Fault-tolerant quantum computation with asymmetric Bacon-Shor codes

Peter Brooks and John Preskill

Fault-tolerant quantum computation
with asymmetric Bacon-Shor codes

Peter Brooks and John Preskill

Institute for Quantum Information, California Institute of Technology, Pasadena, CA 91125, USA

We develop a scheme for fault-tolerant quantum computation based on asymmetric Bacon-Shor codes, which works effectively against highly biased noise dominated by dephasing. We find the optimal Bacon-Shor block size as a function of the noise strength and the noise bias, and estimate the logical error rate and overhead cost achieved by this optimal code. Our fault-tolerant gadgets, based on gate teleportation, are well suited for hardware platforms with geometrically local gates in two dimensions.

I. INTRODUCTION

The theory of fault-tolerant quantum computation [1, 2] has established that noisy quantum computers can operate reliably provided the noise is neither too strong nor too strongly correlated. In a fault-tolerant quantum circuit, carefully designed gadgets process logical qubits protected by quantum error-correcting codes.

Typical fault-tolerant gadgets are designed to work effectively against generic noise without any special structure. But in some physical settings, the noise is expected to be highly biased, with dephasing in the computational basis far more likely than bit flips. This paper addresses how noise bias can be exploited to improve the reliability of fault-tolerant quantum circuits.

Specifically, we have in mind a setting in which, to an excellent approximation, the computational basis states $\{|0\rangle, |1\rangle\}$ are the energy eigenstates for an unperturbed qubit, and single-qubit gates which are diagonal in this basis can be performed by adiabatically adjusting the energy splittings between these states. Similarly, diagonal two-qubit gates are performed by adjusting the energy splittings of the four (approximate) energy eigenstates $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$. Various noise sources may induce fluctuations in these energy spacings, and pulse imperfections may cause gates to be over-rotated or under-rotated; in either case the noisy gate deviates from the ideal gate, but remains diagonal. Other physical processes may induce transitions between energy eigenstates, but we will assume that this non-diagonal noise is far weaker than the dominant diagonal noise.

Our scheme for fighting biased noise is based on asymmetric Bacon-Shor codes [3–5]. These quantum codes combine together a length-m repetition code protecting against dephasing and a length-n repetition code protecting against bit flips; by "asymmetric" we mean $m > n$, so the code works more effectively against dephasing than against bit flips. A complete universal set of fault-tolerant gates can be constructed using only diagonal gates, plus single-qubit measurements in the X eigenstate basis $\{|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)\}$ and preparation of single-qubit $|+\rangle$ states.

In most previous studies of quantum fault-tolerance based on Bacon-Shor codes, these codes have been concatenated to build a hierarchy of codes within codes [6, 7]. Here we will mainly study the performance of a single large block code rather than a concatenated scheme. For the Bacon-Shor code family, used without concatenation, there is no accuracy threshold; rather, for a fixed value of the noise strength and noise bias, there is an optimal block size which achieves the best performance.

Our gadget constructions and analysis extend the results in [8, 9], where the case $n = 1$ was regarded as the bottom layer of a concatenated code. In contrast to [8, 9], we consider the case where the gates are required to be geometrically local in a two-dimensional array. In the geometrically local case, logical blocks can be transported as needed via teleportation, with the fundamental operations still limited to diagonal gates, X measurements, and $|+\rangle$ preparations.

We have obtained upper bounds on the optimal performance and overhead cost of logical gates by deriving analytic formulas, and our estimates are far from tight. We expect that significantly better results could be obtained using numerical simulations assuming independent biased noise, which we have not attempted. In particular, our analysis includes quite conservative estimates of the probability of failure for preparations of cat states that are used for measurements of logical Pauli operators.

Another particularly attractive approach to achieving quantum fault tolerance with geometrically local gates is based on topological codes [10–12], which unlike Bacon-Shor codes have an accuracy threshold and hence can reach arbitrarily low error rates per logical gate. Topological codes can be adapted for optimized protection against biased noise, and they perform well against biased noise even without any such adaptation. Our current results do not conclusively identify noise parameter regimes for which Bacon-Shor codes are clearly superior to topological codes. However, our scheme has several appealing features — for example, the optimal error rate per logical gate is achieved with a relatively modest number of physical qubits per code block, and is not too adversely affected when qubit measurements are noisier than quantum gates. Furthermore, only relatively modest classical computational resources are needed to

interpret error syndromes.

We review Bacon-Shor codes in Sec. II, describe our biased noise model in Sec. III, and construct our fault-tolerant gadgets in Sec. IV. We obtain upper bounds on the performance of our fault-tolerant logical CNOT gate in Sec. V, repeat the analysis in Sec. VI for the case where two-qubit gates are required to be geometrically local, and report numerical values in Sec. VII. In Sec. VIII and IX we discuss the state injection and state distillation procedures needed to complete our fault-tolerant universal gate set.

Fault-tolerant gadgets protecting against biased noise have also been discussed previously in [13–15].

II. BACON-SHOR CODES

Bacon-Shor codes are quantum subsystem codes which are constructed by combining together two quantum repetition codes, one protecting against Z (phase) errors and the other protecting against X (bit flip) errors. We will consider using Bacon-Shor codes to protect quantum information against biased noise, such that Z errors are much more common than X errors; therefore, the length m of the code protecting against Z errors will be longer than the length n of the code protecting against X errors. In this case we say that the code is asymmetric. The code protects one encoded qubit in a block of nm physical qubits.

It is convenient to arrange the physical qubits in a rectangular lattice with n rows and m columns. The code has $(m-1)$ Z-type stabilizer generators, or check operators, which can be measured to determined a syndrome for detecting X errors, and $n-1$ X-type stabilizer generators, which can be measured to determine the syndrome for Z errors. Each Z-type check operator is a weight-$2m$ operator applying Z to each qubit in a pair of adjacent rows, and each X-type check operator is a weight-$2n$ operator applying X to each qubit in a pair of adjacent columns. Because each Z-type check operator "collides" with each X-type check operator at four lattice sites, the Z-type and X-type check operators are mutually commuting and can be measured simultaneously. The logical Pauli operator Z^L acting on the encoded qubit is a tensor product $Z^{\otimes m}$ acting on all the qubits in a (long) row; all rows are equivalent because a product of two such row operators is in the code stabilizer. The operator Z^L collides with each X-type check operator at two sites; therefore it commutes with these check operators and hence preserves the code space. The logical Pauli operator X^L is a tensor product $X^{\otimes n}$ acting on all qubits in a (short) column; again the columns are equivalent, and X^L commutes with the Z-type check operators. The operators Z^L and X^L collide at a single site and hence anticommute.

Assuming that m and n are both odd, the code can correct $(m-1)/2$ Z errors and $(n-1)/2$ X errors. To perform a decoded X^L measurement, we can measure
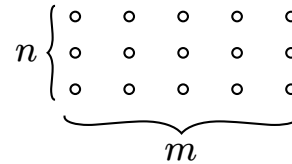

FIG. 1: Bacon-Shor qubits arranged in an $n \times m$ lattice.

all nm qubits in the X basis, compute the parity of the outcomes for each of the m columns, and then perform a majority vote on the m column parities. The decoded measurement fails only if $(m+1)/2$ columns each have at least one Z errors. Similarly, we can perform a decoded Z^L measurement by measuring all qubits in the Z basis, computing the parity of the outcomes for each of the n rows, and then performing a majority vote on the n row parities. This decoded measurement fails only if $(n+1)/2$ rows each have at least one X error.

The code also has a gauge algebra of Pauli operators that commute with the check operators and with the logical operators, though the gauge operators do not necessarily commute with one another. The Z-type gauge algebra is generated by weight-two operators with Z acting on a pair of neighboring qubits in the same column, and the X-type gauge algebra is generated by weight-two operators with X acting on a pair of neighboring qubits in the same row. Thus each X-type check operator is a product of n weight-two gauge operators, and each Z-type check operator is a product of m weight-two gauge operators. Conveniently, then, the error syndrome can be determined by measuring only weight-two gauge operators, and furthermore each of these gauge operators is geometrically local, acting on a pair of neighboring qubits. Measuring an X-type gauge operator disturbs the values of some Z-type gauge operators (and vice versa), but without inflicting any damage on the protected logical subsystem.

III. NOISE MODEL

We consider a local stochastic biased noise model, in which we distinguish between a rate ε for dephasing faults and a (perhaps much lower) rate ε' for general faults in diagonal gates. The ratio $R = \varepsilon/\varepsilon'$ will be called the noise bias. For simplicity we will assume that ε is also the error rate for single-qubit preparations and measurements in most of the formulas we display, though we will treat the case where preparations and measurements are noisier than diagonal gates in Sec. VII C. A dephasing fault, which we will also call a diagonal fault, is a trace-preserving completely positive map such that all Kraus operators are diagonal in the computational basis; for
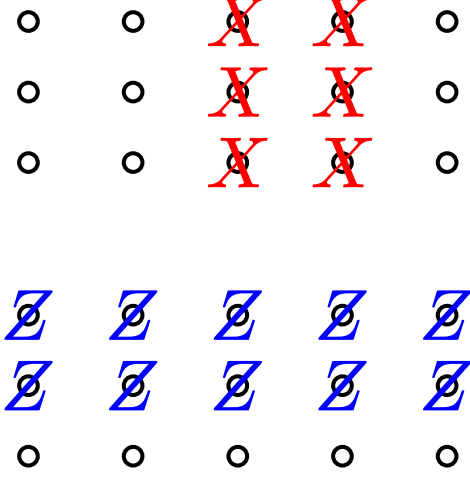
FIG. 2: (Color online) X-type stabilizer operator (top) and Z-type stabilizer operator (bottom)

a general fault, which we will also call a non-diagonal fault, the Kraus operators are unrestricted. We assume that the sum of the probabilities for all fault paths with r dephasing faults and s non-dephasing faults at specified circuit locations is no greater than $\varepsilon^r(\varepsilon')^s$.

IV. FAULT-TOLERANT GADGETS

The construction of fault-tolerant gadgets for universal quantum computation follows Ref. [8]. We implement the logical gate set

$$\mathcal{G}_{\text{CSS}}^L = \{\mathcal{P}_{|+\rangle}^L, \mathcal{P}_{|0\rangle}^L, \mathcal{M}_X^L, \mathcal{M}_Z^L, \text{CNOT}^L\} \qquad (1)$$

using the physical gate set

$$\mathcal{G}_{\text{fund}} = \{\mathcal{P}_{|+\rangle}, \mathcal{M}_X, \text{CZ}\}; \qquad (2)$$

here $\mathcal{M}_X$, $\mathcal{M}_Z$ denote measurements of the Pauli operators X and Z, $\mathcal{P}_{|+\rangle}$, $\mathcal{P}_{|0\rangle}$ denote preparations of the X and Z eigenstates with eigenvalue 1, and CZ denotes the two-qubit controlled phase gate

$$\text{CZ} = \exp\left(-i\frac{\pi}{4}(Z - I) \otimes (Z - I)\right), \qquad (3)$$

which is diagonal in the computational (Z-eigenstate) basis $|00\rangle, |01\rangle, |10\rangle, |11\rangle$ with eigenvalues $(1, 1, 1, -1)$. A diagonal gate like this one can be realized by adiabatically perturbing the energy splittings of the computational basis states, and hence could plausibly have highly biased noise. Note that we have not included $\mathcal{M}_Z$ in the physical gate set Eq. (2), because it will not be needed in our fault-tolerant gadget constructions.

The CZ gate acts symmetrically on the two qubits, and, like any Clifford gate, it can be usefully characterized by how it propagates Pauli errors:

$$\text{CZ} : XI \to XZ, \quad IX \to ZX,$$
$$ZI \to ZI, \quad IZ \to IZ. \qquad (4)$$

Of course, CZ commutes with Z acting on either qubit, and propagates an X acting on either qubit to a Z error acting on the other. These properties are nice if Z errors dominate over X errors and we use a code that corrects Z errors more effectively than X errors. The CZ gates do not propagate the relatively common Z errors, and though X errors do propagate they are transformed to Z errors, which the code handles more readily.

