Limitations in Quantum Computing from Resource Constraints

Marco Fellous-Asiani[1], Jing Hao Chai[1,2] Robert S. Whitney[3], Alexia Auffèves[1], and
Hui Khoon Ng[4,2,5,*]

[1]*Institut Néel, Grenoble, France*
[2]*Centre for Quantum Technologies, National University of Singapore, Singapore*
[3]*Laboratoire de Physique et Modélisation des Milieux Condensés, Université Grenoble Alpes and CNRS,
B.P. 166, Grenoble 38042, France*
[4]*Yale-NUS College, Singapore*
[5]*MajuLab, International Joint Research Unit UMI, 3654, CNRS, Université Côte d'Azur, Sorbonne Université,
National University of Singapore, Nanyang Technological University, Singapore*

Fault-tolerant schemes can use error correction to make a quantum computation arbitrarily accurate, provided that errors per physical component are smaller than a certain threshold and independent of the computer size. However, in current experiments, physical-resource limitations such as energy, volume, or available bandwidth induce error rates that typically grow as the computer grows. We analyse how error correction performs under such constraints and show that the amount of error correction can be optimized, leading to a maximum attainable computational accuracy. We find this maximum for generic situations where noise is scale dependent. By inverting the logic, we provide experimenters with a tool for finding the minimum resources required to run an algorithm with a given computational accuracy. When combined with a full-stack quantum computing model, this provides the basis for energetic estimates of future large-scale quantum computers.

I. INTRODUCTION

With the advent of small-scale quantum computing devices from companies such as IBM and the myriad of software and hardware quantum startups, the interest in building quantum computers is at an all-time high. The latest declaration of quantum supremacy by Google [1] begs the question: How do we make our quantum computers more powerful? The answer is, of course, to have larger quantum computers. But larger also usually means noisier, with more fragile quantum components that can go wrong, leading to more computational errors. The way out of this conundrum is fault-tolerant quantum computation (FTQC), the only known route to scaling up quantum computers while keeping errors in check.

FTQC schemes have been known since the early days of the field [2–8] and are widely reviewed [9–12]. They remain an active field of research, especially in the context of surface codes: see, e.g., Refs. [13–15] or Ref. [16] for an older review. Underlying all FTQC schemes are basic assumptions about the nature of the quantum devices and the noise afflicting them. Many of these assumptions, laid down long before experimental devices came about, were based on general physical expectations not specific to any one implementation. As we learn more about the shape of quantum computers to come, it is important to revisit those assumptions, to update them to properly describe real devices, so that the schemes remain relevant to our progress toward large-scale useful quantum computers.

FTQC tells us how to improve computational accuracy by scaling up the size of the computer. The computational state is carried by more and more physical qubits, encoded using more and more powerful quantum error-correcting (QEC) codes, and thus capable of removing more and more errors. One key assumption is that the physical error probability η—the maximum probability that an error occurs in a physical qubit or gate—remains constant as the computer scales up. Then, so long as the error is below a certain threshold (typically, an error probability per gate of less than 10^{-4}), one can perform more accurate calculations by investing more physical resources to scale up the size of the computer (adding more qubits, gates, etc.). In principle, this can be repeated until computational errors are arbitrarily rare.
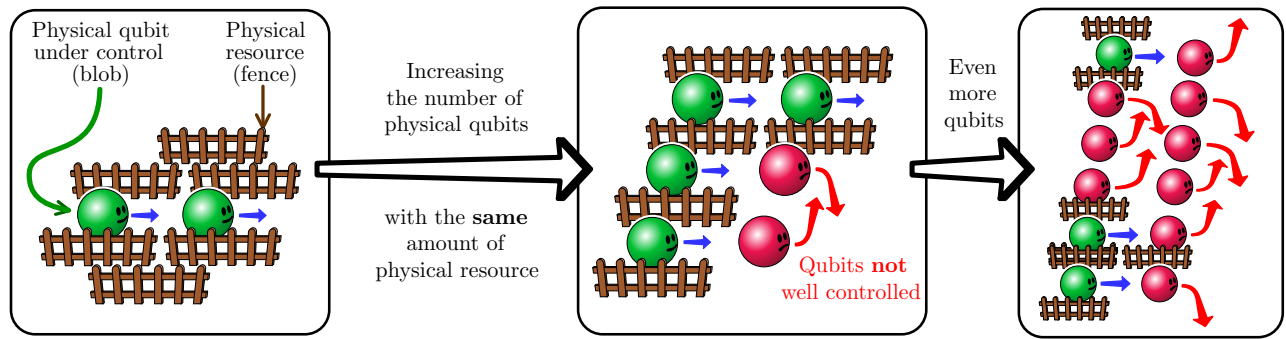
---

*huikhoon.ng@yale-nus.edu.sg

FIG. 1. How resource constraints can lead to increased computational errors. Each gate operation on a physical qubit (blob) requires a certain amount of physical resource (fence) for good control. If the number of physical qubits increases as the computer grows, without a proportionate increase in control resource, errors will increase.

Unfortunately, this key assumption is observed to be violated in most current quantum devices, where η is found to grow with scale. We argue that this is a consequence of resource limitations, so that the resources per physical component (qubit, gate, etc.) shrink as the scale of the computer grows, leading to more errors per gate operation (see Fig. 1). For instance, in ion-trap experiments, the gate fidelity drops rapidly if more and more ions are put into the same trap. This volume constraint is currently a technological bottleneck, even if it is motivating a push toward networked ion traps and flying qubits to communicate between traps (see, e.g., Ref. [17]). Another example is provided by qubits that are coherently controlled, by resonantly addressing their transition. Here, a constraint on the total available energy to perform gates results in lower gate fidelity [18]. Finally, a constraint on the available bandwidth makes the qubit transition frequencies closer and closer as the computer size grows, causing more and more crosstalk between qubits when performing gates [1]. These three typical examples lead to a scale-dependent noise.

If the physical error probability η is scale dependent, so that it grows as the computer scales up, we cannot expect quantum error correction to keep up with the rapid accumulation of errors, so it should come as no surprise that the standard threshold for fault tolerance [2–8]—and its generalizations to correlated and long-range noise [19–28]—no longer apply (see, e.g., Refs. [20,29]). This calls for a revisiting of our expectations for FTQC within this realistic context.

In this work, we examine the consequences on FTQC of growing physical error probability η as the computer scales up. The absence of a threshold means that arbitrarily accurate computation is unattainable but it does not mean that quantum error correction is useless. We find generic situations where a certain amount of error correction is good but too much is bad. Hence, the amount of error correction should be optimized, leading to a maximal

achievable computational accuracy. We provide experimenters with a methodology to estimate this maximum for a given scale-dependent noise and show the importance of adjusting the experimental design to control this scaling. Inverting the perspective allows us to estimate the minimum required resource cost to perform a computation with a given accuracy.

