This is a repository copy of *An efficient magic state approach to small angle rotations*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/107576/

Version: Accepted Version

Article:

Campbell, E. and O'Gorman, J. (2016) An efficient magic state approach to small angle rotations. Quantum Science & Technology, 1. 015007. ISSN 2058-9565

https://doi.org/10.1088/2058-9565/1/1/015007

Reuse

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

An efficient magic state approach to small angle rotations

Earl T. Campbell[1] and Joe O'Gorman[2]
[1]*Department of Physics & Astronomy, University of Sheffield, Sheffield, S3 7RH, United Kingdom.*[*]
[2]*Department of Materials, University of Oxford, Oxford, OX1 3PH, United Kingdom.*

Standard error correction techniques only provide a quantum memory and need extra gadgets to perform computation. Central to quantum algorithms are small angle rotations, which can be fault-tolerantly implemented given a supply of an unconventional species of magic state. We present a low-cost distillation routine for preparing these small angle magic states. Our protocol builds on the work of Duclos-Cianci and Poulin [*Phys. Rev. A* **91**, 042315 (2015)] by compressing their circuit. Additionally, we present a method of diluting magic states that reduces costs associated with very small angle rotations. We quantify performance by the expected number of noisy magic states consumed per rotation, and compare with other protocols. For modest size angles, our protocols offer a factor 24 improvement over the best known gate synthesis protocols and a factor 2 over the Duclos-Cianci and Poulin protocol. For very small angle rotations, the dilution protocol dramatically reduces costs, giving several orders magnitude improvement over competitors. There also exists an intermediary regime of small, but not very small, angles where our approach gives a marginal improvement over gate synthesis. We discuss how different performance metrics may alter these conclusions.

I. INTRODUCTION

Fault-tolerant quantum computing involves a host of resource overheads, entering at different stages of the process. The most widely known cost is that of encoding a logical qubit into many physical qubits, which provides safe storage of quantum information. However, once encoded, a logical qubit only natively supports a limited set of fault-tolerant operations [1]. For high-threshold codes, the native operations belong within the Clifford group. Additional layers of gadgets are required to enable general purpose quantum computation. A corollary of the Solovay-Kitaev theorem [2] is that if the Clifford group is augmented by a non-Clifford T gate, also called the $\pi/8$ gate, the device can efficiently approximate any required quantum algorithm. These T gates can in turn be fault-tolerantly performed by using state injection of high-fidelity magic states, prepared by distillation [3]. However, the Solovay-Kitaev theorem requires a huge number of T gates, and early magic state distillation protocols put a high price on each one. The last decade has seen substantial advances in both these areas. Magic state distillation is now possible at better rates [3–7], reducing the expected cost per high quality T-gate. The number of T-gates needed to approximate some unitary, the so-called T-count, has also been significantly reduced after the discovery of new gate synthesis techniques [8–12].

Nevertheless, fault-tolerant quantum computing remains a monumental challenge, so further resources savings are essential. One suggested route is to circumvent magic states altogether, for instance by using gauge-fixing [13, 14] of subsystem codes like the 3D colour code [15, 16]. However, this route still pays a price for gate-synthesis and the additional qubit cost of 3D codes leads to a cubic scaling of overheads that is also achievable using magic states [17, 18]. Furthermore, all current indications are that colour codes have a much lower error correction threshold [19] than the toric code [17, 20–22]. For now, such low-noise levels appear beyond technological reach, ruling out gauge-fixing in the near-term. Alternatively, one could work with qudits, d-level systems, which are favourable in terms of the cost of magic state distillation [23–26], though in the qudit context little is known about gate synthesis or experimental feasibility.

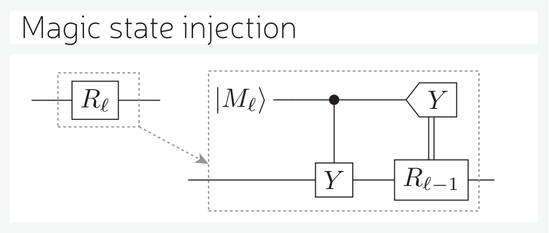

FIG. 1. State-injection using a magic state $|M_\ell\rangle$ to perform a non-Clifford gate R_ℓ. If the Y-basis measurement outcome is "+1", then a correction $R_{\ell-1}$ is needed.

Landahl and Cesare [27] were the first to suggest that the gate synthesis overhead could be reduced by distilling different species of magic states, which provide smaller angle rotations instead of T-gates. Specifically, they considered $\exp(i\pi Z/2^\ell)$ gates for integer ℓ, where $\ell = 3$ gives the T-gate. These gates are part of an important family called the Clifford hierarchy [28], with $\exp(i\pi Z/2^\ell)$ residing in the ℓ^{th} level of the hierarchy and naturally appearing in quantum simulation [29–33] and the quantum Fourier transform. Landahl and Cesare found that for small integer ℓ, distillation was favourable over gate synthesis methods known at the time. Unfortunately, for very small angle rotations $5 \lesssim \ell$, they saw this advantage vanish.

---

[*] earltcampbell@gmail.com

Progress in gate synthesis has significantly eroded the importance of the Landahl and Cesare result, though their general idea endures. Several researchers found efficient protocols that distill Toffoli magic states [34–36]. Duclos-Cianci and Poulin [37] modified a protocol proposed by Meier, Eastin and Knill [4] to prepare magic states that again provide small angle rotations. This recent result was a substantial leap forward in the art of distilling magic states for small angle rotations, making magic state distillation cheaper than gate synthesis in many instances. We revisit the work of Duclos-Cianci and Poulin, and find their protocol can be further refined. Our proposed alternative requires fewer resources per attempt, and achieves superior error suppression with a higher success probability. Both our protocol and that of Duclos-Cianci and Poulin build on the earlier work of Meier, Eastin, and Knill (MEK) [4].

For very small angle rotations, the required magic states become close to stabiliser states. Duclos-Cianci and Poulin observed that in this regime some magic states in the protocol can be supplanted by stabiliser states, reducing resource costs in a certain regime. Instead, we introduce the notion of magic state dilution, which takes magic states for ℓ level rotations, and converts them into a larger number of magic states for a finer rotations with higher ℓ. Error rates are adjusted in the dilution process, and we find it works best at high ℓ. Remarkably, dilution can even reduce noise when used at sufficiently high ℓ.

II. MAGIC STATES MODEL

First we review the Clifford hierarchy and the magic state model. The well known Pauli group $\mathcal{P}$ is the group composed of tensor products of the single qubit Pauli operators. Unitaries mapping the Pauli group to itself are called Clifford unitaries, which again form a group $\mathcal{C} := \{U | UPU^\dagger \in \mathcal{P}, \forall P \in \mathcal{P}\}$. Clifford operations are physical operations composable from Clifford unitaries, measurement of Pauli operators and preparation of their eigenstates (the stabiliser states), and classical feedforward. The magic state model of quantum computation [3] assumes Clifford operations are free resources that can be implemented perfectly, and proceeds to evaluate the cost of non-Clifford operations. Such a model is suitable for logical qubits where the Clifford operations are fault-tolerantly protected, as is common. In our conclusions, we discuss further the validity of counting only non-Clifford resources.

Unitaries outside the Clifford group can fall into other levels of the Clifford hierarchy, defined recursively as

$$\mathcal{C}_\ell := \{U | UPU^\dagger \in \mathcal{P}, \forall P \in \mathcal{C}_{\ell-1}\}, \tag{1}$$

where $\mathcal{C}_1 := \mathcal{P}$. It follows that the Clifford group is the second level, and all higher levels include non-Clifford gates. We will soon see the Clifford hierarchy plays an important operational role in a teleportation process known as state injection [38].

We introduced small angle rotations in the Z basis, but here it is more convenient to work in the Y basis. We define unitaries $R_\ell := \exp(i\theta_\ell Y)$ with $\theta_\ell = \pi/2^\ell$. The R_1 and R_2 gates are Clifford, which are presumed ideal and an inexpensive resource. The R_3 gate is non-Clifford and equivalent to the T gate. For $\ell \geq 3$, the gates are non-Clifford and belong to the ℓ^{th} level of the Clifford hierarchy. Also important here are certain Hermitian operators in the Clifford hierarchy, and we define $H_\ell := R_\ell X R_\ell^\dagger = R_\ell^2 X = R_{\ell-1} X$ and find $H_\ell \in \mathcal{C}_{\ell-1}$. We remark that H_3 equals the Hadamard.

Preparation of non-stabiliser states is also a non-Clifford operation, and these magic states also fall into a natural hierarchy. Recall that stabiliser states are eigenstates of Pauli operators, which are elements of $\mathcal{C}_1$. For example, the computational basis sates $|0\rangle$ and $|1\rangle$ are stabilised by the Z and $-Z$ Pauli operators, respectively. Below we will also make use of the stabiliser states $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$, which satisfy $(\pm X)|\pm\rangle = |\pm\rangle$ for Pauli X. We consider magic states that are eigenstates of the H_ℓ operators, so that

$$|M_\ell\rangle = H_\ell |M_\ell\rangle = R_\ell |+\rangle, \tag{2}$$

$$|\bar{M}_\ell\rangle = (-H_\ell)|\bar{M}_\ell\rangle = R_\ell |-\rangle, \tag{3}$$

which we refer to as ℓ^{th} level magic state states. Such resources can be used to inject R_ℓ rotations into circuits as shown in Fig. (1). The injection procedure is probabilistic, and with probability 1/2 it performs R_ℓ and with probability 1/2 it performs $R_\ell^\dagger$. In the latter case a correction of $R_\ell^2 = R_{\ell-1}$ is needed to get the desired result. Since R_2 is Clifford, the injection process is ensured to terminate within $\ell - 2$ attempts.

Therefore, one can accomplish small angle rotations with a cost that depends on the cost of distilling high-fidelity $|M_\ell\rangle$ states. In contrast, gate synthesis methods prescribe distilling just $|M_3\rangle$ states and finding a gate sequence $R_\ell \approx C_1 T C_2 \ldots T C_n$ in terms of T gates and Cliffords C_j. Although, $|M_\ell\rangle$ states will prove more expensive than $|M_3\rangle$, gate synthesis can require very many $|M_3\rangle$ states, making it more expensive overall. There are many gate synthesis algorithms for finding gate sequences and here we consider the Selinger and Ross [11] algorithm (SR) and probabilistic quantum circuits with fallback (PQF) [39]. SR has the benefit of a freely available software implementation and is provable optimal for unitary synthesis. PQF makes use of ancilla and measurements to achieved the best known performance, about a factor 3 more efficient than SR. Several other methods (see e.g. [9, 40]) exist, but all fall somewhere between SR and PQF in performance. We discuss the applicability of our results to rotations of arbitrarily angles in App. A.

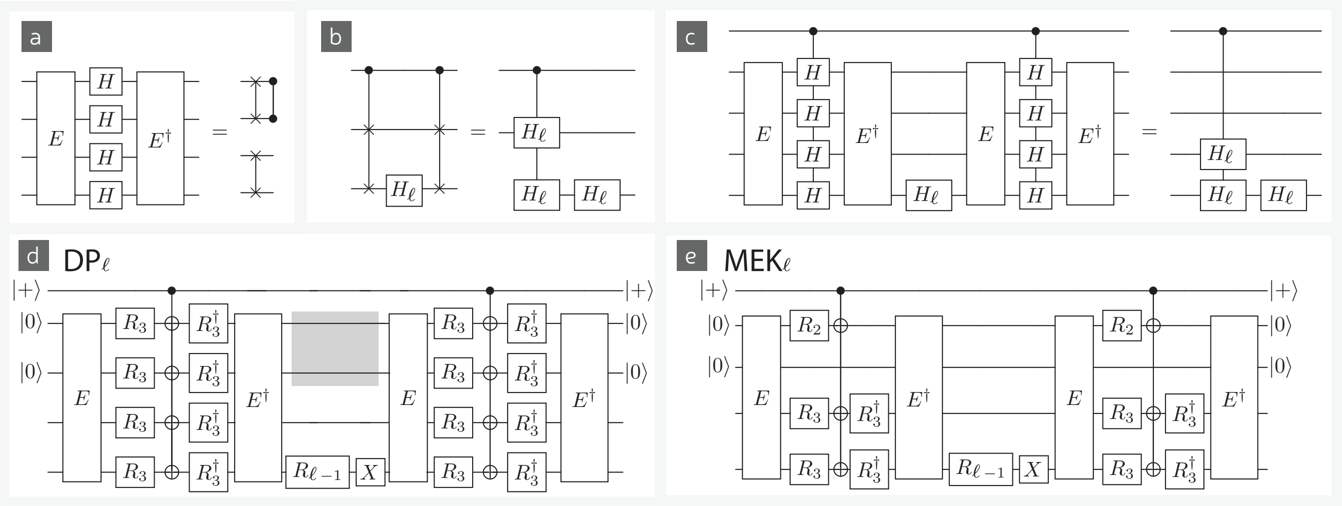

FIG. 2. (a,b,c) Circuit identities and gadgets used to construct distillation protocols. (d) a non-compressed distillation circuit. Adding $|0\rangle\langle 0|$ measurements and preparations within the grey box gives the DP$_\ell$ protocol. (e) the compressed distillation circuit describing MEK$_\ell$. In both (d) and (e) noisy $|M_\ell\rangle$ states are input on the bottom two circuits lines (labelled qubits 3 and 4 in the main text) and all R_ℓ gates with $\ell \geq 3$ are approximately implemented by injection of a noisy $|M_\ell\rangle$ states.