To achieve universal quantum computation, we supplement the logical gate set $\mathcal{G}_{\text{CSS}}^L$ with high-fidelity encoded states $|i\rangle = \frac{1}{\sqrt{2}}(|0\rangle + i|1\rangle)$ and $|T\rangle = \frac{1}{\sqrt{2}}(|0\rangle + e^{i\pi/4}|1\rangle)$, which can be prepared using state injection and distilled using the $\mathcal{G}_{\text{CSS}}^L$ operations [16]. Gates that complete a universal set can then be constructed using these states and the $\mathcal{G}_{\text{CSS}}^L$ gates [17].

A. $\mathcal{M}_X^L$ gadget

To perform a destructive $\mathcal{M}_X^L$ measurement we first perform independent $\mathcal{M}_X$ measurements on all qubits in the code block, and then compute the parity of the n measurement outcomes in each column. Finally a majority vote of the m column parities yields the logical measurement outcome. The result agrees with an ideal measurement if the sum of the number of Z errors in the block and the number of faulty $\mathcal{M}_X$ measurements is no larger than $(m-1)/2$.

B. $\mathcal{M}_Z^L$ gadget

If $\mathcal{M}_Z$ were included in our repertoire of physical gates, then we could perform a destructive $\mathcal{M}_Z^L$ using the dual of the procedure described above for the destructive $\mathcal{M}_X^L$. That is, we could perform independent $\mathcal{M}_Z$ measurements on all qubits in the code block, compute the parity of the m measurement outcomes in each row, and perform a majority vote of the n row parities to obtain the logical measurement outcome. The result agrees with an ideal measurement if the sum of the number of X errors in the block and the number of faulty $\mathcal{M}_Z$ measurements is no larger than $(n-1)/2$.

However, for our fault-tolerant gadget constructions we will require instead a *non-destructive* $\mathcal{M}_Z^L$ measurement, such that the ideal measurement procedure leaves encoded Z^L eigenstates intact. In principle a non-destructive measurement of the row parity could be done by preparing a single ancilla qubit in the state $|+\rangle$, performing m CZ gates in succession acting on the ancilla qubit and the m qubits in the row, and finally measuring

the ancilla qubit in the X basis. However, this procedure is not fault-tolerant for two reasons. First, a single Z error acting on the ancilla qubit can flip the measurement outcome. Second, a single X error acting on the ancilla can propagate multiple times, producing a high-weight Z error acting on the qubits in the row.

This second problem can be addressed by replacing the single-qubit ancilla by an m-qubit cat state $|+^{\text{cat}}\rangle = \frac{1}{\sqrt{2}} \left(|0\rangle^{\otimes m} + |1\rangle^{\otimes m} \right)$, so that each ancilla qubit interacts via a CZ gate with only one data qubit, limiting the error propagation. The fault-tolerant preparation of the cat state is discussed below. After the m CZ gates, the cat state is read out in the basis $|\pm^{\text{cat}}\rangle = \frac{1}{\sqrt{2}} \left(|0\rangle^{\otimes m} \pm |1\rangle^{\otimes m} \right)$ to determine the row parity. This measurement of $X^{\otimes m}$ is performed destructively by measuring each of the m ancilla qubits in the X basis and computing the parity of the outcomes.

To address the first problem (that a single fault can flip the measured parity), the parity measurement is repeated r times for each row, each time with a fresh cat state, and the majority of the results is computed. The repetition provides protection against Z errors in the ancilla, but of course there may also be (relatively rare) X errors acting on the data qubits due to faults during the Z^L measurement circuit. These faults cause a logical X^L error only if X errors occur in a majority of the n rows.

Actually, we find that, depending on the rates for diagonal and non-diagonal faults, it may be advantageous to use a shorter ancilla prepared in a cat state of length p, where $1 \leq p < m$. In that case some ancilla qubits interact with more than one data qubit, increasing the danger of error propagation, but on the other hand errors are less likely to occur during the preparation of the (shorter) cat state. However, using a length-m cat state has the significant advantage that the $\mathcal{M}_Z^L$ measurement can be conveniently executed using geometrically local gates on the two-dimensional lattice, as we will discuss in Sec. VI.

This procedure for measuring Z^L can easily be extended to build gadgets that perform the parity measurements $\mathcal{M}_{ZZ}^L$, $\mathcal{M}_{ZZZ}^L$ on multiple logical qubits.

C. Cat state preparation

The length-p cat state $|+^{\text{cat}}\rangle = \frac{1}{\sqrt{2}} \left(|0\rangle^{\otimes p} + |1\rangle^{\otimes p} \right)$ may be characterized as the simultaneous eigenstate with eigenvalue 1 of $X^{\otimes p}$ and of $p-1$ ZZ operators, each one acting on a pair of neighboring qubits. To prepare the cat state, we start by preparing the product state $|+\rangle^{\otimes p}$ (an eigenstate of $X^{\otimes p}$), and proceed to measure the $p-1$ ZZ operators (each of these measurements is executed using yet another ancilla qubit). If every measurement has the outcome $ZZ = +1$, then the cat state has been prepared successfully, assuming the measurements are flawless. Otherwise, the measurement outcomes provide

a syndrome pointing to X errors in the cat state (each $ZZ = -1$ outcome identifies the endpoint of a chain of X errors). These X errors in the cat state need not be corrected; instead we can keep track of their propagation as the computation proceeds. Specifically, the X errors in the cat state propagate to become Z errors acting on the measured row of the data block, which should be taken into account when we decode a later X^L measurement of that block. Following Knill [17], we say that the measured syndrome is used to update the "Pauli frame" of the computation.

To improve robustness, we add one more (redundant) ZZ measurement acting on the first and last qubit in the cat state. Then an even number among the p ZZ measurements should have the outcome -1 if all measurements are correct; if in fact we find an odd number of -1 outcomes than we reject the syndrome. (Adding further redundant checks would further improve the reliability of the syndrome extraction, though we will not include these.) We conduct r' rounds of syndrome measurement, including rejected rounds, so that the number of accepted rounds may be fewer than r'.

In principle, the cat state Pauli frame could be determined by performing a perfect matching on a two-dimensional graph representing the space-time history of the cat state syndrome measurement [10]. Instead, we will consider a much simpler and less effective method for decoding the syndrome history. Though far from optimal, our procedure has two advantages over the perfect matching algorithm — it requires only very modest (classical) computational resources compared to perfect matching, and it is also much easier to study analytically.

In our unsophisticated scheme, we determine the Pauli frame using whichever syndrome occurs most frequently among the accepted rounds, which we call the "winning" syndrome. (Our analysis does not depend on what rule is used when two or more syndromes tie for the distinction of being most frequent, since we pessimistically assume that the preparation gadget fails catastrophically in the event of such a tie.) The state preparation protocol is deterministic in the sense that ancillas are never discarded, and the number of rounds of syndrome measurement is always r' irrespective of the measurement outcomes.

The complete circuit for measuring the row parity, including repeated measurement of the cat state syndrome, is shown in Fig. 3. In the circuit shown, the cat state is prepared first, and then used to measure the parity of a row. Since the CZ gates commute with each other, we could alternatively perform the ZZ measurements on the cat state *after* the ancilla interacts with the data.

Even if the cat state is "prepared" after its "use," X errors in the ancilla during the preparation can still cause trouble. The X errors cause the cat state syndrome to evolve as it is measured repeatedly, and therefore the syndrome that occurs most often, even if valid when measured, may differ from the syndrome that is applicable when the cat state interacts with the data. The effect of this syndrome "drift" needs to be included in our error
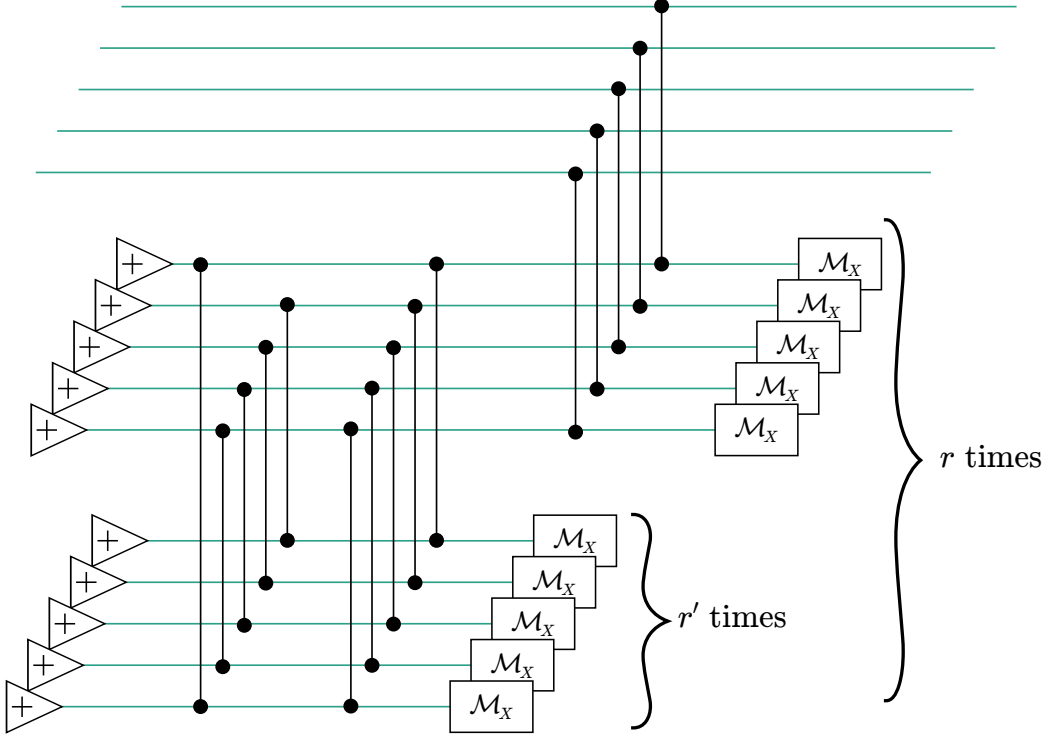
FIG. 3: (Color online) Gadget for a single row of the non-destructive $\mathcal{M}_Z^L$ measurement, where the cat state ancilla length p is equal to the row length m.

analysis.

D. $\mathcal{P}_{|+\rangle}^L$ gadget

The logical state $|+\rangle^L$, for a particular "choice of gauge," is a product of m length-n cat states, one for each column of the block. This state is an eigenstate with eigenvalue 1 of $X^L = X^{\otimes n}$ acting on any column, and of the gauge operator ZZ acting on a pair of neighboring qubits in any column (and hence also a $+1$ eigenstate of all the Z-type check operators). We prepare these m length-n cat states in parallel, using the same procedure for preparing length-p cat states described above. In this case we repeat the syndrome measurement for each column r_+ times.

E. Error correction gadget

Error correction can be performed on a logical block using the "one-bit teleportation" gadget shown in Fig. 4. This circuit transfers the single-qubit state $|\psi\rangle$ from qubit 2 to qubit 1. We prepare qubit 1 in the state $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, perform the joint measurement $\mathcal{M}_{Z^1 Z^2}$ on the two qubits, then perform the single-qubit measurement $\mathcal{M}_{X^2}$. After applying Pauli operators conditioned
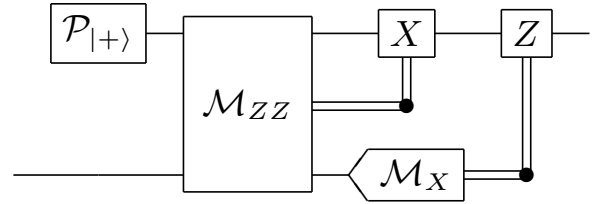

FIG. 4: One-bit teleportation circuit.

on the two measurement outcomes, the result is that the input state $|\psi\rangle$ of qubit 2 is moved to qubit 1. The conditional X and Z gates in this teleportation gadget need not actually be executed; instead the Pauli frame is updated to reflect the measurement outcomes. One-bit teleportation can be performed at the logical level using the fault-tolerant measurement gadgets $\mathcal{M}_{ZZ}^L$ and $\mathcal{M}_X^L$ that we have already described.

