepsilon > 12\,\|O\|_{HS}. \end{cases}$$

This theorem again follows from Schur-Weyl duality for the unitary group and a careful accounting of the moment generating function of $\mathbb{X}_n$.

Theorem 5 shows that the median-of-means estimator in [4] can in principle be replaced by a standard empirical average, as long as one uses the full unitary group as the circuit set. This is akin to earlier such results for the single-qubit Clifford and matchgate groups. However, these earlier statements were a consequence of the fact that the distributions being sampled from are bounded for all n (independently of n, in terms of some separate locality parameter k), and it is hence not surprising that an exponential tail bound can be established. On the other hand, in the case of shadow estimation with the unitary group (or any 3-design), the support of the distribution diverges as $n \to \infty$, making such a statement significantly less trivial. We believe that this exponential tail behavior is fundamentally a property of the full unitary group, making it difficult to achieve in practice.

Clifford group We will now argue that the opposite behavior holds when one averages over the multi-qubit Clifford group instead. By this we mean that for shadow estimation with the Clifford group no "useful" tail bounds are possible. This is a somewhat awkward statement to make, as for any finite number of qubits n the distribution associated with Clifford shadow estimation is bounded on the interval $[-2^n, 2^n]$ (for all input states and observables), so it is always possible to obtain exponential tail bounds that grow exponentially in n. In Theorem 6 below we show that, roughly speaking, one cannot do better (even if $\mathrm{Tr}(O^2)$ is bounded).

Theorem 6. *Consider shadow estimation with the n-qubit Clifford group as circuit set, any n-qubit stabilizer state $\rho = |S\rangle\langle S|$, and the observable $O = |S\rangle\langle S| - 2^{-n}I$, so that $\mathrm{Tr}(O^2) \leq 1$. Suppose that the random variables $\mathbf{X}_n$ satisfies a tail bound of the form*

$$\mathbb{P}(|\mathbf{X}_n - \mathbb{E}(\mathbf{X}_n)| \geq t) \leq A \exp\left(-\frac{t^\beta}{B_n}\right), \tag{7}$$

for constants $A, \beta > 0$ and a positive sequence (B_n). Then we have that $B_n = \tilde{\Omega}(2^{\beta n/4})$.

The key technical tool in this proof is a characterization of the m-th moments of $\mathbb{X}_n$, through the Schur-Weyl duality theory for the Clifford group. Concretely we obtain the following formula, which might be of independent interest:

$$\mathbb{E}(\mathbf{X}_n^m) = (2^n+1)^m \sum_{k=0}^{m} \binom{m}{k} (-1)^{m-k} 2^{-n(m-k)} \prod_{\ell=0}^{k-1} \frac{2^\ell+1}{2^\ell+2^n}. \tag{8}$$

One can see that the moments $\mathbb{E}(\mathbf{X}_n^m)$ of $\mathbf{X}_n$ grow fast with m, and moreover increasingly so as n increases. This is key to the proof of Theorem 6 Note however that for fixed n the growth of the moments levels off when $m \gg n$ (since $\mathbf{X}_n$ is ultimately a bounded random variable). Hence it is natural to discuss the behavior of the moments as we let n tend to infinity for fixed m. In the case of the unitary group (as we saw in the proof of Theorem 5), the limiting moments grow slowly enough with m to uniquely define a limiting random variable, with moments that are the limits of the moments of the random variables at finite n. However, this is not the case for the multi-qubit Clifford group. In the limit we have

$$\lim_{n\to\infty} \mathbb{E}(\mathbf{X}_n^m) = \sum_{k=0}^{m} \binom{m}{k} (-1)^{m-k} \prod_{\ell=0}^{k-1} (2^\ell+1).$$

For $m \geq 6$, the right-hand side this can be lower bounded as follows:

$$\lim_{n\to\infty} \mathbb{E}(\mathbf{X}_n^m) \geq 2^{\frac{m(m-1)}{2}} = \Omega(2^{m^2/2})$$

which shows that the moments grow super exponentially. In fact they grow so fast that any random variable with those moments would have a moment generating function with convergence radius zero (and would hence be genuinely heavy tailed). However the moments grow so fast it is not even clear whether the limiting moments determine a unique probability distribution.

Acknowledgments. We would like to thank Richard Küng for the discussion that kick-started this work. We also thank Harold Nieuwboer and Freek Witteveen for useful conversations. After a preprint of this article was posted, Ref. [25] reported on independent related results. JH is partially supported by the Dutch Research Council (NWO) through the Gravitation grant Quantum Software Consortium, 024.003.037. MW acknowledges support by the European Research Council (ERC) through ERC Starting Grant 101040907-SYMOPTIC, the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2092 CASA - 390781972, the Federal Ministry of Education and Research (BMBF) through project Quantum Methods and Benchmarks for Resource Allocation (QuBRA), and NWO grant OCENW.KLEIN.267.

