Implementation of quantum measurements using classical resources and only a single ancillary qubit

Tanmay Singal,1,2,* Filip B. Maciejewski,2,† and Michał Oszmaniec2,‡

1Department of Analysis, Budapest University of Technology and Economics, 1111 Budapest, Egry József u. 1, Hungary
2Center for Theoretical Physics, Polish Academy of Sciences, Al. Lotników 32/46, 02-668 Warsaw, Poland

We propose a scheme to implement general quantum measurements, also known as Positive Operator Valued Measures (POVMs) in dimension $d$ using only classical resources and a single ancillary qubit. Our method is based on probabilistic implementation of $d$-outcome measurements which is followed by postselection of some of the received outcomes. We conjecture that success probability of our scheme is larger than a constant independent of $d$ for all POVMs in dimension $d$. Crucially, this conjecture implies the possibility of realizing arbitrary nonadaptive quantum measurement protocol on $d$-dimensional system using a single auxiliary qubit with only a constant overhead in sampling complexity. We show that the conjecture holds for typical rank-one Haar-random POVMs in arbitrary dimensions. Furthermore, we carry out extensive numerical computations showing success probability above a constant for a variety of extremal POVMs, including SIC-POVMs in dimension up to 1290. Finally, we argue that our scheme can be favorable for experimental realization of POVMs, as noise compounding in circuits required by our scheme is typically substantially lower than in the standard scheme that directly uses Naimark’s dilation theorem.

Quantum measurements recover classical information stored in quantum systems and, as such, constitute an essential part of virtually any quantum information protocol. Every physical platform has its native measurements that can be realized with relative ease. In many cases, the class of easily implementable measurements contains projective (von Neumann) measurements. However, there are numerous applications [1–9] in which more general quantum measurements, so called Positive-Operator-Valued Measures (POVMs), need to be implemented. Implementation of these measurements requires additional resources. A recent generalization [10] of Naimark’s dilation theorem [11] showed that the most general measurement on $N$ qubits requires $N$ auxiliary qubits, when projective measurements can be implemented on the combined system in a randomized manner.

From the perspective of implementation in near-term quantum devices [12], it is desirable to implement arbitrary POVMs with fewer resources. Particularly, one would like to reduce the number of auxiliary qubits needed to implement a complex quantum measurement. A related problem is to quantify the relative power that generalized measurements in $d$-dimensional quantum systems have with respect to projective measurements in the same dimension. While POVMs appear as natural measurements for a variety of quantum information tasks: quantum state discrimination [13], quantum tomography [14–16], multiparameter metrology [17, 18], randomness generation [19], entanglement [20] and nonlocality detection [21], hidden subgroup problem [22, 23], port-based-teleportation [24–26], to name just a few. It is, however, not clear in general what quantitative advantage the more complex measurements offer over their simpler projective counterparts. This is because of the possibility to realize non-projective quantum measurements via randomization and post-processing of simpler measurements [10, 27–32]. Specifically, taking convex combinations of projective measurements can result in implementation of a priori quite complicated nonprojective POVMs [10, 32].

In this work we advance understanding of the relative power between projective and generalized measurements by focusing on a simpler problem, namely the relation between $d$-outcome POVMs and general (with arbitrary number of outcomes) POVMs acting on a $d$-dimensional Hilbert space $\mathcal{H} \approx \mathbb{C}^d$. We find a strong evidence that general quantum measurements do not offer an asymptotically increasing advantage over $d$-outcome POVMs for general quantum state discrimination problems [13], as $d$ tends to infinity. Specifically, we generalize the method of POVM simulation from [32] based on randomized implementation of restricted-class POVMs, followed by post-processing and postselection (defined later, see also Fig. 1). Here by postselection we mean disregarding certain measurement outcomes and accepting only the selected ones. In [32] it was shown that postselection allows to implement arbitrary POVM on $\mathbb{C}^d$ using only projective measurements and classical resources. This, however, comes with a cost - the method outputs a sample from a target quantum measurement with success probability $q_{\text{succ}} = \frac{1}{d}$. In this work we find that, surprisingly, there exists a protocol that allows to simulate a very broad class of POVMs on $\mathbb{C}^d$ via $d$-outcome POVMs and postselection with success probability $q_{\text{succ}}$ above a constant which is independent on the dimension $d$. Importantly, our construction ensures $d$-outcome POVMs used in the simulation can be implemented using projective measurements in Hilbert space of dimension $2d$. Therefore, our method gives a way to implement quantum measurements on $\mathbb{C}^d$ using only a single auxiliary qubit and projective measurements with constant success probability. We note that there exist schemes implementing arbitrary POVMs on $\mathbb{C}^d$ using a sequence of von Neumann instruments (i.e., a description of quantum measurements which includes post-measurement state of the system) on a system extended by a single auxiliary qubit [33, 34]. Our method is potentially simpler to implement as, in a given round of the experiment, only a single projective measurement has to be realized on the extended system and post-measurement states need not to be considered.

While we do not prove that the success probability $q_{\text{succ}}$ of our scheme is lower bounded by a dimension-independent constant for any POVMs on $\mathbb{C}^d$, we give strong evidence that this is indeed the case. First, we prove that for generic $d$-outcome Haar-random rank-one POVMs in $\mathbb{C}^d$ [35] the success probability is above 6.5% (numerically we observe $\approx 25\%$). We
also support our conjecture by numerically studying specific examples of symmetric informationally complete POVMs (SIC-POVMs) \[36–38\] and for a class of nonsymmetric informationally complete POVMs \[39\] (IC-POVMs), both for dimensions up to 1299. As the dimension increases, we observe that the success probability \(q_{\text{succ}}\) both for SIC-POVMs and IC-POVMs is \(\approx 1/5\). Importantly, if true, our conjecture implies that any non-adaptive measurement protocol can be realized using only single ancillary qubit with a sampling overhead that does not depend on the system size.

Finally, our scheme gives a possibility of more reliable implementation of complicated POVMs in noisy quantum devices. To support this claim, we employ the noise model used in Google’s recent demonstration of quantum computational advantage \[40\]. We make the following comparison between our method and the standard Naimark’s scheme of POVM implementation: for implementing typical random POVMs on \(N\) qubits, the fidelity of circuits which implement our scheme is exponentially higher than for Naimark’s implementation. This is due to the lower number of ancillary qubits required.

**Preliminaries**—We start by introducing notation and the concepts necessary to explain our POVM implementation scheme. We will be studying generalized quantum measurements on \(d\)-dimensional Hilbert space \(\mathcal{H} \approx \mathbb{C}^d\). An \(n\)-outcome POVM, is an \(n\)-tuple of linear operators on \(\mathbb{C}^d\) (usually called effects), i.e., \(\mathbf{M} = (M_1, M_2, \ldots, M_n)\), satisfying \(M_i \geq 0\) and \(\sum_{i=1}^{n} M_i \geq \mathbb{1}\), where \(\mathbb{1}\) is identity on \(\mathbb{C}^d\). A POVM \(\mathbf{P} = (P_1, \ldots, P_n)\) is called projective if all its effects satisfy the following relations: \(P_i P_j = \delta_{ij} P_i\). Measurement of \(\mathbf{M}\) on a quantum state \(\rho\) results in a random outcome \(i\), distributed according to the Born rule \(\text{Pr}(i|\rho, \mathbf{M}) = \text{Tr}(\rho M_i)\). We will denote the set of all \(n\)-outcome POVMs by \(\mathcal{P}(d,n)\). The set \(\mathcal{P}(d,n)\) is convex \[30\]: for \(\mathbf{M}, \mathbf{N} \in \mathcal{P}(d,n)\), and \(p \in [0,1]\) we define \(p\mathbf{M} + (1-p)\mathbf{N}\) to be an \(n\)-outcome POVM with the \(i\)-th effect given by \([p\mathbf{M} + (1-p)\mathbf{N}])_i = p M_i + (1-p) N_i\). A convex mixture \(p\mathbf{M} + (1-p)\mathbf{N}\) can be operationally interpreted as a POVM realized by applying, in a given experimental run, measurements \(\mathbf{M}, \mathbf{N}\) with probabilities \(p\) and \(1-p\) respectively. A POVM \(\mathbf{M} \in \mathcal{P}(d,n)\) is called extremal if it cannot be decomposed as a nontrivial convex combination of other POVMs.

Another classical operation that can be applied to POVMs is classical post-processing \[29, 41\]: given a POVM \(\mathbf{M}\), we obtain another POVM \(\mathbf{Q}(\mathbf{M})\) by probabilistically relabeling the outcomes of the measurement \(\mathbf{M}\). Effects of \(\mathbf{Q}(\mathbf{M})\) are given by \(\mathbf{Q}(\mathbf{M})_i = \sum_j q_{ij} M_j\), where \(q_{ij}\) are conditional probabilities, i.e., \(q_{ij} \geq 0\) and \(\sum_i q_{ij} = 1\). Lastly, postselection, i.e., the process of disregarding certain outcomes can be used to implement otherwise inaccessible POVMs. We say that a POVM \(\mathbf{L} = (L_1, \ldots, L_{n-1})\) simulates a POVM \(\mathbf{M} = (M_1, \ldots, M_n)\) with postselection probability \(q\) if \(L_i = q M_i\) for \(i = 1, \ldots, n\). This nomenclature is motivated by realizing that when we implement \(\mathbf{L}\), then, conditioned on getting the first \(n\) outcomes, we obtain samples from \(\mathbf{M}\). Thus, we can simulate \(\mathbf{M}\) by implementing \(\mathbf{L}\) and post-selecting on non-occurring outcome \(n+1\). The probability of successfully doing so is \(q\) which means that a single sample of \(\mathbf{M}\) is obtained by implementing \(\mathbf{L}\) on average \(1/q\) number of times. The reader is referred to \[32\] for a more detailed discussion of simulation via post-selection.

We will use \(|A|\) to denote the operator norm of a linear operator \(A\), and \([n]\) to denote \(n\)-element set \(\{1, \ldots, n\}\). Moreover, we will use \(\mu_n\) to refer to Haar measure on \(n\)-dimensional unitary group \(U(n)\), and by \(\mathbb{P}_{U \sim \mu}(A)\) we will denote probability of occurrence of an event \(A\) according to this probability measure. Finally, for two positive-valued functions \(f(x), g(x)\) we will write \(f = \Theta(g)\) if there exist positive constants \(c, C > 0\) such that \(c f(x) < g(x) < C f(x)\), for sufficiently large \(x\).

**General POVM simulation protocol**—The following theorem gives a general lower bound on the success probability of simulation of \(n\)-outcome POVMs via measurements with bounded number of outcomes and postselection.

![Diagram](image-url)
Theorem 1. Let $M = (M_1, M_2, \ldots, M_n)$ be an $n$-outcome POVM on $\mathbb{C}^d$. Let $m \leq d$ be a natural number and let $\{X_1, \ldots, X_m\}$ be a partition of $[n]$ into disjoint subsets $X_\gamma$ satisfying $|X_\gamma| \leq m - 1$. Then, there exists a simulation scheme that uses measurements having at most $m$ outcomes, classical randomness and post-selection that implements $M$ with success probability

$$q_{\text{succ}} = \left( \sum_{\gamma=1}^{m} \left| \sum_{i \in X_\gamma} M_i \right| \right)^{-1}.$$  

Furthermore, if rank($M_i$) $\leq 1$, and $m \leq d$, then measurements realizing the scheme can be implemented by projective measurements in dimension $2d$, i.e., using a single auxiliary qubit.

Proof. In what follows we give an explicit simulation protocol that generalizes earlier result from [32, 42] that concerned the case of simulation via dichotomic measurements ($m = 2$). The idea of the scheme is given in Fig. 1. We start by defining, for every element $X_\gamma$ of the partition, auxiliary measurements $N_{X_\gamma}$, each having $m + 1$ outcomes, whose purpose is to "mimick" measurement $M$ for outputs belonging to $X_\gamma$ and collect other (i.e., not belonging to $X_\gamma$) results in the "trash" output labelled by $n + 1$. Effects of $N_{X_\gamma}$ are defined by $N_{X_\gamma}^X = \lambda_i M_i$ for $i \in X_\gamma$, and collect other (i.e., not belonging to $X_\gamma$) results in the "trash" output labelled by $n + 1$. Effects of $N_{X_\gamma}$ are defined by $N_{X_\gamma}^X = \lambda_i M_i$ for $i \in X_\gamma$, and $N_{X_\gamma}^{n+1} = 1 - \lambda_\gamma \sum_{i \in X_\gamma} M_i$, where $\lambda_\gamma = \| \sum_{i \in X_\gamma} M_i \|^{-1}$.