The standard theory of FTQC applies when noise does not grow with scale. It has to be replaced by our theory presented here in two generic situations: (1) if a technological bottleneck stops one from increasing the resources (energy, volume, bandwidth, etc.) in proportion to the number of physical components (qubits, gates, etc.) in the quantum computer; or (2) if one wants to minimize the resources consumed by the quantum computer. We give examples of both situations here; we use our theory to maximize the computational accuracy for the available resources in the first situation and we minimize the resources for a prescribed accuracy in the second.

Our paper is structured as follows. After recalling the basics of FTQC, we present our general strategy. We exemplify it with a toy model that captures the main features of FTQC in the presence of scale-dependent noise. We then focus on three physically motivated situations where resource constraints such as energy, volume, or bandwidth lead to scale-dependent noise and examine the feasibility of FTQC in the limit of large quantum computers. We finally provide first methodological steps toward minimizing the energetic costs to run an algorithm with a given accuracy. This suggests the possibility of a detailed energetic analysis for a full-stack quantum computer—which, however, goes beyond the scope of this paper.

II. ACCURATE QUANTUM COMPUTING

To be concrete, we examine the FTQC scheme of Ref. [7], built on the idea of concatenating a QEC code put forth in earlier works. This has formed the foundation of many subsequent FTQC proposals; our results are hence

applicable to those based on concatenated codes. Such schemes have more well-established and complete theoretical analyses than some of the more recent developments such as surface codes. Hence they are a good starting point for our investigation here.

Universal quantum computation in the scheme of Ref. [7] is built upon the seven-qubit code [30], using seven physical qubits to encode one (logical) qubit of information. We refer to the seven physical qubits used to encode the logical qubit as a "code block," and to gates on the logical qubit as "encoded gates." At the lowest level of protection against errors, which we refer to as "level-1 concatenation," each logical qubit is encoded using the seven-qubit code into one code block and every computational gate is done as an encoded gate on the code blocks. Every encoded gate is immediately followed by a QEC box, comprising syndrome measurements to (attempt to) correct errors in the preceding gate. Faults can occur in any of the physical components—physical qubits and gates—including those in the QEC boxes, so the error correction may not always successfully remove the errors. Faulty components in the QEC box may even add errors to the computer. A critical part of the construction of Ref. [7] is to ensure that the QEC boxes, even when faulty, do not cause or spread errors on the physical qubits in an uncontrolled manner provided that not too many faults occur, a realization of the notion of fault tolerance.

At level-1 concatenation, the ability of the code to remove errors is limited. The seven-qubit code ideally removes errors in at most one of the seven physical qubits in the code block. To increase the QEC power, we raise the concatenation level of the circuit: every physical qubit in the lower concatenation level is encoded into seven physical qubits; and every physical gate is replaced by its seven-physical-qubit encoded version, followed by a QEC box. In this manner, level-k concatenation is promoted to level-$(k + 1)$ concatenation, for $k = 0, 1, 2, \ldots$. The QEC ability of each level of concatenation increases in a hierarchical manner. For example, at level-2 concatenation, every logical qubit is stored in $7^2 = 49$ physical qubits organized into two layers of protection, with the topmost layer comprising seven blocks of seven physical qubits each. Each block of seven physical qubits is protected using the seven-qubit code; the seven blocks of qubits are themselves protected by QEC in the second layer. This logic extends to higher levels of concatenation.

The concatenation endows the overall computational circuit with a recursive structure (see Fig. 2), a crucial ingredient in the proof of the quantum accuracy threshold theorem. The increase in computational resource as the concatenation level grows is beneficial only if the increased noise due to the larger circuit is less than the increased ability to remove errors. This leads to the concept of a fault-tolerance threshold condition. The quantum accuracy threshold theorem gives a prescription for

increasing the accuracy of quantum computation with no more than a polynomial increase in resources, provided that the physical error probability is below a threshold level. Specifically, for the FTQC scheme of Ref. [7], the error probability per *logical* gate at level-k concatenation is upper bounded by

$$p^{(k)} = \frac{1}{B}(B\eta)^{2^k}. \tag{1}$$

Here, η is the *physical* error probability and $p^{(0)} = \eta$. B is a numerical constant, determined by the fault-tolerance scheme, that captures the increase in complexity (number of physical components) of the circuit used to implement a single logical gate as one increases k for increased protection. Equation (1) expresses quantitatively the idea of the accuracy threshold theorem: as long as

$$\eta < \frac{1}{B} \equiv \eta_{\text{thres}}, \tag{2}$$

$p^{(k)}$ decreases as k increases. Equation (2) is the threshold condition, i.e., the physical error probability η in the quantum computer has to be below the threshold level η_{thres} for FTQC to work. The number of physical gates in the circuit that implements the level-k logical gate is $G^{(k)} = A (A')^{k-1}$, where A' and A are integers given by circuit details; the well-known scheme of Ref. [7] has $A = 575$ and $A' = 291$ [31]. From A and A', one counts the number of fault locations, B. A simple overestimate of the integer B is $\binom{A}{2} \simeq 10^5$, with a more careful counting giving an improved value of $B \simeq 10^4$ [7].

The quantum accuracy threshold theorem [2–8] shows that a double-exponential decrease in $p^{(k)}$ with k can be achieved with only an exponential increase in resources,
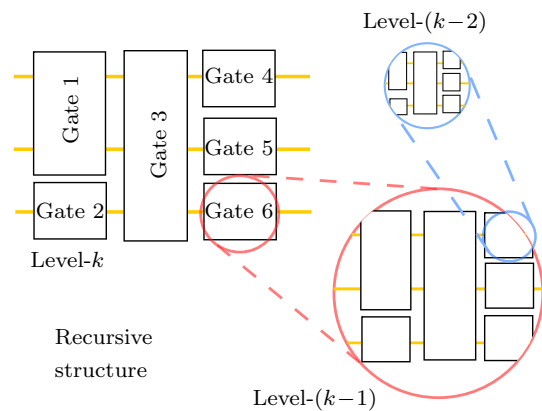

FIG. 2. The accuracy of a quantum computation can be increased by a FTQC scheme that makes use of concatenation and recursive simulation. Circuits are designed to be hierarchical, with high-level gate components built from lower-level components in a self-similar manner.

giving the no-more-than-polynomial increase in resource costs claimed in the accuracy threshold theorem. The theorem assumes that the value of η, the physical error probability, remains constant even as the level of concatenation k increases. However, as mentioned, as k increases and the physical size of the computer grows, current experiments suggest that η also increases. Our goal here is thus to examine how the conclusions on quantum accuracy are modified if η grows with k. Then it is intuitively clear that as k is increased in an attempt to reduce the logical error probability, the underlying noise per physical component increases to thwart that reduction. We will see that there is a maximum k beyond which further concatenation only serves to worsen the computational accuracy.