III. OVERVIEW OF RESOURCE COSTS

Next we summarise the work of Meier, Eastin, and Knill (MEK) [4], upon which Duclos-Cianci and Poulin built their construction. The MEK protocol takes 10 input magic states and with some probability outputs 2 magic states. Therefore, MEK is said to be a 10 → 2 protocol. Accounting for the species of magic states, we call MEK a $10_3 \rightarrow 2_3$ protocol where the subscripts indicate that MEK both inputs and outputs 3rd level magic states. We use DP$_\ell$ to label the Duclos-Cianci and Poulin protocol for distilling ℓ^{th} level magic states. Each round of DP$_\ell$ consumes a cocktail of different input magic states, and we describe it as a

$$\left\{16_3, 2_\ell, 1_{\ell-1}, \left(\frac{1}{2}\right)_{\ell-2}, \left(\frac{1}{4}\right)_{\ell-3}, \dots \left(\frac{1}{2^{\ell-3}}\right)_3\right\} \rightarrow 2_\ell \tag{4}$$

protocol. Again, subscripts show what level magic state is used. We show the expected number of inputs required per attempt, which is sometimes a fraction. The fractional sequence terminates at the third level, because lower levels correspond to Clifford resources and are considered free. The DP$_\ell$ protocol is presented as a direct generalisation of MEK, but one sees that in the case of $\ell = 3$ the DP$_3$ protocol is a $18_3 \rightarrow 2_3$ protocol and so actually needs almost twice as many input states as MEK. Here we construct a streamlined variant of DP$_\ell$, that we call MEK$_\ell$ and is a

$$\left\{8_3, 2_\ell, 1_{\ell-1}, \left(\frac{1}{2}_{\ell-2}\right), \left(\frac{1}{4}\right)_{\ell-3}, \dots \left(\frac{1}{2^{\ell-3}}\right)_3\right\} \rightarrow 2_\ell \tag{5}$$

protocol. We find that MEK$_3$ corresponds precisely to the original MEK, so that MEK$_\ell$ is a more apt generalisation than DP$_\ell$. Not only does MEK$_\ell$ require fewer input resources, it also achieves superior error suppression with a higher success probability.

Both MEK and DP$_\ell$ make use of a simple 4 qubit code. We use E, short for encoder, to denote the Clifford circuit that brings qubits into the codespace and acts on pairs of Pauli operators as

$$(Z_1, X_1) \rightarrow (Z_1Z_2Z_3Z_4, X_1X_3X_4), \tag{6}$$
$$(Z_2, X_2) \rightarrow (X_1X_2X_3X_4, Z_2), \tag{7}$$
$$(Z_3, X_3) \rightarrow (Z_1Z_4, X_1X_3), \tag{8}$$
$$(Z_4, X_4) \rightarrow (X_1X_4, Z_1Z_3). \tag{9}$$

For Pauli operators, we always use the subscript to denote which qubit the operator acts upon. Preparing the first two qubits in the state $|0\rangle$ and running the encoder will yield the code stabilised by $Z_1Z_2Z_3Z_4$ and $X_1X_2X_3X_4$. The logical state in the codespace is determined by the initial states of the last two qubits. Having defined the codespace, we now turn to describing DP$_\ell$ in more detail, where we will also discuss the important features of the 4-qubit code.

IV. UNCOMPRESSED DP$_\ell$ PROTOCOL

The protocol DP$_\ell$ is illustrated in Fig. (2d). Throughout, we label the top wire as the control qubit c, and the subsequent qubits are labelled 1 to 4. We summarise the main steps of the protocol here

1. Prepare qubits c, 1 and 2, in stabiliser states $|+\rangle$, $|0\rangle$ and $|0\rangle$, respectively;

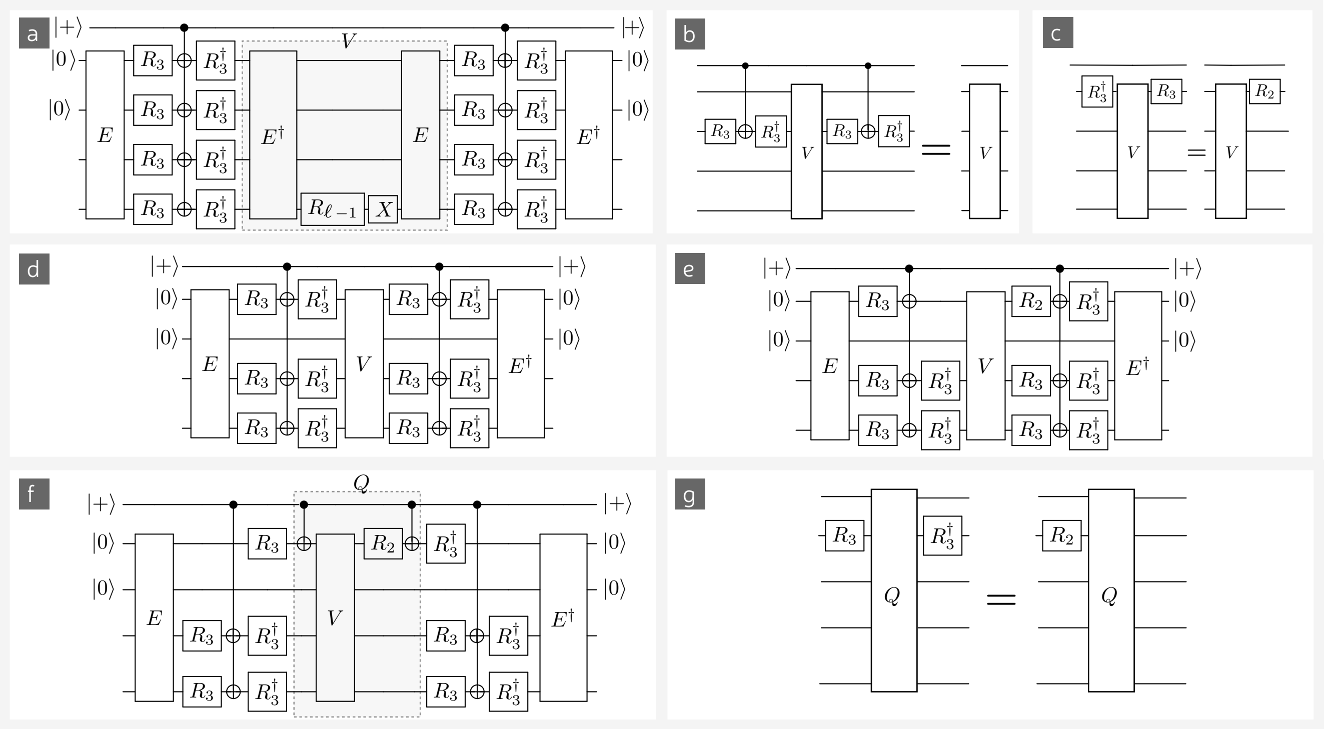

FIG. 3. Circuit identities and reductions used to obtain MEK_ℓ. (a) circuit before compression, with subcircuit V labelled. (b) and (c) show properties of V. (d) shows partially compressed circuit after applying identity (b). (e) further compressed circuit after applying identity (c). In (f) we identify subcircuit Q, with (g) showing a property of Q. Applying (g) to (f) yields the final compressed circuit, shown in Fig. (2e).

2. Prepare qubits 3 and 4 in noisy $|M_\ell\rangle$ states with ϵ_ℓ error;
3. Perform the circuit gates shown in Fig. (2d): the R_3 gates are achieved by state injection using eight $|M_3\rangle$ states, and $R_{\ell-1}$ is achieved with state injection resulting in overall $\eta_{\ell-1}$ error;
4. Qubit c is measured in X basis, and we continue if the outcome is "$+1$" and otherwise declare FAILURE;
5. Qubits 1 and 2 are measured in Z basis, and we declare SUCCESS if the outcome is "$+1$" and otherwise declare FAILURE;
6. If successful, qubits 3 and 4 are output as $|M_\ell\rangle$ states of higher fidelity.

Note that step 3 can include some additional postselected measurements highlighted in Fig. (2d). If included these measurements detect some additional errors, but even without these measurements the protocol quadratically suppresses noise. For simplicity we will assume they are not performed. We next review some of the basic intuition behind why DP_ℓ works, which is closely related to properties of the 4 qubit code used.

The codespace provides both important transversality features and keeps the protocol protected against certain faults. Regarding transversality, a global Hadamard $H \otimes H \otimes H \otimes H$ will preserve the code space, implementing a logical SWAP between the two encoded logical bits. More generally, one can verify that $E(H \otimes H \otimes H \otimes H)E^\dagger$ acts as shown in Fig. (2a). We see that even without fixing the first two qubits to $|0\rangle$, this circuit implements a swap and a phase-swap (a swap combined with a phase gate). Furthermore, implementing controlled Hadamards within the encoding will realise controlled versions of the swap and phase-swap. Next, we note that for any Hermitian unitary, such as H_ℓ, we have that conjugation by controlled-swaps will produce a controlled $H_\ell \otimes H_\ell$ as shown in Fig. (2b). We combine this observation with the transversality properties of the code to get the identity of Fig. (2c). An ancilla on the control of the controlled $H_\ell \otimes H_\ell$ is used to measure the $H_\ell \otimes H_\ell$ observable. If the desired magic state is an eigenstate of H_ℓ, then measuring $H_\ell \otimes H_\ell$ allows us to detect a single error between two noisy $|M_\ell\rangle$ states input on the bottom two circuit lines. In this sense, the codespace has provided a SWAP gadget for distilling noisy $|M_\ell\rangle$ magic states.

Although the controlled-Hadamard rotations and H_ℓ rotation may be non-Clifford, noisy magic states can be used to implement these operations. To see this recall that $H = R_3 X R_3^\dagger$, and so a control-Hadamard can be implemented by a control-X gate sandwiched between R_3 and $R_3^\dagger$. In turn, R_3 and $R_3^\dagger$ can each be implemented at the cost of a single $|M_3\rangle$ magic state. Given noisy $|M_3\rangle$ magic states, we implement noisy R_3 gates. This brings us to the second role played by the error-correction code. The R_3 gates are performed on qubits within an error correction code that can detect a single qubit error, and so we will detect a single error in any $|M_3\rangle$ magic states. Let us explain this point in more detail. When using state injection, if a magic state carries an error, then it will result in $Y \cdot R_3$ instead of R_3, and so there is an additional Y acting on one of the four qubits. Inside the encoding, the state is an eigenstate of $X \otimes X \otimes X \otimes X$, but a Y error will cause the state to become an eigenstate of $-X \otimes X \otimes X \otimes X$. At the end of the circuit, we decode and measure, which is equivalent to measuring the $X \otimes X \otimes X \otimes X$ observable. Since we postselect on all $+1$ outcomes, any error on a single $|M_3\rangle$ will be detected. In contrast, the central $R_{\ell-1}$ rotation occurs outside the protection of the codespace when the logical qubits have been decoded onto single qubits. Therefore, the protocol will not be robust against failure of this gate, and so this rotation must be high fidelity and we herein call it the pivotal rotation. Nevertheless, we can construct good distillation protocols provided magic states used in the pivotal rotation have already been distilled to a much higher fidelity than all other elements of the circuit. These high fidelity resources will be costly, but the protocol remains efficient because resources for performing $R_{\ell-1}$ are less valuable than the $|M_\ell\rangle$ states we are distilling. We are erecting a pyramid of distillation protocols, with resources from lower in the Clifford hierarchy fuelling distillation at higher levels.

V. COMPRESSED MEK$_\ell$ PROTOCOL

Our main contribution is to show that this circuit can be compressed into Fig. (2e). This cancels several R_3 rotations to reduce the number of magic states consumed. The steps of the protocol roughly follows those numerically listed in the previous section, except step 3 now uses the circuit of Fig. (2e), and we use only 8 magic states to inject the R_3 gates. It is important that the circuit retains its fault-tolerance properties through the compression process. That is, even when compressed the noisy R_3 gates still occur within the four-qubit error correction code, and so we still expect to detect the failure of any single R_3 gate. In App. B, we present a full noise analysis that rigorously confirms this intuition.