F. CNOT^L gadget

The gadget shown in Fig. 5 applies a fault-tolerant "teleported" CNOT^L logical operation, up to a known

Pauli error determined by the measurement outcomes [8]. Aside from executing the logical gate, the gadget also achieves error correction by measuring both incoming blocks and preparing fresh outgoing blocks.

To better understand how the gadget works, it is helpful to divide it into two parts. One part is the "one-bit teleportation" gadget described in Sec. IV E, which performs error correction on a single block. The other component of the CNOT gadget is shown in Fig. 6, where again the Pauli frame updates are explicitly indicated. We may verify that this circuit realizes a CNOT gate by checking that it propagates Pauli operators according to the rule

$$\text{CNOT}: XI \to XX, \quad IX \to IX,$$
$$ZI \to ZI, \quad IZ \to ZZ. \quad (5)$$

We denote the measurement outcomes of $\mathcal{M}_X$ and $\mathcal{M}_{ZZZ}$ as $(-1)^a$ and $(-1)^b$ respectively. Since $|+\rangle$ is an eigenstate of X with eigenvalue $+1$, XI acting on the input qubits is equivalent to XIX acting on the three qubits in the circuit, which commutes with $\mathcal{M}_{ZZZ}$ and hence becomes XX acting on the output qubits. Likewise, IX acting on the input qubits is equivalent to IXX, which commutes with $\mathcal{M}_{ZZZ}$ and becomes $(-1)^a IX$ acting on the output qubits using the measurement outcome $IXI = (-1)^a$. ZI acting on the input qubits commutes with $\mathcal{M}_{ZZZ}$ and trivially maps to ZI acting on the output qubits, while IZ acting on the input qubits becomes IZI, which maps to $(-1)^b ZIZ$ using the measurement outcome $ZZZ = (-1)^b$. Hence we have found

$$\text{CNOT}: XI \to XX, \quad IX \to (-1)^a IX,$$
$$ZI \to ZI, \quad IZ \to (-1)^b ZZ,$$
$$(6)$$

which is corrected to Eq.(5) by applying Z^a to both output qubits and X^b to the second qubit. When we combine the CNOT gate with one-bit teleportation of the first qubit, the measurements $\mathcal{M}_{ZZ}$ and $\mathcal{M}_{ZZZ}$ commute and hence can be done in either order.

Four different versions of the CNOT gadget are shown in Fig. 5 Let us denote the blocks appearing in the figure as blocks 1, 2, 3, 4, counting from the top down. Version (a) of the CNOT gadget contains the preparations and measurements $\mathcal{P}_{|+\rangle}[1]$, $\mathcal{P}_{|+\rangle}[4]$, $\mathcal{M}_{ZZ}[12]$, $\mathcal{M}_{ZZZ}[234]$, $\mathcal{M}_X[2]$, $\mathcal{M}_X[3]$. This gadget has some arbitrary features which are exploited to construct the other versions. Since $ZZ[12]ZZZ[134] = ZZZ[234]$, we extract the same parity information if $\mathcal{M}_{ZZZ}[234]$ is replaced by $\mathcal{M}_{ZZZ}[134]$; also, since $\mathcal{M}_{ZZ}$ acts symmetrically on its two measured blocks, we may permute the blocks, in effect replacing $\mathcal{P}_{|+\rangle}[1]$ and $\mathcal{M}_X[2]$ by $\mathcal{P}_{|+\rangle}[2]$ and $\mathcal{M}_X[1]$, thus obtaining version (b). Similarly, since $\mathcal{M}_{ZZZ}$ acts symmetrically, we may replace $\mathcal{P}_{|+\rangle}[4]$ and $\mathcal{M}_X[3]$ by $\mathcal{P}_{|+\rangle}[3]$ and $\mathcal{M}_X[4]$, thus obtaining versions (c) and (d). In all these realizations of the gadget, the upper incoming block is the control block and the lower incoming block is the

target block; by replacing $\mathcal{M}_{ZZ}[12]$ and $\mathcal{M}_{ZZZ}[234]$ by $\mathcal{M}_{ZZZ}[123]$ and $\mathcal{M}_{ZZ}[34]$ we can reverse the order of the CNOT.

This freedom to choose the incoming control block to be either block 1 or block 2, and to choose the incoming target block to be either block 3 or block 4, will be useful when we consider in Sec. VI how to realize the CNOT gadget using geometrically local gates. No matter how we configure the gadget, one block participates in both $\mathcal{M}_{ZZ}$ and $\mathcal{M}_{ZZZ}$. However, for versions (a) and (c) in Fig. 5, the twice measured block is an incoming (control) data block, while for versions (b) and (d) the twice measured block is an outgoing (control) data block.

The CNOT^L gadget is characterized by four parameters: the length p of the cat states used in the Z-type measurements and the number of times measurements are repeated (r_+ for the $|+\rangle^L$ preparation, r' for the cat state preparation, and r for ZZ^L and ZZZ^L measurements). In principle we could choose different values of these parameters in the measurements $\mathcal{M}^L_{ZZ}$ and $\mathcal{M}^L_{ZZZ}$, but in practice we have found it optimal or near optimal to use the same parameters for both.

G. Scheduling

If geometrically nonlocal gates are allowed, and fresh ancilla qubits are always available on demand, then we can schedule the CZ gates so that no data qubit or ancilla qubit is idle during any time step. We assume that a preparation, measurement, or CZ gate can be performed in a single time step. Suppose, for example, that we use the gadget design in Fig. 5b, where the outgoing control data block participates in both $\mathcal{M}^L_{ZZ}$ and $\mathcal{M}^L_{ZZZ}$, and let us also assume that the [12] cat states have length $2m$ while the [234] cat states have length $3m$, matching the weight of the measured operators. We label the r ancilla registers used to measure a row operator r times by $a \in \{1, 2, 3, \dots, r\}$ for the ZZ^L measurement and by $b \in \{1, 2, 3, \dots, r\}$ for the ZZZ^L measurement. Suppose that $\mathcal{P}^L_{|+\rangle}[2]$ is completed in time step -1 and that $\mathcal{P}^L_{|+\rangle}[3]$ is completed in time step 0. Then the CZ gates are scheduled as follows: The [12] ancilla a interacts with block 2 in time step $a - 1$ and with block 1 in time step a, while the [234] ancilla b interacts with blocks 2, 3, and 4 in time state b. Hence processing of the incoming blocks (1 and 4) begins in time step 1, and processing of the outgoing blocks (2 and 3) ends in time step r.

The "preparation" of the [234] ancilla b begins in time step $b + 1$, but the "preparation" of the [12] ancilla a is "staggered." The syndrome measurement for the first m qubits in ancilla a (those that interact with block 1) begins in time step $a + 1$, while the syndrome measurement for the last m qubits (those that interact with block 2) begins one step earlier (step a). Despite this one time step delay, no qubits are ever idle — the ancilla qubit used to measure syndrome bit $Z[m]Z[m+1]$ in cat state a interacts with cat state qubit $m+1$ in time step a and
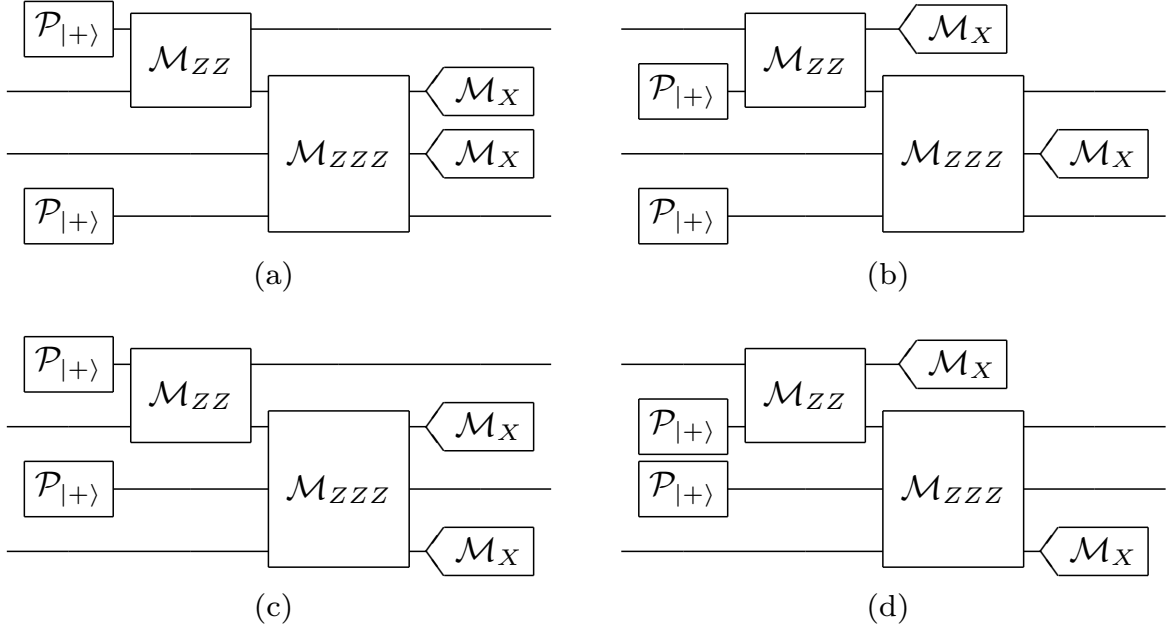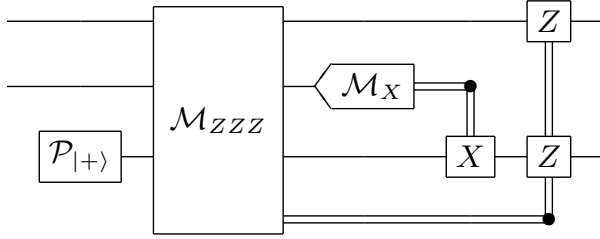
FIG. 5: Four versions of the CNOT gadget.

FIG. 6: Teleported CNOT gate.

interacts with cat state qubit m in time step $a+1$. Meanwhile, the ancilla qubit used to measure the syndrome bit $Z[1]Z[2m]$ in cat state a interacts with cat state qubit $2m$ in time step $a+1$ and with cat state qubit 1 in time step $a+2$, *etc.*

If instead we use the CNOT gadget design in Fig. 5a, where the incoming control data block participates in both $\mathcal{M}_{ZZ}^L$ and $\mathcal{M}_{ZZZ}^L$, the processing of the data blocks is not perfectly synchronized. If $\mathcal{M}_{ZZ}^L$ is performed first, then the incoming and outgoing control blocks are one time step ahead of the incoming and outgoing target blocks, while if $\mathcal{M}_{ZZZ}^L$ is performed first it is the other way around. Thus if the incoming control block is either one step ahead or one step behind the incoming target block, we can perform the CNOT without leaving any qubits idle, while maintaining the one-step lag between the blocks. But if the two incoming blocks are synchronized, then using the gadget in Fig. 5a one block or the other would have to wait for one time step before the processing of the block begins.

H. Injection by teleportation

To complete a universal set of protected gates, we will use the magic state distillation procedure introduced by Bravyi and Kitaev [16]. First we inject non-Clifford states into the code block using teleportation, then we improve the fidelity of these encoded states via distillation. The distillation protocol is discussed further in Sec. IX; here we briefly describe the state injection procedure, with further details postponed until Sec. VIII.

The state injection makes use of the "one-bit teleportation" circuit depicted in Fig. 4. To inject a state into the Bacon-Shor block, first the state $|\psi\rangle$ is prepared using a single physical qubit, while a logical block is prepared in the state $|+\rangle^L$. Then the weight-$(m+1)$ operator $Z^L Z$ is measured on the single qubit and a single row of the block, and finally the single qubit is measured in the X basis. This procedure prepares the block in the state $|\psi\rangle^L$, up to a logical Pauli operator known from the measurement outcomes.

There are other possible injection procedures, but this one is particularly simple. A single fault in the circuit can cause failure.

V. EFFECTIVE ERROR STRENGTH

The CNOT gadget may produce a logical error if any one of its four measurements deviates from the ideal case. To estimate the failure probability for the CNOT gadget, we must enumerate the ways in which these measurements might fail.