III. EFFECT OF SCALE-DEPENDENT NOISE

We examine the consequences of a k-dependent physical error probability $\eta^{(k)}$, illustrating it first with a toy model, before analyzing the more realistic situation where a constraint on the total resource available for the computation leads to a shrinking amount of resource per physical gate as the computer scales up. The general effect of a k-dependent physical error probability can be summarized in the schematic of Fig. 3. We also show how to obtain the maximum computational accuracy available for a given model of scale-dependent error.

A. Toy model

We first illustrate this with a simple model in which $\eta^{(k)} = \eta^{(0)}(1 + ck)$, for $k = 0, 1, 2, \ldots$, where $\eta^{(k)}$ is the physical error probability per gate in a computer large enough to perform level-k concatenation. Here, $c \geq 0$ and $\eta^{(0)} \geq 0$ are constants governed by the physical system in question. Although this is a toy model, one can think of it as the affine approximation of any $\eta^{(k)}$ function with weak k dependence, expanded about $\eta^{(k)} = \eta^{(0)}$ (see also the long-range noise with $z = d$ in Table I of the Supplemental Material [32]). For this $\eta^{(k)}$, Eq. (1) gives

$$p^{(k)} = \frac{1}{B}\left[B\eta^{(0)}(1 + ck)\right]^{2^k} = p_0^{(k)}(1 + ck)^{2^k}. \quad (3)$$

Here, we define

$$p_0^{(k)} \equiv \frac{1}{B}\left(B\eta^{(0)}\right)^{2^k}, \quad (4)$$

which would be the value of the error probability per logical gate if the error probability per physical gate were k independent. If $c = 0$, $p^{(k)}/p^{(k-1)} = p_0^{(k)}/p_0^{(k-1)} < 1$ as long as $\eta^{(0)} < 1/B$, as in Eq. (2); if $c > 0$, the multiplicative factor $(1 + ck)$ grows with k so, eventually, $p^{(k+1)} > p^{(k)}$ for k beyond some $k_{\max}$ value. Figure 4(a) shows an example of how $p^{(k)}$ varies as k increases, for different c
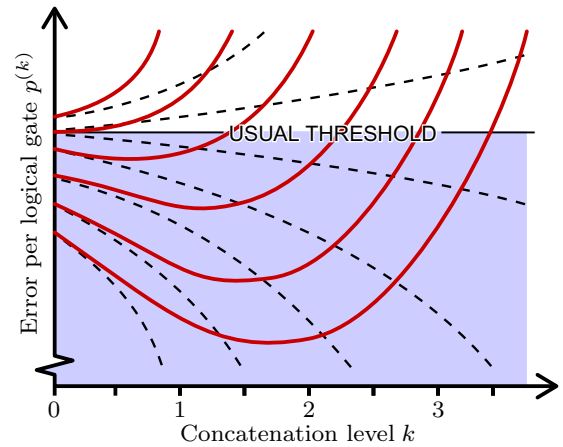

FIG. 3.   The black dotted curves are a schematic of the conventional situation, where the physical error probability η is independent of the scale—the concatenation level k— of the computer. The red solid curves are a schematic of the consideration in this work, where η grows with k (each red curve is for a different value of $p^{(0)} \equiv \eta$). If η is k independent, standard FTQC analysis says that the error per logical gate $p^{(k)}$ can be brought as close to 0 as desired by increasing k, provided that one starts below the threshold (solid horizontal line) at $k = 0$. If η depends on k, even if one starts below the threshold, $p^{(k)}$ eventually turns around for large enough k. $p^{(k)}$ cannot reach 0; there is a maximum concatenation level and a further increase in k only increases the logical error. All examples in this work have at most one minimum in each red curve but, in general, there can be multiple minima.

values. As long as $c > 0$, $p^{(k)}$ decreases (if at all) before rising again, above some $k_{\max}$ value.

Corresponding to this maximum useful level of concatenation, $k_{\max}$, is the minimum attainable error probability, $p_{\min} \equiv p^{(k_{\max})}$, giving the limit to computational accuracy attainable for given values of c and $B\eta^{(0)}$, quantities that give information about the noise scaling and the fault-tolerance overheads. Figure 4(b) shows the $k_{\max}$ values for different c and $B\eta^{(0)}$ values [cf. Fig. 3(b)]. Clearly, $k_{\max}$ decreases as c grows (stronger k dependence). Current experiments have $B\eta^{(0)} \gtrsim 1$; for example, the IBM Quantum Experience system has $\eta^{(0)} \gtrsim 10^{-3}$, giving $B\eta^{(0)} \gtrsim 10$ for $B = 10^4$. In near- to middle-term experiments, we expect $B\eta^{(0)}$ to not be far below 1, i.e., the error probability is just below the $c = 0$ threshold value [see Eq. (2)]. In this case, Fig. 4(b) suggests that one quickly loses the advantage of concatenating to higher levels even for small c values. In fact, for encoding to be helpful at all, i.e., for $k = 1$, we must have $p^{(1)} = B[\eta^{(0)}]^2(1 + c)^2 < p^{(0)} = \eta^{(0)}$, which in turn requires
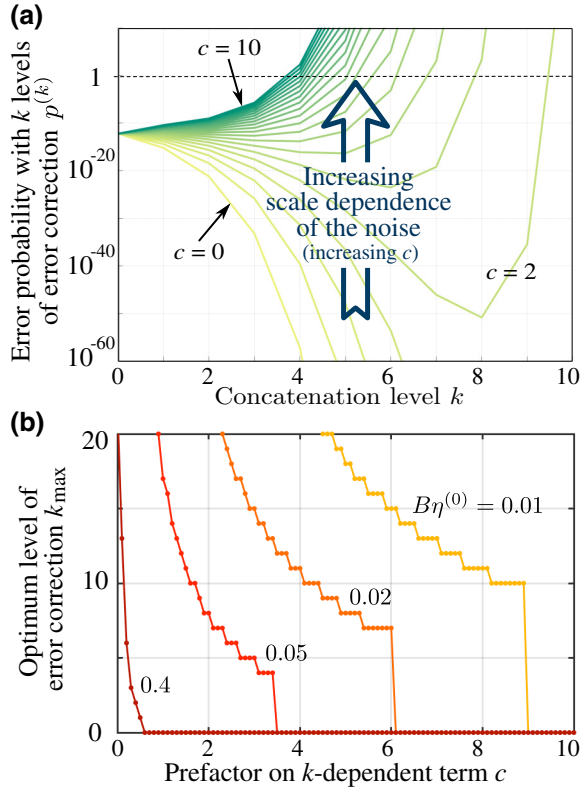
$$c < \frac{1}{\sqrt{B\eta^{(0)}}} - 1. \quad (5)$$

FIG. 4. (a) An example of how $p^{(k)}$ varies as the concatenation level k increases, for the affine model with $c = 0$ (lightest color), $0.5, 1, \ldots, 10$ (darkest color), $B = 10^4$, and $\eta^{(0)} = 5 \times 10^{-6}$, for which $B\eta^{(0)} = 0.05$, far below the threshold for $c = 0$. All curves, apart from that for $c = 0$, turn around for large enough k. (b) The maximum k value, $k_{\max}$, such that $p^{(k+1)} < p^{(k)}$, with $p^{(0)} \equiv \eta^{(0)}$, the unencoded error probability, for different values of $B\eta^{(0)} < 1$ and $c \in \{0, 0.1, 0.2, \ldots, 10\}$. In every case, $k_{\max}$ eventually falls to zero for large enough c, i.e., it is better to have no encoding.

