# Advantages and Limitations of Quantum Routing

Aniruddha Bapat,[1,2,3,4,*] Andrew M. Childs,[2,5,6,†] Alexey V. Gorshkov,[2,3,‡] and Eddie Schoute[2,5,6,7,§]

[1] *Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*
[2] *Joint Center for Quantum Information and Computer Science, NIST/University of Maryland, College Park, Maryland 20742, USA*
[3] *Joint Quantum Institute, NIST/University of Maryland, College Park, Maryland 20742, USA*
[4] *Department of Physics, University of Maryland, College Park, Maryland 20742, USA*
[5] *Institute for Advanced Computer Studies, University of Maryland, College Park, Maryland 20742, USA*
[6] *Department of Computer Science, University of Maryland, College Park, Maryland 20742, USA*
[7] *Computer, Computational, and Statistical Sciences Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*

(Received 1 July 2022; accepted 5 January 2023; published 1 February 2023)

The SWAP gate is a ubiquitous tool for moving information on quantum hardware, yet it can be considered a classical operation because it does not entangle product states. Genuinely quantum operations could outperform SWAP for the task of permuting qubits within an architecture, which we call *routing*. We consider quantum routing in two models: (i) allowing arbitrary two-qubit unitaries, or (ii) allowing Hamiltonians with norm-bounded interactions. We lower bound the circuit depth or time of quantum routing in terms of spectral properties of graphs representing the architecture interaction constraints, and give a generalized upper bound for all simple connected $n$-vertex graphs. In particular, we give conditions for a superpolynomial classical-quantum routing separation, which exclude graphs with a small spectral gap and graphs of bounded degree. Finally, we provide examples of a quadratic separation between gate-based and Hamiltonian routing models with a constant number of local ancillas per qubit and of an $\Omega(n)$ speedup if we also allow fast local interactions.

DOI: 10.1103/PRXQuantum.4.010313

## I. INTRODUCTION

Scalable quantum architectures are expected to have geometrically constrained interactions [1–6]. Unlike conversions between gate sets, which introduce only a logarithmic overhead due to the Solovay-Kitaev theorem [7], architecture connectivity can introduce a polynomial overhead from the cost of simulating nonlocal interactions. For example, a unitary implementation of a CNOT gate on the ends of an $n$-qubit one-dimensional (1D) chain requires time $\Omega(n)$ by a signaling argument. This raises a natural question: how do we implement general nonlocal operations in minimal depth under architectural constraints?

Depending on the implementation, it may be possible to reconfigure the architecture connectivity. For example, neutral atoms and ions can be physically moved in the architecture [8,9], and optical switches can be reconfigured to arbitrary connectivity [4]. It is also possible to have application-specific connectivity such as in optical quantum computers [10]. Instead, we study hardware-agnostic methods for fixed connectivity.

A natural such method for implementing nonlocal gates is to first permute qubits within the architecture. We call the task of implementing an arbitrary given permutation of qubits via operations on neighboring qubits *routing*. Routing generalizes well-studied tasks, such as state transfer [11–17] and state reversal (or mirroring) [18–22]. By exploring limits on routing, we also explore limits on information transfer and entanglement generation. In particular, the entanglement across a bipartition can be increased by routing, e.g., by starting with a maximally entangled qubit pair in the left partition and sending one qubit across the bipartition. Therefore, bounds on routing relate to

*ani@lbl.gov
†amchilds@umd.edu
‡gorshkov@umd.edu
§eschoute@lanl.gov

bounds on entanglement capacity [23–26] and have further relations to Lieb-Robinson bounds [27].

A common implementation of routing uses SWAP gates to implement permutations [28–31]. We call this approach *classical routing* since a separable state acted on by SWAP gates cannot become entangled. In routing, we represent architecture connectivity by a simple connected graph, $G$. Classical routing algorithms have been developed using a variety of techniques, including shortest path algorithms [28,32–36], sorting algorithms [37–41], routing on a spanning tree [42], and exhaustive search [29,30]. In fact, classical routing is equivalent to the *routing via matchings* problem in classical computer science [31]. *Routing via matchings* is NP complete [43,44], but efficient algorithms exist for special cases of architecture connectivity such as paths, complete graphs, trees, and graph products that capture practical architectures, such as grids [31,45,46]. A natural lower bound on classical routing arises from small vertex cuts in the architecture [45].

In this work, we explore the extent to which genuinely quantum operations can accelerate routing in what we call *quantum routing*. The relative power of the models we explore is depicted in Fig. 1. In Sec. II, we introduce gate-based quantum routing with arbitrary two-qubit unitaries and Hamiltonian (quantum) routing by Hamiltonian evolution with norm-bounded interactions.

A key observation is that routing can distribute entanglement across a bipartition of the system. Thus, by lower bounding the circuit depth to create entanglement across vertex cuts in the architecture, we derive bounds on quantum routing and improve bounds on classical routing. The same argument does not apply in the continuous-time setting of the Hamiltonian model. However, we show a similar (but weaker) lower bound from a lower bound on the time to create entanglement across small edge cuts in the architecture that improves on a previous bound [47] by constant factors.