One potential source of trouble is the preparation of cat states and the logical $|+\rangle^L$ states. Recall that in these preparation gadgets a syndrome is measured repeatedly, and that the "winning" (*i.e.*, most frequently occurring) syndrome measurement result is used to update the Pauli frame. This winning syndrome might differ from the syndrome that would have been inferred in an ideal gadget because of repeated Z errors that induce errors in the syndrome measurement, because of X errors that cause the syndrome to evolve between rounds, or because of some combination of the two. We say that the preparation gadget "succeeds" if the winning syndrome matches the syndrome that could have been obtained by an ideal measurement at some stage during the multiple rounds of syndrome measurement, and if furthermore fewer than half of the cat state qubits are afflicted by X errors during the preparation process. Otherwise the preparation gadget "fails."

Failure of a preparation gadget can be catastrophic, because the inferred Pauli frame may differ from the actual Pauli frame by a high-weight X-type error. If it occurs in a cat state, this high-weight X error could propagate to the data, causing a logical Z^L error. If it occurs in the preparation of a logical block, the error could flip the value of a subsequent Z-type logical measurement.

On the other hand, even if a cat state preparation succeeds, the winning syndrome might not coincide perfectly with the ideal syndrome that applies to the cat state at the time it interacts with the data. The syndrome may drift due to X errors that accumulate during the many rounds of syndrome measurement. However, the inferred Pauli frame differs from the ideal Pauli frame by a number of X errors that is no larger than the number of non-diagonal faults that occur inside the preparation gadget. Though the inferred Pauli frame may not be exactly right, it is not catastrophically wrong unless there are many X errors in the circuit.

Aside from errors in the measured cat state syndrome, a $\mathcal{M}_{ZZ}^L$ or $\mathcal{M}_{ZZZ}^L$ could be faulty because of Z errors acting on the cat state qubits, which flip the outcomes of row measurements. Because each row measurement is repeated r times, multiple errors are required for a row measurement to fail. Alternatively, an X error acting on a data qubit could flip the outcome of a row measurement. Because a majority vote is performed on n row measurement outcome, again multiple errors are required for the logical Z-type measurement to fail. An $\mathcal{M}_X^L$ could be faulty because of Z errors on data qubits, which flip outcomes of subsequent X measurements on those qubits. Because a majority vote is performed on the m column parities, multiple errors are required for the $\mathcal{M}_X^L$ to fail.

In estimating the failure probability for $\mathcal{M}_X^L$, $\mathcal{M}_{ZZ}^L$, or $\mathcal{M}_{ZZZ}^L$, we need to take into account possible faults in circuit locations preceding the CNOT gadget, which might cause these measurements to deviate from their ideal outcomes. In our fault-tolerant scheme, a logical CNOT gate is either preceded by teleportation into the code block or by another CNOT gate; in the worst case, each input block to the CNOT gadget is an output block from an immediately preceding CNOT gadget. Thus, in the preceding gadget, each block is prepared in the state $|+\rangle^L$, and then subjected to either $\mathcal{M}_{ZZ}^L$ and $\mathcal{M}_{ZZZ}^L$ (if it is the control block of the CNOT gadget) or just $\mathcal{M}_{ZZZ}^L$ (if it is the target block of the CNOT gadget).

A. Measurement failure

We will assume that failure of any of its preparation gadgets will cause the CNOT to fail, and we denote by P_{err}^* the failure probability for a logical measurement due to a cause other than failure of a preparation gadget. Thus, the failure probability for the CNOT can be bounded as

$$
\begin{aligned}
P_{\mathrm{err}}(\mathrm{CNOT}^L) \ \leq \ & P_{\mathrm{err}}^*(\mathcal{M}_{ZZ}^L) + P_{\mathrm{err}}^*(\mathcal{M}_{ZZZ}^L) + 2P_{\mathrm{err}}^*(\mathcal{M}_X^L) \\
& + 4P_{\mathrm{err}}(\mathcal{P}_{|+\rangle}) + P_{\mathrm{err}}(\mathcal{P}_{|ZZ\text{-cat}\rangle}) + P_{\mathrm{err}}(\mathcal{P}_{|ZZZ\text{-cat}\rangle}) \\
& + 2\left(P_{\mathrm{err}}(\mathcal{P}_{|ZZ\text{-cat}\rangle}) + P_{\mathrm{err}}(\mathcal{P}_{|ZZZ\text{-cat}\rangle}) \right).
\end{aligned}
\tag{7}
$$

The last term accounts for preparation errors that may have occurred in a $\mathcal{M}_{ZZ}^L$ or $\mathcal{M}_{ZZZ}^L$ immediately preceding the current CNOT gadget, acting on either one of the two input blocks to the current gadget. In the worst case, both incoming blocks were control blocks in the preceding CNOT gadgets, and were therefore subjected to both $\mathcal{M}_{ZZ}^L$ and $\mathcal{M}_{ZZZ}^L$. The coefficient of $P_{\mathrm{err}}(\mathcal{P}_{|+\rangle})$ is 4 rather than 2 for a similar reason — the logical measurement outcomes in the current CNOT gadget may differ from ideal outcomes due to errors in the preparation of logical blocks either in the current gadget or in one of the two preceding gadgets.

Before considering the probability of a preparation failure, let us estimate P_{err}^* for each of the three types of logical measurements. For an $\mathcal{M}_X^L$ measurement to fail, more than half of the m length-n columns must each have at least one Z error. In the worst case, each qubit in the block participates in $3r + 2r_+ + 2$ circuit locations: a preparation, a measurement, $2r_+$ CZ gates contained in the preparation of the logical $|+\rangle^L$ state, r CZ gates in the logical Z measurement in the current gadget, and $2r$ CZ gates in two logical Z measurements in the preceding gadget. Each of these locations could be faulty with probability $\varepsilon + \varepsilon'$, since either a diagonal of non-diagonal fault could produce a phase error.

For the moment, let us assume for simplicity that the cat state has length $2m$ for the ZZ^L measurement and length $3m$ for the ZZZ^L measurement, so that no ancilla qubit interacts with more than one data qubit. (We will consider the case where the cat state is shorter in Sec. V C.) Because the preparation of the cat state actually occurs *after* the ancilla interacts with the data, we do *not* need to worry about X errors arising in the cat state preparation propagating directly to the data. We *do* need to worry that the Pauli frame inferred from the "winning" syndrome after r' rounds of syndrome measurement actually differs from the ideal Pauli frame. However, assuming that the syndrome is decoded correctly (as it will be if the preparation succeeds), a Pauli frame X error afflicts a qubit only if at least one X error acted on that qubit during the multi-round syndrome measurement. The probability of such a Pauli-frame X error on any qubit in the cat state is therefore bounded above by $2r'\varepsilon'$, since two CZ gates act on each qubit in each round of syndrome measurement, and the syndrome measurement is repeated r' times in the cat state preparation. Because each row measurement is repeated r times, the total probability that a row measurement fails due to a Pauli-frame X error in a cat state is bounded above by $2rr'\varepsilon'$. Furthermore, in the worst case there are three logical Z^L measurements between the preparation and measurement of the block, in each of which a cat state interacts with the data.

Combining together the probability per qubit of a Z error caused directly by a fault with the probability of a Z error that propagates from the cat state, we conclude that the error probabilities for the X^L measurements can be bounded as

$$P_{\text{err}}^*(\mathcal{M}_X^L) \leq \binom{m}{\frac{m+1}{2}} \left[n(2r_+ + 3r + 2)(\varepsilon + \varepsilon') + 6nrr'\varepsilon' \right]^{(m+1)/2}. \tag{8}$$

As explained in Sec. IV G, there are no time steps in which "resting" qubits are subject to "storage errors."

For the $\mathcal{M}_{ZZ}^L$ to fail, more than half of the n weight-$2m$ row parity measurements must have errors. An X error acting on any one of the $2m$ qubits in the same row of the two blocks, either before or during the measurement, could flip the parity of the row. Before the $\mathcal{M}_{ZZ}^L$ is completed, each qubit participates in $2r_+$ CZ gates during the $\mathcal{P}_{|+\rangle}$ operation, r CZ gates during the $\mathcal{M}_{ZZ}^L$, and in the worst case another $2r$ CZ gates during the $\mathcal{M}_{ZZZ}^L$ and $\mathcal{M}_{ZZ}^L$ in the preceding CNOT gadget. Thus the probability per qubit of a non-diagonal error is $(2r_+ + 3r)\varepsilon'$. There might also be an X-type Pauli frame error in the $|+\rangle^L$ preparation, but we do not have to count that separately; if the preparation "succeeds," then a Pauli-frame X error afflicts a qubit in the block only if a non-diagonal fault damaged that same qubit during the preparation circuit.

In the absence of non-diagonal faults, the row parity measurement might still fail because at least one Z error acts on the cat state in each of $(r+1)/2$ of the repeated measurements. If the cat state has length p, then in each repetition of the measurement there are $2m + 2p + 2r'p$ circuit locations ($2m$ CZ gates coupling the ancilla to the data, plus p qubit preparations, p qubit measurements and $2pr'$ CZ gates for the cat state preparation). We conclude that the failure probability can be bounded as

$$P_{\text{err}}^*(\mathcal{M}_{ZZ}^L) \leq \binom{n}{\frac{n+1}{2}} \left[2m(2r_+ + 3r)\varepsilon' + \binom{r}{\frac{r+1}{2}} [(2m + 2p + 2pr')(\varepsilon + \varepsilon')]^{(r+1)/2} \right]^{(n+1)/2}. \tag{9}$$

Similarly,

$$P_{\text{err}}^*(\mathcal{M}_{ZZZ}^L) \leq \binom{n}{\frac{n+1}{2}} \left[3m(2r_+ + 4r)\varepsilon' + \binom{r}{\frac{r+1}{2}} [(3m + 2p + 2pr')(\varepsilon + \varepsilon')]^{(r+1)/2} \right]^{(n+1)/2}. \tag{10}$$

(The $\mathcal{M}^L_{ZZZ}$ is preceded by a $\mathcal{M}^L_{ZZ}$ acting on one of its input blocks inside the current CNOT gadget, plus additional Z-type logical measurements in the preceding CNOT gadgets acting on another one of its input blocks.)

B. Cat state preparation failure

Now it remains to estimate the probability of failure for the cat state and the logical state $|+\rangle^L$. Recall that the preparation fails if the syndrome measurement is faulty in every round that yields the "winning" syndrome outcome. The faults can be either non-diagonal or diagonal, but there must be at least two faults; otherwise only one syndrome bit would be wrong, the result would fail the parity test, and the syndrome would be rejected. (A round with a single non-diagonal fault might be accepted if the fault flips a syndrome bit while also simultaneously applying an X error to one of the cat state qubits, which flips a second syndrome bit measured in the following step. However, in this case the measured syndrome matches the actual cat state syndrome at the end of the round, and therefore is not counted as a failure.)

Suppose that the winning syndrome occurs t times, and that there are u additional rounds that each contain at least one fault (some of these rounds might be rejected). Suppose that non-diagonal faults occur in s of the r' rounds of syndrome measurement; these faults can alter the syndrome. There are $r' - t - u$ rounds without any faults, and the number of distinct syndromes detected in these rounds is at most $s + 1$.

Now we can use the pigeonhole principle to obtain a lower bound on t, expressed in terms of r', u, and s. There are at most $s + 1$ valid syndromes that can occur in syndrome measurement rounds that have no faults. Combining these with the winning syndrome, there are at most $s + 2$ possible syndromes that can occur in the $r' - u$ accepted rounds which either have no faults or produce the winning outcome. Of these $s + 2$ syndromes, the winning syndrome must occur at least as many times as any other syndrome; hence

$$t \geq \left\lceil \frac{r' - u}{s + 2} \right\rceil, \tag{11}$$

where $\lceil x \rceil$ denotes the smallest integer greater than or equal to x.