We then define a probability distribution $\{ q_{\text{succ}} \}_{\gamma=1}^{m}$. The simulation of $M$ is realized by considering a convex combination of $N_{X_\gamma}$, according to this distribution: $L = \sum_{\gamma=1}^{m} q_{\text{succ}} N_{X_\gamma}$. An explicit computation shows that we have $L_i = q_{\text{succ}} M_i$, for $i \in [n]$ and therefore $L$ simulates the target measurement $M$ with success probability $q_{\text{succ}}$.

Finally, each of the measurements $N_{X_\gamma}$ comprising $L$ has at most $|X_\gamma| + 1$ nonzero effects and therefore they can be implemented with POVMs with at most $m$ outcomes. From the standard Naimark scheme of implementation of POVMs (c.f. [11]) we see that the dimension needed to implement a POVM $N_{X_\gamma}$ via projective measurements equals at most $m$ and $m+1$ ranks of effects of $N_{X_\gamma}$. In the case of rank-one $M$ and $|X_\gamma| \leq m - 1$ this sum for each $N_{X_\gamma}$ is at most $d + m - 1 \leq 2d$, which completes the proof.

Crucially, we recall that an arbitrary quantum measurement on $\mathbb{C}^d$ can be implemented by a convex combination of rank-one POVMs having at most $d^2$ outcomes followed by suitable post-processing [27, 30]. This implies that our protocol facilitates the simulation of any POVM on $\mathbb{C}^d$ using only a single ancillary qubit – first by decomposing the target POVM into a convex combination of rank-one $\leq d^2$-outcome measurements, and subsequently applying Theorem 1 to each of them.

Importantly, the standard Naimark’s implementation of a general POVM would require appending an extra system of dimension $d$ (which can be realised by log$_2 d$ ancillary qubits) and carrying out a global projective measurement. Our simulation protocol greatly reduces this requirement on the dimension cost of implementing $M$ with the possible downside being the probabilistic nature of the scheme. The success probability $q_{\text{succ}}$ depends on the choice of the partition $\{X_\gamma\}_{\gamma=1}^{m}$, and finding the optimal one (for a given bound on the size of $X_\gamma$) is in general a difficult combinatorial problem.

Conjecture. For arbitrary extremal rank-one POVM $M = (M_1, \ldots, M_n)$ on $\mathbb{C}^d$, there exists a partition $\{X_\gamma\}_{\gamma=1}^{m}$ of $[n]$ satisfying $|X_\gamma| \leq d - 1$ such that the corresponding value of success probability $q_{\text{succ}}$ from Eq. (1) is larger than a positive constant independent of $d$.

Let us explore the intriguing conceptual consequences of the validity of this conjecture. First, consider a general nonadaptive measurement protocol that utilizes some quantum measurement $M$ on $\mathbb{C}^d$. Such a protocol consists of $S$ independent measurement rounds of a quantum state $\rho$ resulting in outcomes $i_1, i_2, \ldots, i_S$ distributed according to the probability distribution $p(i|M, \rho) = \text{tr}(M_i \rho)$. This experimental data is then processed to solve a specific problem at hand. If we can simulate any arbitrary $M$ (see comment below proof of Theorem 1) via POVMs that can be implemented using only a single auxiliary qubit with probability $q$, which is independent of the dimension $d$, then this means that we can, on average, exactly reproduce the implementation of the above protocol for $qS$ of the total $S$ rounds. Importantly, we also know in which rounds the simulated protocol was successful, so we know which part of the output data generated by our simulation comes from the target distribution. Crucially, the above considerations are completely oblivious to the figure of merit and the structure of the problem that measurements of $M$ aim to solve.

For many quantum information tasks, losing only a constant fraction of the measurement rounds is not prohibitive and hence, assuming the validity of the conjecture, our POVM simulation scheme offers a way to significantly reduce quantum resources needed for said POVM’s implementation. Such exemplary tasks include quantum state tomography [16], quantum state discrimination [13], multi-parameter quantum metrology [17, 18] or port-based teleportation [24], and will be explored by us in future works.

Our simulation protocol and the above conjecture are also relevant from the perspective of POVM simulability [10, 32, 43] that attracted a lot of attention recently in the context of resource theories [44–50]. Namely, the maximal post-selection probability, $q^{(m)}(M)$, with which a target POVM $M$ on $\mathbb{C}^d$ can be simulated using strategies utilizing randomized POVMs with at most $m$ outcomes, quantifies how far $M$ is from the set of $m$-outcome simulable POVMs in $\mathbb{C}^d$, denoted by $S_m$. Moreover, $q^{(m)}(M)$ imposes bounds on the so-called white noise critical visibility $t^{(m)}(M)$ [10] and the robustness $R^{(m)}(M)$ [44] against simulation via POVMs from $S_m$. Here by critical visibility we mean a parameter $t^{(m)}(M)$ associated with a minimal amount of white noise that ensures that noisy version of $M$ belongs to subset $S_m$, namely

$$t^{(m)}(M) := \max \{ t \mid \Phi_t(M) \in S_m \},$$

where
where $\Phi_t (M)$ is a POVM with effects $\Phi_t (M_i) := t M_i + (1 - t) \mathbb{1}$. By robustness $R^{(m)} (M)$ with respect to $\mathbb{S}_m$, we mean the minimal amount of mixing of $M$ with a POVM from $\mathbb{S}_m$ so that the resulting POVM belongs to $\mathbb{S}_m$, i.e.,

$$R^{(m)} (M) := \min \left\{ s \mid \exists K \text{ s.t. } \frac{M + sK}{1 + s} \in \mathbb{S}_m \right\}. \quad (3)$$

Now, the above quantities are bounded with the success probability of our scheme via (see Appendix A):

$$q^{(m)} (M) \leq q^{(m)} (M), \quad R^{(m)} (M) \leq \frac{1}{q^{(m)} (M)} - 1. \quad (4)$$

Importantly, we note that the robustness $R^{(m)} (M)$ has an appealing operational interpretation: it is also expressible as the relative advantage that $M$ offers over any POVM in $\mathbb{S}_m$ for a state discrimination task [44]:

$$R^{(m)} (M) = \max_{\mathcal{E}} \frac{P_{\text{succ}} (\mathcal{E}, M)}{\max_{\mathcal{N} \in \mathbb{S}_m} P_{\text{succ}} (\mathcal{E}, N)} - 1, \quad (5)$$

where $\mathcal{E} = \{(q_i, \sigma_i)\}_{i=1}^n$ is an ensemble of quantum states, and $P_{\text{succ}} (\mathcal{E}, M)$ is the probability for the minimum error discrimination of the states from $\mathcal{E}$ with $M$. Now, from the second inequality in (4) and the (conjectured) constant lower bound on $q^{(d)}$ we get a surprising conclusion: general POVMs on $\mathbb{C}^d$ do not offer asymptotically increasing (with $d$) advantage over $d$-outcome simulable measurements for general quantum state discrimination problems.

**Haar Random POVMs**—We want to qualitatively understand how $q_{\text{succ}}$ depends on $n$, the number of POVM outcomes used in the simulation $m$, and the dimension $d$. To make the problem feasible we turn to study Haar-random POVMs on $\mathbb{C}^d$. Quantum measurements comprising this ensemble can be realized by a construction motivated by Naimark’s extension theorem: (i) attach to $\mathbb{C}^d$ an ancillary system $\mathbb{C}^a$ so that the composite system is $n$-dimensional: $\mathbb{C}^d \otimes \mathbb{C}^a \approx \mathbb{C}^n$, (ii) apply on this composite system a random unitary $U$ chosen from the Haar measure $\mu_n$ in $U(\mathbb{C}^n)$, and (iii) measure the composite system in the computational basis. Effects of this measurement $M^U$ are given by $M^U_i = \text{tr}_{\mathbb{C}^a} (\mathbb{1} \otimes |0\rangle\langle0| U^\dagger i|0\rangle\langle0| U)$, where $|0\rangle\langle0|$ is a fixed state on $\mathbb{C}^a$. Haar-random POVMs were introduced first in [23] in the context of the hidden subgroup problem and are a special case of a more general family of random POVMs studied recently in [35]. Measurements $M^U$ are extremal for almost all $U \in U(n)$. Furthermore, all extremal rank-one POVMs in $\mathbb{C}^d$ are of the form $M^U$ for some $U \in U(n)$, and $n \in \{d, d + 1, \ldots, d^2\}$. Hence, Haar-random POVMs form an ensemble consisting of extremal non-projective measurements, making them a natural test-bed for studying the performance of our simulation algorithm.

**Theorem 2** (Success probability of the implementation of Haar-random POVMs). Let $n \in \{d, \ldots, d^2\}$, $m \leq d$. Let $M^U$ denote a rank-one $n$-outcome Haar-random POVM on $\mathbb{C}^d$. Let $q_{\text{succ}}^{(m)} (M^U)$ denote the success probability of implementing $M^U$ via $m$-outcome measurements as in Eq. (1) for the standard partition of $[n]$, i.e., $X_1 = \{1, \ldots, m - 1\}$, $X_2 = \{m, m + 1, \ldots, 2m - 2\}$, etc. We then have

$$\mathbb{P}_{U \sim \mu_n} \left( q_{\text{succ}}^{(m)} (M^U) \geq \Theta \left( \frac{m}{d} \right) \right) \to 1, \text{ as } d \to \infty. \quad (6)$$

Moreover, let $q^{(m)} (M^U)$ be the maximal success probability of implementing $M^U$ with postselection via convex combination of $m$-outcome measurements using any simulation protocol. We then have

$$\mathbb{P}_{U \sim \mu_n} \left( q^{(m)} (M^U) \leq \Theta \left( \frac{m}{d} \log(d) \right) \right) \to 1, \text{ as } d \to \infty. \quad (7)$$

The above result shows that when simulating Haar-random POVMs on $\mathbb{C}^d$ with $m$-outcome measurements in our scheme, the success probability scales as $\frac{m}{d}$. Furthermore, Eq. (7) shows the optimality of our method up to a factor logarithmic in $d$. Specifically, we obtain the following crucial result: when $m = d$, with overwhelming probability over the choice of random $U \in \mathbb{U}(d^2)$, $q_{\text{succ}}^{(d)} (M^U)$ is above 6.74%. Below we sketch the proof for Theorem 2. We provide a complete proof in Appendix C, with expressions for finite $d$, for bounds in Eq. (6) and (7).

**Sketch of Proof.** An explicit computation shows that for any subset $X \subset [n]$, we have $\| \sum_{i \in X} M^U_i \| = \| U_X \|^2$, where $U_X$ is a $d \times |X|$ matrix, obtained by choosing the first $d$ rows of $U$, and then taking from the resulting matrix those columns with indices in $X$. With this we analyze the statistical behaviour of $q_{\text{succ}}^{(m)} (M^U)$ in the regime $d \to \infty$ using tools from random matrix theory. Specifically, the proof relies on the phenomenon of concentration of measure [51] on the unitary group $U(n)$ equipped with the Haar measure and distance induced by the Hilbert-Schmidt norm. It shows that as $n \to \infty$, Lipschitz-continuous random variables on $U(n)$ are with high probability close to their Haar-averages - this is captured by large deviation bounds (also known as concentration inequalities), that upper bound the probability that a random variable take values drastically different form its Haar-average.

In order to prove Eq. (6), we choose $\| U_X \|$ as the random variable to which we apply the machinery of concentration of measure. An upper bound to its Haar-average is obtained by performing a discrete optimization over an $\varepsilon$-net of an $m - 1$-dimensional complex sphere. Since the concentration inequality is true for all subsets $X$ in the partition of $[n]$, the union bound shows that $\sum_X \| \sum_{i \in X} M^U_i \|$ also exhibits concentration of measure, which gives Eq. (6).

In order to prove Eq. (7), we invoke the inequality in Eq. (4), and use it to upper bound $q^{(m)}$ with the robustness $R^{(m)} (M^U)$ of a random POVM $M^U$ with respect to $m$-outcome simulable POVMs in $\mathbb{C}^d$. Using the interpretation of robustness in the context of state-discrimination (see Eq. (5)), we lower bound it by constructing a specific ensemble of quantum states obtained by rescaling the effects of $M^U$. In this way, a lower bound on the robustness (hence an upper bound on the success probability) becomes a function of the matrix elements $|U_{ij}|^2$ of the Haar-random unitary $U$. Finally, we prove a concentration of measures inequality for this resulting function, by again invoking the
used fiducial states from a catalogue in Ref \[53\] for simulation scheme by computing values of effects of symmetric POVMs numerically from a single fiducial implemented with a single auxiliary qubit (this corresponds to set-

FIG. 2. Success probability $q_{\text{succ}}$ as a function of dimension $d$ of the Hilbert space for $d^2$-outcome measurements. Results are shown for Weyl-Heisenberg SIC-POVMs (green stars), non-symmetric IC-POVMs (blue dots), and random POVMs (magenta triangles) for dimensions upto 1299. For each dimension, we plot the maximum of $q_{\text{succ}}$ (computed according to the Eq. (1)), which was obtained from random $\leq 24$ partitions. For random POVMs, in each dimension, we generate $10$ to $500$ random POVMs (lower number for higher dimensions) and plot the minimum $q_{\text{succ}}$ across them. For IC-POVMs, the measurement operators are specified by a single parameter $\alpha$ which we keep at a fixed value across all dimensions (see Appendix E for details).