Here we show how to compress the distillation protocol, removing unnecessary non-Clifford gates. We start with the protocol shown in Fig. (2d) and through a series of circuit reductions arrive at Fig. (2e). Taking Fig. (2d), we identify a subcircuit V shown in Fig. (3a) inside a shaded box. Next we establish two properties of V, which are illustrated in Figs. (3b) and (3c). Algebraically, the V circuit is simply $V = E\exp(i\theta_{\ell-1}Y_4)X_4E^\dagger$, and using Eqs. (6) we see that $V = \exp(-i\theta_{\ell-1}Y_1Z_3X_4)Z_1Z_3$. This acts trivially on the second qubit and control qubit, entailing the circuit identity of Fig. (3b). Going further, we notice that $VY_1 = -Y_1V$ and so $\exp(i\theta_3Y_1)V\exp(-i\theta_3Y_1) = \exp(i2\theta_3Y_1)V$. Using $2\theta_3 = \theta_2$, we conclude $\exp(i\theta_3Y_1)V\exp(-i\theta_3Y_1) = \exp(i\theta_2Y_1)V$. Recall that $\exp(i\theta_2Y_1)$ is the R_2 gate acting on qubit 1, and so we have the identity shown in Fig. (3c).

Applying the identity Fig. (3b), shows that Fig. (3d) is equivalent to the original circuit. This observation has eliminated 4 non-Clifford gates. Next, we apply the identity of Fig. (3c), to obtain Fig. (3e). Since R_2 is Clifford, we have removed 2 further non-Cliffords from the circuit.

Next, we group together a new collection of circuit elements Q shown in dashed box of Fig. (3f). Algebraically, Q is

$$\begin{aligned} Q &= C^X_{c,1}\exp(i\theta_2Y_1)VC^X_{c,1} \qquad (10)\\ &= C^X_{c,1}\exp(i\theta_2Y_1)\exp(-i\theta_{\ell-1}Y_1Z_3X_4)Z_1Z_3C^X_{c,1}\\ &= \exp(i\theta_2Z_cY_1)\exp(-i\theta_{\ell-1}Z_cY_1Z_3X_4)Z_cZ_1Z_3 \end{aligned}$$

where we have used C^X for control-X gates and their conjugation relations $C^X_{c,1}X_cC^X_{c,1} = X_cX_1$ and $C^X_{c,1}Z_1C^X_{c,1} = Z_cZ_1$. From this expression for Q we can again see that $QY_1 = -Y_1Q$, which entails that $\exp(-i\theta_3Y_1)Q\exp(i\theta_3Y_1) = Q\exp(i2\theta_3Y_1) = Q\exp(i\theta_2Y_1)$. This demonstrates the identity of Fig. (3g). Applying this identity to Fig. (3f), yields the final representation of MEK$_\ell$ as shown in Fig. (2e).

VI. MEASURING PERFORMANCE

This section introduces several definitions and notations used to describe the performance of protocols, and quantify their cost.

A. Quantifying noise

If ρ is a noisy $|\psi\rangle$ state then we say it has error rate ϵ where $\epsilon := \frac{1}{2}||\rho - |\psi\rangle\langle\psi|||_1$ and $||X||_1 = \mathrm{tr}[\sqrt{XX^\dagger}]$ is the Schatten 1-norm. Unless otherwise stated, we assume diagonal noise so that ρ is diagonal in the same basis as $|\psi\rangle$, and for such states one can show $\epsilon = 1 - \langle\psi|\rho|\psi\rangle$. The protocols considered here use a cocktail of input resources. We use ϵ_ℓ and ϵ_3 to denote the input error rates of the noisy $|M_\ell\rangle$ and $|M_3\rangle$ states used.

Let U be any rotation of the form $U = \exp(i\theta Y)$ and $\mathcal{U}$ be the associated ideal channel. We consider this channel with Y noise

$$\mathcal{E}(\rho) = (1-\eta)U\rho U^\dagger + \eta YU\rho U^\dagger Y^\dagger, \qquad (11)$$

and say the channel has error rate η. This can be easily to related to other noise metrics. If d is a unitarily invariant distance measure on channels then $d(\mathcal{U}, \mathcal{E}) = d(\mathbb{1}, \mathcal{U}^{-1} \circ \mathcal{E})$ where $\mathbb{1}$ is the identity channel. The composite channel $\mathcal{U}^{-1} \circ \mathcal{E}$ is a simple Pauli noise channel

$$(\mathcal{U}^{-1} \circ \mathcal{E})(\rho) = (1-\eta)\rho + \eta Y \rho Y^{\dagger}. \tag{12}$$

A widely used distance measure is derived from the diamond norm [2], so that $d_{\diamond}(\mathcal{U}, \mathcal{E}) := \frac{1}{2}||\mathcal{U} - \mathcal{E}||_{\diamond}$, and for Pauli noise of the above type it is well-known that $d_{\diamond}(\mathcal{U}, \mathcal{E}) = \eta$.

The diamond norm distance of channels is closely related to the 1-norm distance on states. If a noisy magic state with ϵ error (as measured by 1-norm distance) is used to inject a rotation, then the resulting noisy rotation will have error not exceeding ϵ (as measured by the diamond norm). Furthermore, diagonal noise on states results in diagonal noise on the rotation. Primarily, we investigate this diagonal noise, but in App. B show that generic noise is also suppressed by MEK$_\ell$.

B. Distillation cost

The output error from MEK$_\ell$ is always measured as the error on a single output qubit (ignoring correlations) and we find

$$\delta_\ell(\epsilon_3, \epsilon_\ell, \eta_{\ell-1}) \sim 8\epsilon_3^2 + \epsilon_\ell^2 + \frac{1}{4}\eta_{\ell-1} + \ldots, \tag{13}$$

$$P_{\mathrm{suc}}(\epsilon_3, \epsilon_\ell, \eta_{\ell-1}) \sim 1 - 8\epsilon_3 - 2\epsilon_\ell - \frac{1}{2}\eta_{\ell-1} + \ldots. \tag{14}$$

In App. F, we provide the full expressions as calculated by explicit simulation. For $\eta = 0$ and $\epsilon_3 = \epsilon_\ell$, these expressions are the same as those found by Meier, Eastin and Knill in their analysis of MEK.

Within the context of a single distillation round the performance is independent of the level ℓ, and is solely a function of the noise of the input states ϵ_ℓ, ϵ_3 and η. When we ask how much the input states cost, we find this can increase with ℓ. Next, we consider many distillation rounds and combine all performance metrics into a single quantity, the expected resource cost $\mathfrak{C}(M_\ell, \delta)$ to distill a $|M_\ell\rangle$ state of δ error. Lower δ can require more rounds of distillation, which drives up costs. For our protocol we use that the cost is

$$\mathfrak{C}(M_\ell, \delta_\ell) = \frac{2\mathfrak{C}(M_\ell, \epsilon_\ell) + 8\mathfrak{C}(M_3, \epsilon_3) + \mathfrak{C}(R_{\ell-1}, \eta_{\ell-1})}{2P_{\mathrm{suc}}(\epsilon_3, \epsilon_\ell, \eta_{\ell-1})}, \tag{15}$$

where δ_ℓ obeys Eqs. (13) and (F1). In our analysis, we optimise the cost over many thousands of possible combinations of input resources by brute force search. Notice that we capture the cost of the pivotal rotation as $\mathfrak{C}(R_\ell, \eta_\ell)$, which will in turn depend on the cost of magic states used to implement the rotation as

$$\mathfrak{C}(R_\ell, \eta_\ell) = \mathfrak{C}(M_\ell, \epsilon_\ell) + \frac{1}{2}\mathfrak{C}(R_{\ell-1}, \eta_{\ell-1}), \tag{16}$$

where

$$\eta_\ell = \frac{1}{2}\epsilon_\ell + \frac{1}{2}(\epsilon_\ell(1-\eta_{\ell-1}) + (1-\epsilon_\ell)\eta_{\ell-1}). \tag{17}$$

For the lowest non-Clifford levels, $\mathfrak{C}(M_3, \delta_3)$ are found by minimising over different combinations of protocols including Bravyi-Haah [5], MEK$_3$ [4] and using the 15 qubit Reed-Muller code [3]. Recall Clifford operations are considered free and perfect so that $\mathfrak{C}(M_2, 0) = 0$ and $\mathfrak{C}(R_2, 0) = 0$. Throughout we assume that raw initial non-Clifford states can be prepared with a fidelity that is independent of ℓ, and these have unit cost, so that $\mathfrak{C}(M_\ell, \epsilon_{\mathrm{raw}}) = 1$. This last assumption is warranted in light of the results of Li [41].

C. Gate-synthesis cost

In general, gate synthesis techniques have an inherent cost that we denote as $\mathfrak{T}(U, \epsilon_{\mathrm{GS}})$, where ϵ_{GS} is the precision of the synthesis. This T-count assumes perfect $|M_3\rangle$ magic states are available. Also accounting for the cost of distilled $|M_3\rangle$ states, the full cost is

$$\mathfrak{C}^{GS}(R_\ell, \eta_\ell) = \mathfrak{T}(R_\ell, \epsilon_{\mathrm{GS}}) \cdot \mathfrak{C}(M_3, \epsilon_3) \tag{18}$$

where

$$\eta_\ell \simeq \epsilon_{\mathrm{GS}} + \mathfrak{T}(R_\ell, \epsilon_{\mathrm{GS}}) \cdot \epsilon_3. \tag{19}$$

Notice that gate synthesis has an inherent error ϵ_{GS} and will also fail if one of the $\mathfrak{T}$ noisy states fail. In our analysis, we present data points for combinations of actual instances of gate synthesis using the SR protocol. The SR data is obtained using the Gridsynth package [42]. While SR is optimal for unitary synthesis, lower overheads can be using PQF, which uses ancilla, measurements and feed forward. For PQF, we use the approximation

$$\begin{aligned} \mathfrak{T}^{PQF}(U, \epsilon_{\mathrm{GS}}) \simeq & \log_2(\sqrt{2}\epsilon_{\mathrm{GS}}^{-1}) \\ & + 4\log_2(\log_2(\sqrt{2}\epsilon_{\mathrm{GS}}^{-1}) + 1.187, \end{aligned} \tag{20}$$

which is the lowest currently known overhead for gate synthesis. We show in App. B that the notion of ϵ used in the gate synthesis literature is comparable to our definition.

VII. RESULTS FOR SMALL ℓ

We postpone the technical details on noise analysis until App. C, and here present results demonstrating the benefits of MEK$_\ell$.

A. Comparison with DP protocol

There are three aspects to the comparison: the number of resources required per attempt, the success probability and how the protocols suppress errors. Regarding resources, each round of MEK$_\ell$ uses 8 fewer $|M_3\rangle$

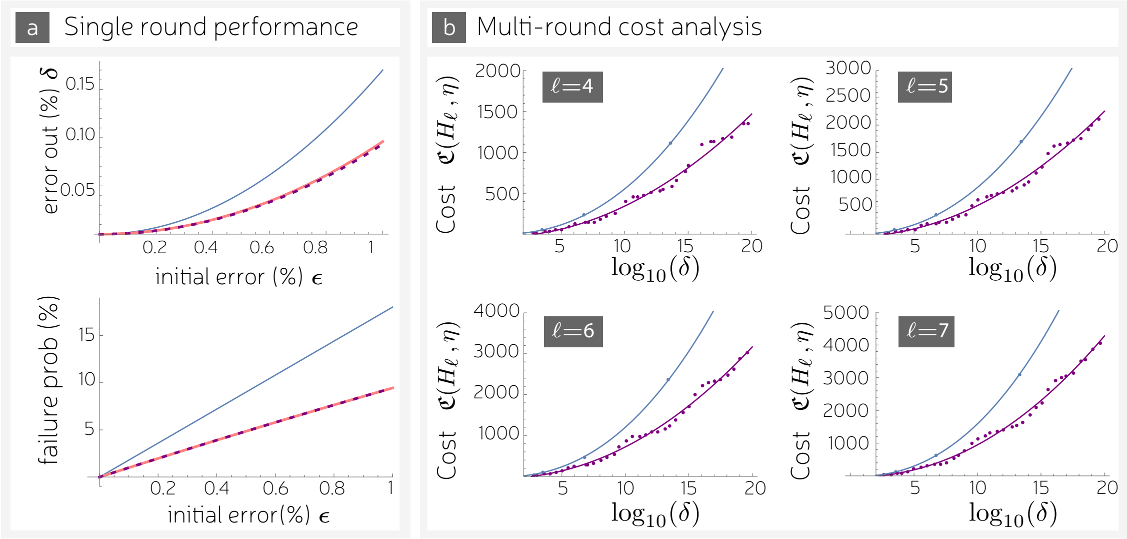

FIG. 4. Comparison of the resource cost of distillation using MEK_ℓ and DP_ℓ. (a) shows performance when $\epsilon_3 = \epsilon$, $\epsilon_\ell = \epsilon$ and $\delta = \epsilon^2$. These are benchmarked against the standard MEK_3 protocol (pink), where the curves for MEK_ℓ (purple and dotted) are barely distinguishable from MEK_3, whereas DP_ℓ (blue) performs much worse. The curves for DP_ℓ are based on leading order approximations [43]. (b) shows the full resource cost of MEK_ℓ protocol (purple) and DP_ℓ (blue) of distilling a $|M_\ell\rangle$ state of final error of δ using resources with an initial error of 1%. Data for DP_ℓ taken from Table 1 of Ref. [37]. Lines are fitted functions of the form $\mathfrak{C} = a\log(\delta)^b + c$.

states than a round of DP_ℓ. In Figs. (5a) and (5b) we fix $\epsilon_\ell = \epsilon_3 = \epsilon$ and $\eta_{\ell-1} = \epsilon^2$ and compare the performance of MEK_ℓ and DP_ℓ, both benchmarked against MEK_3. We remark that $\eta_{\ell-1}$ must be set significantly lower than other errors as the protocol can not detect noise in the pivotal rotation. In this context, MEK_ℓ is barely indistinguishable from MEK_3. This is expected as the protocols perform identically when $\eta_{\ell-1} = 0$, and since $\eta_{\ell-1} = \epsilon^2$ is very small we only observe a very slight difference between MEK_ℓ and MEK_3. In contrast, DP_ℓ performs worse despite consuming more resources.