To bound $P_{\text{err}}(\mathcal{P}_{|\text{cat}\rangle})$, we sum over s and u in each cat state preparation step, estimating the number of possible fault histories using the upper bound Eq. (11) on t. In the first of the t winning rounds, a particular winning syndrome is found, which differs in at least two bits from the actual syndrome in the beginning of that round. Then this same syndrome is found again in each of the remaining winning rounds. The sum over all possible winning syndromes, weighted by their probabilities, is bounded above by the probability that at least two measured syndrome bits are faulty in the first of the t winning rounds. Each ZZ measurement is performed using one $|+\rangle$ preparation, two CZ gates, and one X measurement; therefore the probability of error in the measurement of a single syndrome bit is bounded above by $4\varepsilon + 2\varepsilon'$, and the probability that at least two syndrome bits are in error is bounded above by $\binom{p}{2}(4\varepsilon + 2\varepsilon')^2$.

In each of the s rounds that contain X errors, we must sum over all the possible X-error patterns that can occur in that round. The sum over all X-error patterns, weighted by the probabilities, is bounded above by the probability that at least one nondiagonal fault occurs in that round. Since the round contains $2p$ CZ gates, this probability is in turn bounded above by $2p\varepsilon'$.

Once the X-error pattern has been chosen in each round that contains X errors, we know the actual syndrome at the beginning of each round. And once the winning syndrome is chosen in the first winning round, we know which syndrome bits must have errors in each of the remaining $t-1$ winning rounds. If the cat state preparation fails, then by definition at least two syndrome bits have errors in each of these rounds; hence each winning round after the first has a probability weight bounded above by $(4\varepsilon + 2\varepsilon')^2$.

Taking into account that the rounds with non-diagonal faults can be chosen in at most $\binom{r'}{s}$ ways, and enumerating the ways to choose which t rounds produce the winning syndrome and which u additional rounds have faults, we obtain

$$P_{\text{err}}(\mathcal{P}_{|\text{cat}\rangle}) \leq nr \sum_{s=0}^{r'} \sum_{u=0}^{r'} \binom{r'}{s} \binom{r'}{u+t} \binom{u+t}{u} \binom{p}{2} (4\varepsilon + 2\varepsilon')^{2t} (4p\varepsilon + 2p\varepsilon')^u (2p\varepsilon')^s, \tag{12}$$

and similarly

$$P_{\text{err}}(\mathcal{P}_{|+\rangle}) \leq m \sum_{s=0}^{r_+} \sum_{u=0}^{r_+} \binom{r_+}{s} \binom{r_+}{u+t} \binom{u+t}{u} \binom{n}{2} (4\varepsilon + 2\varepsilon')^{2t} (4n\varepsilon + 2n\varepsilon')^u (2n\varepsilon')^s, \tag{13}$$

where t denotes $\lceil \frac{r'-u}{s+2} \rceil$. The prefactor nr in Eq. (12) arises because we use n length-p cat states in each mea-

surement, and each measurement is repeated r times. The prefactor m in Eq. (13) arises because the encoded state $|+\rangle^L$ is a product of m length-n cat states.

We should add another contribution to the failure probability, because if half of the qubits (or more) have X errors, we might decode the cat-state syndrome incorrectly. A syndrome of the repetition code points to two possible X error patterns, one low weight and one high weight. We always assume the low-weight interpretation is correct, so if the high-weight interpretation is actually correct, then a Pauli-frame error $X^{\otimes p}$ is applied to the cat state (or worse if the cat state qubits are used multiple times). The additional contribution, then, is bounded by the probability that each of $\lceil p/2 \rceil$ qubits in the cat state are each hit by X errors at least once in at least one of the cat state preparations. This upper bound is

$$ nr \binom{p}{\lceil p/2 \rceil} (2r'\varepsilon')^{\lceil p/2 \rceil} . \tag{14} $$

The Kraus operator arising from a general non-diagonal fault could be a coherent linear combination of Pauli operators; therefore the pattern of X errors acting on the cat state may be a coherent superposition of many possibility. For convenience and clarity, we assumed in the above discussion that non-diagonal errors are actually Pauli errors, so that X-error patterns can be assigned probabilities rather than amplitudes. But the argument also applies to more general non-diagonal faults. — by expanding each Kraus operator in terms of Pauli operators, and noting that all syndrome bits are measured in the X basis, we can bound the probability for each syndrome measurement outcome in terms of ε and ε' as above.

C. Shorter ancillas

Cat states are used in measurements of the weight-$2m$ logical operator ZZ^L and the weight-$3m$ logical operator ZZZ^L. In the derivation of Eq. (8) we assumed that the cat states have length $2m$ and $3m$ respectively, so that each of the cat state's qubits participates in just one CZ gate during the measurement. But we have found that the CNOT gadget may be more reliable if a shorter cat state is used instead.

If the length of the cat state is at least m, then each cat-state qubit interacts with only one qubit in each encoded block. But if the length is $p < m$, then a single cat-state qubit may interact with as many as

$$ R = \lceil m/p \rceil \tag{15} $$

data qubits in a single block, and hence an X error on that cat-state qubit may propagate to produce Z errors in R distinct columns of one block. If k of the cat state's p qubits have X errors, these could generate Z errors in kR columns of the block; thus an X^L measurement on the block could fail if there are

$$ L(k) = (m+1)/2 - Rk \tag{16} $$

additional columns that each contain at least one Z error.

In each measurement, each data block interacts with the ancilla for R consecutive time steps, where in each step at most p CZ gates act on the data block. The repeated measurements can be staggered as described in [8], to avoid time steps in which data qubits are idle.

We use the index $i \in \{1, 2, 3, \ldots, p\}$ to label the positions of the p qubits in the cat state, and design the measurement circuits so that the cat-state qubit i interacts with the same set of data qubits in each one of the measured code blocks. Since the ZZ^L measurement is performed on all n rows of each block, and also repeated r times, inside a $\mathcal{M}^L_{ZZ}$ there are at most $2nrR$ CZ gates that act on a cat-state qubit at position i (not counting CZ gates in the cat state preparation step), any of which could have a non-diagonal fault. Similarly, inside a $\mathcal{M}^L_{ZZZ}$ there are at most $3nrR$ CZ gates that act on a cat-state qubit at position i.

The $\mathcal{M}^L_X$ inside a CNOT gadget is preceded by $\mathcal{M}^L_{ZZZ}$ or $\mathcal{M}^L_{ZZ}$ within the CNOT, and in the worst case by both $\mathcal{M}^L_{ZZZ}$ and $\mathcal{M}^L_{ZZ}$ in the preceding gadget, so at most there are $8nrR$ CZ gates in these measurements where a non-diagonal fault could disturb cat-state qubit i. An X error acting on cat-state qubit i might also occur because of an error in the winning syndrome in one of the cat-state preparation steps, which can occur only if a faulty CZ gate acts on qubit i at least once during the r' repeated syndrome measurements. Noting that a damaged cat-state qubit in position i could cause Z errors in R columns of the encoded block, and that k damaged qubits in the length-p cat state can be chosen in $\binom{p}{k}$ ways, we conclude that if length-p cat states are used then Eq. (8) should be replaced by

$$ P^*_{\text{err}}(\mathcal{M}^L_X) \leq \sum_{k=0}^{k_{\max}} \binom{p}{k} [8nrR\varepsilon' + 6nrr'\varepsilon']^k \binom{m}{L(k)} [n(2r_+ + 3r + 2)(\varepsilon + \varepsilon')]^{L(k)} , \tag{17} $$

where

$$ k_{\max} = \lceil (m+1)/2R \rceil . \tag{18} $$

VI. GEOMETRICALLY LOCAL CIRCUITS

In our analysis so far, we have assumed that the two-qubit CZ gates can be performed on any pair of qubits,

with an error rate independent of the distance between the qubits. Now we will consider how the analysis is modified if CZ gates can be performed only on neighboring pairs of qubits.

There are a variety of possible architectures for fault-tolerant quantum computing using Bacon-Shor codes. To be concrete, we will consider an effectively one-dimensional arrangement, in which logical qubits are encoded in a ribbon of physical qubits with constant width in the vertical direction and length in the horizontal direction proportional to the total number of logical qubits. Each logical qubit lives in a "bi-block" — a pair of Bacon-Shor code blocks, one storing the data processed by the computation, and the other used as an auxiliary block for teleporting gates. If desired, the data block can be shuffled back and forth between the left and right sides of the bi-block via a "one-bit-teleportation" circuit consisting of a $|+\rangle^L$ preparation, a ZZ^L measurement, and an X^L measurement, as in Fig. 4.

Each $n \times m$ logical block is interlaced with an $n \times (2m-1)$ array of ancilla qubits, as in Fig. 7. Of the $2m-1$ ancilla qubits in a row, $m-1$ are used to read out the outcome of a ZZ measurement performed on the qubit's two horizontal neighbors, preparing an m-qubit cat state in the row. In addition, there is a column of n ancilla qubits at the boundary between two adjacent blocks, allowing the cat states to be extended to length $2m$ or $3m$ as needed for the measurements $\mathcal{M}^L_{ZZ}$ and $\mathcal{M}^L_{ZZZ}$ in the CNOT gadget. In the $|+\rangle^L$ preparation gadget, $n-1$ ancilla qubits in a column are used to read out the outcome of a ZZ measurement on the qubit's two vertical neighbors, preparing an n-qubit cat state in a column.
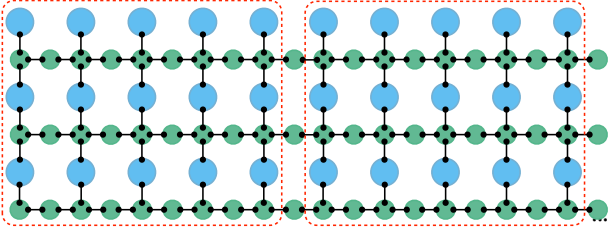

FIG. 7: (Color online) Arrangement of Bacon-Shor qubits for geometrically local computation for 3×5 code block. Large blue circles indicate data qubits, small green circles ancilla qubits. Solid lines indicate locations between which CZ gates may be performed. Dotted lines separate qubits into code blocks. The ribbon may be extended arbitrarily far to the right and left with additional code blocks.

A CNOT gate is performed on a neighboring pair of logical bi-blocks using one of the variants of the CNOT gadget depicted in Fig. 5. Let us denote the bi-block on the left as bi-block 1 and the bi-block on the right as bi-block 2. If the data in bi-block 1 is on the right side of the bi-block and the data in bi-block 2 is on the left side, then we may use the CNOT gadget design in Fig. 5a

to perform a CNOT gate in either direction (*i.e.* with either logical qubit as the control and the other as the target) using cat states of length $2m$ and $3m$, with the output data from the gate appearing on the left side of bi-block 1 and the right side of bi-block 2. Otherwise, we may use one of the alternative gadget designs discussed in Fig. 5 to deal with the cases where the data in bi-block 1 is on the left side and/or the data in bi-block 2 is on the right side. Whichever design we choose, the CNOT gadget flips the data from one side to the other in both bi-blocks, and only the $\mathcal{M}^L_{ZZZ}$ measurement reaches across the boundary between the two bi-blocks.

Note that with this method for performing a logical CNOT gate using geometrically local gates there is no need to swap the positions of pairs of physical qubits. The CZ gate is the only two-qubit gate used at the physical level, so that the assumption of highly biased noise remains physically plausible. Bacon-Shor blocks can be swapped using circuits of logical CNOT gates.

To complete a universal gate set we also need to be able to inject noisy non-Clifford states into code blocks which can then be purified by state distillation. This state injection can be performed using one-bit teleportation — the state $|\psi\rangle$ is prepared using a single physical qubit, while a logical block is prepared in the state $|+\rangle^L$, then the weight-$(m+1)$ operator $Z^L Z$ is measured on the single qubit and a single row of the block, and finally the single qubit is measured in the X basis. This procedure prepares the block in the state $|\psi\rangle^L$, up to a logical Pauli operator known from the measurement outcomes. To perform this task in a bi-block, we prepare $|+\rangle^L$ on (say) the left side of the bi-block, and build a length-$(m+1)$ cat state using m of the ancilla qubits that accompany a row of the block on the left side, and one additional ancilla qubit on the right side. The single-qubit state $|\psi\rangle$ is prepared using one physical qubit on the right side of the bi-block which is adjacent to the $(m+1)$st qubit of the cat state, allowing the measurement of $Z^L Z$ to be complete using local CZ gates.