Numerical results—We tested the performance of our POVM simulation scheme by computing $q_{\text{succ}}$ for SIC-POVMs \[36-38\], IC-POVMs \[39\] and for Haar-random $d^2$-outcome POVMs. We focused on simulation strategies via POVMs that can be implemented with a single auxiliary qubit (this corresponds to setting $m = d$ in Theorem 1). For every dimension, we generated effects of symmetric POVMs numerically from a single fiducial pure state via transformations $X_i Z_j$, where $i, j \in [0, d - 1]$ and $X, Z$ are $d$-dimensional analogues of Pauli X and Z operators. For IC-POVMs we used a one-parameter family of fiducial states $|\psi_\alpha\rangle$ described in Ref \[39\] for the specific value $\alpha = \frac{1}{2} (1 + i)$ (we remark that POVMs originating from other values of $\alpha$ exhibited a similar behaviour). For SIC-POVMs we used fiducial states from a catalogue in Ref \[53\] for $d < 100$ and states in higher dimension (up to $d = 1299$), which were provided to us by Markus Grassl in a private correspondence. The construction of random POVMs is described in Appendix E.

Results of our numerical investigation are given in Fig 2. For every considered measurement, the success probability was obtained via direct maximization over only $\leq 24$ random partitions of $[d^2]$. The graph shows that with increasing dimension, $q_{\text{succ}}$ approaches $\approx 25\%$ for SIC POVMs and random POVMs, while for IC it is above $\approx 20\%$ even up to $d = 1299$.

Noise Analysis—Let us now discuss the effects of experimental imperfections on practical implementation of our scheme for generic POVMs. The quantum circuits implementing Haar-random POVMs can be considered generic random circuits. The simplest noise model often adopted for such circuits (see Ref. \[54\]) is a global completely depolarizing channel described by a “visibility” parameter $\eta$. In what follows we assume that this noise is going to affect implementation of circuits used to realize a target POVM $M$ (either via Naimark’s construction or via our method). This noise acts in the following way on effects of $n$-outcome POVM: $M_i \rightarrow M_i^{\eta} := \eta M_i + (1 - \eta) \frac{1}{n}$ (see Section D of the Appendix for details).

To quantitatively compare noisy and ideal implementation of a POVM we use Total-Variation Distance $d_{TV}(p(M|\rho), p(N|\rho)) := \frac{1}{2} \sum_{i=1}^n |p(i|\rho, M) - p(i|\rho, N)|$ between probability distributions $p(M|\rho)$ and $p(N|\rho)$ obtained when $\rho$ is measured by $M$ ($N$). In particular, we will be interested in the worst-case distance, i.e., TVD maximized over quantum states $\rho$, which can be interpreted as measure of statistical distinguishability of $M$ and $N$ (without using entanglement \[55\]). This notion of distance is used to benchmark quality of quantum measurements on near-term devices \[56-58\].

The following result, proven in Section D of the Appendix, gives a lower bound for the average worst-case distance between ideal and noisy implementation of Haar-random POVMs.

**Proposition 1.** Let $M^U$ be a Haar-random $n$-outcome rank-one POVM on $\mathbb{C}^d$ and let $M^{U,\eta}$ be its noisy implementation with effects $(M^{U,\eta})_i = \eta M_i + (1 - \eta) \frac{1}{n}$. We then have

$$\left\langle \max_{\rho} \left( d_{TV} (p(M^U|\rho), p(M^{U,\eta}|\rho)) \right) \right\rangle_{\text{Haar}} \geq (1 - \eta) c_n,$$

where $c_n = (1 - \frac{1}{n})^n \approx \frac{1}{e}$. (8)

To make qualitative comparison between our and standard (i.e., based on Naimark’s dilation theorem) implementation of POVMs, we use noise model used in Google’s recent demonstration of quantum advantage \[40\]. Assuming that main source of errors are multiple two-qubit gates, we get that dominating term in visibility is exponentially decaying function: $\eta = \eta(r_2, g_2) \approx \exp(-r_2 g_2)$, where $r_2$ is two-qubit error rate and $g_2$ is the number of two-qubit gates needed to construct a given circuit. Now recall that for implementation of $d^2$-outcome POVM using Naimark’s dilation, one needs to implement circuits on the Hilbert space with doubled number of qubits $2N$ (we assume $d = 2N$), while our post-selection scheme requires only a single additional qubit, hence the target space has only $N + 1$ qubits. We note that for implementation of generic circuits on $2N$ qubits, the theoretical lower bound \[59\] for needed number of CNOT gates is $g_2^{\text{Naimark}} = \Theta(4^N) = \Theta(16N)$, while our scheme gives the scaling $g_2^{\text{post}} = \Theta(4^N)$.

Finally, combining the above considerations with Proposition 1, we get expected worst-case distance between ideal and noisy Naimark implementation of generic $d^2$-outcome measurement is lower bounded by $\approx (1 - \exp(-\Theta(16N)))^{-1}$, which corresponds to $\eta^{\text{Naimark}} = \exp(-\Theta(16N))$. We compare this to the quality of probability distribution $p_{\text{post}}(M|\rho)$ generated by the noisy version of our simulation scheme which is based on implementation of projective measurements on $N + 1$ (not $2N$) qubits and hence incurring noise with $\eta^{\text{post}} \approx \exp(-\Theta(4^N))$. In Appendix D we show that postselection step in our scheme does not significantly affect the quality of pro-
duced samples by proving that for typical Haar random $M^U$
\[
d_{\text{TV}} (p(M^U | \rho), p_{\text{post}}(M^U | \rho)) \leq C (1 - \exp(-\Theta(4^N))),
\]
where $C$ is an absolute constant. Therefore, for generic measurements, implementation via our scheme will be affected by much lower noise than in the case of Naimark’s. We expect that similar behaviour (i.e., amount of noise in our scheme compared to Naimark’s dilation) should be exhibited also for more realistic noise models – the high reduction of the dimension of the Hilbert space is, reasonably, expected to highly reduce the noise.

Discussion and open problems— Aside from their practical relevance, our results shed light onto the question whether POVMs are more powerful (in quantum information tasks requiring sampling) than projective measurements. Indeed, since typical POVMs in $\mathbb{C}^d$ can be implemented using $d$-outcome measurements, it suggests (and if our conjecture is true, then it implies) that, if there exists a gap in the relative usefulness (quantified for example via robustness), then it is between projective measurements and $d$-outcome POVMs. Moreover, the surprisingly high value of $q_{\text{succ}}^{(d)}$ will likely have potential applications to nonlocality. First, it significantly limits (due to inequality (4)) the amount of local depolarizing noise that can be tolerated in schemes for generation secure quantum randomness using extremal $d^2$-outcome measurements [19, 60]. We also anticipate that our results can be used to construct new local models for entangled quantum states that undergo general POVM measurement (by using techniques similar to those of [10, 61]).

We conclude with giving directions for future research. First, naturally, is to verify whether our conjecture is true. The difficulty in proving it comes from the combinatorial nature of the optimization problem in Eq. (1) - it is difficult to analytically find the optimal partition of $[n]$ that maximizes $q_{\text{succ}}$ for a target POVM $M$. Effects of Haar random POVMs have similar properties - in particular, they have (on average) equal operator norms - this symmetry allowed us to study them analytically. However, general POVMs can be highly unbalanced (in the sense of having effects whose operator norms can vary significantly) and suitable strategies need to be devised to tackle such situations. Second, it is desirable to devise an algorithmic method which, when given the circuit description of some POVM, returns the circuits needed to implement it with postselection. Another direction is to identify and quantify the real-time implementation costs of randomisation and post-processing, and how these cost considerations can be taken into account for suitable modifications of the scheme. Finally, it would be interesting to see if the success probability is connected to other properties of POVMs – for instance, their entanglement cost [62].

Data availability The data obtained in numerical simulations is available from authors upon request.

Code availability The code used to obtain numerical simulations is available from authors upon request.

Acknowledgements We are sincerely grateful to Markus Grassl for fruitful discussions and for sharing with us the numerical form of fiducial kets of SIC POVMs for high dimensions. We thank Zbigniew Puchała for the discussions at the initial stage of this project and Michal Horodecki for suggesting potential application of our scheme in PBT. The authors acknowledge the financial support by TEAM-NET project co-financed by EU within the Smart Growth Operational Programme (contract no. POIR.04.04.00-00-17C1/18-00). A portion of this work was done while TS was in Fudan university, and TS acknowledges support from the National Natural Science Foundation of China (Grant No. 11875110) and Shanghai Municipal Science and Technology Major Project (Grant No. 2019SHZDZX01).

Author Contributions TS had a leading role in proving Theorem 2, Proposition 1 and many auxiliary technical results. FBM carried out numerical simulations and proved results concerning noise robustness of POVM implementation methods. MO contributed with the main idea of the project, proved Theorem 1 and supervised the other parts project. All authors equally contributed to writing the manuscript equally.

Competing interests The Authors declare no Competing Financial or Non-Financial Interests.

* tanmaysingal@gmail.com
† filip.b.maciejewski@gmail.com
‡ oszmaniec@cft.edu.pl
[1] E. Knill and R. Laflamme, Phys. Rev. A 55, 900 (1997).
[2] H. J. Briegel, D. E. Browne, W. Dür, R. Raussendorf, and M. Van den Nest, Nature Physics 5, 19 (2009).
[3] N. Gisin and R. Thew, Nature Photonics 1, 165 (2007).
[4] J. A. Bergou, Journal of Modern Optics 57, 160 (2010), https://doi.org/10.1080/09500340903477756.
[5] S. L. Braunstein and C. M. Caves, Phys. Rev. Lett. 72, 3439 (1994).
[6] S. L. Braunstein, C. M. Caves, and G. Milburn, Annals of Physics 247, 135 (1996).
[7] G. Tóth and I. Apellaniz, Journal of Physics A: Mathematical and Theoretical 47, 424006 (2014).
[8] S. Pirandola, B. R. Bardhan, T. Gehring, C. Weedbrook, and S. Lloyd, Nature Photonics 12, 724 (2018).
[9] C. L. Degen, F. Reinhard, and P. Cappellaro, Rev. Mod. Phys. 89, 035002 (2017).
[10] J. Bae and L.-C. Kwek, Journal of Physics A: Mathematical and Theoretical 48, 083001 (2015).
[11] J. A. Preskill, Quantum 2, 79 (2018).
[12] J. Bae and L.-C. Kwek, Journal of Physics A: Mathematical and Theoretical 48, 2171 (2004), https://doi.org/10.1063/1.1737053.
[13] M. van den Nest, Nature Physics 1, 190501 (2017).
[14] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, IEEE Transactions on Information Theory 63, 5628 (2017).
[15] S. Ragy, M. Jarzyna, and R. Demkowicz-Dobrzański, Phys. Rev. A 94, 052108 (2016), arXiv:1608.02634 [quant-ph].
[16] M. Szczysyńska, T. Baumgratz, and A. Datta, Advances in Physics: X 1, 621 (2016), https://doi.org/10.1080/23746149.2016.1230476.
[17] J. Preskill, Quantum 2, 79 (2018).
[18] J. Bae and L.-C. Kwek, Journal of Physics A: Mathematical and Theoretical 48, 083001 (2015).
[19] R. Derka, V. Buz’ek, and A. K. Ekert, Phys. Rev. Lett. 80, 1571 (1998).
[20] J. M. Renes, R. Blume-Kohout, A. J. Scott, and C. M. Caves, Journal of Mathematical Physics 45, 2171 (2004), https://doi.org/10.1063/1.1737053.
[21] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, IEEE Transactions on Information Theory 63, 5628 (2017).
[22] S. Ragy, M. Jarzyna, and R. Demkowicz-Dobrzański, Phys. Rev. A 94, 052108 (2016), arXiv:1608.02634 [quant-ph].
[23] J. Bae and L.-C. Kwek, Journal of Physics A: Mathematical and Theoretical 48, 083001 (2015).
[24] R. Derka, V. Buz’ek, and A. K. Ekert, Phys. Rev. Lett. 80, 1571 (1998).
[25] J. M. Renes, R. Blume-Kohout, A. J. Scott, and C. M. Caves, Journal of Mathematical Physics 45, 2171 (2004), https://doi.org/10.1063/1.1737053.
[26] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, IEEE Transactions on Information Theory 63, 5628 (2017).
[27] S. Ragy, M. Jarzyna, and R. Demkowicz-Dobrzański, Phys. Rev. A 94, 052108 (2016), arXiv:1608.02634 [quant-ph].
[28] M. Szczysyńska, T. Baumgratz, and A. Datta, Advances in Physics: X 1, 621 (2016), https://doi.org/10.1080/23746149.2016.1230476.
[29] A. Acín, S. Pirtonio, T. Vértesi, and P. Wittek, Phys. Rev. A 93, 040102 (2016).
Appendix

We collect here technical results that are used in the main part of the paper, as well as more detailed descriptions of some of the presented concepts. In Section A, we discuss a relation between success probability of our implementation scheme, and a resource-theoretic quantities – visibility and robustness of POVMs. In Section B we explain concentration of measure for general random variables on probability spaces, especially for the special cases of the unitary group $U(n)$ and the $(n-1)$-complex sphere. The contents of this section should be treated as preliminaries for further sections. The proofs technical version of Theorem 2 are provided in Sections C. In Section D we describe in more detail the effects that completely depolarizing noise has on the implementation of quantum measurements. Finally, in Section E we provide details of numerical simulations presented in the main text.