We also consider the full cost of performing many rounds of MEK_ℓ and DP_ℓ. Because the cost of the pivotal rotation increases with ℓ, we now see a variation in performance with ℓ. Our results are shown for $\ell = 4, 5, 6, 7$ in Fig. (5c) and compare favourably against results reported by Duclos-Cianci and Poulin. Roughly, we observe a factor 1/2 reduction in cost, providing a clear cut case for using our compressed MEK_ℓ protocol rather than the original DP_ℓ proposal. Given that all component metrics (resources per round, failure probability and error out) are very favourable towards MEK_ℓ, one may have expected a more dramatic reduction in cost over DP_ℓ. One explanation is that both MEK_ℓ and DP_ℓ require 1 very high fidelity pivotal rotation, which is very costly, and this shared cost limits the extent to which MEK_ℓ can outperform DP_ℓ.

Duclos-Cianci and Poulin discuss a potential improvement to their scheme that uses larger code sizes. These larger code blocks would encode, and hence distill, $2m$ copies of $|M_\ell\rangle$ and consume $(8m + 8)$ copies of $|M_3\rangle$ and m pivotal rotations. In the large m regime, the ratio of input to output resources (the rate) becomes comparable to MEK_ℓ. However, larger code blocks involve more complex circuits and typically do not suppress noise as effectively. Furthermore, moving to large block codes will substantially increase the failure probability, which scales at least linearly with the block size. The exact performance of this large block protocol is unclear and DP presented a rough estimate based on leading order approximations. For large block codes the number of possible undetected errors of high weight can grow at a combinatorial rate and so it is unclear how accurate leading order approximations will be. In contrast, MEK_ℓ offers the improved rate without any of these drawbacks and MEK_ℓ will also outperform the large block variant of DP_ℓ. For a precise comparison, one must move beyond leading order approximations and presently no such analysis is available for the large block code variant of DP_ℓ.

B. Comparison with gate synthesis protocol for modest ℓ

Here, we present our own analysis of gate synthesis for modest $\ell = 3, \ldots, 7$, with higher levels discussed in following sections. Fig. 4 shows a comparison of MEK_ℓ against SR and PQF. For both SR and PQF, the resource cost has little dependence on ℓ. In contrast, the resource cost of MEK_ℓ increases gradually with ℓ. Fig. (4a) reports the magic state cost $\mathfrak{C}(M_\ell, \epsilon_\ell)$. However, for a fair comparison to gate synthesis we must benchmark against the cost of implementing a R_ℓ gate, which is $\mathfrak{C}(R_\ell, \eta_\ell)$ in our notation and shown in Fig. (4b).

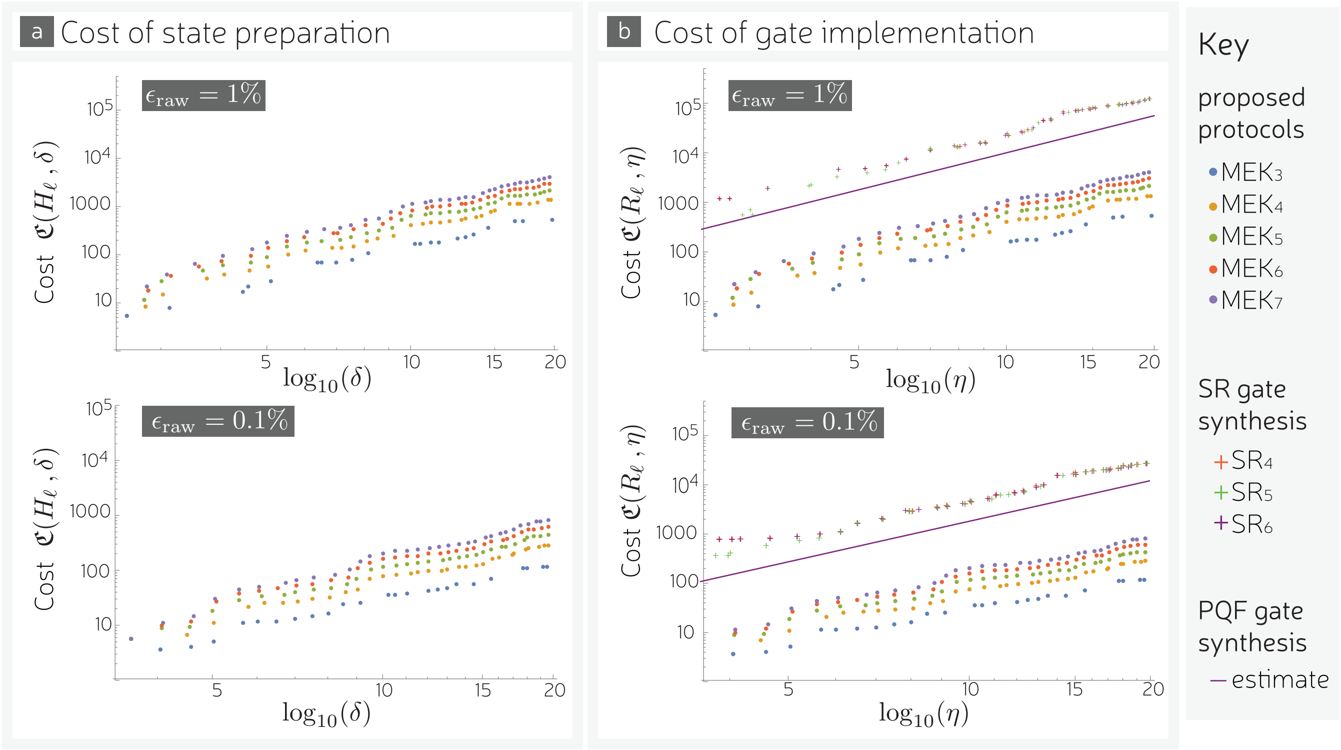

FIG. 5. The resource cost of MEK$_\ell$. (a) shows cost $\mathfrak{C}(|M_\ell\rangle, \delta)$ of preparing a magic state $|M_\ell\rangle$ at error rate δ. (b) shows cost $\mathfrak{C}(R_\ell, \eta)$ of performing a non-Clifford R_ℓ at error rate η. For comparison, (b) also shows the cost of gate-synthesis using the protocol of Ref. [11]. Both (a) and (b) use initial error rates of $\epsilon = 0.01$ and $\epsilon = 0.001$ for axial rotations of angle $\theta_\ell = \pi/2^\ell$ with $\ell = 3, 4, \dots 7$. Smaller rotations and other gate-synthesis protocols are considered later.

We consider the cases of raw error rates of $\epsilon_{\rm raw} = 0.01$ and $\epsilon_{\rm raw} = 0.001$. The case of higher raw noise is an important benchmark as it has been widely studied [4, 5, 37], and here we see improvements over gate synthesis. For instance, at $\delta = 10^{-15}$ we find PQF is ~ 22.5 times more costly than MEK$_6$. However, the lower raw noise regime $\epsilon_{\rm raw} = 0.001$ is in many ways more interesting. In this regime, the advantage of using MEK$_\ell$ is further improved by a slight margin. For instance, at target error rate $\delta = 10^{-15}$ we find PQF is ~ 24 times more costly than MEK$_6$, which is a larger factor than in the high noise regime. This widening gap between gate synthesis and MEK$_\ell$ is seen for all ℓ and δ. This increase in gap is intuitive because the distillation cost drops with $\epsilon_{\rm raw}$, but the T-count of gate synthesis is independent of $\epsilon_{\rm raw}$.

While the high noise regime is widely studied, we next argue that the lower noise regime is also more realistic. Underneath magic state factories is a layer of quantum error correction. The highest known thresholds for error correction are $\sim 1\%$ for the toric code [17, 20–22] and $\sim 3\%$ for Knill's model of postselected quantum computation [44]. To prevent astronomical overheads, physical gates must be comfortably below the threshold, and so we assume all physical gates have infidelities well below 1%. Because preparing raw magic states will involve several physical gates, it has often been assumed that $\epsilon_{\rm raw}$ will be higher than the physical gate error rate, maybe even an order of magnitude higher. However, Li [41] has shown that we can probabilistically prepare raw magic states at infidelities of about half the physical control-X gate error, assuming control-X failure is the dominant noise mechanism. We remind the reader that we assume logical level control-X gates are ideal, but beneath the hood of error correction we have very noisy physical control-X gates. All this indicates that $\epsilon_{\rm raw} = 0.001$ is a feasible regime, more plausible than $\epsilon_{\rm raw} = 0.01$. As such, following subsections focus on the low-noise regime.

VIII. MAGIC STATE DILUTION

As we ascend the Clifford hierarchy, we expect the cost of our protocol to increase. However, for sufficiently small angles the associated magic states become very close to stabiliser states, and we should be able to exploit this to reduce resource costs. In the analysis of Duclos-Cianci and Poulin, they replaced noisy $|M_\ell\rangle$ states with a stabiliser state whenever ℓ exceeded 8. This enabled resource costs to eventually drop with ℓ. The $|+\rangle$ state can be considered a noisy $|M_\ell\rangle$ state, though with coherent (non-diagonal) noise. We calculate the noise of this

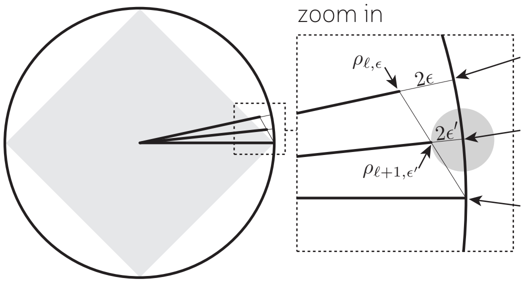

FIG. 6. A geometric representation of the dilution protocol in a cross-section of the Bloch sphere. Note that two points separated by a geometric distance of d in the Bloch sphere, correspond to operators with $d/2$ distance in 1-norm. The grey diamond shows the set of stabiliser states. Given a $\rho_{\ell,\epsilon}$ state and a pure $|+\rangle$ state we can prepare mixtures on the line between these points. This line intersects the line for states of the form $\rho_{\ell,\epsilon'}$. We see that for sufficiently large ϵ, the resulting ϵ' can be reduced $\epsilon' \leq \epsilon$. Furthermore, we show a $2\epsilon'$ radius ball about the point $|M_{\ell+1}\rangle$ to highlight that both $\rho_{\ell,\epsilon}$ state and $|+\rangle$ are further than $2\epsilon'$ away, and so these states would provide a worse approximation of $|M_{\ell+1}\rangle$ than their mixture.

stabiliser state with respect to $|M_\ell\rangle$ and find

$$\begin{aligned}\frac{1}{2}||(|+\rangle\langle+| - |M_\ell\rangle\langle M_\ell|)||_1 &= |\sin(\theta_\ell)| \qquad (21)\\ &\sim \theta_\ell = \frac{\pi}{2^\ell},\end{aligned}$$

where the last line gives the small angle approximation. Therefore, the resource becomes free whenever $\pi/(2^\ell)$ is smaller than the target error rate.

We present an alternative solution that keeps us within the framework of diagonal noise and ensures rapid decrease of costs whenever $\theta_\ell^2/\sqrt{2}$ is smaller than the required error rate. This ℓ cutoff is quadratically smaller than that of Duclos-Cianci and Poulin. Let ρ and ρ' be two resource states with costs $\mathfrak{C}$ and $\mathfrak{C}'$. If we choose to generate ρ with probability λ and ρ' with probability $1-\lambda$, then we have the random mixture $\lambda\rho + (1-\lambda)\rho'$. Furthermore, the expected cost is $\lambda\mathfrak{C} + (1-\lambda)\mathfrak{C}'$. We consider mixing a noisy $|M_\ell\rangle$ state with a $|+\rangle$ to obtain a good approximation of a noisy $|M_{\ell+1}\rangle$ state. We say the state $|M_\ell\rangle$ has been diluted, since this allows a source of $|M_\ell\rangle$ states to provide, on average, a greater number of noisy $|M_{\ell+1}\rangle$ states. We will see that while dilution may increase noise, there are practically relevant regimes where dilution reduces noise.