If we wish to build a two-dimensional architecture, we may stack horizontal ribbons of logical qubits on top of one another. To perform CNOT gates on pairs of logical bi-blocks that are stacked vertically, we need to be able to perform $\mathcal{M}^L_{ZZZ}$ using cat states shared by vertically stacked blocks. For this purpose we can use ancilla qubits in neighboring bi-blocks to establish a "channel" linking the cat states neighboring corresponding rows of the vertically stacked blocks.

We note that in the case of unbiased noise, a two-dimensional architecture might be realized more simply, using symmetric Bacon-Shor codes. In that case we can build a CNOT gate using either the gadget in Fig. 5 or its Hadamard dual, with X and Z interchanged. The primal gadget uses horizontal cat states to execute CNOT gates on horizontally neighboring blocks, while the dual gadget uses vertical (dual) cat states to execute CNOT gates on vertically neighboring blocks.

For the geometrically local case, our analysis needs to

be modified in several ways. As already indicated, we will not be able to choose the length p of the cat state to optimize the gadget. Rather, the locality constraint requires that the length of the cat state match the weight of the measured operator ($p = 2m$ for $\mathcal{M}_{ZZ}^L$ and $p = 3m$ for $\mathcal{M}_{ZZZ}^L$). Another change is that we must now consider the consequences of storage errors acting on idle qubits. Previously, we assumed that the off-line preparation of cat states can be scheduled so that the cat states are always available immediately when needed. But now there is only one set of m ancilla qubits to accompany each row of a data block. Each time a cat state is measured we must prepare a new one, and the preparation involves a syndrome measurement repeated r' times. In each round of syndrome measurement, all qubits except the ones at the ends of the cat state participate in two ZZ measurements, and because the same ancilla qubits are used for two different ZZ measurements, four time steps are required for each one (a preparation, two CZ gates, and an X measurement). Therefore, the data waits for $8r'$ time steps before the cat state is ready to interact with the

data. While the data qubits wait for the cat state to be ready, they may be subject to storage errors. We must estimate how these storage errors contribute to the failure probability of the logical measurements in the CNOT gadget.

We will denote by ε_s the probability per time step of a diagonal storage error and by ε_s' the probability per time step of a non-diagonal storage error. How many storage steps are included depends on the gadget design, which in turn depends on whether the data is on the left or right side of the incoming bi-blocks. In the worst case, the measurement $\mathcal{M}_X^L$ is performed on a block that undergoes four logical Z-type measurements between its preparation and measurement, two in the current gadget plus another two in the immediately preceding gadget, and for each of these measurements the data waits for $8r'$ steps as the cat state is prepared. If each Z-type logical measurement is repeated r times, then the total contribution to the Z error probability per qubit due to the identity gates in the circuit is $32rr'(\varepsilon_s + \varepsilon_s')$. Therefore, Eq. (8) is modified to become

$$P_{\mathrm{err}}^*(\mathcal{M}_X^L) \le \binom{m}{\frac{m+1}{2}} [n(2r_+ + 3r + 2)(\varepsilon + \varepsilon') + 32nrr'(\varepsilon_s + \varepsilon_s') + 8nrr'\varepsilon']^{(m+1)/2}. \tag{19}$$

Furthermore, a non-diagonal storage fault on any of the $2m$ data qubits can flip the outcome of $\mathcal{M}_{ZZ}^L$. This fault can occur in any of the $8r'$ time steps while a cat state is prepared in either the $\mathcal{M}_{ZZ}^L$ or one of the preceding measurement. Similarly, a non-diagonal storage fault

on any of the $3m$ data qubits can flip the outcome of $\mathcal{M}_{ZZZ}^L$, where his fault can occur in any of the $8r'$ time steps while the cat state is prepared in either the $\mathcal{M}_{ZZZ}^L$ or one of the preceding measurements. Thus Eq. (9,10) are replaced by

$$P_{\mathrm{err}}^*(\mathcal{M}_{ZZ}^L) \le \binom{n}{\frac{n+1}{2}} \left[2m(2r_+ + 3r)\varepsilon' + 2m(24rr'\varepsilon_s') + \binom{r}{\frac{r+1}{2}}[(2m + 2p + 2pr')(\varepsilon + \varepsilon')]^{(r+1)/2} \right]^{(n+1)/2},$$

$$P_{\mathrm{err}}^*(\mathcal{M}_{ZZZ}^L) \le \binom{n}{\frac{n+1}{2}} \left[3m(2r_+ + 4r)\varepsilon' + 3m(32rr'\varepsilon_s') + \binom{r}{\frac{r+1}{2}}[(3m + 2p + 2pr')(\varepsilon + \varepsilon')]^{(r+1)/2} \right]^{(n+1)/2}. \tag{20}$$

Note that the number of storage locations in the circuit might be reduced by adding additional ancilla sites, so that more measurements can be performed in parallel, and/or by combining together measurement locations with immediately following preparation locations.

Geometric locality also requires that we remove the redundant cat state measurements that "wrap around"

the code block; Thus there is no syndrome parity check; any syndrome outcome is accepted, and a single diagonal fault can generate a syndrome error. There are also storage locations where the cat state qubits wait while ancilla qubits are prepared and measured. Because of these changes, Eq. (12,13) are replaced by

$$P_{\mathrm{err}}(\mathcal{P}_{|\mathrm{cat}\rangle}) \le nr \sum_{s=0}^{r'} \sum_{u=0}^{r'} \binom{r'}{s} \binom{r'}{u+t} \binom{u+t}{u} [4p\varepsilon + 2p\varepsilon' + 4p(\varepsilon_s + \varepsilon_s')]^{t+u} (2p\varepsilon' + 4p\varepsilon_s')^s \tag{21}$$

and

$$P_{\mathrm{err}}(\mathcal{P}_{|+\rangle}) \leq m \sum_{s=0}^{r_+} \sum_{u=0}^{r_+} \binom{r_+}{s} \binom{r_+}{u+t} \binom{u+t}{u} \left[4n\varepsilon + 2n\varepsilon' + 4n(\varepsilon_s + \varepsilon_s')\right]^{t+u} \left(2n\varepsilon' + 4n\varepsilon_s'\right)^s. \qquad (22)$$

VII. RESULTS FOR LOGICAL CSS GATES

A. Unrestricted gates

With our analytic upper bound on the effective error strength $\varepsilon^{(1)}$ of the CNOT gadget, we can find the optimum choice of the code block size $n \times m$ as well as the four CNOT parameters r, r', r_+ and p, for any choice of the error strengths ε and ε'. We did this by brute force search over the parameter space. Fig. 8 plots the results of this optimization for five choices of the bias $R = \varepsilon/\varepsilon'$.

These results were obtained by optimizing the size of a Bacon-Shor code block which is used by itself, rather than as part of a more complex concatenated code. Alternatively, we may consider using our Bacon-Shor gadgets at the bottom level of a concatenated coding scheme. In particular, we may estimate the accuracy threshold for biased noise achieved by such concatenated codes, as was done in Ref. [8] for the special case of the $n = 1$ Bacon-Shor code (i.e., the repetition code). We have extended the analysis of the accuracy threshold performed in [8] to more general Bacon-Shor codes. However, for bias above 10^3 we found no improvement over the threshold estimate in [8], because the $n = 1$ code turns out to provide the best threshold value. For values of the bias between 1 and 10^3, the optimal value of n turns out to be greater than 1, and for that range of the bias we found modest improvements in the accuracy threshold estimate compared to [8].

Using asymmetric Bacon-Shor codes, we can exploit the noise bias to improve the number of physical gates needed to construct a fault-tolerant CNOT gate with a specified logical error rate. Fig. 9 shows this overhead factor and also indicates the dimensions of the optimal code block. This plot includes a comparison with the performance of concatenated codes surveyed in [7] for the case of unbiased noise. For highly biased noise, asymmetric Bacon-Shor codes achieve a much lower logical error rate with a smaller number of physical two-qubit gates, compared to the performance of these previous constructions for unbiased noise.

B. Geometrically local gates

Optimizing the parameters for the case of geometrically local gates, assuming the storage error rate is negligible, we obtain the results displayed in Fig. 10 and Fig. 11. Enforcing locality significantly weakens the performance of our constructions — at a bias of 10^4 and dephasing error strength of $\varepsilon = 10^{-4}$, the optimal effec-

tive error strength we can achieve increases from 10^{-20} to 10^{-9}, while requiring roughly 8 times as many gates. Therefore, in the geometrically local case our estimated logical error rate for asymmetric Bacon-Shor codes and highly biased noise is roughly similar to what can be achieved using surface codes with a similar number of logical gates, while disregarding the bias [11, 12]. The most important reason that nonlocal gates prove to be advantageous is that if geometry is ignored then the cat state used in (for example) the measurement $\mathcal{M}_{ZZZ}^L$ can be chosen to be much shorter than the length-$3m$ cat state used in our geometrically local construction. The number of physical qubits per block required to achieve the optimal logical error rate stays relatively small, as indicated in Fig. 12.

Our upper bound on the logical error rate is likely to be far from optimal, and we also expect that the performance could be substantially improved by using a more sophisticated (but harder to analyze) method like perfect matching for decoding the cat state syndrome history. Suitable lattice deformations may also make it more feasible to reduce the length of cat states substantially, yielding further enhancements in performance for the case of geometrically local gates. Such improvements might make asymmetric Bacon-Shor codes more competitive relative to surface codes.

C. Geometrically local gates and measurement bias

In some experimental settings measurements are noisier than gates, in which case we say the noise has "measurement bias." We may consider a noise model in which the dephasing error rate ε in diagonal gates exceeds the rate ε' for nondiagonal faults, while in addition the error rate $\varepsilon_{\mathcal{M}}$ for single-qubit measurements exceeds ε. (We also continue to assume that ε is the error rate for single-qubit preparations.) As shown in Fig. 13, our geometrically local asymmetric Bacon-Shor code gadgets are somewhat robust against increasing measurement bias, because our gadgets contain considerably more gates than measurements (or preparations). Relative to the case $\varepsilon_{\mathcal{M}} = \varepsilon$, the performance of the logical CNOT gate is not much affected as the measurement bias $\varepsilon_{\mathcal{M}}/\varepsilon$ rises to about 5. In contrast, surface-code gadgets, which contain a higher number of measurements relative to the number of gates, are more sensitive to measurement bias.
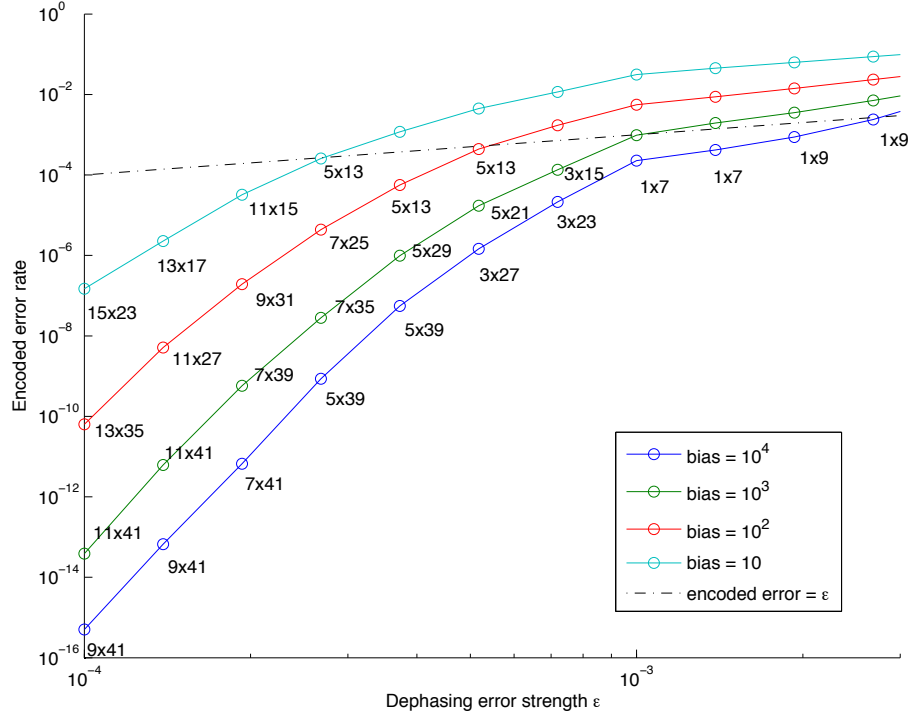
FIG. 8: (Color online) Optimal CNOT logical error rate versus physical error rate for various values of the bias. Labels represent the $n \times m$ dimensions of the Bacon-Shor code block. For blocks with a single row, one can use the code studied in [8] which is a special case of the asymmetric Bacon-Shor code.