For the benefit of the reader, in table below we explain the notation used in the Appendix

| Symbol | Explanation |
|--------|-------------|
| $d$    | dimension of principal system. |
| $n$    | Number of outcomes of a target POVM. |
| $m$    | Number of outcomes of POVMs which we simulate target POVM with. |
| $\mathbb{C}^d$ | Hilbert space of our principal system of study. |
| $\mathbb{C}^n$ | Hilbert space of extended system. |
| $\rho, \sigma, \text{etc.}$ | General mixed states on quantum system. |
| $M$, $N$, etc | Quantum measurements on our principal system. |
| $q_{\text{succ}}$ | Success probability of simulating a measurement using method specified in Theorem 1 of the main text. |
| $R^{(m)}(M)$ | Robustness of a POVM $M$ with respect to $\mathbb{S}_m$. |
| $[n]$, $[d]$, etc. | The set $\{1, 2, \ldots, n\}$. Similarly for $[d]$. |
| $X, Y$, etc. | Subsets of $[n]$. |
| $U(n)$ | Unitary group of $n \times n$ complex matrices. |
| $U$, $W$, etc. | An $n \times n$ unitary matrix. |
| $U_{X, W_X}$, etc. | A truncation of $n \times n$ unitary matrix $U$, occurring at the intersection of rows in $[d]$ and columns in $X$. |
| $P_{U_{\sim \mu_n}(E)}$ | Probability of some event $E$. |
| $\langle f \rangle_{\text{Haar}} = \frac{1}{d} \int_{U(n)} f(U) d\mu_n(U)$ | Integral (expectation value) of function $f$ on unitary group $U(n)$ with respect to the Haar measure. |
| $\| \| \psi \| \|$ | Vector norm of state vector $|\psi\rangle$. |
| $\| A \|_2$ | Operator norm of a linear operator $A$. |
| $d_{\text{TV}}(p, q)$ | Total variational distance between probabilities $p$ and $q$. |

TABLE I. Notation used in the Appendix

Appendix A: Relation between $q^{(m)}$, and critical visibility $t^{(k)}$, and robustness $R^{(k)}$.

Let $\mathbb{S}_m \subset \mathcal{P}(d, n)$ denote all $m$-outcome simulable POVMs and let $\Phi_t$ denote the depolarising channel $\Phi_t(X) := tX + (1 - t)\frac{1}{d} \mathbf{1}$. Its action naturally extends to POVMs, via action on individual effects: $\Phi_t(M_i) := tM_i + (1 - t)\frac{1}{d} \mathbf{1}$. For any $n$-outcome POVM $M$ the visibility with respect to $\mathbb{S}_m$

$$t^{(m)}(M) := \max \{ t \mid \Phi_t(M) \in \mathbb{S}_m \}.$$  \hspace{1cm} (A.1)

The robustness $R^{(m)}(M)$ of measurement $M$ with respect to $\mathbb{S}_m$ is defined via

$$R^{(m)}(N) := \min \left\{ s \mid \exists K \text{ s.t. } \frac{N + sK}{1 + s} \in \mathbb{S}_m \right\}.$$  \hspace{1cm} (A.2)

Let $q^{(m)}(M)$ be the largest success probability with which $M$ can be simulated via $m$-outcome POVMs. It follows that

$$L = (q^{(m)}(M), M_1, q^{(m)}(M), M_2, \ldots, q^{(m)}(M), M_n, (1 - q^{(m)}(M)) \mathbf{1})$$  \hspace{1cm} (A.3)

can be simulated via $m$-outcome POVMs. It follows that a POVM $\Phi_{q^{(m)}(M)}(M) \in \mathbb{S}_m$. The inequalities claimed in the main text (cf. Eq.(4))

$$q^{(m)}(M) \leq t^{(m)}(M), \quad R^{(m)}(M) \leq \frac{1}{q^{(m)}(M)} - 1.$$  \hspace{1cm} (A.4)

follow directly from definitions of $t^{(m)}(M)$ and $R^{(m)}(M)$ respectively.
Appendix B: Preliminaries

In this Part we provide some basic theoretical background that will be used in Lemmas 1, 2, 4, 5, 6, and Theorems 5 and 6. In Subsection B 1, we introduce the notion of concentration of measure, which will be used extensively for proving the aforementioned lemmas and theorems. Related concepts like Lipshitz constants of functions and log-Sobolev inequalities and log-Sobolev constants are also explained alongside. The metric spaces which we use in this work are the unitary group $U(n)$ (with metric induced by Hilbert-Schmidt inner product), and the $(n - 1)$-complex sphere $S^{n-1}_C$, with the metric it inherits from $C^n$. The Haar-measure on $U(n)$ and the uniform measure on $S^{n-1}_C$ will be introduced in subsections B 2 and B 3 respectively, and the corresponding log-Sobolev constants also mentioned.

1. Concentration of Measure: Lipshitz constants and log-Sobolev inequalities

We start by recalling notions of Lipshitz constants and log-Sobolev inequalities. Let $(\mathcal{X}, d)$ be a metric space, and let $f : \mathcal{X} \to \mathbb{R}$ be a real function on $\mathcal{X}$. We say that $f$ is $L$-Lipshitz on $\mathcal{X}$ with respect to the metric $d$, if

$$|f(x) - f(y)| \leq L d(x, y), \text{ for all } x, y \in \mathcal{X}. \quad (B.1)$$

Now let $\mu$ be a probability measure on $(\mathcal{X}, d)$, and let function $f$ be such that the length of the gradient of $f$ can be defined at any point $x$ in $\mathcal{X}$, namely

$$|\nabla f|(x) := \limsup_{y \to x} \frac{|f(x) - f(y)|}{d(x, y)}. \quad (B.2)$$

Then for any such function, the following concentration inequalities hold

$$\int d\mu(x) \exp \left( \lambda \left( f(x) - \int d\mu(x) f(x) \right) \right) \leq \exp \left( \frac{CL^2 \lambda^2}{2} \right), \text{ for all } \lambda \in \mathbb{R}, \quad (B.3)$$

$$\Pr_{x \sim \mu} \left( f(x) \geq \int d\mu(x) f(x) + t \right) \leq \exp \left( -\frac{t^2}{2CL^2} \right), \text{ for } t \geq 0, \quad (B.4)$$

where $C$ is called the log-Sobolev constant of $\mu$ with respect to the metric $d$ of $\mathcal{X}$. We note that the inequality (B.4) can be derived from (B.3) (see Theorem 5.39, in [51]). We refer the reader to [51] for more details on log-Sobolev inequalities.

2. Haar-measure on $U(n)$

The group of $n \times n$ unitary matrices $U(n)$ is endowed with the well known probability measure known as the Haar-measure. It follows that for any integrable function $f$ on $U(n)$, its expectation value with respect to the Haar measure is invariant under the following operations

$$\int d\mu_n(U) f(U) = \int d\mu_n(U) f(WU) = \int d\mu_n(U) f(UW) = \int d\mu_n(U) f(U^{-1}), \quad (B.5)$$

where $W$ is an arbitrary fixed unitary in $U(n)$. $U(n)$ inherits a metric from the Hilbert-Schmidt inner product on the space of $n \times n$ complex matrices. The distance between two unitaries $U, W$ with respect to the Hilbert-Schmidt metric is

$$||U - W||_2 = \sqrt{\text{tr} ((U - W)^\dagger (U - W))}. \quad (B.6)$$

The following Theorem then gives the log-Sobolev constant for the Haar measure with respect to the Hilbert-Schmidt metric (table 5.4 in [51]).

**Theorem 3.** [51] The log-Sobolev constant for the Haar measure on the unitary group $U(n)$ with the Hilbert-Schmidt metric is $\frac{6}{n}$. 
3. Uniform measure on \( S_{C}^{n-1} \)

The complex \((n-1)\)-sphere \( S_{C}^{n-1} \) is defined as
\[
S_{C}^{n-1} = \{ |x\rangle \in \mathbb{C}^n \mid \langle x|x\rangle = 1 \}.
\] (B.7)

For any \( n \times n \) unitary \( U \), the unitary action \(|x\rangle \rightarrow U|x\rangle \) is norm-preserving. Thus, the Haar-measure of \( U(n) \) endows a rotationally invariant probability measure on \( S_{C}^{n-1} \) in the following way: fix some arbitrary \(|x\rangle \in S_{C}^{n-1}\), then for Haar-random \( U \), \(|z\rangle = U|x\rangle \) is a random variable in \( S_{C}^{n-1} \), endowed with a probability measure called the uniform probability measure on \( S_{C}^{n-1} \). In particular, one can choose \(|x\rangle\) to be a standard basis vector \(|e_i\rangle\), which tells us that when \( U \) is Haar-random, then it’s columns are distributed with the uniform measure on \( S_{C}^{n-1} \). The uniform probability measure on \( S_{C}^{n-1} \) has a log-Sobolev constant with respect to the usual norm-induced metric on \( S_{C}^{n-1} \) (see table 5.4 in [51]: note that \( S_{C}^{n-1} \approx S^{2n-1} \), which is the \((2n-1)\)-sphere in \( \mathbb{R}^{2n} \)).

**Theorem 4.** [51] The log-Sobolev constant for the uniform measure on the complex \((n-1)\)-sphere, \( S_{C}^{n-1} \) is
\[
\frac{1}{2n-1}.
\]

Let \( \{ |e_j\rangle \}^{n}_{j=1} \), denote the standard basis for \( \mathbb{C}^n \). Each vector \(|\psi\rangle \in S_{C}^{n-1} \) can be mapped to an \( n\)-probability vector as follows:
\[
|\psi\rangle \rightarrow \mathbf{p}, \text{ where } p_i = |\langle e_i|\psi\rangle|^2.
\] (B.8)

Imposing the uniform measure on \( S_{C}^{n-1} \), converts \( p_i \) into a random variable on interval \([0, 1]\). Denote \( p_i \) by \( x \), the probability density of this random variable is given by [52]
\[
p(x) = (n - 1)(1 - x)^{n-2}, \text{ where } 0 \leq x \leq 1.
\] (B.9)

It is easy to see that the expectation value of \( x \) is \( \frac{1}{n} \). Also, the distribution of \( x \) is given by \( \mathbb{P}(x \geq y) = (1 - y)^{n-1} \) and it follows that
\[
\mathbb{P}(x \geq y) \leq \exp\left(-(n-1)y\right).
\] (B.10)

4. Haar-random POVMs

In this subsection we recall the construction of rank-one Haar random POVMs. An \( n\)-outcome, rank-one POVM \( M^U \) on \( \mathbb{C}^d \) can be constructed from Haar-random unitary \( U \in U(n) \) using the following steps

1. Extend the principal system \( \mathbb{C}^d \) to a larger system \( \mathbb{C}^n \) using an ancillary system, which is prepared in a fixed state \(|0\rangle\).
2. Rotate the composite system by the unitary \( U \) in \( U(n) \).
3. Measure the composite system in a computational basis \( \{ |e_i\rangle \}^{n}_{i=1} \).