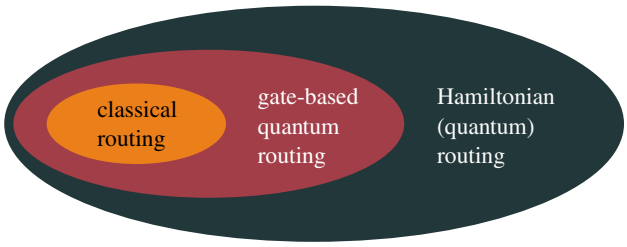


FIG. 1. The relative power of routing models considered in this work. Most prior work considers only using SWAP gates for routing, which we call *classical routing*. In this work, we explore the additional routing power provided by genuine quantum operations. We consider increasingly more powerful quantum routing models: first, allowing arbitrary two-qubit gates in *gate-based quantum routing*, and then allowing continuous Hamiltonian evolution in *Hamiltonian (quantum) routing*.

We show lower bounds on routing by proving lower bounds on state preparation in the respective models. The circuit depth to distribute entanglement in the gate-based model is lower bounded by the maximum matching size in the edge boundary of a vertex cut. In the Hamiltonian model, the evolution time is lower bounded in terms of edge cuts. Our state-preparation lower bounds generalize earlier analyses for lattices [48–50] to general graphs.

Next, in Sec. III, we prove a generalized upper bound on classical routing. This rules out a superpolynomial separation between classical and Hamiltonian routing conditioned on the spectral gap of the Laplacian and the degree distribution of the vertices. In particular, our results rule out a superpolynomial separation for interaction graphs of bounded (constant) degree, a common feature of practical quantum architectures [1,51,52]. It is often easy to prove a polynomial lower bound on classical routing on such architectures, implying that any quantum routing procedure will also have polynomial overhead in the worst case. Naively, routing thus rules out subpolynomial-depth quantum algorithms in such architectures (unlike the choice of gate set, which introduces at most polylogarithmic overhead), showing that polynomial overhead can only be avoided by designing specialized quantum algorithms for architectures that avoid worst-case routing behavior.

While we can rule out a superpolynomial separation in some cases, we are not aware of even a superconstant speedup of Hamiltonian routing over classical routing for any family of graphs. In Sec. IV, we give two such examples in strengthened routing models. The first is an $\Omega\left(\sqrt{n}\right)$ factor speedup in a strengthened routing model with one local ancilla per qubit. The second, a $\Theta(n)$ speedup, follows from allowing fast local interactions, which give an asymptotically optimal gate-based routing algorithm and an asymptotically optimal Hamiltonian routing algorithm if allowed one local ancilla per qubit. Without assumptions, we observe that our bounds on Hamiltonian routing cannot exclude a superconstant speedup on a star graph architecture; however, it is quite possible that tighter bounds can be proven.

## II. QUANTUM ROUTING

In this section, we introduce quantum routing and prove lower bounds dependent on graph expansion properties. We model the architectural constraints by a simple graph $G$ on $n$ vertices, with the qubits represented by the vertex set $V(G)$ and the allowed interactions between qubits by the edge set $E(G)$.

### A. Gate-based quantum routing

First, we consider routing in the gate-based model of quantum computation. Analogous to the (classical) routing number [45] [see also Eq. (61) in Sec. III], we define

the *gate-based quantum routing number* $\mathrm{qrt}(G)$ as

$$\mathrm{qrt}(G) := \max_{\pi} \mathrm{qrt}(G, \pi), \tag{1}$$

where $\pi$ is a permutation of the qubits and $\mathrm{qrt}(G, \pi)$ is the minimum depth of a unitary circuit that implements the permutation $\pi$ while respecting the architecture constraints $G$, i.e., only having two-qubit gates [53] acting along the edges $E(G)$. In this model, single-qubit gates are free since they can be absorbed into adjacent two-qubit gates.

We briefly prove a diameter lower bound on gate-based quantum routing. The diameter of a graph $G$ is

$$\mathrm{diam}(G) := \max_{u,v \in V(G)} d(u, v), \tag{2}$$

where $d(u, v)$ is the (shortest) distance between vertices $u$ and $v$.