We use $\rho_{\ell,\epsilon}$ for a $|M_\ell\rangle$ state with ϵ diagonal noise, so that

$$\rho_{\ell,\epsilon} = (1-\epsilon)|M_\ell\rangle\langle M_\ell| + \epsilon|\bar{M}_\ell\rangle\langle\bar{M}_\ell|. \qquad (22)$$

Preparing $\rho_{\ell,\epsilon}$ costs $\mathfrak{C}(M_\ell,\epsilon)$ resources. The principle result of this section the relation

$$\rho_{\ell+1,\epsilon'} = \lambda\rho_{\ell,\epsilon} + (1-\lambda)|+\rangle\langle+|, \qquad (23)$$

where

$$\lambda = \frac{1}{2(1-\epsilon)}, \qquad (24)$$

$$\epsilon' = \frac{1}{2}\left(1 - (1-2\epsilon)\left[\frac{\cos(\theta_\ell)}{(1-\epsilon)}\right]\right). \qquad (25)$$

The dilution provides $\rho_{\ell+1,\epsilon'}$ at a cost $\lambda\mathfrak{C}(M_\ell,\epsilon)$, which is half the cost of the $\rho_{\ell,\epsilon'}$ state when ϵ is small. Remarkably, dilution can even decrease the error rate, provided ℓ is large enough. Specifically, if ℓ is large enough that the fraction in square bracket exceeds 1, then we have $\epsilon' \leq \epsilon$. A simpler sufficient condition is $\theta_\ell \leq \sqrt{2\epsilon}$, which can be used to predict a transition in the performance of dilution at

$$\ell_c = \log_2\left(\frac{\pi}{\sqrt{2\epsilon}}\right). \qquad (26)$$

In this regime the cost is halved, and iterating this process decreases costs exponentially with ℓ. However, even when ℓ is does not satisfy the above condition, an increase in error rate may still make dilution more efficient than distillation. We give a proof of Eq. (23) in App. E with the underlying geometric intuition presented in Fig. (6).

IX. RESULTS FOR HIGHER ℓ

Here we discuss the performance of MEK$_\ell$ combined with dilution at performing small angle rotations, beyond $\ell = 7$. Our results are generated iteratively. After having compiled a list of achievable costs $\mathfrak{C}(M_\ell,\epsilon)$ and $\mathfrak{C}(R_\ell,\eta)$ for different values of ϵ, we next built lists for $\ell{+}1$. The first step is to build a list of $\mathfrak{C}(M_{\ell+1},\epsilon')$ derived from $\mathfrak{C}(M_\ell,\epsilon)$ using the dilution protocol in the previous section. Next, we considered different combinations of input states into the MEK$_\ell$, allowing for using diluted states without any further distillation or inputting diluted states into MEK$_\ell$. The results are presented in Fig. (7) as a function of ℓ, showing how resource costs scale with decreasing angle.

For the target error rates $10^{-10}, 10^{-15}, 10^{-20}$ we can see three clear regions, with the middle region absent for the 10^{-5} plot. First, the resource cost increases roughly linearly with ℓ. Next, there is a transition where the gradient becomes gentler. Lastly, at some cutoff the cost starts to fall exponentially with ℓ. In this last regime, we rely solely on diluting magic states. This exponential cliff was predicted by the analysis in the previous section, and is labeled in the plots by ℓ_c. The behaviour of the middle region is also due to dilution. Here we typically find that one round of MEK$_\ell$ is used, with the input noisy $|M_\ell\rangle$ states produced by dilution, as opposed to using two or more rounds of MEK$_\ell$ on a raw resource of error rate ϵ_{raw}. Since dilution is less effective at low ℓ, in this regime we are completely reliant on MEK$_\ell$ and here observe the most rapid increase in costs.

Our results are also presented against the cost of two gate-synthesis methods SR and PQF. We see that for small ℓ, and large $\ell > \ell_c$ there is a significant gain over both gate-synthesis methods by over an order of magnitude. In the intermediate regime, our protocol still outperforms gate-synthesis but approaches a similar order of magnitude.

X. CONCLUSIONS

We have proposed a protocol for distilling magic states providing small angle rotations at a lower cost than the previous proposal of Duclos-Cianci and Poulin. The cost, as measured by number of raw magic states used, is less than best current gate synthesis techniques. For modest size rotation of angle $\pi/(2^6)$, we saw our protocol operates ~ 24 times better than gate synthesis. Assessing this improvement, a factor 12 can be attributed to the innovations of Duclos-Cianci and Poulin, with our compression of the protocol providing the additional factor 2. To perform smaller angle rotations, the resource cost gradually increases until the resources become close to stabiliser states. To maintain an advantage over gate-synthesis this phenomenon must be exploited. We proposed a magic state dilution protocol, which converts one magic state into a larger number suitable for a smaller angle rotation. This exponentially suppresses resource costs for angles $\theta_\ell \leq \sqrt{2\epsilon}$. In this regime, resource costs are many orders of magnitude lower than gate-synthesis.

Throughout, we have quantified cost by raw magic states consumed. This is the standard metric in every paper that has introduced a new protocol for magic state distillation. Other quantities of interest are the total number of qubits used (space cost) and depth of circuits used (time costs). These full resource costs include stabiliser and Clifford costs, but are highly sensitive to the specific architecture used. The fully costed performance of Bravyi-Haah and Reed-Muller magic state distillation have been considered for surface code architectures, both in a braiding picture [7] and using transversal logical gates [18]. Such full resource assessments are significant undertakings that followed the initial proposals, and so lie beyond our present scope. However, all such analyses to date have found that protocols with lower T-counts also have lower full resource costs. Although, in Ref. [7] they noted that significant improvements in protocol T counts can lead to more modest improvements of full resource costs. This occurs because the surface code incurs a polylogarthmic overhead in ϵ^{-1} that grows more rapidly than overheads due to magic state distillation or gate-synthesis. That is, in a fully costed analysis the surface code overhead dwarfs all other overheads. Though this peculiarity could disappear if future developments provided a more efficient and practical alternative over the surface code.

Here we have considered only single-qubit gate-synthesis, but in parallel work there has been progress on unifying magic state distillation with gate-synthesis for a class of multi-qubit circuits [45, 46]. We close by remarking that the fields of magic state distillation and gate synthesis have rapidly evolved in the last several years, giving good reason to be hopeful that advancement in these areas will continue to bring quantum computation ever closer to reality.

XI. ACKNOWLEDGEMENTS

This work was supported by the EPSRC (grant EP/M024261/1) and collaboration between the authors was facilitated by the NQIT quantum hub. We thank Bryan Eastin and Adam Meier for assistance with reproducing the simulation of MEK. We thank Guillaume Duclos-Cianci and Mark Howard for comments on the manuscript. Gate synthesis calculations used the Gridsynth package, and we thank Peter Selinger and Neil J. Ross for writing this software and making it available. We thank the developers of the Qcircuit package used in generation of figures. All our analysis is available in the Supplementary material as the Mathematica source files and also in pdf format.

Appendix A: Arbitrary angle rotations

We focus on small angle rotations, but some of our techniques also apply to larger rotations. In particular, not only is R_ℓ in the ℓ^{th} level of the Clifford hierarchy, but R_ℓ^m belongs to the same level for any integer m. For the compressed MEK$_\ell$ protocols described we can replace $R_\ell \rightarrow R_\ell^m$ throughout and they perform identically, and so R_ℓ^m has the same resource cost as R_ℓ.

If we require some rotation with an angle ϕ that is not a rational fraction of π, then it will not fall within any level of the Clifford hierarchy. Rather, we may resort to inexact synthesis and find a $n\theta_\ell$ that is sufficiently close to ϕ. The larger we choose ℓ, the better the approximation, but the higher the resource cost. Specifically, for all ϕ and ℓ we can always find an n such that $|\phi - n\theta_\ell| \leq \pi/2^\ell$, which improves exponentially with ℓ. To prove this, we observe that for fixed ℓ the set of reachable angles $\{n\theta_\ell | 1 \leq n \leq 2^{\ell+1}\}$ are equally spaced on the interval $[0, 2\pi]$. Therefore, the reachable angles have gaps between them of distance $\pi/2^\ell$. The angle ϕ furthest from any angle in the reachable set will sit halfway between two reachable angles, and so this worst case ϕ lies $\pi/2^{\ell+1}$ away from the nearest reachable angle. For instance, to ensure $|\phi - n\theta_\ell| \leq 10^{-10}$ we can use $\ell = 35$. It is straightforward to verify that high precision in angle ensures closeness (upto a constant factor) in diamond norm distance between the corresponding channels.

However, for large ℓ using MEK$_\ell$ becomes increasingly costly with ℓ. For large ℓ, this is overcome by using magic state dilution. This protocol relies on $|M_\ell\rangle$ being close to the $|+\rangle$ stabiliser state to reduce overheads, and so the above comments about replacing $R_\ell \rightarrow R_\ell^m$ do not apply

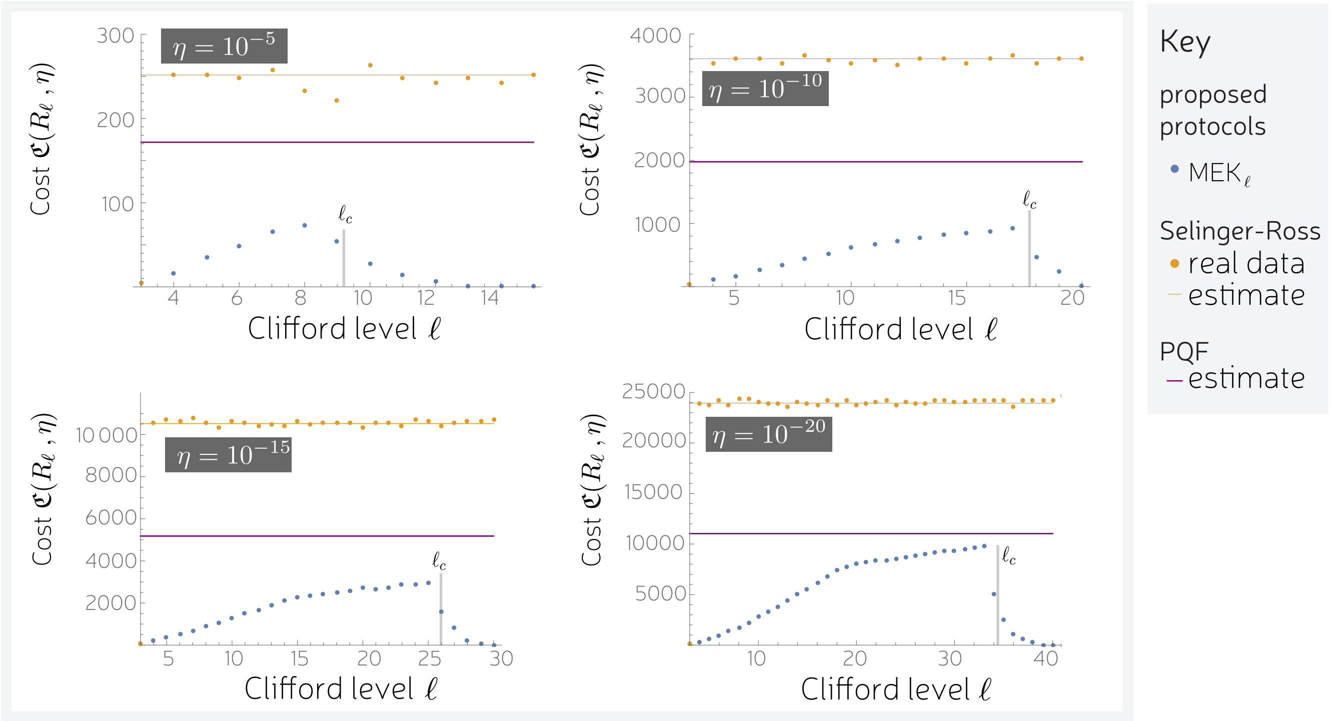

FIG. 7. The resource cost R_ℓ rotations using MEK$_\ell$ distillation combined with magic state dilution with $\epsilon_{\mathrm{raw}} = 0.1\%$. Each plot is for a different final error rate η, and is against the Clifford level ℓ beginning with the first nonClifford level $\ell = 3$. We compare perform against gate-synthesis methods, which perform mostly independently of ℓ. For the Selinger-Ross (SR) gate synthesis protocol [11], we show exact data extracted from the gridsynth application and also the analytically proven typical performance. We also show the typical performance for probabilistic quantum circuits with fallback (PQF) [39]. All results here assume $\epsilon_{\mathrm{raw}} = 0.001$.

to dilution. Therefore, the resource costs presented in Sec. IX only apply when ϕ is close to some θ_ℓ.

Appendix B: Precision in gate-synthesis

Across the gate-synthesis literature, different notions of precision are used. Here we discuss how these relate to the diamond norm measure of noise. We assume throughout that all unitaries are in $SU(2)$ and so have determinant 1.

In Def 7.1 of Ref. [11], SR state that they quantify precision using the spectral norm $\epsilon_{SR} := ||U - V||_\infty$. Let $W = U^\dagger V$ and have eigenvalues $\exp(i\varphi)$ and $\exp(-i\varphi)$ with $\varphi \geq 0$, so that φ is small whenever U and V are close. It follows that $||U - V||_\infty = |1 - \exp(i\varphi)| = 2|\sin(\frac{\varphi}{2})| \sim \varphi$.