If $B\eta^{(0)} = 0.8$, say, this amounts to the requirement that $c \lesssim 0.1$, so that a very weak dependence on k is necessary for even one level of encoding to help at all in reducing the error probability.

B. General case

More generally, consider the physical error probability growing as a monotonic function of k: $\eta^{(k)} = \eta^{(0)} f(k)$, with $f(k) \geq 1$ monotonically growing with k and $f(0) = 1$. Then, the error probability per logical gate is $p^{(k)} = p_0^{(k)} f(k)^{2^k}$, where Eq. (4) gives $p_0^{(k)}$. Treating k as a continuous variable, let us assume there is only one minimum, which we define as $k = k_{\text{st}}$ (with "st" for stationary point). Then the minimal attainable error will occur at an integer $k_{\max}$, which is one of the two integers nearest to the minimum k_{st}, so the minimal error will occur at $k_{\max} \leq k_{\text{st}} + 1$. If there are multiple minima, we define k_{st} as the minimum

with the largest k. *A priori*, we do not know which minimum will be the best but we still know that $k_{\max} \leq k_{\text{st}} + 1$. Combining this with a little algebra for k_{st} yields

$$k_{\max} < 1 + f^{-1}\left(\frac{1}{B\eta_0}\right), \quad (6)$$

with $f^{-1}(\cdot)$ the inverse of $f(\cdot)$. If $f(k \to \infty)$ is finite, $p^{(k)}$ can be made arbitrarily small only if $\eta^{(0)} < [Bf(k \to \infty)]^{-1}$. However, in many cases (such as the above toy model and the example in the following section), one has $f(k \to \infty) \to \infty$. Then, the minimum $p^{(k)}$ will occur at finite $k_{\max}$, no matter how small $\eta^{(0)}$ is. In this case, one can never attain arbitrarily small logical error probability by concatenating further.

IV. EXAMPLES OF RESOURCE CONSTRAINTS

We now give examples of how specific physical-resource constraints can lead to the scale-dependent noise discussed in the previous section. In the first example, the constraints lead to a scale-dependent *local noise* on each physical component, so the above theory applies directly. In the second example, the constraints lead to scale-dependent *crosstalk* between qubits, which can be mapped to the above theory, using a mapping in Refs. [20,25].

A. Resource constraints affecting local noise

This section considers local noise on each qubit $\eta^{(k)}$ scaling with a total number of physical components $N(k)$ that grows exponentially with k. Let us assume that adding a level of concatenation involves replacing each physical component by D physical components, so $N(k) = D^k$. For the noise, we take $\eta^{(k)} \propto N(k)^{\beta}$ for some positive constant exponent β, so

$$\eta^{(k)} = \eta^{(0)} D^{\beta k}. \quad (7)$$

There could be various origins for such a scaling. A common one is total resource constraint. One expects the resources needed to maintain a given quality of physical gate operations to scale with N, so a constraint on the *total* available resource will result in a fall in the resource per physical component as the computer scales up. This gives a consequential drop in the quality of the gate or, equivalently, a rise in the physical error probability η.

The error probability per logical gate is then $p^{(k)} = p_0^{(k)} D^{\beta 2^k k}$, where Eq. (4) gives $p_0^{(k)}$. Going from $(k-1)$ to k levels of concatenation reduces the logical error probability when $p^{(k)}/p^{(k-1)} < 1$. This is satisfied for the model of Eq. (7) only when $k \leq k_{\max}$, where $k_{\max}$ is the largest positive integer satisfying (see Sec. A of the Supplemental

Material [32])

$$k_{max} < -\ln\left(B\eta^{(0)}D^{\beta}\right)\big/\ln\left(D^{\beta}\right). \qquad (8)$$

If no positive integer k_{max} satisfies this inequality, then $k_{max} = 0$ and concatenation is not useful at all. This is because concatenation is useful only if $p^{(1)} < \eta^{(0)}$, which requires

$$\eta^{(0)} < B^{-1}D^{-2\beta}. \qquad (9)$$

This is often a much more stringent condition than $\eta^{(0)} < \eta_{thres}$ in Eq. (2). For example, if the noise scales with the number of gates in a concatenated FTQC scheme, we can take $N(k) = G^{(k)} \equiv A(A')^{k-1}$ [see the paragraph following Eq. (2) above]. This means that we set $D = A'$, where $A' = 291$ as in Ref. [7]. One sees that, for $\beta = 1$, concatenation is useful only if $\eta^{(0)} < B^{-1}A^{-2} \sim 10^{-9}$, which is 10^5 times smaller than the usual threshold η_{thres}.

This condition is so stringent because D is so large. If the noise scales with a different physical parameter (number of qubits, number of wires, or similar), the value of D will be different but it will still typically be large. Equation (9) then makes it clear that the larger D is for a given parameter, the more important it is to minimize the scaling of the noise with that parameter (i.e., to minimize β).

The minimal attainable error probability per logical gate is given by taking $k = k_{max}$ in the above formula for $p^{(k)}$. For fixed system parameters (D, B, η_0, β), this $p^{(k_{max})}$ is easily found by taking $p^{(k)}$ for different integers k to see which is smallest. However, to see its dependence on those parameters, Sec. A of the Supplemental Material [32] gives algebraic formulas for upper and lower bounds on $p^{(k_{max})}$.

B. Resource constraints affecting crosstalk

A common problem in existing prototype quantum computers is crosstalk between qubits. This is an example of a more general problem of nonlocal non-Markovian noise, usually called long-range correlated noise. To treat this, we follow Refs. [20,25] and define H_{ij} as the arbitrary (and potentially noisy) unwanted interaction between physical qubits i and j. This interaction could be direct or it could be mediated by other degrees of freedom (which one traces out). In the latter case, it could be non-Markovian, meaning that it can account for interactions mediated by sub-Ohmic, Ohmic, or super-Ohmic baths [33]. One then defines the error strength

$$\Delta = \max_i \left(\sum_{j=1}^{N} ||H_{ij}|| \right) \qquad (10)$$

for a computer containing N physical qubits. References [20,25] have shown that $t_0\Delta$ is a good measure of the error

per gate, where t_0 is the duration of the slowest physical gate, although it should not be interpreted directly as the error probability per gate; see, e.g., Ref. [34]. They have then shown that fault tolerance holds when $t_0\Delta < (2e^{2+1/e}B^2)^{-1} \sim 10^{-9}$ for $N \to \infty$.