[1] E. Magesan, J. M. Gambetta, and J. Emerson, Scalable and robust randomized benchmarking of quantum processes, Physical Review Letters **106**, 180504 (2011).

[2] J. Helsen, I. Roth, E. Onorati, A. H. Werner, and J. Eisert, General framework for randomized benchmarking, PRX Quantum **3**, 020357 (2022).

[3] J. Eisert, D. Hangleiter, N. Walk, I. Roth, D. Markham, R. Parekh, U. Chabaud, and E. Kashefi, Quantum certification and benchmarking, Nature Reviews Physics **2**, 382 (2020).

[4] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, Nature Physics **16**, 1050 (2020).

[5] The name being derived from a more theoretical proposal due to Aaronson [26].

[6] A. Zhao, N. C. Rubin, and A. Miyake, Fermionic partial tomography via classical shadows, Physical Review Letters **127**, 110504 (2021).

[7] H.-Y. Hu, S. Choi, and Y.-Z. You, Classical shadow tomography with locally scrambled quantum dynamics, arXiv preprint arXiv:2107.04817 (2021), 2107.04817.

[8] W. J. Huggins, B. A. O'Gorman, N. C. Rubin, D. R. Reichman, R. Babbush, and J. Lee, Unbiasing fermionic quantum Monte Carlo with a quantum computer, Nature **603**, 416 (2022).

[9] G. Struchalin, Y. A. Zagorovskii, E. Kovlakov, S. Straupe, and S. Kulik, Experimental estimation of quantum state properties from classical shadows, PRX Quantum **2**, 010307 (2021).

[10] A. Elben, S. T. Flammia, H.-Y. Huang, R. Kueng, J. Preskill, B. Vermersch, and P. Zoller, The randomized measurement toolbox, Nature Reviews Physics 10.1038/s42254-022-00535-2 (2022).

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

[12] J. J. Wallman and S. T. Flammia, Randomized benchmarking with confidence, New Journal of Physics **16**, 103032 (2014).

[13] J. Helsen, J. J. Wallman, S. T. Flammia, and S. Wehner, Multiqubit randomized benchmarking using few samples, Physical Review A **100**, 032304 (2019).

[14] The quantum channel $\mathcal{F}$ is also called a *frame operator*. While it need not be invertible in general, it is invertible for all circuit sets considered in this Letter (as well as those considered in all other shadow tomography protocols that we are aware of).

[15] G. Lugosi and S. Mendelson, Mean estimation and regression under heavy-tailed distributions: A survey, Foundations of Computational Mathematics **19**, 1145 (2019).

[16] See Supplemental Material at [URL will be inserted by publisher] for detailed proofs of some technical results.

[17] B. Collins, S. Matsumoto, and J. Novak, The Weingarten calculus, Notices of the AMS **69**, 734 (2022).

[18] D. Gross, S. Nezami, and M. Walter, Schur-Weyl duality for the Clifford group with applications: Property testing, a robust Hudson theorem, and de Finetti representations, Communications in Mathematical Physics , 1 (2021).

[19] H. Zhu, R. Kueng, M. Grassl, and D. Gross, The Clifford

group fails gracefully to be a unitary 4-design, arXiv preprint arXiv:1609.08172 (2016), 1609.08172.

[20] J. Helsen, J. J. Wallman, and S. Wehner, Representations of the multi-qubit Clifford group, Journal of Mathematical Physics **59**, 072201 (2018).

[21] F. Montealegre-Mora and D. Gross, Rank-deficient representations in the theta correspondence over finite fields arise from quantum codes, Representation Theory **25**, 193 (2021).

[22] J. Haferkamp, F. Montealegre-Mora, M. Heinrich, J. Eisert, D. Gross, and I. Roth, Quantum homeopathy works: Efficient unitary designs with a system-size independent number of non-clifford gates, arXiv preprint arXiv:2002.09524 (2020), 2002.09524.

[23] H. Qassim, H. Pashayan, and D. Gosset, Improved upper bounds on the stabilizer rank of magic states, Quantum **5**, 606 (2021).

[24] This is a subject of active research, and thus this scaling might improve further.

[25] Y. Zhou and Q. Liu, Performance analysis of multi-shot shadow estimation, arXiv preprint arXiv:2212.11068 (2022).

[26] S. Aaronson, Shadow tomography of quantum states, in *Proceedings of the 50th Annual ACM SIGACT STOC* (2018) pp. 325–338.