If $\mathcal{U}$ and $\mathcal{V}$ are channels associated with U and V, then Refs. [47, 48] show the inequality

$$\epsilon_{GS} := \frac{1}{2}||\mathcal{U} - \mathcal{V}||_\diamond \leq ||U - V||_\infty \sim \varphi. \tag{B1}$$

We now consider a lower bound. For any ρ with $||\rho||_1 = 1$ we have by definition that

$$\frac{1}{2}||\mathcal{U} - \mathcal{V}||_\diamond \geq \frac{1}{2}||(\mathcal{U} \otimes \mathbb{1})\rho - (\mathcal{V} \otimes \mathbb{1})\rho||_1. \tag{B2}$$

Setting $\rho = |+\rangle\langle+| \otimes |+\rangle\langle+|$, we find that

$$\begin{aligned} \frac{1}{2}||\mathcal{U} - \mathcal{V}||_\diamond &\geq \frac{1}{2}|1 - \exp(i2\varphi)| \\ &= |\sin(\varphi)| \sim \varphi \end{aligned} \tag{B3}$$

Therefore, in the small φ limit we have $\epsilon_{GS} = \epsilon_{SR}$.

In the work on PQF, they measure precision [49] as follows

$$\epsilon_{PQF} := \sqrt{1 - \frac{1}{2}\mathrm{tr}(U^\dagger V)}. \tag{B4}$$

This evaluates to $\epsilon_{PQF} = \sqrt{2}|\sin(\varphi/2)| \sim \varphi/\sqrt{2}$. Therefore, this quantity is $\sqrt{2}$ smaller than our diamond norm measure, so that $\epsilon_{PQF} = \epsilon_{GS}/\sqrt{2}$. This adjustment is shown in Eq. (B), which carries a $\sqrt{2}$ correction relative to the result of Ref. [39]. Although we have carefully accounted for differences in precision measures, this small constant factor adjustment in error rate makes a negligible difference to resource overheads.

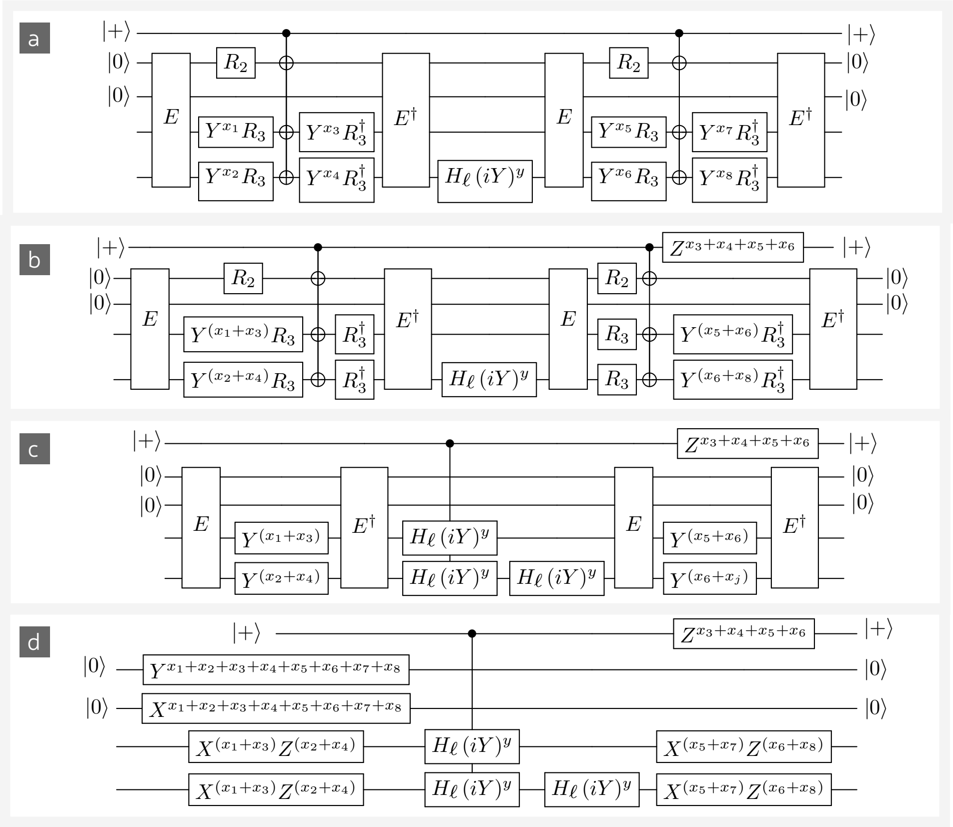
a
b
c
d
$|+\rangle$
$|0\rangle$
E
$E^\dagger$
R_2
$Y^{x_1}R_3$
$Y^{x_2}R_3$
$Y^{x_3}R_3^\dagger$
$Y^{x_4}R_3^\dagger$
$H_\ell\,(iY)^y$
$Y^{x_5}R_3$
$Y^{x_6}R_3$
$Y^{x_7}R_3^\dagger$
$Y^{x_8}R_3^\dagger$
$Z^{x_3+x_4+x_5+x_6}$
$Y^{(x_1+x_3)}R_3$
$Y^{(x_2+x_4)}R_3$
$R_3^\dagger$
R_3
$Y^{(x_5+x_6)}R_3^\dagger$
$Y^{(x_6+x_8)}R_3^\dagger$
$Y^{(x_1+x_3)}$
$Y^{(x_2+x_4)}$
$Y^{(x_5+x_6)}$
$Y^{(x_6+x_j)}$
$Y^{x_1+x_2+x_3+x_4+x_5+x_6+x_7+x_8}$
$X^{x_1+x_2+x_3+x_4+x_5+x_6+x_7+x_8}$
$X^{(x_1+x_3)}Z^{(x_2+x_4)}$
$X^{(x_5+x_7)}Z^{(x_6+x_8)}$

FIG. 8. Propagation noise terms around distillation circuit.

Appendix C: Noise analysis

This section shows how the circuit operates when the non-Clifford gates (R_3 and $R_{\ell-1}$) are noisy due to imperfect magic states. When a R_ℓ gate fails due to noise, it is followed by an additional Y rotation. There are 8 locations that errors can occur on R_3 gates and we define a binary vector $x = (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8)$ that records whether a Y error occurs at a particular R_3 gate. We use the binary variable $y = 0, 1$ to track an error in the pivotal rotation $R_{\ell-1}$. When the $R_{\ell-1}$ gate fails, we have a rotation $H_\ell(iY)$ rather than H_ℓ. Notice the complex phase i, which makes no physical difference to the unitary. However, aspects of the proof rely on H_ℓ being Hermitian, and the additional phase keeps it Hermitian. The resulting random circuit is shown in Fig. (8a).

We pull some Y noise operators through the control phase gates, which can cause Z noise on the control as shown in Fig. (8b). The central portion of the circuit now contains no noise terms (except y) and so we can replace it with a logical control ($H_\ell \otimes H_\ell$) or its equivalent when $y = 1$. This yields Fig. (8c).

Next, we pull the Y noise backwards through the encoder E. We already know how E acts on Pauli operators from which we conclude that the inverse action satisfies

$$E^\dagger : Y_3 \rightarrow Y_1 X_2 Z_3 Z_4, \tag{C1}$$

$$E^\dagger : Y_4 \rightarrow Y_1 X_2 X_3 X_4. \tag{C2}$$

Similarly random unitaries $Y_3^\alpha Y_4^\beta$ map as

$$E^\dagger : Y_3^\alpha Y_4^\beta \rightarrow Y_1^{\alpha+\beta} X_2^{\alpha+\beta} (Z_3^\alpha X_3^\beta)(Z_4^\alpha X_4^\beta) \tag{C3}$$

Applying this rule to our circuit yields Fig. (8d).

From Fig. (8d), we see that to obtain the correct measurement outcome on qubit 1 or 2 requires that

$$|x| := x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 = 0 \pmod 2, \tag{C4}$$

which detects any single error on the R_3 gates. This shows that the error correction properties of our original circuit have survived the compression.

From this circuit we can get some intuition for the leading order error terms presented in Eqs. 13. First consider the failure probability, $P_{\text{fail}} = 1 - P_{\text{suc}}$. If $|x| := x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 = 1 \pmod 2$, then we detect an error and have a failure. The leading order contribution is single error processes, and there are 8 such errors, so this contributes $8\epsilon_3$ to P_{fail}. Otherwise, if $x = (0, 0, \ldots 0)$ and the pivotal rotation also works, then we perform a perfect projection onto the $+H_\ell \otimes H_\ell$

subspace. This projection detects an errors, and the protocol declares failure if one of the two noisy $|M_\ell\rangle$ states is faulty, which occurs with probability $2\epsilon_\ell$. Lastly, we consider if everything works except the pivotal rotation. The pivotal rotation fails with probability $\eta_{\ell-1}$, but not all this contributes to the failure probability. When the pivotal rotation fails, and the rest of the circuit works correctly, it results in the operation shown in Fig. (2c), but with the replacement $H_\ell \rightarrow iH_\ell Y$. Algebraically, this operation is

$$\begin{aligned} V &= \frac{1}{2}\left[\mathbb{1} + (iH_\ell Y) \otimes (iH_\ell Y)\right]\left[(iH_\ell Y) \otimes \mathbb{1})\right] \qquad \text{(C5)} \\ &= \frac{-i}{2}\left[(YH_\ell) \otimes \mathbb{1} + \mathbb{1} \otimes (YH_\ell)\right] \end{aligned}$$

Applying this channel to $|M_\ell\rangle|M_\ell\rangle$, and noting that $Y|M_\ell\rangle = |\bar{M}_\ell\rangle$, we have

$$|\psi\rangle = V|M_\ell\rangle|M_\ell\rangle = \frac{-i}{2}(|\bar{M}_\ell\rangle|M_\ell\rangle + |M_\ell\rangle|\bar{M}_\ell\rangle) \qquad \text{(C6)}$$

Therefore, the probability of an even parity measurement, combined with pivotal rotation failure, is $\eta_{\ell-1}\langle\psi|\psi\rangle$. Recalling, $\langle M_\ell|\bar{M}_\ell\rangle = 0$, we find $\langle\psi|\psi\rangle = 1/2$. Therefore, the failure probability gains a contribution of $\frac{1}{2}\eta_{\ell-1}$. This covers all the leading order processes that contribute to a detected failure.

Next we consider leading order processes that go undetected, but output an erroneous state $|\bar{M}_\ell\rangle$. The protocol outputs a two qubit state, and we trace out the second qubit and evaluate the error rate on the first qubit. Switching the first and second output qubit will yield the same result. First we consider pairs of errors in the R_3 gates. There are 28 such pairs, all satisfying $|x| = 0 \pmod 2$. However, the parity measurement must yield the correct outcome also. Since we are only considering leading order errors, we can assume the noisy $|M_\ell\rangle$ states are error free. Therefore, to obtain the $|+\rangle$ on the control ancilla, there cannot be a Z flip on the control, and so $x_3 + x_4 + x_5 + x_6 = 0 \pmod 2$. This cuts the number of undetected error pairs down to 14. Not all undetected error pairs lead to a logical error on the first qubit. For $x_1 = x_2 = 1$ and $x_7 = x_8 = 1$ we see that there is a logical $Y \otimes Y$ error, and so both these processes contribute. For the other 12 undetected errors, the parity projection becomes deformed so that it projects onto a state that on average has 1/2 overlap with $|M_\ell\rangle$. Let us expand on this notion of a deformed parity projection by considering the case $x_1 = x_5 = 1$, with other combinations proving similar. This causes an $X \otimes X$ rotation both before and after the parity projection, so that instead of projecting onto $H_\ell \otimes H_\ell$ we project onto $H_\ell^\perp \otimes H_\ell^\perp$ where $H_\ell^\perp := XH_\ell X$, which we call a deformed projection. How much this particular process contributes varies with θ_ℓ, but a lengthily evaluation over all such error pairs shows the average contribution is 1/2. This totals the error contribution from failed R_3 gates to $(2 + \frac{12}{2})\epsilon_3^2 = 8\epsilon_3^2$. If the R_3 gates do not fail, but instead we have a perfect parity projection, then a logical error occurs if the projection is applied to $|\bar{M}_\ell\rangle|\bar{M}_\ell\rangle$, which occurs with probability ϵ_ℓ^2. Lastly, we contemplate when the pivotal rotation cairres an error. We see from Eqn. (C6) that $|\langle\bar{M}_\ell, M_\ell|\psi\rangle| = 1/4$ and so the this leads to a logical error with probability $\frac{1}{4}\delta_{\ell-1}$. We remark again that the leading order contribute here is linear, and not quadratically suppressed, and so the pivotal rotation must be high fidelity to enable distillation.

Appendix D: Generic noise

Throughout the main text, we assumed diagonal noise as described in Sec. VI A. Here we show that MEK$_\ell$ is robust against generic noise, making use of the 1-norm and diamond norm to measure error rates. Specifically, our proof will demonstrate that MEK$_\ell$ noise is quadratically reduced, converging towards zero. Our proof will provide an upper bound on the output error rate, rather than the exact expression.