Let us denote by \( P^U \) a rank-1 \( n \)-outcome projective measurement on the composite system, whose effects are given by
\[
P_i^U = U^\dagger |e_i\rangle\langle e_i| U, \text{ for } i \in [n].
\] (B.11)

Now if the ancillary system is prepared in state \(|0\rangle\langle 0|\), then performing the above measurement on the composite system, implements on original system \( \mathbb{C}^d \) a rank-1 \( n\)-outcome measurement \( M^U \) with effects given by \( M_i^U = \text{tr}_B (I \otimes |0\rangle\langle 0|) U^\dagger P_i^U U \). Importantly, the matrix elements of \( M_i \) can be related to the matrix elements of \( U \) via
\[
( M_j)_{ij} = U^*_{ji} U_{jt}.
\] (B.12)

Finally, when \( U \) is distributed according to the Haar measure on \( U(n) \), then a POVM \( M^U \) also becomes a random variable. This is called a Haar-random POVM.

Appendix C: Proof of Theorem 2

In this section we prove the Theorem 2 concerning bounds on the success probability of implementation of Haar-random POVMs with postselection. The first three subsections contain auxiliary lemmas needed in the proof of the main result which we provide in Section C.4. From now on, unless stated otherwise, we denote by \( X \) a subset of \([n]\) such that \(|X| = m\), by \( U \) a \( n \times n \) unitary matrix, and by \( U_X \) a truncation of unitary \( U \), occurring at the intersection between rows in \([d]\) and columns in \( X \). Furthermore, \( \{ |e_i\rangle \}^n_{i=1} \) is a standard orthonormal basis in \( \mathbb{C}^n \) and by \( P = \sum_{i=1}^n |e_i\rangle\langle e_i| \) we denote a projector onto the space of its first \( d \) components.
1. Lipschitz constants for functions used in proof of Theorem 2

We first bound Lipschitz constants for some functions which will be used latter.

**Lemma 1.** The function $U \rightarrow ||U||$ is 1-Lipschitz on $U(n)$ with respect to the Hilbert-Schmidt metric.

**Proof.** Let $U, W$ be two $n \times n$ unitaries, such that $U \neq W$. Then
\[
\frac{||U|| - ||W||}{||U - W||_2} \leq \frac{||U - W||}{||U - W||_2} \leq \frac{||U - W||_2}{||U - W||_2} \leq 1.
\] (C.1)

**Lemma 2.** For any $|z\rangle$ in $S_{n-1}^n$, the function $|z\rangle \rightarrow ||P|z\rangle||$ is 1-Lipschitz.

**Proof.** Let $|z_1\rangle, |z_2\rangle \in S_{n-1}^n$, such that $|z_1\rangle \neq |z_2\rangle$. Then
\[
\frac{||P|z_1\rangle - ||P|z_2\rangle||}{|||z_1\rangle - |z_2\rangle||} \leq \frac{||P(|z_1\rangle - |z_2\rangle)||}{|||z_1\rangle - |z_2\rangle||} \leq ||P|| = 1.
\] (C.2)

2. Upper bound to the Haar-averaged norm of truncations of unitary matrices

The following auxiliary results allow us to upper bound expected value of the operator norm of truncations of Haar random unitaries.

**Lemma 3.** Let $S_X \subset S_{n-1}^n$ be defined as
\[
S_X = \{|a\rangle \in S_{n-1}^n \mid \langle e_i|a\rangle = 0, \forall i \notin X\},
\] (C.3)
Let $E_X$ be an $\epsilon$-net for $S_X$. Then $\langle e_i|P|x\rangle = 0$ for $i \geq d + 1$ for all $|x\rangle \in \mathbb{C}^n$, and we have
\[
||U_X|| \leq \frac{1}{1 - \epsilon} \left( \max_{|x\rangle \in E_X} ||PU|x\rangle|| \right).
\] (C.4)

**Proof.** From the singular value decomposition of $U_X$, we get that
\[
||U_X|| = \max_{|a\rangle \in S_X} ||PU|a\rangle|| = ||PU|\tilde{a}\rangle||,
\] (C.5)
where $|\tilde{a}\rangle \in S_X$ is the (or is a) vector at which the maximization in equation (C.5) is attained.

Now to discretize the optimization in equation (C.5), we optimize over $E_X$ instead, and we note that then there exists $|\tilde{x}\rangle \in E_X$ such that $|||\tilde{x}\rangle - |\tilde{a}\rangle|| \leq \epsilon$. Hence we get that $||PU|\tilde{a}\rangle|| \leq ||PU|\tilde{x}\rangle|| + \epsilon ||U_X||$, which gives us
\[
||U_X|| \leq \frac{1}{1 - \epsilon} ||PU|\tilde{x}\rangle|| \leq \frac{1}{1 - \epsilon} \left( \max_{|x\rangle \in E_X} ||PU|x\rangle|| \right),
\] (C.6)
for all $0 < \epsilon < 1$.

**Lemma 4.** We have the following upper bound for expected value of the norm of truncation of the unitary matrix
\[
\langle||U_X||\rangle_{\text{Haar}} \leq c \left( 1 + \sqrt{\frac{2m}{d}} \right) \sqrt{\frac{d}{n}}, \text{ where } c \approx 1.92.
\] (C.7)

Additionally, when $m = d - 1$,
\[
\langle||U_X||\rangle_{\text{Haar}} \leq c \sqrt{\frac{d - 1}{n}}, \text{ where } c \approx 3.86.
\] (C.8)
Remark 1. The proof of Lemma 4 is inspired by the proof of equation (18) and Theorem 7 in [63] (please see Section 2 of the appendix in [63]). In Remark 2 below we briefly explain the differences between the proof presented here and the proof in [63].

Proof. Let |z| ∈ S^{n-1} and define the function |z| → \| P |z| \|. This function is 1-Lipschitz on S^{n-1} (Lemma 2). Define S_X as in equation (C.3). Now fix some |x| ∈ S_X. Let U ∈ U(n) be Haar-random, and let |z| = U |x|. Then |z| is uniformly distributed on S^{n-1} (see Subsection B 3). Thus the function |z| → \| P |z| \| satisfies the following log-Sobolev inequality with a constant C = \frac{1}{\sqrt{2n-1}} with respect to the uniform measure on S^{n-1} (see Subsection B 3)

\[
\int d\mu_n(U) \exp (\lambda (\| PU |x| \| - A)) \leq \exp \left( \frac{\lambda^2}{2(2n-1)} \right), \quad \forall \lambda \in \mathbb{R}. \tag{C.9}
\]

where A := \langle \| PU |x| \| \rangle_{\text{Haar}}. Since \exp (-\lambda A) is independent of the integrating variable, we get

\[
\int d\mu_n(U) \exp (\lambda \| PU |x| \| ) \leq \exp \left( \frac{\lambda^2}{2(2n-1)} + \lambda A \right). \tag{C.10}
\]

First we prove that A ≤ \frac{\sqrt{d}}{n}. Using the well-known result \langle |U_{ij}|^2 \rangle_{\text{Haar}} = \frac{1}{n}, one obtains

\[
\langle \| PU |x| \| ^2 \rangle_{\text{Haar}} = \sum_{i=1}^{n} \langle |U_{ij}|^2 \rangle_{\text{Haar}} = \frac{d}{n}, \tag{C.11}
\]

where we chose |x| = |e_j| for some j ∈ X. Now note that \langle \| PU |x| \| ^2 \rangle_{\text{Haar}} ≥ (\| PU |x| \| )^2_{\text{Haar}}. Hence we get

\[
\int d\mu_n(U) \exp (\lambda \| PU |x| \| ) \leq \exp \left( \frac{\lambda^2}{2(2n-1)} + \lambda \sqrt{\frac{d}{n}} \right). \tag{C.12}
\]

Now let E_X be an ε-net for S_X. Then we sum the inequality (C.12) over all |x| ∈ E_X, and we get

\[
\sum_{|x| \in E_X} \int d\mu_n(U) \exp (\lambda \| PU |x| \| ) \leq |E_X| \exp \left( \frac{\lambda^2}{2(2n-1)} + \lambda \sqrt{\frac{d}{n}} \right). \tag{C.13}
\]

For each U ∈ U(n) there is some |x_U| ∈ E_X, such that

\[
\| PU |x_U| \| = \max_{|x| \in E_X} \| PU |x| \|. \tag{C.14}
\]

It is not difficult to see that U → \| PU |x_U| \| is a continuous function, which implies that \exp (\lambda \| PU |x_U| \| ) is integrable on U(n). Thus we get

\[
\int d\mu_n(U) \exp (\lambda \| PU |x_U| \| ) \leq \sum_{|x| \in E_X} \langle \exp (\lambda \| PU |x| \| ) \rangle_{\text{Haar}} \leq |E_X| \exp \left( \frac{\lambda^2}{2(2n-1)} + \lambda \sqrt{\frac{d}{n}} \right). \tag{C.15}
\]

Since the exponential function is convex, Jensen’s inequality can be applied in equation (C.15), which gives

\[
\exp \left( \lambda \int d\mu_n(U) \| PU |x_U| \| \right) \leq |E_X| \exp \left( \frac{\lambda^2}{2(2n-1)} + \lambda \sqrt{\frac{d}{n}} \right). \tag{C.16}
\]

Now taking the (natural) logarithm (and assuming that λ > 0) we get

\[
\int d\mu_n(U) \| PU |x_U| \| \leq \frac{1}{\lambda} \left( \log |E_X| + \frac{\lambda^2}{2(2n-1)} + \lambda \sqrt{\frac{d}{n}} \right). \tag{C.17}
\]

Since the inequality (C.17) is valid for all λ > 0, we directly minimize the RHS over λ, and we get

\[
\langle \max_{|x| \in E_X} \| PU |x| \| \rangle_{\text{Haar}} \leq \sqrt{\frac{2 \log |E_X|}{2n-1}} + \sqrt{\frac{d}{n}}. \tag{C.18}
\]
which is obtained at the value $\lambda = \sqrt{2(2n - 1) \log |E_X|}$. Note that we have used equation (C.14) in the LHS of equation (C.18).

There's a well-known theorem (see, e.g., [51, 64]) that an \(\epsilon\)-net for \(S_X\) has at most \((1 + 2/\epsilon)^{2m}\) points. This gives us an upper bound for \(|E_X|\), which, inserted into RHS of (C.18) gives

\[
\left< \max_{|x| \in E_X} \| PU |x| \| \right>_{\text{Haar}} \leq \sqrt{\frac{4m \log (1 + 2/\epsilon)}{2n - 1}} + \sqrt{\frac{d}{n}}. \tag{C.19}
\]

Finally, using the inequality (C.4) with the inequality (C.19) we get

\[
\| U_X \| \leq \frac{1}{1 - \epsilon} \left( \sqrt{\frac{4m \log (1 + 2/\epsilon)}{2n - 1}} + \sqrt{\frac{d}{n}} \right), \quad \text{for } \epsilon \in (0, 1), \tag{C.20}
\]

which is valid for any \(\epsilon \in (0, 1)\). Now recall that in our scheme we are interested in case when \(m \leq d - 1\), which allows to rewrite the above inequality as

\[
\| U_X \| \leq \frac{\sqrt{\log (1 + 2/\epsilon)}}{1 - \epsilon} \left( 1 + \sqrt{\frac{2m}{d}} \right) \sqrt{\frac{d}{n}}, \quad \text{for } \epsilon \in (0, 1), \tag{C.21}
\]

where we have used the fact that \(1 < \log (1 + 2/\epsilon)\) for any \(\epsilon \in (0, 1)\) and we assume that \(n\) is large. With this approximation it is possible to perform minimization over \(\epsilon\), which gives us the inequality (C.7). Note that the result of minimization will generally depend on the relative values of \(d\) and \(m\), and so for special case \(m = d - 1\) we get inequality (C.8).

\(\square\)

**Remark 2.** There are two differences between the proof that we gave above, and the proof for equation (18) in Theorem 7 of [63]. Firstly, the goal of Lemma 4 is to find an upper bound to \(\left< \| U_X \| \right>_{\text{Haar}},\) while in [63], the upper bound being sought is for \(\left< \max_{|x|,|y|} \| U_{X,Y} \| \right>_{\text{Haar}},\) where \(X, Y \in [n]\), such that \(|X| = d\) and \(|Y| = m\), and \(U_{X,Y}\) is the \(d \times m\) truncation of \(U\) lying at the intersection between rows in \(X\) and columns in \(Y\). For this purpose, the optimization in [63] is over an \(\epsilon\)-net whose cardinality is \(\binom{n}{d} \binom{m}{d} \left( 1 + \frac{2}{\epsilon} \right)^{2(d+m)}\). The second difference is that we use the equation (C.4) for the optimization, whereas in [63], they used \(\| U_{X,Y} \| = \max_{|x|,|y|} \Re \langle x| U |y \rangle\), where \(|x| \in E_X\), which is an \(\epsilon\)-net for \(S_X\), and \(|y| \in E_Y\), which is an epsilon-net for \(S_Y\). Our reason for choosing equation (C.4) is that it allows us to obtain a lower upper bound in inequality (C.7) and (C.8). This is important because this upper bound is closely associated with the success probability, as can be seen in the proof of Theorem 5.