Here, in contrast, we consider cases where Δ diverges for $N \to \infty$, violating the condition for fault tolerance in Refs. [20,25]. This growth of Δ with N will often occur due to resource constraints. A simple example would be a constraint on the physical volume of the quantum computer, which is a current limitation in qubit technologies based on ion traps [35]. Then the density of qubits must scale like N. If each qubit has unwanted interactions with all other qubits within a given radius, one would have $\Delta \propto N$. A second example—relevant to multiple technologies—is a bandwidth constraint, i.e., a limit on the available range of transition frequencies of the qubits. This is particularly a problem for existing superconducting and ion-trap qubit technologies. There, each two-qubit gate corresponds to a different transition frequency and a given gate is performed by sending a driving signal (typically a microwave signal) into the quantum computer with the frequency of that gate. This means that the driving signal for a two-qubit gate between qubits n and m will also cause an unwanted interaction H_{ij} for pairs of qubits i and j with frequencies too close to that for n and m [36]. In some technologies, all qubits feel the driving signal. The number of qubits suffering this unwanted interaction then grows with the number of qubits in any given window of transition frequencies, which grows like N. In this case, $\Delta \propto N$. Of course, clever engineering may well reduce the scaling of Δ with N, so we prefer to consider $\Delta \propto N^{\beta}$ with some β [37].

We now study how the physics depends on the scaling of Δ with N. By taking $\Delta \propto N^{\beta}$, we have $\Delta \propto D^{\beta k}$ for k levels of concatenation, where the number of physical qubits increases by a factor of D with each level of concatenated error correction. We then use the method of Refs. [20,25], which involves taking all results in Sec. III above and replacing $\eta^{(k)}$ by $e^{1+1/(2e)}\sqrt{2t_0\Delta}$, where $t_0\Delta \propto D^{\beta k}$ (see Sec. B of the Supplemental Material [32]). Defining $\Delta_L^{(k)}$ as the upper bound on effective long-range correlated noise between *logical* qubits performing a given algorithm with k levels of concatenation, Sec. B of the Supplemental Material [32] shows that

$$t_0\Delta_L^{(k)} = \frac{\left(2e^{2+1/e}B^2 t_0\Delta^{(0)}\right)^{2^k}}{2e^{2+1/e}B^2} D^{\beta 2^k k}. \qquad (11)$$

Here, $\Delta^{(0)}$ is the magnitude of the long-range noise between physical qubits performing the same algorithm without error correction (so its logical qubits are its physical qubits). Equation (11) gives an overestimate of the true error, but no better bound exists at present. Thus the k that minimizes this bound gives the best existing bound on

the achievable accuracy in the presence of such noise. For example, if $\Delta^{(0)}$ is the crosstalk between *physical* qubits in a computer performing a given calculation without error correction, then $\Delta_L^{(k)}$ is an upper bound on the crosstalk between *logical* qubits in a computer performing the same calculation with k levels of concatenation. Since Eq. (11) has the same k dependence as in Sec. IV A, all results there and in Sec. A of the Supplemental Material [32] hold for this long-range noise, upon replacing B by $2e^{2+1/e}B^2$. As a quick estimate, one can set $D \sim A' = 291$ as done in the local-noise example; a more careful calculation [38] confirms that D is of order A'.

The good news is that this shows that error correction can reduce errors to a certain extent, even for noise too long ranged to have a fault-tolerance threshold. The bad news is that one requires $t_0\Delta^{(0)} < \left[2e^{2+1/e}B^2D^{2\beta}\right]^{-1}$ for error correction to be useful in reducing crosstalk between qubits. This is always tiny: it is of order 10^{-9} for small β and of order 10^{-13} for $\beta = 1$. If one achieves noise as weak as this, one can already do huge quantum calculations without worrying about errors.

In this section, we treat scale-dependent crosstalk induced by resource constraints but our conclusions apply to long-range correlated noise of any origin. One example is unwanted long-range interactions, which decay with the distance r between qubits i and j, so that $H_{ij} \propto (1/r)^z$. Reference [20] has considered this example in a d-dimensional lattice of qubits for $z > d$ but we treat longer-ranged noise ($z \leq d$) in Sec. B of the Supplemental Material [32], for which $\Delta \propto N^{(1-z/d)}$. This example is then the same as those treated above, with $\beta = (1 - z/d)$.

C. Energy constraint for resonant gates performing Shor's algorithm

As a concrete example of the situation described in Sec. IV A (with $\beta = 1$), we examine a resource constraint in a specific type of quantum gate implementation, performing a specific quantum algorithm. We take the gate implementation to be resonant qubit gates and assume that there are limited energy resources available to perform a given computation (see also a related early analysis in Ref. [29]). We take the desired computation to be Shor's algorithm and investigate how big a computation can be performed (with a given calculational accuracy) when there are limited energy resources.

We consider qubits embedded in waveguides, i.e., a continuum of electromagnetic modes prepared at zero temperature. Gates are activated by resonant propagating light pulses with a well-defined average energy or, equivalently, an average photon number $\overline{n}_g$ [see Fig. 5(a)]. This describes the situation in superconducting circuits [39,40] and integrated photonics [41]. It is also the paradigm of quantum networks and light-matter interfaces, with successful implementations in atomic qubits [42]. Here, we
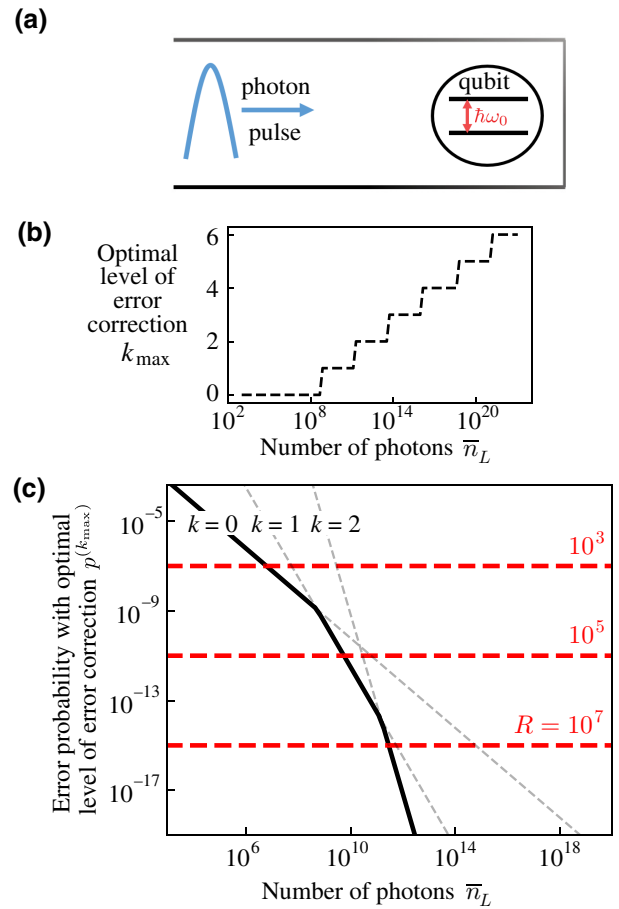