We first consider MEK$_\ell$ as a quantum channel $\mathcal{E}$ acting on two noisy $\rho_{\ell,\epsilon_\ell}$ states, such that

$$\epsilon_\ell := \frac{1}{2}||\rho_{\ell,\epsilon_\ell} - \rho_\ell||_1. \qquad \text{(D1)}$$

where here $\rho_{\ell,\epsilon_\ell}$ may not be diagonal in the $\{|M_\ell\rangle, |\bar{M}_\ell\rangle\}$ basis. For brevity, we use $\rho_\ell := \rho_{\ell,0} = |M_\ell\rangle\langle M_\ell|$. Below, we also use $\Delta_\ell := \rho_{\ell,\epsilon_\ell} - \rho_\ell$ where $||\Delta_\ell||_1 = 2\epsilon_\ell$.

The protocol uses 8 ρ_{3,ϵ_3} states, which by Clifford twirling arguments can be forced to have purely diagonal noise. We use $x = \{x_1, \ldots x_8\}$ to label define Pauli errors from the ρ_{3,ϵ_3} states, so that

$$\rho_{3,\epsilon_3}^{\otimes 8} = \sum_{x \in \mathbb{Z}_2^8} p_x Y[x] \rho_3^{\otimes 8} Y[x], \qquad \text{(D2)}$$

where $Y[x] := \otimes_j Y_j^{x_j}$ and

$$p_x := \epsilon_3^{|x|}(1 - \epsilon_3)^{8-|x|}, \qquad \text{(D3)}$$

with $|x| := \sum_j x_j$. For a given error configuration x, the protocol implements $\mathcal{E}_x$. We decompose this as $\mathcal{E}_x = \mathcal{E}_x'' \circ \mathcal{P}_{\ell-1} \circ \mathcal{E}_x'$, where $\mathcal{P}_{\ell-1}$ is the noisy pivotal rotation, $\mathcal{E}_x'$ is the circuit before the pivotal rotation and $\mathcal{E}_x''$ is the circuit afterwards. Throughout, we use the symbol $\circ$ to denote composition of channels. Lastly we trace out one of the qubits, using the partial trace map tr_a, to give the single qubit output. The trace is over either qubit $a = 1$ or $a = 2$, and our analysis will be independent of this choice.

The complete channel is

$$\mathcal{E} = \sum_x p_x \mathrm{tr}_a \circ \mathcal{E}_x'' \circ \mathcal{P}_{\ell-1} \circ \mathcal{E}_x'. \qquad \text{(D4)}$$

Next, we use that since $\frac{1}{2}||\mathcal{P}_{\ell-1} - \mathcal{U}_{\ell-1}||_\diamond = \eta_{\ell-1}$, we know $\mathcal{P}_{\ell-1} = \mathcal{U}_{\ell-1} + \mathcal{D}$, where $||\mathcal{D}||_\diamond = 2\eta_{\ell-1}$. Note that

$\mathcal{D}$ is not necessarily a positive map. This entails

$$\mathcal{E} = \left(\sum_x p_x \mathrm{tr}_a \circ \mathcal{E}''_x \circ \mathcal{U}_{\ell-1} \circ \mathcal{E}'_x \right) + \left(\sum_x p_x \mathrm{tr}_a \circ \mathcal{E}''_x \circ \mathcal{D} \circ \mathcal{E}'_x \right). \tag{D5}$$

Recall from Eq. (C4) that the terms $\mathcal{E}''_x \circ \mathcal{U}_{\ell-1} \circ \mathcal{E}'_x$ vanish whenever x has an odd number of 1s. For the nonvanishing x values, we split the first bracket into two components, representing the $x = (0, 0, \dots, 0)$ term and then the nonzero $|x| = 2, 4, 6, 8$ terms

$$\begin{aligned} \mathcal{E} =& p_0 \mathrm{tr}_a \circ \mathcal{E}''_0 \circ \mathcal{U}_{\ell-1} \circ \mathcal{E}'_0 \\ &+ \left(\sum_{|x|=2,4,6,8} p_x \mathrm{tr}_a \circ \mathcal{E}''_x \circ \mathcal{U}_{\ell-1} \circ \mathcal{E}'_x \right) \\ &+ \eta_{\ell-1} \left(\sum_x p_x \mathrm{tr}_a \circ \mathcal{E}''_x \circ \mathcal{D} \circ \mathcal{E}'_x \right). \end{aligned} \tag{D6}$$

We next introduce some new notation to simplify this expression to

$$\mathcal{E} = \mathcal{A} + \mathcal{B} + \mathcal{C}. \tag{D7}$$

where our new maps are

$$\mathcal{A} = \mathrm{tr}_a \circ p_0 \mathcal{E}''_0 \circ \mathcal{U}_{\ell-1} \circ \mathcal{E}'_0, \tag{D8}$$

$$\mathcal{B} = \mathrm{tr}_a \circ \left(\sum_{|x|=2,4,6,8} p_x \mathcal{E}''_x \circ \mathcal{U}_{\ell-1} \circ \mathcal{E}'_x \right), \tag{D9}$$

$$\mathcal{C} = \mathrm{tr}_a \circ \sum_x p_x \mathcal{E}''_x \circ \mathcal{D} \circ \mathcal{E}'_x. \tag{D10}$$

One can easily verify the channels satisfy

$$||\mathcal{A}||_\diamond \leq p_g := p_0 = (1-\epsilon_3)^8, \tag{D11}$$

$$||\mathcal{B}||_\diamond \leq p_b := \sum_{x=2,4,6,8} \binom{8}{x} \epsilon_3^{|x|} (1-\epsilon_3)^{8-|x|}, \tag{D12}$$

$$||\mathcal{C}||_\diamond \leq 2\eta_{\ell-1}. \tag{D13}$$

Later we also make use of (assuming $\epsilon_3 \leq 0.01$) the following

$$1 - 8\epsilon_3 \leq p_g \leq 1 \tag{D14}$$

$$28\epsilon_3^2 - 168\epsilon_3^3 \leq p_b \leq 28\epsilon_3^2 \tag{D15}$$

We next consider the effect of these channels individually before composing the results together.

The $\mathcal{A}$ map represents a perfect parity projection with no errors, so that

$$\mathcal{A}(\rho_\ell \otimes \rho_\ell) = p_g \mathrm{tr}_a[\rho_\ell \otimes \rho_\ell] = p_g \rho_\ell. \tag{D16}$$

We consider its action on a noisy state $\rho_{\ell,\epsilon_\ell} = \rho_\ell + \Delta_\ell$. In the $|M_\ell\rangle$ basis we have

$$\Delta_\ell = \epsilon_\ell \begin{pmatrix} -a & b \\ b^* & a \end{pmatrix}. \tag{D17}$$

From $||\Delta_\ell|| = 2\epsilon_\ell$, we deduce $0 \leq |a|^2 + |b|^2 \leq 1$. Performing the parity projection on single error terms yields

$$\mathcal{A}(\rho_\ell \otimes \Delta_\ell) = -p_g a \epsilon_\ell \mathrm{tr}_a[(\rho_\ell \otimes \rho_\ell)] = -p_g \epsilon_\ell a \rho_\ell \tag{D18}$$

$$\mathcal{A}(\Delta_\ell \otimes \rho_\ell) = -p_g a \epsilon_\ell \mathrm{tr}_a[-(\rho_\ell \otimes \rho_\ell)] = -p_g a \epsilon_\ell \rho_\ell \tag{D19}$$

Lastly for the double error term we have

$$\mathcal{A}(\Delta_\ell \otimes \Delta_\ell) = p_g a^2 \epsilon_\ell^2 [\rho_\ell + (Y \rho_\ell Y)] \tag{D20}$$

Combining Eq. (D16), Eq. (D18) and Eq. (D20), we have that

$$\mathcal{A}(\rho_\ell \otimes \rho_\ell) = p_g[(1 - a\epsilon_\ell)^2 \rho_\ell + (a\epsilon_\ell)^2 (Y \rho_\ell Y)], \tag{D21}$$

where $-1 \leq a \leq 1$.

The $\mathcal{B}$ map represents when the pivotal rotation works perfectly, but at least two ρ_{ℓ,ϵ_3} states cause an error. By elementary norm properties we have

$$||\mathcal{B}(\rho_{\ell,\epsilon_\ell} \otimes \rho_{\ell,\epsilon_\ell})||_1 \leq ||\mathcal{B}||_\diamond \cdot ||\rho_{\ell,\epsilon_\ell} \otimes \rho_{\ell,\epsilon_\ell}||_1 \leq p_b. \tag{D22}$$

We define $\rho_{\mathcal{B}} := \mathcal{B}(\rho_\ell \otimes \rho_\ell)$, and since $\mathcal{B}$ is a completely positive channel we have that $\mathrm{tr}[\rho_{\mathcal{B}}] = ||\rho_{\mathcal{B}}||_1$.

Lastly, the $\mathcal{C}$ map represents a failed pivotal rotation. We define $\rho_{\mathcal{C}} := \mathcal{C}(\rho_{\ell,\epsilon_\ell} \otimes \rho_{\ell,\epsilon_\ell})$ and use norm properties to deduce $||\rho_{\mathcal{C}}||_1 \leq 2\eta_{\ell-1}$. Since $\mathcal{C}$ is not a positive map, all we know regarding the trace is that $-2\eta_{\ell-1} \leq \mathrm{tr}[\rho_{\mathcal{C}}] \leq 2\eta_{\ell-1}$.

Putting all these components together we have that

$$\begin{aligned} \mathcal{E}[\rho_{\ell,\epsilon_\ell} \otimes \rho_{\ell,\epsilon_\ell}] =& p_g[(1 - a\epsilon_\ell)^2 \rho_\ell + (a\epsilon_\ell)^2 (Y \rho_\ell Y)] \\ &+ \rho_{\mathcal{B}} + \rho_{\mathcal{C}} \end{aligned} \tag{D23}$$

Taking the trace gives the success probability

$$P_{\mathrm{suc}} = p_g(1 - 2a\epsilon_\ell + 2a^2\epsilon_\ell^2) + \mathrm{tr}[\rho_{\mathcal{B}}] + \mathrm{tr}[\rho_{\mathcal{C}}]. \tag{D24}$$

The smallest value this can take is when $\mathrm{tr}[\rho_{\mathcal{B}}] = 0$, $\mathrm{tr}[\rho_{\mathcal{C}}] = -2\eta_{\ell-1}$, and $a = 1$. This gives the rigorous, albeit pessimistic bound, that

$$P_{\mathrm{suc}} \geq p_g(1 - 2\epsilon_\ell + 2\epsilon_\ell^2) - 2\eta_{\ell-1}. \tag{D25}$$

We renormalise to obtain

$$\rho^{\mathrm{out}} = \frac{1}{P_{\mathrm{suc}}} \mathcal{E}[\rho_{\ell,\epsilon_\ell} \otimes \rho_{\ell,\epsilon_\ell}]. \tag{D26}$$

The output error rate $\delta_\ell := \frac{1}{2}||\rho^{\mathrm{out}} - \rho_\ell||_1$ is then

$$\begin{aligned} \delta_\ell = \frac{1}{2P_{\mathrm{suc}}} \Bigg|\Bigg| & [p_g(1 - a\epsilon_\ell)^2 - P_{\mathrm{suc}}]\rho_\ell \\ & + p_g(a\epsilon_\ell)^2 (Y \rho_\ell Y) + \rho_{\mathcal{B}} + \rho_{\mathcal{C}} \Bigg|\Bigg|_1. \end{aligned} \tag{D27}$$

Using the triangle inequality we obtain

$$\begin{aligned} \delta_\ell \leq& \frac{|p_g(1 - a\epsilon_\ell)^2 - P_{\mathrm{suc}}| + p_g|a\epsilon_\ell|^2 + ||\rho_{\mathcal{B}}|| + ||\rho_{\mathcal{C}}||}{2P_{\mathrm{suc}}} \\ \leq& \frac{p_g a^2 \epsilon_\ell^2 + ||\rho_{\mathcal{B}}|| + ||\rho_{\mathcal{C}}||}{P_{\mathrm{suc}}} \\ \leq& \frac{p_g a^2 \epsilon_\ell^2 + p_b + 2\eta_{\ell-1}}{P_{\mathrm{suc}}} \end{aligned} \tag{D28}$$

The worst case scenario is again that $a = 1$, so that

$$\delta_\ell \leq \frac{p_g \epsilon_\ell^2 + p_b + 2\eta_{\ell-1}}{p_g(1 - 2\epsilon_\ell + 2\epsilon_\ell^2) - 2\eta_{\ell-1}}. \tag{D29}$$

Using the worst case bound on p_g and p_b from Eq. (D14), we have

$$\delta_\ell \leq \frac{\epsilon_\ell^2 + 28\epsilon_3^2 + 2\eta_{\ell-1}}{(1 - 8\epsilon_3^2)(1 - 2\epsilon_\ell + 2\epsilon_\ell^2) - 2\eta_{\ell-1}} \sim \epsilon_\ell^2 + 28\epsilon_3^2 + 2\eta_{\ell-1}, \tag{D30}$$

where the last approximation gives the leading order terms, showing quadratic suppression in ϵ_ℓ and ϵ_3. This completes our proof that MEK$_\ell$ suppresses generic noise.