3. Lemmas needed for the proof of upper bound on \(q^{(m)}(M^{U})\)

**Lemma 5.** Let \(w^U_j := \sum_{i=1}^d |U_{ij}|^2\) for \(j \in [n]\). Then we have

\[
P_{U \sim \mu_n} \left( \max_{X \subset [n], |X| \leq m} \sum_{j \in X} w^U_j \geq \frac{2dm \log n}{n} (1 + \epsilon) \right) \leq \frac{e d}{n^1 + 2\epsilon}, \quad \text{where } \epsilon \in \left( \frac{1}{2 \log n}, \frac{n}{2 \log n} - 1 \right). \tag{C.22}
\]

**Proof.** Consider the event

\[
E := \left\{ U \in U(n) \left| \max_{X \subset [n], |X| \leq m} \sum_{j \in X} w^U_j \geq \frac{r d m \log n}{n} (1 + \epsilon) \right\}, \tag{C.23}
\]

where \(r\) is a constant, that will be determined later to get a decent concentration. The event \(E\) implies that there exists some \(i \in [d]\) and some \(j \in [n]\) such that the following event is true:

\[
E_{ij} := \left\{ U \in U(n) \left| |U_{ij}|^2 \geq \frac{r \log n}{n} (1 + \epsilon) \right\}, \tag{C.24}
\]

Hence we have

\[
E \subset \bigcup_{i \in [d], j \in [n]} E_{ij}. \tag{C.25}
\]
Now we note that for \( \epsilon \in \left(-1, \frac{n}{r \log n} - 1\right) \) and \( y = \frac{r(1 + \epsilon) \log n}{n} \), from inequality (B.10) we have
\[
\mathbb{P}_{U \sim \mu_n} \left( E_{ij} \right) \leq \frac{e}{n^{r(1+\epsilon)}},
\]
where we used the fact that \( 1 < \exp\left( \frac{r(1+\epsilon) \log n}{n} \right) < e \). Using the union bound gives
\[
\mathbb{P}_{U \sim \mu_n} \left( \bigcup_{i \in [d], j \in [n]} E_{ij} \right) \leq \sum_{i \in [d], j \in [n]} \mathbb{P}_{U \sim \mu_n} (E_{ij}) \leq \frac{c d n}{n^{r(1+\epsilon)}}.
\]
Note that for the probabilities appearing on the RHS of the inequality (C.27) to be meaningful, it’s necessary to revise the interval for \( \epsilon \) as follows.
\[
\frac{1}{r \log n} < \epsilon < \frac{n}{r \log n} - 1,
\]
provided that \( r \) is chosen so that \( n^r \geq n d \). The maximum value of \( d \) in terms of \( n \) is when \( d = n \). Thus we choose \( r = 2 \), which proves the lemma.

\[\text{Lemma 6.}\] Let \( w^U_j := \sum_{i=1}^d |U_{ij}|^2 \) for \( j \in [n] \). Then
\[
\min_{U \in \mathbb{U}(n)} \sum_{j=1}^n (w^U_j)^2 = \frac{d^2}{n}.
\]
\[\text{Proof:}\] Since \( \sum_{j=1}^n w^U_j = d \), we get that \( (\frac{w_1}{d}, \frac{w_2}{d}, \ldots, \frac{w_d}{d}) \) (where we dispense with the superscript \( U \)) is an \( n \)-probability vector. For any \( n \)-probability \( p \), consider the function \( p \to \sum_{j=1}^n p_j^2 \) is Schur-convex \([65]\) and therefore its minimum value is
\[
\min_p \sum_{j=1}^n p_j^2 = \frac{1}{n},
\]
where minimization goes over all \( n \)-probability vectors, and is attained at the uniform distribution, \( p = (\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}) \). Finally, to prove the lemma we note that for the Fourier matrix \( F \), with elements
\[
F_{jl} = \frac{1}{\sqrt{n}} \omega^{(j-1)(l-1)}, \text{ where } \omega = \exp \frac{2\pi i}{n}, \text{ and } j, l \in [n]
\]
we have \( \sum_{j=1}^n (w^F_j)^2 = \frac{d^2}{n} \).

### 4. Technical version Theorem 2 in the main text

Now we are ready to provide a technical version of the first part of the Theorem 2 from the main text. Since the methods used in the proofs of inequalities (6) and (7) comprising Theorem 2 differ, we formulated two auxiliary technical theorems (Theorem 5 and Theorem 6 below), each covering one of the inequalities.

\[\text{Theorem 5.}\] [Technical formulation of inequality (6) from Theorem 2] Let \( n \in \{d, \ldots, d^2\} \), \( m \leq d \). Let \( \mathcal{M}^U \) denote a rank-one \( n \)-outcome Haar-random POVM on \( \mathbb{C}^d \). Let \( q_{\text{suc}}^{(m)}(\mathcal{M}^U) \) denote success probability of implementing \( \mathcal{M}^U \) via \( m \)-outcome measurements as in Eq. (1) for the standard partition \( X_1 = \{1, \ldots, m - 1\} \), \( X_2 = \{m, m + 1, \ldots, 2m - 2\}, \) etc., of \( [n] \). We then have
\[
\mathbb{P}_{U \sim \mu_n} \left( q_{\text{suc}}^{(m)}(\mathcal{M}^U) \geq c \frac{\gamma}{(1 + \sqrt{\gamma})^2} (1 - \epsilon) \right) \geq 1 - \frac{n}{m - 1} \exp \left( -a \left( 1 + \sqrt{\gamma} \right)^2 \frac{d^2}{\epsilon^2} \right),
\]
where \( 0 < \epsilon < \frac{\sqrt{\gamma} - 1}{2}, \gamma = \frac{2(m-1)}{d}, c \approx 6.79 \times 10^{-2} \) and \( a \approx 0.307. \) Furthermore, for special case \( m = d \), we have
\[
\mathbb{P}_{U \sim \mu_n} \left( q_{\text{suc}}^{(d)}(\mathcal{M}^U) \geq c (1 - \epsilon) \right) \geq 1 - \frac{n}{d - 1} \exp \left( -a \frac{d^2}{\epsilon^2} \right), \text{ where } 0 < \epsilon < 1,
\]
where \( c \approx 6.74 \times 10^{-2} \) and \( a \approx 1.79. \)
Remark 3. One can directly obtain an upper bound for the $m = d$ case, by evaluating the RHS of inequality (C.32) for $m = d1$. But in that case the success probability one gets is $4.65\%$, which is lower than the success probability in inequality (C.33) (which is $(6.74\%)$. Thus, a separate derivation for (C.33) is warranted.

Proof. Let $U_j$ be a truncation of $U$, occurring at the intersection between rows in $[d]$ and columns in $X_γ$. Using Lemma 4 we obtain the following upper bound to $\langle \| U_j \| \rangle_{\text{Haar}}$.

\[
\langle \| U_j \| \rangle_{\text{Haar}} \leq c' \left( 1 + \sqrt{\frac{d}{n}} \right) \sqrt{\frac{d}{n}},
\]

where $c' \approx 1.92$, and $γ = \frac{2(m-1)}{d}$. For the case $m = d$, the upper bound is simpler:

\[
\langle \| U_j \| \rangle_{\text{Haar}} \leq c' \sqrt{\frac{d-1}{n}},
\]

where $c' \approx 3.85$. To simplify the presentation, define

\[
A := c' \left( 1 + \sqrt{\gamma} \right) \sqrt{\frac{d-1}{n}}.
\]

From Lemma 1 it follows that the function $U \rightarrow \| U_j \|$ is 1-Lipschitz on $U(n)$ with respect to the Hilbert-Schmidt metric. Therefore, the function satisfies the following concentration inequality (see Subsection B 1)

\[
\mathbb{P}_{U \sim \mu_n} \left( \| U_j \| \geq A + t \right) \leq \exp \left( -\frac{nt^2}{12} \right), \forall t \geq 0,
\]

where we have used the fact that

\[
\left\{ U \in U(n) \left| \| U_j \| \geq A + t \right. \right\} \subseteq \left\{ U \in U(n) \left| \| U_j \| \geq \langle \| U_j \| \rangle_{\text{Haar}} + t \right. \right\}, \forall t \geq 0.
\]

By defining

\[
E_j := \left\{ U \in U(n) \left| \| U_j \|^2 \geq (A+t)^2 \right. \right\}.
\]

we can rewrite the inequality (C.37) as

\[
\mathbb{P}_{U \sim \mu_n} \left( E_j \right) \leq \exp \left( -\frac{nt^2}{12} \right), \forall t \geq 0.
\]

Suppose $U$ be such that it satisfies: $\sum_{j=1}^{n} \| U_j \|^2 \geq \alpha(A+t)^2$. This implies that for at least one $j \in [n], U \in E_j$. Using $\alpha \leq \frac{n}{m-1}$, we obtain

\[
\mathbb{P}_{U \sim \mu_n} \left( \sum_{j=1}^{n} \| U_j \|^2 \geq \frac{n}{m} (A+t)^2 \right) \leq \frac{n}{m} \exp \left( -\frac{nt^2}{12} \right), \forall t \geq 0,
\]

where we have used the union bound inequality on the event $\bigcup_{j=1}^{n} E_j$. When $U$ satisfies the inequality $\sum_{j=1}^{n} \| U_j \|^2 \geq \frac{n}{m} (A+t)^2$ then using the fact that $1 - t/A \geq (A+t)^{-2}$ when $0 \leq t/A \leq \frac{\sqrt{5} - 1}{2}$, we get that the success probability of our scheme is bounded by

\[
q^{(m)}_{\text{learn}}(M^U) \leq \frac{m}{n A^2} \left( 1 - \frac{t}{A} \right), \text{ where } 0 \leq \frac{t}{A} \leq \frac{\sqrt{5} - 1}{2}.
\]

Finally, by taking $\epsilon := \frac{t}{A}$, and using equation (C.36), the event (C.42) can be rewritten as

\[
q^{(m)}_{\text{learn}}(M^U) \leq c \frac{\gamma}{(1 + \sqrt{\gamma})^2} \left( 1 - \epsilon \right), \text{ where } 0 < \epsilon < \frac{\sqrt{5} - 1}{2},
\]

where $c = \frac{1}{2 \sqrt{\gamma}} \approx 0.136$. By plugging this into equation (C.12) we get the inequality (C.32). For the special case when $m = d$, we follow the same reasoning as above, starting from inequality (C.35), and then obtaining (C.33).
Theorem 6 (Technical formulation of inequality (7) from Theorem 2). Let \( n \in \{d, \ldots, d^2\} \), \( m \leq d \). Let \( M^U \) denote a rank-one \( n \)-outcome Haar-random POVM on \( \mathbb{C}^d \). Let \( q^{(m)}(M^U) \) be the maximal success probability of implementing \( M^U \) with postselection via convex combination of \( m \)-outcome measurements. We then have

\[
\mathbb{P}_{U \sim \mu_n} \left( q^{(m)}(M^U) \right) \leq \frac{2m \log n}{d} \left( 1 + \epsilon \right) \geq 1 - \frac{e d}{n^{1+2\epsilon}}, \quad \text{where} \quad \frac{1}{2 \log n} < \epsilon < \frac{n}{2 \log n} - 1. \tag{C.44}
\]

Remark 4. Theorem 6 is meaningful only for values of \( d, m \) and \( n \) such that \( 2m \log n < d \). Moreover, inequality (7) is reproduced by setting \( \epsilon = 1 \) in Eq. (C.44).