VIII. STATE INJECTION

Having analyzed the performance of our fault-tolerant CSS gates, we now turn to the state injection and distillation protocols needed to complete a universal set of fault-tolerant gates. An arbitrary single-qubit state $|\psi\rangle$ can be injected into the Bacon-Shor block using the "one-bit teleportation" circuit depicted in Fig. 4. Here a logical qubit is prepared in the state $|+\rangle^L$, the two-qubit measurement $\mathcal{M}_{ZZ}^L$ is performed jointly on an unprotected qubit and the Bacon-Shor block using a length-$(m+1)$ cat state, and finally the X-basis measurement $\mathcal{M}_X$ is performed on the unprotected qubit. To estimate the error in the state injection step, we should consider all the ways in which the outcomes of the measurements $\mathcal{M}_{ZZ}^L$ and $\mathcal{M}_X$ might deviate from their ideal values.

In the state distillation circuit, this injection step is directly followed by a CNOT gate. Hence some sources of error in the teleportation circuit need not be attributed to the state injection step, as they are already included in our error estimate for the CNOT gate that follows. In particular, the possibility of failure in the preparation of $|+\rangle^L$ or a cat state is included in Eq.(7), and conservatively at that, since we assumed there that the cat states had length $2m$ or $3m$ rather than $m+1$. Here we assume that the number of repetitions r' of the cat state syndrome measurement, and the number of syndrome measurement repetitions r_+ in the $|+\rangle^L$ state preparation,

match the number of repetitions in the following CNOT gate. If so, we may bound the injection error probability $P_{\mathrm{err}}(\mathcal{J})$ by

$$P_{\mathrm{err}}(\mathcal{J}) \leq \varepsilon_\psi + P_{\mathrm{err}}^*(\mathcal{M}_{ZZ}^L) + P_{\mathrm{err}}^*(\mathcal{M}_X), \qquad (23)$$

where ε_ψ bounds the probability of an error in the preparation of the unprotected state $|\psi\rangle$ (we presume that this error is not necessarily diagonal in the Z basis).

We assume that $\mathcal{M}_{ZZ}^L$ is repeated r times to improve its reliability. In the first measurement we can time the preparation of the unprotected qubit and the cat state so that there are no storage errors prior to step in which the cat state interacts with the data. But in subsequent measurements storage errors on the data may accumulate during the $8r'$ time steps while the cat state is prepared for the next round. To estimate $P_{\mathrm{err}}^*(\mathcal{M}_X)$, we note that each of the $m+1$ cat-state qubits is acted upon by two CZ gates during each of r' rounds of cat-state syndrome measurement, where a nondiagonal fault in any of these gates could result in a Pauli-frame X error. Furthermore, a storage fault acting on the unprotected qubit, a fault in final X measurement of the unprotected qubit, or a fault in the CZ gate that couples the unprotected qubit to the cat state could cause $\mathcal{M}_X$ to fail. Therefore we find
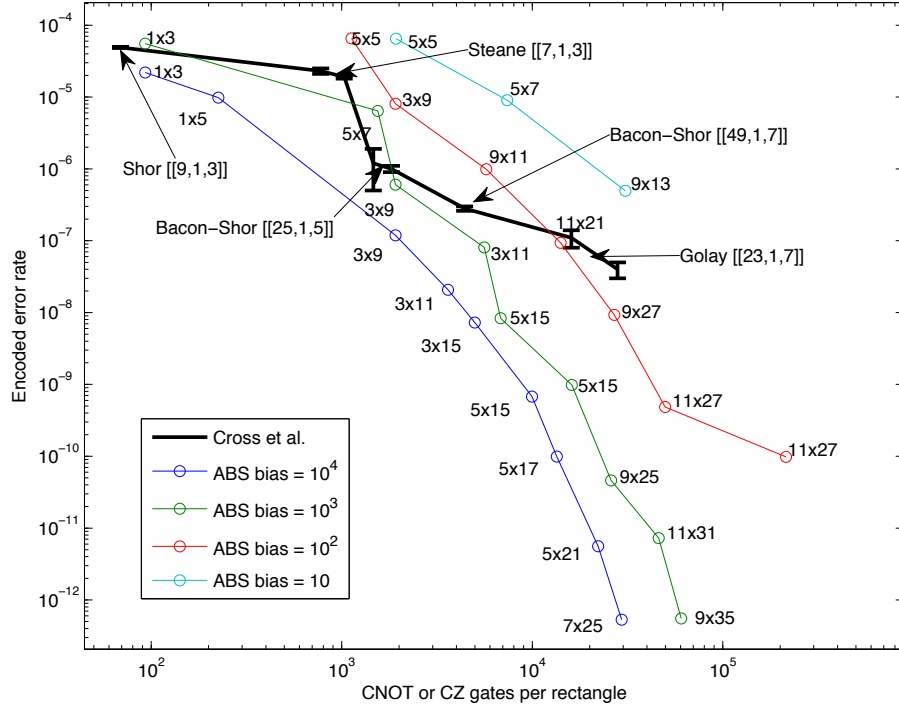
FIG. 9: (Color online) Optimal CNOT logical error rate versus required number of physical gates for $n \times m$ asymmetric Bacon-Shor codes, for a physical error rate $\varepsilon = 10^{-4}$ at various values of the bias, plotted alongside results for codes studied in [7] with no bias.
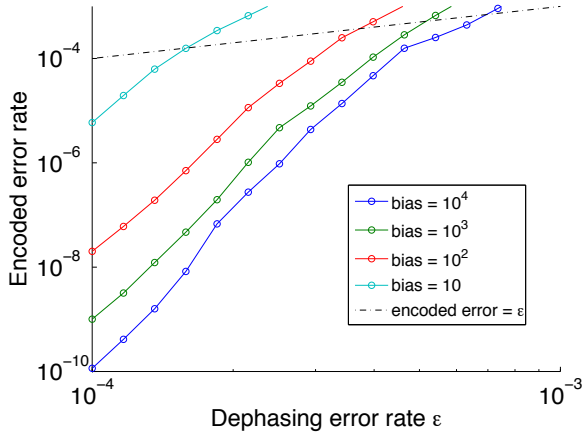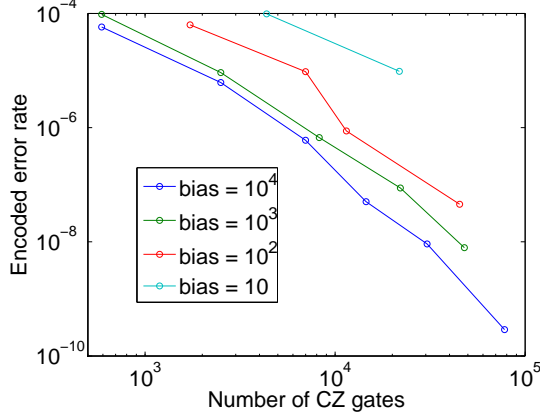

FIG. 10: (Color online) Optimal CNOT logical error rate versus physical error rate for various values of the bias, where two-qubit physical gates are assumed to be geometrically local.

$$P_{\mathrm{err}}^{*}(\mathcal{M}_X) \leq (r+1)(\varepsilon + \varepsilon') + 8(r-1)r'(\varepsilon_s + \varepsilon_s') + 2(m+1)rr'\varepsilon'. \qquad (24)$$

A nondiagonal error acting on any data qubit can flip the outcome of $\mathcal{M}_{ZZ}^{L}$. Otherwise, a diagonal fault in each of

at least $(r+1)/2$ measurement rounds could cause the $\mathcal{M}_{ZZ}^{L}$ to fail. Since in each round there are $m+1$ gates

coupling the qubits to the data, $(m + 1)r'$ CZ gates in the cat state preparation, as well as $m + 1$ single-qubit preparations and measurements, we find

FIG. 11: (Color online) Optimal CNOT logical error rate versus required number of physical gates for geometrically local asymmetric Bacon-Shor codes at various values of the bias, for a physical error rate $\varepsilon = 10^{-4}$

$$P^*_{\text{err}}(\mathcal{M}^L_{ZZ}) \leq r\varepsilon' + 8r'(r - 1)\varepsilon'_s + m(2r_+ + r)\varepsilon' + \binom{r}{\frac{r+1}{2}}[(m + 1)(r' + 3)(\varepsilon + \varepsilon')]^{(r+1)/2}. \tag{25}$$

Fig. 14 bounds the injection error probability assuming the values of m, r and r' chosen to optimize the geometrically local CNOT gadget. For large biases the injection error probability is around 1%, well below the thresholds for successful distillation, as will be seen in Sec. IX.

By choosing $r = 1$ (no repetition of the ZZ^L measurement), we can avoid storage locations, and the leading contribution to $P_{\text{err}}(\mathcal{J})$ linear in ε (assuming $\varepsilon' \ll \varepsilon$) is

$$P_{\text{err}}(\mathcal{J}) = 3\varepsilon + (m + 1)(r' + 3)\varepsilon + \cdots . \tag{26}$$

Increasing to $r = 3$ entails increasing the sensitivity to storage errors, while providing better protection against diagonal errors in CZ gates:

$$P_{\text{err}}(\mathcal{J}) = 5\varepsilon + 16r'\varepsilon_s + \cdots , \tag{27}$$

which might be a significant improvement if $\varepsilon_s \ll \varepsilon$.

The probability of error in state injection depends on m and also on r', the number of times the cat-state syndrome measurement is repeated, which we assume matches the value of r' in the CNOT^L gate that follows the state injection step. We note, though, that the value of r' might increase gradually as state injection proceeds. If there are multiple rounds of state distillation, we might be willing to accept a larger CNOT^L error rate $P_{\text{err}}(\text{CNOT}^L)$ in early rounds where the error in the distilled state is higher, with $P_{\text{err}}(\text{CNOT}^L)$ declining in later rounds as the state's purity improves. Adjustment of the Bacon-Shor block size, and hence of $P_{\text{err}}(\text{CNOT}^L)$, is easy to incorporate in our circuit constructions, as there is no need for the control and target blocks in the CNOT^L circuit to be of equal size. Using a smaller code in early rounds may save on overhead, but more importantly reducing the value of m and r' used in the first round improves the error $P_{\text{err}}(\mathcal{J})$ of the initially
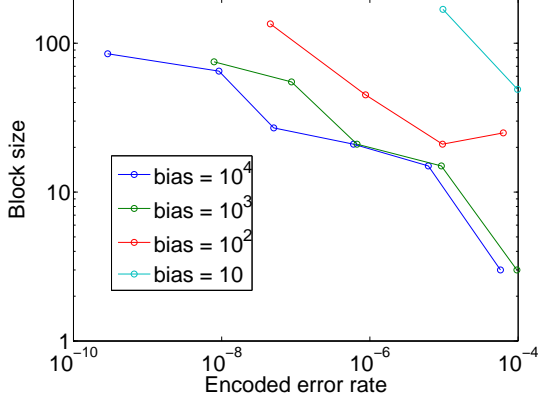
FIG. 12: (Color online) Bacon-Shor block size vs optimal (in number of gates) CNOT logical error rate for geometrically local asymmetric Bacon-Shor codes at various values of the bias, for a physical error rate $\varepsilon = 10^{-4}$.

injected state. However, for simplicity, we did not invoke this strategy of gradually increasing the block size in the analysis of state distillation reported in Sec. IX.