FIG. 5. (a) A qubit-in-waveguide schematic: the photon pulse for gate operations. (b) k_{max} as a function of number of photons per logical gate $\overline{n}_L = \overline{n}_{tot}/R^2$ (see text). (c) The lowest possible error per logical gate $p^{(k_{max})}$ as a function of $\overline{n}_L = \overline{n}_{tot}/R^2$, for Shor's factoring algorithm. The horizontal red dashed lines correspond to the different target (p_{err}) values for different R values.

maximize the accuracy of a computation for a given energy constraint. Then, by inverting the logic, we use this to find the minimal energy budget necessary to realize a specific computation with desired accuracy. For our illustrative goals, we treat only single-qubit gates subjected to noise from spontaneous emission. In doing so, we neglect dephasing noise. While this is a fair approximation for atomic qubits, it is more demanding for solid-state qubits but within eventual reach of superconducting circuits and spin qubits.

The dynamics of the qubit follow the Lindblad equation, $\dot{\rho} = -(i/\hbar)[H(t), \rho] + \mathcal{D}(\rho)$, with the total Hamiltonian $H(t) = H_0 + H_D(t)$. Here, $H_0 \equiv -(1/2)\hbar\omega_0\sigma_z$ is the bare Hamiltonian of the qubit, with $\sigma_z \equiv |0\rangle\langle0| - |1\rangle\langle1|$. We assume that the gates are designed as resonant driving, in the rotating-wave approximation (RWA), described by the Hamiltonian $H_D(t) \equiv (\hbar/2)\Omega h(t) \left(|0\rangle\langle1|e^{i\omega_0 t} + |1\rangle\langle0|e^{-i\omega_0 t}\right)$, with the Rabi frequency $\Omega \ll \omega_0$. We take $h(t)$

as a square function, nonzero only for the duration of the gate: $h(t) = 1$ for $t \in [0, \tau]$ and 0 otherwise. The unitary evolution induced by this Hamiltonian is a rotation around the x axis of the Bloch sphere by an angle $\theta = \Omega\tau$. This rotation is our single-qubit gate. In addition, the Lindblad dissipator $\mathcal{D}(\rho) \equiv \gamma[\sigma_-\rho\sigma_+ - (1/2)\{\rho, \sigma_+\sigma_-\}]$ accounts for the spontaneous emission in the waveguide. Here, $\sigma_- \equiv |0\rangle\langle 1|$ is the lowering operator, $\sigma_+ = \sigma_-^\dagger = |1\rangle\langle 0|$ the raising operator, and γ the spontaneous emission rate.

Note that Ω and γ are not independent, typical of waveguide quantum electrodynamics, where the driving and the relaxation take place through the same one-dimensional electromagnetic channel. Spontaneous emission events while the driving Hamiltonian is turned on cause errors in the gate implementation. Their impact is reduced if the qubit is driven faster, i.e., if the Rabi frequency is larger. Conversely, the Rabi frequency is related to the mean number of photons inside the driving pulse through $\Omega = (4\gamma/\theta)\bar{n}_g$ (see Sec. C of the Supplemental Material [32]). In principle, pulses containing more photons induce better gates, with perfect gates for an infinite number of photons. However, if one designs the gates to work within RWA (to avoid the complicated pulse-shaping issues that come with finite counter-rotating terms), then there cannot be too many photons: $\bar{n}_g \ll \omega_0/\gamma$. Then, the remnant noise, for a $\theta = \pi$ gate, has physical error probability (see Sec. C of the Supplemental Material [32]) $\eta = (\pi^2/16)(1/\bar{n}_g)$, with a minimal noise of order γ/ω_0.

We now assume a constraint on the total number of photons $\bar{n}_{tot}$ available to run the whole computation. As we show below, taking this constraint into account allows us to minimize the resource needed, for a target level of tolerable computational error. At level-k concatenation, the number of physical gates needed to implement a computation with L (logical) gates is $LG^{(k)} = LA^k$. Assuming a distinct pulse for each gate, the number of photons available per physical gate, given the total energetic constraint, is $\bar{n}_g = \bar{n}_{tot}(LA^k)^{-1}$. Thus, the physical error probability for the $\theta = \pi$ gate acquires an exponential k dependence:

$$\eta^{(k)} = \frac{\pi^2}{16}\frac{LA^k}{\bar{n}_{tot}}. \tag{12}$$

Thus this is a concrete example corresponding to the $\beta = 1$ case in Eq. (7) above.

To better grasp the consequences of this k-dependent η, which we take as the generic behavior for all gates, we consider carrying out Shor's factoring algorithm [43]. Shor's factoring algorithm is touted as the reason why the RSA public-key encryption system will be insecure when large-scale quantum computers become available. The current RSA key length is $R = 2048$ bits. The exponential speed-up of Shor's algorithm over known classical methods comes from the fact that we can do the discrete Fourier transform on an R-bit string using $O(R^2)$ gates

on a quantum computer (see, e.g., Ref. [10]), compared to $O(R2^R)$ gates on a classical computer. The discrete Fourier transform gives a period-finding routine within the factoring algorithm, the only step that cannot be done efficiently classically. Thus, to run Shor's algorithm, one needs $L \sim R^2$ nonidentity logical quantum gates for the discrete Fourier transform. The exact number of computational gates, including the identity gate operations, which can be noisy, for the full Shor's algorithm depends on the chosen circuit design and architecture. We will take the lower limit of $L \sim R^2$ in what follows. The concatenation values we find below are thus likely optimistic estimates.

A standard strategy is to demand that the computation runs correctly with probability $P_{target} > 1/2$; once this is true, the computation can be repeated to exponentially increase the success probability toward 1. For $P_{target} = 2/3$, with L logical gates, each with error probability p_{err} ($\ll 1$), we require $(1 - p_{err})^L > P_{target} = 2/3$, giving a target error probability per logical gate of $p_{err} \lesssim (3L)^{-1}$.

Figure 5(b) presents k_{max}, the maximum concatenation level, as an increasing function of the photon budget per logical gate $\bar{n}_L = \bar{n}_{tot}/R^2$ (approximately $\bar{n}_{tot}/L$) (note that $\bar{n}_L = G^{(k)}\bar{n}_g = A^k\bar{n}_g$). For fixed total resource, in this case, the photon budget per logical gate, as the concatenation level increases, the available photon count per physical component falls and we recover the behavior observed in earlier sections, giving a finite k_{max} and, consequently, a limit to the computational accuracy. The solid bold black line in Fig. 5(c) gives the corresponding minimum attainable error per logical gate as a function of $\bar{n}_L$.