Proof. Let \( S_m \) be the set of all \( n \)-outcome POVMs simulable by quantum measurements with at most \( m \)-outcomes. Let \( M \) be arbitrary \( n \)-outcome POVM on \( \mathbb{C}^d \). To establish inequality (C.44) we shall use the following inequality between \( q^{(m)}(M) \) and the robustness \( R^{(m)}(M) \) (cf. Appendix A)

\[
q^{(m)}(M) \leq \frac{1}{R^{(m)}(M) + 1}. \tag{C.45}
\]

The robustness \( R^{(m)}(M) \) has an operational interpretation: it can be expressed via the maximal relative advantage that \( M \) can offer over all over all possible POVMs in quantum state diecrimination \( S_m \) (see Theorem 2, in [44]):

\[
R^{(m)}(M) = \max_{\varepsilon} \frac{P_{\text{succ}}(E, M)}{\max_{N \in S_m} P_{\text{succ}}(E, N)} - 1, \tag{C.46}
\]

where \( E = \{(q_i, \sigma_j)\}_{i=1}^n \) is an \( n \)-element ensemble of quantum states, and \( P_{\text{succ}}(E, M) \) (\( P_{\text{succ}}(E, N) \)) is the success probability for the minimum error discrimination of the states with the POVM \( M \) (or \( N \) respectively). For a given \( M \), we construct the following ensemble of states:

\[
E_M := \left\{ (q_j, \sigma_j) \text{ } \left| \text{ } q_j = \frac{1}{d \tr M_j}, \text{ } \sigma_j = \frac{1}{\tr(M_j^2)} \right. \right\}_{j=1}^n. \tag{C.47}
\]

Note that the convexity of \( S_m \) implies that \( \max_{N \in S_m} P_{\text{succ}}(E, N) \) is maximized on the extremal points of \( S_m \), which implies

\[
\max_{X \in \mathbb{C}^n, |X| = m} P_{\text{succ}}(E, N) \leq \max_{X \in \mathbb{C}^n, |X| = m} \sum_{j \in X} q_j, \tag{C.48}
\]

By using inequality (C.45), and the fact that \( E_M \) is a particular ensemble of quantum states (and that not-necessarily corresponding to the optimal value of the maximization in (C.46)). The obtain

\[
q^{(m)}(M) \leq \frac{\max_{X \in \mathbb{C}^n, |X| = m} \sum_{j \in X} \tr M_j}{\sum_{j=1}^n \tr M_j^2}. \tag{C.49}
\]

Now let \( M \) be a rank-one \( n \)-outcome POVM, whose \( j \)-th effect takes the form \( M_j = w_j \langle \psi_j | \psi_j \rangle \), where \( \langle \psi_j | \psi_j \rangle = 1 \), for all \( j \) and \( w_j = \tr M_j \). For this choice of \( M \), we get

\[
q^{(m)}(M) \leq \frac{\max_{X \in \mathbb{C}^n, |X| = m} \sum_{j \in X} w_j}{\sum_{j=1}^n w_j^2}. \tag{C.50}
\]

If \( M = M^U \) then we have \( w_j = \sum_{i=1}^d |U_{ij}|^2 \), and \( w_j \langle e_i | \psi_i \rangle = U_{ij} \), for \( i \in [d] \), and \( j \in [n] \). From Lemma 6, it is seen that the minimum value of \( \sum_{j=1}^n w_j^2 \) for any \( n \times n \) unitary \( U \) is \( \frac{d^2}{m} \). Thus we get

\[
q^{(m)}(M) \leq \frac{n}{d^2} \left( \max_{X \in \mathbb{C}^n, |X| = m} \sum_{j \in X} w_j \right). \tag{C.51}
\]

When \( U \) is distributed according to the Haar measure, then we can use inequality (C.51) from Lemma 5, which proves the theorem. □
Appendix D: Effects of depolarizing noise on the implementation of POVMs

In this section we provide detailed description of some concepts which appeared in the "Noise analysis" section of the main text. This includes description of how completely depolarizing noise on the level of quantum circuits propagates into POVMs implemented via two schemes – Naimark’s dilation (Section D 1) and the scheme introduced in this work. In Section D 2 we give a proof of Proposition 1 from the main text. We finish this section by providing some details and brief motivation behind the used noise model.

1. Depolarizing noise in Naimark’s dilation

In the main text, we described how the depolarizing noise acts on the quantum measurements. However, as already noted in Subsection B 4, in actual implementations to perform change of basis required by Naimark’s dilation, one usually implements (adjoint) unitaries acting on the states, i.e.,

\[
\text{tr} \left( \rho \otimes |0\rangle \langle 0 | U P_i U^\dagger \right) = \text{tr} \left( U^\dagger \rho \otimes |0\rangle \langle 0 | U P_i \right),
\]

where we used the fact that \( \text{tr} (P_i) = 1 \) for each rank-1 projector \( P_i \). Hence we see that performing perfect measurement \( \{ P_i \} \) on noisy quantum state \( \rho \) is experimentally equivalent to having a perfect state \( \eta \) and performing noisy quantum measurement with effects distorted as \( P_i \rightarrow \eta P_i + (1 - \eta) \frac{1}{n} \). Now since effects of (ideal) target POVM \( M \) are given (via Naimark’s dilation) by \( M^U = \text{tr}_B \left( \mathbb{1} \otimes |0\rangle \langle 0 | U P_i U^\dagger \right) \), we get that after the action of the noise channel, the effects of our target POVM are distorted as: \( M^U_i \rightarrow \eta M^U_i + (1 - \eta) \frac{1}{n} \), where \( \mathbb{1} \) is the identity operator on \( \mathbb{C}^d \). This motivates defining depolarized version \( M^{U,\eta}_i \) of the measurement \( M \) with effects given by

\[
M^{U,\eta}_i := \eta M^U_i + (1 - \eta) \frac{1}{n}.
\]

2. Proof of Proposition 1 in the main text

We start by stating the formal definition of Total-Variation Distance (TVD) followed by reiterated Proposition 1 from the main text.

**Definition 1** (Total variation distance). Let \( \mathbf{p} \) and \( \mathbf{q} \) be two \( n \)-probabilities, with \( i \)-th outcomes \( p_i \) and \( q_i \) respectively. Then the total variation distance between \( \mathbf{p} \) and \( \mathbf{q} \) is defined as

\[
d_{TV} (\mathbf{p}, \mathbf{q}) := \frac{1}{2} \sum_{j=1}^{n} | p_i - q_i |.
\]

**Theorem 7.** (Proposition 1 of the main text) Let \( M^{U,\eta}_i \) be the noisy implementation of Haar-random POVM \( M^U \) associated with unitary \( U \) (see Subsection B 4), with effects given by

\[
M^{U,\eta}_i := \eta M^U_i + (1 - \eta) \frac{1}{n}.
\]
where \( \eta \in [0,1] \) Then we have

\[
\left\langle \max_{\rho} d_{\text{TV}} \left( p \left( M^U \vert \rho \right), p \left( M^{U,\eta} \vert \rho \right) \right) \right\rangle_{\text{Haar}} \geq (1 - \eta) c_n ,
\] (D.7)

where \( p \left( M^U \vert \rho \right) \) (\( p \left( M^{U,\eta} \vert \rho \right) \)) is a probability distribution obtained via Born’s rule when measurement \( M^U \) \( M^{U,\eta} \) is performed on the state \( \rho \), and

\[
c_n = \left( 1 - \frac{1}{n} \right)^n \approx \frac{1}{e} .
\] (D.8)

**Proof.** The completely depolarising noise model specified in equation (D.6), allows to relate the probabilities \( p \left( M^U \vert \rho \right) \) and \( p \left( M^{U,\eta} \vert \rho \right) \) for any quantum state \( \rho \) as follows

\[
p \left( M^{U,\eta} \vert \rho \right) = \eta p \left( M^{U,\eta} \vert \rho \right) + (1 - \eta) p^\eta ,
\] (D.9)

where \( p^\eta \) is a uniform probability distribution over \( n \) outcomes. After basic manipulations, this gives that the total variation distance between \( p \left( M^U \vert \rho \right) \) and \( p \left( M^{U,\eta} \vert \rho \right) \) is

\[
d_{\text{TV}} \left( p \left( M^U \vert \rho \right), p \left( M^{U,\eta} \vert \rho \right) \right) = (1 - \eta) d_{\text{TV}} \left( p \left( M^U \vert \rho \right), p^\eta \right) ,
\] (D.10)

where, explicitly,

\[
d_{\text{TV}} \left( p \left( M^U \vert \rho \right), p^\eta \right) = \frac{1}{2} \sum_{j=1}^n \left| \text{tr} \left( \rho M_j^U \right) - \frac{1}{n} \right| .
\] (D.11)

Hence it turns out that total-variation distance between distributions generated by ideal and completely-depolarized version of \( M \) is proportional to a distance between the original distribution \( p \left( M^U \vert \rho \right) \) and completely random distribution \( p^\eta \). This observation will greatly simplify further considerations. Namely, recall that we are interested in bounding (expected value of) the worst-case (over quantum states) error in TVD of LHS of Eq. (D.11). We now see that it is equivalent to providing bound for the RHS of Eq.(D.11), which is easier to handle.

To start, recall that the matrix elements of \( M_j \) are directly related to the matrix elements of Naimark’s unitary \( U \) via \( M_j^U = |U_{ij}|^2 \) (see Subsection B 4). Now, since maximal value of any function is lower-bounded by any of the function’s value, by choosing input state \( \rho = |e_i\rangle \langle e_i| \) for some \( i \in [d] \), and inserting it into equation (D.11), we get that

\[
\max_{\rho} d_{\text{TV}} \left( p \left( M^U \vert \rho \right), p^\eta \right) \geq \frac{1}{2} \sum_{j=1}^n \left| \langle e_i \vert M_j^U \vert e_i \rangle - \frac{1}{n} \right| = \frac{1}{2} \sum_{j=1}^n \left| |U_{ij}|^2 - \frac{1}{n} \right| .
\] (D.12)

Thus the expected value is lower bounded as follows.

\[
\left\langle \max_{\rho} d_{\text{TV}} \left( p \left( M^U \vert \rho \right), p^\eta \right) \right\rangle_{\text{Haar}} \geq \left\langle \frac{1}{2} \sum_{j=1}^n \left| |U_{ij}|^2 - \frac{1}{n} \right| \right\rangle_{\text{Haar}} .
\] (D.13)

The permutational invariance of the Haar measure implies that

\[
\left\langle \frac{1}{2} \sum_{j=1}^n \left| |U_{ij}|^2 - \frac{1}{n} \right| \right\rangle_{\text{Haar}} = \frac{n}{2} \left\langle \left| |U_{ij}|^2 - \frac{1}{n} \right| \right\rangle_{\text{Haar}} .
\] (D.14)

Now we note that since \( U \) is Haar-random, the \( |U_{ij}|^2 \) has the same distribution as \( x \) from Eq.(B.9), i.e., \( p(x) = (n - 1) (1 - x)^{n-2} \). This allows to perform integration as

\[
\left\langle \left| |U_{ij}|^2 - \frac{1}{n} \right| \right\rangle_{\text{Haar}} = \int_0^1 dx \left| x - \frac{1}{n} \right| p(x) = \frac{2}{n} \left( 1 - \frac{1}{n} \right)^n .
\] (D.15)

Thus we get

\[
\left\langle \max_{\rho} d_{\text{TV}} \left( p \left( M^U \vert \rho \right), p^\eta \right) \right\rangle_{\text{Haar}} \geq \left( 1 - \frac{1}{n} \right)^n \approx 1/e .
\] (D.16)

Combining inequality (D.16) with equation (D.10) proves the theorem. \( \square \)
3. Depolarizing noise in implementation with post-selection

In this part we study how global depolarizing noise acting affects the quality of our POVM implementation scheme involving postselection. Recall that our scheme implements a measurement

\[ \mathbf{N} = (qM_1, \ldots, qM_n, (1-q)\mathbb{1}) \quad (D.17) \]

where \( M \) is a target POVM (which we assume consist of rank one effects) and \( q \) is a success probability of the implementation. The above measurement is realized as a convex mixture of \( m \)-outcome measurements (for simplicity we assume here that \( m - 1 \) divides \( n \) as

\[ \mathbf{N} = \sum_{\gamma} p_{\gamma} \mathbf{N}_{\gamma} \quad (D.18) \]

where each \( \mathbf{N}_{\gamma} \) has \( n + 1 \) formal outcomes, such that

\[ N_i^\gamma = \begin{cases} \lambda_\gamma M_i & \text{if } i \in \gamma, \\ 1 - \lambda_\gamma \sum_{i \in X_\gamma} M_i & \text{if } i = n + 1, \\ 0 & \text{if } i \in [n] \setminus X_\gamma, \end{cases} \]

where \( X_\gamma \) is subset of \( |X_\gamma| \leq m - 1 \) outcomes and probability distribution \( \{p_{\gamma}\} \) is defined by

\[ p_{\gamma} = \frac{q}{\lambda_\gamma}, \quad \lambda_\gamma = \|\sum_{i \in X_\gamma} M_i\|^{-1}, \quad q = \left( \sum_{\gamma=1}^{n} \frac{1}{\lambda_\gamma} \right)^{-1} = \left( \sum_{\gamma=1}^{n} \|\sum_{i \in X_\gamma} M_i\|^{-1} \right). \quad (D.20) \]

Each of the measurements \( \{\mathbf{N}_{\gamma}\} \) is implemented via Naimark’s dilation theorem (i.e. projective POVM on extended Hilbert space). As explained in the main text, if the target POVM \( M \) is rank one, and \( m \leq d \) then POVMs \( \mathbf{N}_{\gamma} \) can be implemented using Hilbert space of dimension \( m - 1 + d \leq 2d =: d_{\text{tot}} \). Now, due to the noise, the effects of the implemented POVM are distorted as

\[ N_i^\gamma \to \eta N_i^\gamma + (1-\eta) \frac{1}{d_{\text{tot}}}, \quad \text{for } i \notin \gamma \cup \{n + 1\}. \quad (D.21) \]

Therefore, in the presence of the assumed noise model our protocol effectively implements a POVM \( \mathbf{N}^\eta := \sum_{\gamma} p_{\gamma} \mathbf{N}_{\gamma}^\eta \), where by \( \mathbf{N}^\eta_{\gamma,\eta} \) we denoted indicate noisy veriat of POVM \( \mathbf{N}_{\gamma} \), with effects given in \( (D.21) \).