We remark that though quadratic the prefactor of ϵ_3 is 3.5 times larger than in Eq. (13) and Eq. (F1), and the prefactor of $\eta_{\ell-1}$ is 8 times larger. However, these increased prefactors are mainly artefacts of approximations made in the proof. We report that we also found a proof with a leading order approximation $\epsilon_\ell^2 + 8\epsilon_3^2 + \eta_{\ell-1}$, though the proof is much longer.

Appendix E: Dilution proof

Here we verify the dilution expression in Eq. (23). As an intermediately step in our proof, we begin by considering the mixture $\frac{1}{2}\left(\rho_\ell + |+\rangle\langle+|\right)$. The pure magic state ρ_ℓ can be expanded in the Pauli basis as

$$\begin{aligned} \rho_\ell &= \frac{1}{2}(\mathbb{1} + H_\ell) && \text{(E1)} \\ &= \frac{1}{2}\left(\mathbb{1} + \cos(\theta_{\ell-1})X + \sin(\theta_{\ell-1})Z\right), \end{aligned}$$

and $|+\rangle\langle+| = \frac{1}{2}(\mathbb{1} + X)$. Therefore, in the Pauli basis

$$\frac{(\rho_\ell + |+\rangle\langle+|)}{2} = \frac{1}{2}\left(\mathbb{1} + q(\cos(\phi)X + \sin(\phi)Z)\right), \tag{E2}$$

where we introduce the variables q and ϕ as follows

$$\begin{aligned} q &= \frac{1}{2}\sqrt{(\cos(\theta_{\ell-1}) + 1)^2 + \sin(\theta_{\ell-1})^2}, && \text{(E3)} \\ \sin(\phi) &= \frac{\sin(\theta_{\ell-1})}{2q}. \end{aligned}$$

Using standard trigonometric identities, we find

$$\begin{aligned} q &= \cos(\theta_\ell), && \text{(E4)} \\ \phi &= 2\theta_{\ell-1} = \theta_\ell. \end{aligned}$$

This entails

$$\begin{aligned} \frac{\rho_\ell + |+\rangle\langle+|}{2} &= \cos(\theta_\ell)\rho_{\ell+1} + (1 - \cos(\theta_\ell))\frac{\mathbb{1}}{2}, && \text{(E5)} \\ &= \rho_{\ell+1, \sin^2(\theta_{\ell+1})}. \end{aligned}$$

where in the last line we have used $(1 - \cos(\theta_\ell))/2 = \sin^2(\theta_{\ell+1})$. Let us recap. By 50-50 mixing a pure magic state $|M_\ell\rangle$ with a pure $|+\rangle$ state, we obtain a noisy $|M_{\ell+1}\rangle$ state with error rate $\sin^2(\theta_{\ell+1})$. Next, we extend the proof to account for noise on the initial magic state.

Given a noisy magic state $\rho_{\ell,\epsilon}$, this can be decomposed as

$$\rho_{\ell,\epsilon} = (1 - 2\epsilon)\rho_\ell + \epsilon\mathbb{1}. \tag{E6}$$

Mixing this with the $|+\rangle$ stabiliser state by an amount λ, gives

$$\lambda\rho_{\ell,\epsilon} + (1-\lambda)|+\rangle\langle+| = \lambda(1-2\epsilon)\rho_\ell + (1-\lambda)|+\rangle\langle+| + \lambda\epsilon\mathbb{1}. \tag{E7}$$

We choose λ so that the coefficients of ρ_ℓ and $|+\rangle\langle+|$ are equal, which requires

$$\lambda = \frac{1}{2(1-\epsilon)}. \tag{E8}$$

This yields

$$\begin{aligned} &\lambda\rho_{\ell,\epsilon} + (1-\lambda)|+\rangle\langle+| && \text{(E9)} \\ &\quad = \frac{(1-2\epsilon)}{1-\epsilon}\left[\frac{\rho_\ell + |+\rangle\langle+|}{2}\right] + \frac{\epsilon\mathbb{1}}{2(1-\epsilon)}. \end{aligned}$$

We can now apply Eq. (E5) to the terms in the square brackets, so that

$$\begin{aligned} &\lambda\rho_{\ell,\epsilon} + (1-\lambda)|+\rangle\langle+| && \text{(E10)} \\ &\quad = \frac{\cos(\theta_\ell)(1-2\epsilon)}{1-\epsilon}\rho_{\ell+1,0} + \frac{(1-\cos(\theta_\ell))(1-2\epsilon) + \epsilon}{2(1-\epsilon)}\mathbb{1}, \end{aligned}$$

which can be more compactly written as

$$\lambda\rho_{\ell,\epsilon} + (1-\lambda)|+\rangle\langle+| = \rho_{\ell+1, \frac{1-q'}{2}}, \tag{E11}$$

where

$$q' = \cos(\theta_\ell)\frac{1-2\epsilon}{1-\epsilon} = \cos(\theta_\ell)\frac{1-2\epsilon}{1-\epsilon}. \tag{E12}$$

The most relevant quantity to report is the error rate $\epsilon' = (1-q')/2$, which follows immediately from the above as given by Eq. (25).

Appendix F: Simulation results

Using Mathematica, we symbolically simulate the circuit in Fig. (8d). Full details are available in the Supplementary material (see MEKL_simulation.nb).

We find the output states are again diagonal in the $\{|M_\ell\rangle, |\bar{M}_\ell\rangle\}$ basis. The main text quoted the leading order error and success probability as Eqs. (13), and use we present the exact results here

$$\delta_\ell = \frac{1}{P_{\mathrm{suc}}}\Bigg(8\epsilon_3^2 + \epsilon_\ell^2 + \frac{1}{4}\eta_{\ell-1} - 2\eta_{\ell-1}\epsilon_3 + 6\eta_{\ell-1}\epsilon_3^2 - 48\epsilon_3^3 - 8\eta_{\ell-1}\epsilon_3^3 + 136\epsilon_3^4 + 4\eta_{\ell-1}\epsilon_3^4 - 224\epsilon_3^5 + 224\epsilon_3^6 - 128\epsilon_3^7 + 32\epsilon_3^8 + \epsilon_\ell^2 - \eta_{\ell-1}\epsilon_\ell^2 - 8\epsilon_3\epsilon_\ell^2 + 8\eta_{\ell-1}\epsilon_3\epsilon_\ell^2 + 24\epsilon_3^2\epsilon_\ell^2 - 24\eta_{\ell-1}\epsilon_3^2\epsilon_\ell^2 - 32\epsilon_3^3\epsilon_\ell^2 + 32\eta_{\ell-1}\epsilon_3^3\epsilon_\ell^2 + 16\epsilon_3^4\epsilon_\ell^2 - 16\eta_{\ell-1}\epsilon_3^4\epsilon_\ell^2\Bigg), \tag{F1}$$

$$P_{\mathrm{suc}} = 448\epsilon_3^5 - 448\epsilon_3^6 + 256\epsilon_3^7 - 64\epsilon_3^8 + 1/2\eta_{\ell-1}(1-2\epsilon_3)^4(1-2\epsilon_\ell)^2 + 2\epsilon_\ell - 2\epsilon_\ell^2 + 64\epsilon_3^3(2 - \epsilon_\ell + \epsilon_\ell^2) - 32\epsilon_3^4(9 - \epsilon_\ell + \epsilon_\ell^2) + 8\epsilon_3(1 - 2\epsilon_\ell + 2\epsilon_\ell^2) - 8\epsilon_3^2(5 - 6\epsilon_\ell + 6\epsilon_\ell^2). \tag{F2}$$

[1] B. Eastin and E. Knill, Phys. Rev. Lett. **102**, 110502 (2009).

[2] A. Y. Kitaev, A. Shen, and M. N. Vyalyi, *Classical and quantum computation*, Vol. 47 (American Mathematical Society Providence, 2002).

[3] S. Bravyi and A. Kitaev, Phys. Rev. A **71**, 022316 (2005).

[4] A. M. Meier, B. Eastin, and E. Knill, Quant. Inf. and Comp. **13**, 195 (2013).

[5] S. Bravyi and J. Haah, Phys. Rev. A **86**, 052329 (2012).

[6] C. Jones, Phys. Rev. A **87**, 042305 (2013).

[7] A. G. Fowler, S. J. Devitt, and C. Jones, Scientific reports **3**, 1939 (2013).

[8] V. Kliuchnikov, D. Maslov, and M. Mosca, Physical review letters **110**, 190502 (2013).

[9] A. Paetznick and K. M. Svore, Quantum Information & Computation **14**, 1277 (2014).

[10] D. Gosset, V. Kliuchnikov, M. Mosca, and V. Russo, Quantum Information & Computation **14**, 1261 (2014).

[11] N. J. Ross and P. Selinger, arXiv preprint arXiv:1403.2975 (2014).

[12] M. Amy and M. Mosca, arXiv preprint arXiv:1601.07363 (2016).

[13] J. T. Anderson, G. Duclos-Cianci, and D. Poulin, Phys. Rev. Lett. **113**, 080501 (2014).

[14] H. Bombin, arXiv preprint arXiv:1311.0879 (2013).

[15] H. Bombin and M. A. Martin-Delgado, Physical review letters **97**, 180501 (2006).

[16] H. Bombin, R. W. Chhajlany, M. Horodecki, and M. A. Martin-Delgado, New Journal of Physics **15**, 055023 (2013).

[17] R. Raussendorf, J. Harrington, and K. Goyal, New Journal of Physics **9**, 199 (2007), quant-ph/0703143.

[18] J. O'Gorman and E. T. Campbell, arXiv preprint arXiv:1605.07197 (2016).

[19] B. J. Brown, N. H. Nickerson, and D. E. Browne, Nat Commun **7** (2016).

[20] C. Wang, J. Harrington, and J. Preskill, Annals of Physics **303**, 31 (2003).

[21] R. Raussendorf and J. Harrington, Phys. Rev. Lett. **98**, 190504 (2007).

[22] A. G. Fowler, A. C. Whiteside, A. L. McInnes, and A. Rabbani, Physical Review X **2**, 041003 (2012).

[23] H. Anwar, E. T. Campbell, and D. E. Browne, New Journal of Physics **14**, 063006 (2012).

[24] E. T. Campbell, H. Anwar, and D. E. Browne, Phys. Rev. X **2**, 041021 (2012).

[25] E. T. Campbell, Phys. Rev. Lett. **113**, 230501 (2014).

[26] H. Dawkins and M. Howard, Physical review letters **115**, 030501 (2015).

[27] A. J. Landahl and C. Cesare, arXiv preprint arXiv:1302.3240 (2013).

[28] D. Gottesman and I. L. Chuang, Nature **402**, 390 (1999).

[29] S. Lloyd, Science **273**, 1073 (1996).

[30] D. Poulin, M. B. Hastings, D. Wecker, N. Wiebe, A. C. Doherty, and M. Troyer, arXiv preprint arXiv:1406.4920 (2014).

[31] C. J. Trout and K. R. Brown, International Journal of Quantum Chemistry **115**, 1296 (2015).

[32] D. Wecker, M. B. Hastings, N. Wiebe, B. K. Clark, C. Nayak, and M. Troyer, Physical Review A **92**, 062318 (2015).

[33] M. B. Hastings, D. Wecker, B. Bauer, and M. Troyer, Quantum Information & Computation **15**, 1 (2015).

[34] C. Jones, Physical Review A **87**, 022328 (2013).

[35] B. Eastin, Physical Review A **87**, 032321 (2013).

[36] A. Paetznick and B. W. Reichardt, Phys. Rev. Lett. **111**, 090505 (2013).

[37] G. Duclos-Cianci and D. Poulin, Phys. Rev. A **91**, 042315 (2015).

[38] D. Gottesman and I. Chuang, Nature **402**, 390 (1999).

[39] A. Bocharov, M. Roetteler, and K. M. Svore, Physical Review A **91**, 052317 (2015).

[40] A. Bocharov, M. Roetteler, and K. M. Svore, Phys. Rev. Lett. **114**, 080502 (2015).

[41] Y. Li, New Journal of Physics **17**, 023037 (2015).

[42] The Gridsynth package is available at http://www.mathstat.dal.ca/ selinger/newsynth/.

[43] Duclos-Cianci, private communication.

[44] E. Knill, Nature **434**, 39 (2005).

[45] E. T. Campbell and M. Howard, arXiv preprint arXiv:1606.01906 (2016).

[46] E. T. Campbell and M. Howard, arXiv preprint arXiv:1606.01904 (2016).

[47] D.-S. Wang, D. W. Berry, M. C. de Oliveira, and B. C. Sanders, Phys. Rev. Lett. **111**, 130504 (2013).

[48] D.-S. Wang and B. C. Sanders, New Journal of Physics **17**, 043004 (2015).

[49] Private communication with Alex Bocharov.