V. MINIMIZING THE RESOURCE COSTS OF AN ALGORITHM

One can turn our results for resource constraints around, to enable us to answer the following question: What are the minimum resources needed for a target computational accuracy, sufficient for a given problem?

As an illustration, we answer this question for Shor's algorithm for an R-bit string in Sec. IV C. The number of gates in the algorithm grows with R, requiring a smaller p_{err} for the algorithm to be successful. This demands a larger photon budget to implement the algorithm using resonant gates. For the parameters in Fig. 5(c), $R = 10^3$ requires no concatenation and the minimum photon budget is $\bar{n}_L = 10^6$; for $R = 10^5$, we need $k = 1$ and $\bar{n}_L = 10^9$; for $R = 10^7$, we need $k = 2$ and $\bar{n}_L = 10^{11}$. Recall that the gates are assumed to be designed within the RWA; hence $\omega_0 \gg \gamma\bar{n}_g$. For $R = 10^3$, this translates to the condition $\omega_0 \gg \gamma\bar{n}_g = \gamma\bar{n}_L/A^0 = 10^7\gamma$; for $R = 10^7$, we need $\omega_0 \gg 10^6\gamma$ for $A = 575$ (this A is from Ref. [7]). These conditions are attained for atomic qubits. They are within reach of future generations of superconducting qubits, where $\gamma \sim 10$ Hz for qubit frequency $\omega_0 \sim 10$ GHz. Today, the best

coherence time for a superconducting qubit is within the millisecond range: $\gamma \sim 1$ kHz [44,45].

Our analysis also provides an estimate of the energy needed to run the gates involved in Shor's algorithm, namely, $E_{tot} \sim \hbar\omega_0 L\bar{n}_L(R)$. For $R = 10^3$, with the above photon budget of 10^6, this translates into $E_{tot} \sim 1$ pJ. Taking into account the parallelization of the computation (see Sec. D of the Supplemental Material [32]), this corresponds to a typical power consumption of about 1 pW, while the $R = 10^7$ case requires only 10 nW of power.

Thus for a realistic constraint on the photon power to perform quantum gate operations, error correction would allow one to perform large quantum computations. This is a surprising and positive conclusion, when one considers that the constraint causes scale-dependent errors for which there is no fault-tolerance threshold. It clearly shows that the absence of a threshold is not necessarily a significant impediment to using error correction in quantum computing.

Above, we calculated the energy of photons arriving at the qubits for a gate operation. However, that energy is a fraction of the energy for signal generation, because there is typically an attenuator between the signal generator and the qubit. The job of this attenuator is to absorb thermal photons to avoid perturbing the qubit. However, this means that it also absorbs most of the photons in the signal sent to perform the gate operation. To calculate the signal-generation energy from the above discussion, one must multiply those results by the attenuation factor. Unfortunately, the attenuation depends on design choices beyond the discussion here (such as the temperature at which the signal generation occurs). Furthermore, there is a large cryogenic energy budget for keeping the qubits and attenuators cool, which depends on the photon energy absorbed by the attenuator. Elsewhere [38], we will perform a full energetic optimization of all of these interlinked components, along with other critical components of a quantum computer, in a *full-stack* analysis of a large-scale quantum computer.

Nevertheless, the above example does show that the fundamental ingredient in the minimization of energy, or any other resource, is the calculation of the scale dependence of the noise that occurs when that resource is constrained. To go from this to a full-stack analysis is largely an issue of optimizing cryogenics, control circuitry (including signal generation), and the quantum algorithm for the calculation in question.

VI. CONCLUSIONS

Many quantum computing technologies currently exhibit physical gate errors that grow with the size and complexity of the quantum computer. Standard fault-tolerance discussions do not apply in such situations and no rigorous threshold exists. Despite this, we show that a certain amount of error correction can increase computational accuracy but that this accuracy decreases again with too much error correction. We show how to find the amount of error correction that optimizes this accuracy. For concreteness, we consider fault tolerance based on concatenated seven-qubit codes here. However, our approach could be applied to other fault-tolerance schemes, including the currently popular surface codes [46], where we also expect scale dependence of noise on physical components to lead to situations where a little error correction is good but too much is bad.

We explore the optimization of computational accuracy in the presence of resource constraints, whereby the resources per physical component go down as the number of components grows. We do this in increasing levels of practical relevance, from a simple toy example, to physical qubits in waveguides. We identify some cases, such as reasonable energy constraints for gate operations, where optimization gives a maximum accuracy good enough for large quantum computations. In other cases, such as volume or bandwidth constraints causing long-range crosstalk between qubits, error correction is only useful against such crosstalk when the error strength per physical gate is already so small (ranging from 10^{-9} to 10^{-13}) that one could perform huge quantum computations without any error correction.

Our analysis suggests three priorities for experimenters working toward useful quantum computers: (1) that they should try to characterize the scale dependence of the errors for their technology; (2) that they should strive to make this scale dependence as weak as possible; and (3) that they should reduce the physical error probability significantly below the standard threshold. Point (3) is necessary for standard fault tolerance to work well but it becomes critical when errors are scale dependent. In this context, the optimization in this work will enable experimenters to see the size of quantum computation that can be achieved with their error magnitude and scale dependence. In our upcoming work [38], we further address these points through a full-stack model of a quantum computer build from the theory presented here.

ACKNOWLEDGMENTS

———————

[1] F. Arute *et al.*, Quantum supremacy using a programmable superconducting processor, Nature **574**, 505 (2019).

[2] P. W. Shor, in *Proceedings of 37th Conference on Foundations of Computer Science* (1996), p. 56, arXiv:quant-ph/9605011.

[3] A. Yu. Kitaev, Quantum computations: Algorithms and error correction, Russian Math. Surveys **52**, 1191 (1997).

[4] D. Gottesman, Ph.D. thesis, California Institute of Technology (1997).

[5] E. Knill, R. Laflamme, and W. H. Zurek, Resilient quantum computation: Error models and thresholds, Proc. R. Soc. London. Ser. A: Math. Phys. Eng. Sci. **454**, 365 (1998).

[6] J. Preskill, Reliable quantum computers, Proc.: Math. Phys. Eng. Sci. **454**, 385 (1998).

[7] P. Aliferis, D. Gottesman, and J. Preskill, Quantum accuracy threshold for concatenated distance-3 codes, Quantum Info. Comput. **6**, 97 (2006).

[8] D. Aharonov and M, Ben-Or, Fault-tolerant quantum computation with constant error rate, SIAM J. Comput. **38**, 1207 (2008).

[9] D. Gottesman, in *Proceedings of Symposia in Applied Mathematics: Quantum Information Science and Its Contributions to Mathematics*, edited by Samuel J. Lomonaco Jr. (American Mathematical Society, 2010), Vol. 68, p. 13, Eprint arXiv:0904.2557.

[10] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, Cambridge, UK, 2010).