IX. STATE DISTILLATION

So far we have seen how, using asymmetric Bacon-Shor codes, to perform fault-tolerant encoded versions of the operations in the "CSS set" $\mathcal{G}_{\mathrm{CSS}}^{L}$: the CNOT gate, preparations of the encoded states $|0\rangle^{L}$ and $|+\rangle^{L}$, and the measurements $\mathcal{M}_{X}^{L}$, $\mathcal{M}_{Z}^{L}$. To perform fault-tolerant universal quantum computation, we will need in addition to prepare high-fidelity encoded versions of the states

$$|+i\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + i|1\rangle\right),$$

$$|T\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + e^{i\pi/4}|1\rangle\right). \tag{28}$$

Using a $|+i\rangle$ ancilla state and CSS operations, we can teleport the Clifford group gates $Q = \exp\left(i\frac{\pi}{4}X\right)$ and $S = \exp\left(-i\frac{\pi}{4}Z\right)$, which suffice for generating the full Clifford group. Using a $|T\rangle$ ancilla, the S gate, and CSS operations, we can teleport $T = \exp\left(-i\frac{\pi}{8}Z\right)$, completing
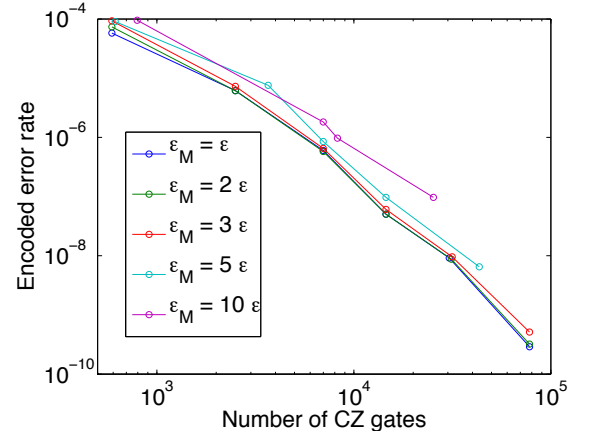

FIG. 13: (Color online) Optimal CNOT logical error rate versus required number of gates for geometrically local asymmetric Bacon-Shor codes, assuming a physical error rate $\varepsilon = 10^{-4}$, bias 10^{4}, and various rates of the measurement error rate $\varepsilon_{\mathcal{M}}$.

a universal gate set.

To prepare these encoded ancilla states, we first prepare noisy versions of the encoded states, and then use a distillation protocol to generate the needed high-fidelity versions of these states [16]. To distill the $|+i\rangle$ ancillas we need only CSS operations, and can use a circuit based on the [[7,1,3]] Steane code; it takes 7 noisy $|+i\rangle$ $(|-i\rangle)$ ancillas and produces one clean $|-i\rangle$ $(|+i\rangle)$ ancilla. The $|T\rangle$ distillation protocol uses both CSS operations and $|+i\rangle$ ancillas and is based on Reed-Muller codes [16]; it produces one clean $|T\rangle$ ancilla from 15 noisy ones. Other, more efficient, distillation protocols have been proposed recently [18–20].

In the original analysis [16] of the performance of these state distillation protocols, it was assumed that the CSS operations were perfect. Here we want to consider the case in which the CSS operations, protected by an optimally chosen Bacon-Shor code, are themselves noisy. We provide more details about our analysis in [21]; state distillation using noisy CSS operations has also been discussed previously in, for example, Refs. [11, 12, 22].

When the CSS operations are perfect, we can attain ancilla states with arbitrarily low error by performing a sufficient number of rounds of distillation, assuming that the fidelity of the initial ancilla with the desired ideal state exceeds a threshold value. However, when the CSS operations are noisy, the error rate $\varepsilon_{\mathrm{CSS}}$ for the CSS operations sets a nonzero floor on the error in the output ancilla states — once this floor is reached, further rounds of distillation produce no further improvement, because the error in the output is dominated by the errors in the distillation circuit rather than the errors in the input ancilla states. We have designed fault-tolerant distillation circuits that minimize this noise floor, and have found the corresponding lower bounds on the output error for distillation of both $|+i\rangle$ and $|T\rangle$ states [21]. For $|+i\rangle$ dis-
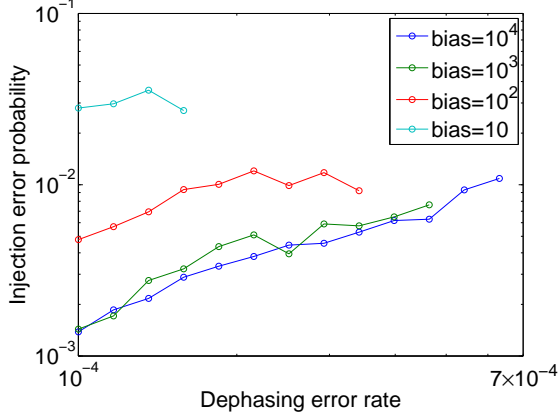
FIG. 14: (Color online) Injection error probability for the optimal geometrically local Bacon-Shor codes found in Fig. 10, with the probability of an error on the unprotected state $|\psi\rangle$, ε_ψ, set to be equal to the dephasing error probability ε.

FIG. 15: (Color online) Output error probability after n levels of $|T\rangle$ distillation, starting from $\varepsilon_{\text{in}} = 0.15$, for various values of ε_{css}. The $|+i\rangle$ ancillas are assumed to be maximally distilled, so that $\varepsilon_{|+i\rangle} = 4\varepsilon_{\text{css}}$. Eventually the output probability levels off to around $8\varepsilon_{\text{css}}$.

tillation, the best achievable error rate for our procedure is

$$\varepsilon_{|+i\rangle} \geq 4\varepsilon_{\text{css}} + O(\varepsilon_{\text{css}}^2), \tag{29}$$

and for $|T\rangle$ distillation it is

$$\varepsilon_{|T\rangle} \geq 8\varepsilon_{\text{css}} + O(\varepsilon_{\text{css}}^2). \tag{30}$$

A single round of the noisy distillation protocol improves the ancilla error probability substantially, provided that the CSS error rate ε_{css} in the distillation circuit is small compared to the input error probability ε_{in} of the noisy ancilla states, and also that ε_{in} is safely below the (quite high) threshold error rate for the ideal distillation protocol. The convergence of the ancilla error rate to the floor $\approx 8\varepsilon_{\text{css}}$ is illustrated for $|T\rangle$ distillation in Fig. 15, for the case where the error rate of the input ancillas is $\varepsilon_{\text{in}} = 0.15$.

For initial ancillas with errors that are very close to the threshold of the ideal distillation protocol, it can in principle take many rounds of distillation to reach the floor set by the CSS error rate. On the other hand, if the initial ancillas are not too noisy, then it will take just a few rounds of distillation to reach this floor. In Fig. 16, the erro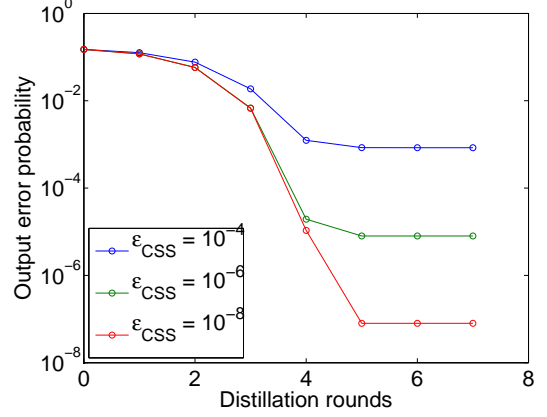r rate of the output ancilla is plotted as a function of the number of rounds of distillation, where the input error rate is the injection error probability estimated in Fig. 14. For the parameters shown, only two rounds of distillation suffice to reach the floor set by the logical CSS error rate. This relatively modest overhead cost for executing non-Clifford logical quantum gates is one of the advantages of our fault-tolerant scheme.

Acknowledgments

FIG. 16: (Color online) Output error probability after n levels of $|T\rangle$ distillation, with the values of ε_{in} and ε_{css} taken from the optimal values of Fig. 14, for $\varepsilon = 10^{-4}$ and various biases. Only two rounds of distillation are required in each case to reach the floor.

[1] P. Shor, "Fault-tolerant quantum computation," in *Foundations of Computer Science, 1996. Proceedings., 37th Annual Symposium on*, pp. 56–65, IEEE. 1996.

[2] D. Gottesman, "An introduction to quantum error correction and fault-tolerant quantum computation," in *Quantum Information Science and Its Contributions to Mathematics, Proceedings of Symposia in Applied Mathematics*, vol. 68, p. 13. 2009.

[3] P. W. Shor, "Scheme for reducing decoherence in quantum computer memory," *Phys. Rev. A* **52** no. 4, (Oct, 1995) R2493–R2496.

[4] D. Bacon, "Operator quantum error-correcting subsystems for self-correcting quantum memories," *Physical Review A* **73** no. 1, (Jan, 2006) 012340, `arXiv:quant-ph/0506023`.

[5] J. Napp and J. Preskill, "Optimal Bacon-Shor codes," *arXiv preprint arXiv:1209.0794* (2012) .

[6] P. Aliferis and A. Cross, "Subsystem fault tolerance with the Bacon-Shor code," *Physical review letters* **98** no. 22, (2007) 220502.

[7] A. W. Cross, D. P. DiVincenzo, and B. M. Terhal, "A comparative code study for quantum fault-tolerance," *ArXiv e-prints* (Nov., 2007) , `arXiv:0711.1556 [quant-ph]`.

[8] P. Aliferis and J. Preskill, "Fault-tolerant quantum computation against biased noise," *Phys. Rev. A* **78** no. 5, (Nov, 2008) 052331.

[9] P. Aliferis, F. Brito, D. P. Di Vincenzo, J. Preskill, M. Steffen, and B. M. Terhal, "Fault-tolerant computing with biased-noise superconducting qubits: a case study," *New Journal of Physics* **11** no. 1, (Jan., 2009) 013061, `arXiv:0806.0383 [quant-ph]`.

[10] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, "Topological quantum memory," *arXiv preprint quant-ph/0110143* (2001) .

[11] R. Raussendorf, J. Harrington, and K. Goyal, "Topological fault-tolerance in cluster state quantum computation," *New Journal of Physics* **9** (2007) 199.

[12] A. Fowler, A. Stephens, and P. Groszkowski, "High-threshold universal quantum computation on the surface code," *Physical Review A* **80** no. 5, (2009) 052312.

[13] I. Gourlay and J. F. Snowdon, "Concatenated coding in the presence of dephasing," *Phys. Rev. A* **62** no. 2, (Jul, 2000) 022308.

[14] Z. W. E. Evans, A. M. Stephens, J. H. Cole, and L. C. L. Hollenberg, "Error correction optimisation in the presence of X/Z asymmetry," *ArXiv e-prints* (Sept., 2007) , `arXiv:0709.3875 [quant-ph]`.

[15] A. M. Stephens, Z. W. E. Evans, S. J. Devitt, and L. C. L. Hollenberg, "Asymmetric quantum error correction via code conversion," *Phys. Rev. A* **77** no. 6, (June, 2008) 062335, `arXiv:0708.3969 [quant-ph]`.

[16] S. Bravyi and A. Kitaev, "Universal quantum computation with ideal Clifford gates and noisy ancillas," *Phys. Rev. A* **71** no. 2, (Feb, 2005) 022316.

[17] E. Knill, "Quantum computing with realistically noisy devices," *Nature* **434** (Mar., 2005) 39–44, `arXiv:quant-ph/0410199`.

[18] A. M. Meier, B. Eastin, and E. Knill, "Magic-state distillation with the four-qubit code," *ArXiv e-prints* (Apr., 2012) , `arXiv:1204.4221 [quant-ph]`.

[19] S. Bravyi and J. Haah, "Magic state distillation with low overhead," *ArXiv e-prints* (Sept., 2012) , `arXiv:1209.2426 [quant-ph]`.

[20] C. Jones, "Multilevel distillation of magic states for quantum computing," *ArXiv e-prints* (Oct., 2012) , `arXiv:1210.3388 [quant-ph]`.

[21] P. Brooks, "State distillation with noisy Clifford gates." Forthcoming.

[22] T. Jochym-O'Connor, Y. Yu, B. Helou, and R. Laflamme, "The robustness of magic state distillation against errors in Clifford gates," *Arxiv preprint arXiv:1205.6715* (2012) .