We are interested in bounding the distance between target distribution \( \{p(i|M,\rho)\}_{i=1}^{n} \) and the post-selected distribution from noisy POVM \( \mathbf{N}^\eta \), i.e., the distance,

\[ d_{TV} \left( \mathbf{p}(M|\rho), \mathbf{p}^{\text{noisy}}_{\text{post}}(M|\rho) \right) = \frac{1}{2} \sum_{i=1}^{n} |p(i|M,\rho) - \frac{p(i|\mathbf{N}^\eta_{\gamma},\rho)}{p(i < n + 1|\mathbf{N}^\eta_{\gamma},\rho)}| \quad (D.22) \]

where we have used

\[ p^{\text{noisy}}_{\text{post}}(i|M,\rho) = \frac{p(i|\mathbf{N}^\eta_{\gamma},\rho)}{p(i < n + 1|\mathbf{N}^\eta_{\gamma},\rho)}. \quad (D.23) \]

Let \( \gamma(i) \) to denote the label of the subset of outcomes to which \( i \) belongs. Consequently we have

\[ p_{\gamma(i)} = p_{\gamma} \text{ for all } i \in X_\gamma. \quad (D.24) \]

We note that for \( i \in \gamma \) we have

\[ p(i|\mathbf{N}^\eta_{\gamma},\rho) = p_{\gamma(i)} p(i|\mathbf{N}^\eta_{\gamma,\eta},\rho) = p_{\gamma(i)} \left( \eta p(i|\mathbf{N}^\eta_{\gamma},\rho) + (1-\eta) \frac{1}{d_{\text{tot}}} \right) = \eta q p(i|M,\rho) + (1-\eta) \frac{p_{\gamma(i)}}{d_{\text{tot}}}, \]

where we used the fact that \( p_{\gamma(i)} p(i|\mathbf{N}^\eta_{\gamma},\rho) = q p(i|M,\rho) \). To understand the behaviour of Eq. \( (D.22) \) we need to calculate how probability of postselection changes due to the noise. Using the fact that subsets \( \gamma \) are disjoint end employing \( (D.25) \) we obtain

\[ p(i < n + 1|\mathbf{N}^\eta_{\gamma},\rho) = \sum_{\gamma=1}^{\alpha} p_{\gamma} \sum_{i \in X_\gamma} p(i|\mathbf{N}^\eta_{\gamma,\eta},\rho) = \eta q + (1-\eta) \frac{|X_\gamma|}{d_{\text{tot}}}, \quad (D.25) \]
where we defined $\langle |X_\gamma| \rangle := \sum_{\gamma=1}^{\alpha} p_\gamma |X_\gamma|\rangle$ and used the fact that $\sum_{\gamma} p_\gamma |M, \rho\rangle = 1$.

Now we rewrite the Eq. (D.22) as

$$\frac{1}{2 p(i < n + 1|N^q, \rho)} \sum_{i=1}^{n} |p(i < n + 1|N^q, \rho) p(i|M, \rho) - p(i|N^q, \rho)|, \quad (D.27)$$

We calculate each of the summands explicitly using Eq. (D.25) and Eq. (D.26) and obtain

$$|p(i < n + 1|N^q, \rho) p(i|M, \rho) - p(i|N^q, \rho)| = \frac{(1-\eta)}{d_{tot}} |p(i|M, \rho) \langle |X_\gamma| \rangle - p_\gamma(i)|, \quad (D.28)$$

Using the bound $|a - b| \leq |a| + |b|$ and summing over $i$ we obtain

$$\frac{(1-\eta)}{d_{tot}} \sum_{i=1}^{n} |p(i|M, \rho) \langle |X_\gamma| \rangle - p_\gamma(i)| \leq \frac{(1-\eta)}{d_{tot}} \sum_{i=1}^{n} (p(i|M, \rho) \langle |X_\gamma| \rangle + p_\gamma(i)) = \frac{2(1-\eta)}{d_{tot}} \langle |X_\gamma| \rangle. \quad (D.29)$$

Consider our scheme for the special choice $n = d^2$ and $m = d + 1$, hence $d_{tot} = 2d$ and $|X_\gamma| = d$ for all $\gamma$. This gives $\frac{2\langle |X_\gamma| \rangle}{d_{tot}} = 1$. Combining this with the inequality in Eq. (D.29) and the Eq. (D.27) yields that for our scheme we have

$$d_{TV}(p(M|\rho), p_{\text{post}}^{\text{noisy}}(M|\rho)) \leq \frac{1}{2} \frac{(1-\eta)}{p(i < n + 1|N^q, \rho)} = \frac{1}{2} \frac{(1-\eta)}{\eta q + (1-\eta)} \leq (1-\eta) \max\left\{\frac{1}{2q}, 1\right\}. \quad (D.30)$$

For Haar-random rank-one POVMs we have $q_{\text{succ}}(M^U) > c$ (see Theorem 2), where $c$ is an absolute constant. Combining this with the fact that for generic unitaries on $2N$ qubits we have $\eta_{\text{post}}^{\text{exp}}(-\Theta(4^N))$ we obtain the assertion made in the main text, i.e., that for typical Haar-random $d^2$-outcome POVMs $M^U$ we have

$$d_{TV}(p(M^U|\rho), p_{\text{post}}^{\text{noise}}(M^U|\rho)) \leq C(1-\exp(-\Theta(4^N))). \quad (D.31)$$

4. Noise model details

In the main text and in previous subsections, we adopted a very simple noise model parametrized by only single number – visibility $\eta$. The main motivation for that choice was the fact that since we consider mostly generic Haar-random POVMs, the circuits which implement them can be considered random, and that such model was considered in Google’s recent demonstration of computational advantage (which used random circuits) [40]. In Ref. [40], authors consider $\eta$ of the following form [54]

$$\eta = \exp(-r_1 g_1 - r_2 g_2 - N(r_p + r_m)), \quad (D.32)$$

where $r_1$, $r_2$ are respectively the error rates for single and two-qubit gates, $g_1$, $g_2$ are number of single-qubit and two-qubit gates, $N$ is the total number of qubits in the circuit, and $r_p$ and $r_m$ are SPAM (state preparation and measurement) errors. As indicated in the main text, since generic circuits require number of two-qubit gates scaling exponentially with the system size, we considered faulty two-qubit gates as the main error source. We note, however, that in the above model it is in fact assumed that readout noise can be effectively treated as uncorrelated and identical. In presence of the measurement noise cross-talk, the more realistic noise model should be considered (see, for example, recent works [57, 58]).

Appendix E: Numerical results

1. IC and SIC POVMs

   a. Informationally complete measurements covariant with respect to $\mathbb{Z}_d \times \mathbb{Z}_d$

   To explain how we construct informationally complete (IC) measurements, let us first recall that a POVM is called covariant with respect to a group, if all of the measurement operators can be obtained from some fiducial vector by the action of that group. Hence if one has a way of constructing that fiducial vector and the unitary representation of chosen group, one can easily generate all of the effects of covariant measurement. In this work we use the explicit construction from Ref. [39] which shows how to obtain fiducial vector for the POVM covariant with respect to $\mathbb{Z}_d \times \mathbb{Z}_d$ (which can be thought of as finite-dimensional analogue of Weyl-Heisenberg
group), where $d$ is the dimension of the system. Such POVM has $d^2$ rank-1 effects and is shown to be informationally-complete [39]. A fiducial vector is constructed as

$$|\psi_\alpha\rangle = \sqrt{\frac{1-|\alpha|^2}{1-|\alpha|^{2d}}} \sum_{i=0}^{d-1} \alpha^i |i\rangle,$$

(E.1)

where $\alpha$ is a parameter characterizing the POVM and has to fulfill condition $0 < |\alpha| < 1$. Now vectors defining other effects of that POVM are obtained as

$$|\psi_{m,n}\rangle = \frac{1}{\sqrt{d}} U_{m,n} |\psi_\alpha\rangle,$$

(E.2)

where $U_{m,n}$ is a (projective) unitary representation of $\mathbb{Z}_d \times \mathbb{Z}_d$ given by

$$U_{m,n} = \sum_{k=0}^{d-1} \exp \left( \frac{2\pi i}{d} km \right) |k\rangle \langle k \oplus n|,$$

(E.3)

with $m, n \in [0, d-1]$ and $\oplus$ is addition modulo $d$. See Ref. [39] for more details. In our simulations we arbitrarily choose the free parameter to be $\alpha = \frac{1}{2} (1 + i)$. We note that we checked a few other instances of this parameter and we did not observe quantitative differences in the probability of success of POVMs simulation using our scheme.

### b. Symmetric Informationally Complete measurements

The measurement is called symmetric if its effects have equal pairwise Hilbert-Schmidt scalar products. The search for symmetric and informationally complete (SIC) measurements is an active area of research [38] and even existence of SICs in arbitrary dimension $d$ is an open problem. To date, SIC POVMs have been found either numerically or analytically for a restricted collection of dimensions [36, 66, 67]. SIC POVMs are, similarly to IC, represented by a single fiducial vector and we generate other measurement operators from that vector by the action of $\mathbb{Z}_d^d \times \mathbb{Z}_d^d$ group (we note that all SIC POVMs found to date are covariant with respect to some group, and the most of them covariant to $\mathbb{Z}_d^d \times \mathbb{Z}_d^d$ group).

In this work, the POVMs in dimensions $d \in [2, 100]$ have been downloaded from database [53] maintained by Christopher A. Fuchs, Michael C. Hoang, and Blake C. Stacey. The POVMs for dimensions

$$[100, 193] \cup \{194, 195, 201, 204, 224, 228, 255, 259, 288, 292, 323, 327, 364, 399, 403, 489, 528, 725, 844, 1155, 1299\}$$

(E.4)

were provided by Markus Grassl in private correspondence.

### 2. Haar-random POVMs

In this work, we are interested in generating Haar random $d$-dimensional POVMs with $d^2$ outcomes. A straightforward method to do so would be to generate Haar-random $d^2 \times d^2$ unitary matrix and take its $d^2 \times d$ submatrix as defining such POVM. However, generation of random matrices quickly becomes unfeasible – due to large amount of memory required, we were not able to generate such matrices for high $d$. As a workaround, instead of generating random $d^2 \times d^2$ unitary matrices, we generated random $d^2 \times d$ isometries. To do so, we implemented the following algorithm.

1. Generate $d$ iid random complex Gaussian vectors of size $d^2$ – call them $\{v_i\}_{i=1}^d$.
2. Construct a Gramian matrix $G$ of those vectors as

$$G_{ij} = \langle v_i | v_j \rangle.$$

(E.5)

3. Perform LDL decomposition of the Gramian matrix as

$$G = L \sqrt{D} \sqrt{D}^*,$$

(E.6)

where $L$ is lower-triangular and $D$ diagonal.
4. Define \( R = \left( \sqrt{D} L^* \right)^{-1} \) and construct new set of vectors as

\[
e_k = \sum_i R_{i,k} v_i .
\]

(E.7)

It follows that \( \{e_k\}_{k=1}^d \) forms an orthonormal set of \( d^2 \)-dimensional random vectors. Hence those vectors can be used to construct a \( d^2 \times d \) isometry.

5. To construct a POVM one simply looks at rows of this isometry as a set of \( d^2 \) vectors of dimension \( d \). Since the matrix is an isometry, it follows that those rows define rank-1 effects of \( d^2 \)-outcome random POVM.