[11] R. Raussendorf, Key ideas in quantum error correction, Philos. Trans. Royal Soc. A **370**, 4541 (2012).

[12] S. J. Devitt, W. J. Munro, and K. Nemoto, Quantum error correction for beginners, Rep. Prog. Phys. **76**, 076001 (2013).

[13] D. K. Tuckett, S. D. Bartlett, S. T. Flammia, and B. J. Brown, Fault-Tolerant Thresholds for the Surface Code in Excess of 5% under Biased Noise, Phys. Rev. Lett. **124**, 130501 (2020).

[14] B. J. Brown, A fault-tolerant non-Clifford gate for the surface code in two dimensions, Sci. Adv. **6**, eaay4929 (2020).

[15] J. Pablo Bonilla Ataides, D. K. Tuckett, S. D. Bartlett, S. T. Flammia, and B. J. Brown, The XZZX surface code, Nat. Commun. **12**, 1 (2021).

[16] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, Surface codes: Towards practical large-scale quantum computation, Phys. Rev. A **86**, 032324 (2012).

[17] C. Monroe and J. Kim, Scaling the ion trap quantum processor, Science **339**, 1164 (2013).

[18] J. Ikonen, J. Salmilehto and M. Möttönen, Energy-efficient quantum computing, Npj. Quantum Inf. **3**, 17 (2017).

[19] B. M. Terhal and G. Burkard, Fault-tolerant quantum computation for local non-Markovian noise, Phys. Rev. A **71**, 012336 (2005).

[20] D. Aharonov, A. Kitaev, and J. Preskill, Fault-Tolerant Quantum Computation with Long-Range Correlated Noise, Phys. Rev. Lett. **96**, 050504 (2006).

[21] E. Novais and H. U. Baranger, Decoherence by Correlated Noise and Quantum Error Correction, Phys. Rev. Lett. **97**, 040501 (2006).

[22] E. Novais, E. R. Mucciolo, and H. U. Baranger, Hamiltonian formulation of quantum error correction and correlated noise: Effects of syndrome extraction in the long-time limit, Phys. Rev. A **78**, 012314 (2008).

[23] P. Aliferis and J. Preskill, Fault-tolerant quantum computation against biased noise, Phys. Rev. A **78**, 052331 (2008).

[24] H. K. Ng and J. Preskill, Fault-tolerant quantum computation versus Gaussian noise, Phys. Rev. A **79**, 032318 (2009).

[25] J. Preskill, Sufficient condition on noise correlations for scalable quantum computing, Quant. Inf. Comput. **13**, 181 (2013).

[26] E. Novais and E. R. Mucciolo, Surface Code Threshold in the Presence of Correlated Errors, Phys. Rev. Lett. **110**, 010502 (2013).

[27] A. G. Fowler and J. M. Martinis, Quantifying the effects of local many-qubit errors and nonlocal two-qubit errors on the surface code, Phys. Rev. A **89**, 032316 (2014).

[28] P. Jouzdani, E. Novais, I. S. Tupitsyn, and E. R. Mucciolo, Fidelity threshold of the surface code beyond single-qubit error models, Phys. Rev. A **90**, 042315 (2014).

[29] J. Gea-Banacloche, Some implications of the quantum nature of laser fields for quantum computations, Phys. Rev. A **65**, 022308 (2002).

[30] A. M. Steane, Error Correcting Codes in Quantum Theory, Phys. Rev. Lett. **77**, 793 (1996).

[31] Both A and A' are independent of the nature of the errors and so independent of the system size. The integer $A' = 291$ is the number of physical gate operations in the "Rec" of a controlled-NOT (CNOT) and $A = 575$ is the number in its "exRec." "Rec" is the encoded CNOT plus the following QEC box, while "exRec" is the encoded CNOT plus the preceding and following QEC boxes (see Ref. [7]).

[32] See the Supplemental Material at http://link.aps.org/supplemental/10.1103/PRXQuantum.2.040335 for technical details on various aspects of this work. These include bounds on $p^{(k_{max})}$, calculations for long-range correlated noise, specifics of resonant qubit gates, and a discussion of the energetic bill for running an algorithm.

[33] One simply requires that the bath spectrum has cut-offs that ensure that $||H_{ij}||$ is not divergent.

[34] P. Aliferis, in *Quantum Error Correction* (Cambridge University Press, Cambridge, United Kingdom, 2013), p. 126. Section 5.2.3.1 reviews the meaning of the error strength for local non-Markovian noise; a similar argument applies for long-range correlations [20].

[35] More generally, volume constraints will naturally occur for any technology requiring two-qubit gates between arbitrary qubits, if such gates become inaccurate at long distances.

[36] This mechanism would also mean that single-qubit gates cause local noise on other qubits with nearby frequencies. This has an effect as in Sec. IV A, so we do not consider it further here and we focus on the effect of H_{ij}.

[37] For example, if driving signals do not affect all qubits equally, machine-learning methods can be used to choose qubit frequencies to minimize crosstalk [47], thereby reducing β.

[38] M. Fellous-Asiani *et al.*, (to be published).

[39] P. Krantz, M. Kjaergaard, F. Yan, T. P. Orlando, S. Gustavsson, and W. D. Oliver, A quantum engineer's guide to superconducting qubits, Appl. Phys. Rev. **6**, 021318 (2019).

[40] B. Peropadre, J. Lindkvist, I.-C. Hoi, C. M. Wilson, J. J. Garcia-Ripoll, P. Delsing, and G. Johansson, Scattering of coherent states on a single artificial atom, New J. Phys. **15**, 035009 (2013).

[41] P. Lodahl, S. Mahmoodian, and S. Stobbe, Interfacing single photons and single quantum dots with photonic nanostructures, Rev. Mod. Phys. **87**, 347 (2015).

[42] A. Reiserer and G. Rempe, Cavity-based quantum networks with single atoms and optical photons, Rev. Mod. Phys. **87**, 1379 (2015).

[43] P. Shor, in *2013 IEEE 54th Annual Symposium on Foundations of Computer Science* (IEEE Computer Society, Los Alamitos, California, USA, 1994), p. 124.

[44] M. Kjaergaard, M. E. Schwartz, J. Braumüller, P. Krantz, J. I.-J. Wang, S. Gustavsson, and W. D. Oliver, Superconducting qubits: Current state of play, Annu. Rev. Condens. Matter Phys. **11**, 369 (2020).

[45] A. P. Sears, *Extending Coherence in Superconducting Qubits: From Microseconds to Milliseconds* (Yale University, 2013).

[46] For such schemes, k will be replaced by the scale of the error correction (i.e., a measure of the number of physical components in the error correction). Then one needs to replace our Eq. (1) with a relation, or numerical estimate, for the scaling of the logical error strength with k, which is not yet available for some schemes.

[47] P. V. Klimov, J. Kelly, J. M. Martinis, and H. Neven, The Snake Optimizer for Learning Quantum Processor Control Parameters, arXiv:2006.04594 (2020).