**Theorem II.1.** *For any simple graph G,*

$$\mathrm{qrt}(G) \geq \mathrm{diam}(G). \tag{3}$$

*Proof.* Consider two vertices $u, v \in V(G)$ at a distance $\mathrm{diam}(G)$ and a circuit $\mathcal{C}$ of two-qubit unitaries with depth $D$ acting on $G$. Any local operator acting on $u$ evolved in the Heisenberg picture under $\mathcal{C}$ will have no support on vertices further than distance $D$. In order to swap $u$ and $v$, all of the support of that Heisenberg-evolved operator must be on $v$, which implies $D \geq \mathrm{diam}(G)$. Therefore, $\mathrm{qrt}(G) \geq \mathrm{diam}(G)$. ■

To prove a lower bound on gate-based quantum routing, we relate routing to the task of generating entanglement. We can quantify the entanglement of a pure state $\rho$ on a bipartite joint system $X\bar{X}$, consisting of the subsystems $X$ and $\bar{X}$, by the von Neumann entropy of the reduced density operator $\rho_X := \mathrm{Tr}_{\bar{X}}(\rho)$, defined as

$$S(\rho_X) := -\mathrm{Tr}(\rho_X \log \rho_X). \tag{4}$$

[The function $\log(x)$ denotes the logarithm base 2 unless specified otherwise. We denote the natural logarithm by $\ln(x)$.] We refer to the von Neumann entropy as "the entropy" and denote $S_X(\rho) := S(\rho_X)$. For completeness, we list some elementary properties of the entropy that will be useful later and can be easily verified.

**Lemma II.2.** *For a state $\rho$ on a joint system $X\bar{X}$, the following statements about the entropy hold:*

1. *If $\rho$ is a pure state, then the entropy is symmetric, i.e.,*

$$S_X(\rho) = S_{\bar{X}}(\rho). \tag{5}$$

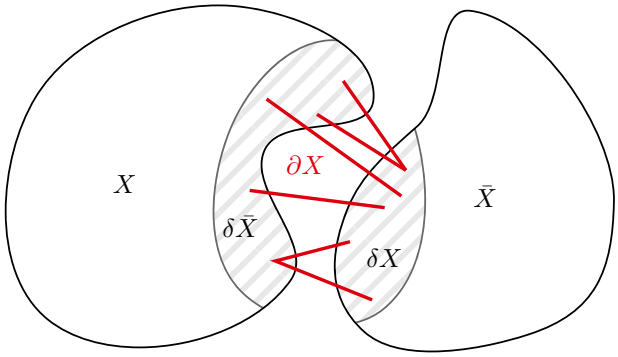


FIG. 2. A graph can be partitioned into two sets of vertices $X$ and $\bar{X}$. The vertex boundary $\delta X$ of $X$ is the set of vertices outside of $X$ that are directly connected to $X$, and similarly for $\delta\bar{X}$ and $\bar{X}$. The edge boundary $\partial X$ (red) of $X$ (and $\bar{X}$) is the set of edges that connect $X$ to $\bar{X}$.

2. *The entropy is invariant under change of basis, i.e.,*

$$S(U\rho U^\dagger) = S(\rho). \tag{6}$$

3. *The entropy is invariant under local unitaries $U_X$ on $X$ and $U_{\bar{X}}$ on $\bar{X}$, i.e.,*

$$S_X((U_X \otimes U_{\bar{X}})\rho(U_X \otimes U_{\bar{X}})^\dagger) = S_X(\rho). \tag{7}$$

The change in entropy of the reduced state on $X \subseteq V(G)$ by a unitary respecting the constraints of the interaction graph $G$ can be bounded by a quantity proportional to the vertex boundary of $X$. The set $X$ and its *vertex complement* $\bar{X} := V(G) \setminus X$ define a vertex partition of $G$; see Fig. 2. By the invariance of the entropy under local unitaries (part 3 of Lemma II.2), we need consider only the unitary acting across the partition. In particular, the unitary must act on the *vertex boundary*

$$\delta X := \{v \in \bar{X} \mid \{u, v\} \in E(G),\ u \in X\}, \tag{8}$$

which forms a vertex cut in $G$. We formalize this bound on the change in entropy in the following lemma derived from the small total entangling property [54].

**Lemma II.3.** [Small total entangling (STE)]. *For a unitary $U$ acting nontrivially only on the joint subsystem $\delta X \delta\bar{X}$, the change in the entropy of any state $\rho$ is bounded by*

$$|S_X(U\rho U^\dagger) - S_X(\rho)| \leq 2\min(|\delta X|, |\delta\bar{X}|). \tag{9}$$

*Proof.* We consider a purification system $R$ such that $\mathrm{Tr}_R(|\psi\rangle\langle\psi|) = \rho$ for some pure state $|\psi\rangle$ on the joint

system $X\bar{X}R$. For any subsystem $Y$ of $X\bar{X}$,

$$S_Y(|\psi\rangle) := S_Y(\mathrm{Tr}_R(|\psi\rangle\langle\psi|)). \tag{10}$$

The unitary $U$ is a local unitary on the joint subsystem $\delta X \delta\bar{X}$, so Lemma II.2 implies

$$\begin{aligned}|S_X(U\rho U^\dagger) - S_X(\rho)| &= |S_X(U|\psi\rangle) - S_X(|\psi\rangle)| \\ &= |S_{X|\delta X}(U|\psi\rangle) - S_{X|\delta X}(|\psi\rangle)|,\end{aligned} \tag{11}$$

where we use the conditional quantum entropy, $S_{X|\delta X}(|\psi\rangle) = S_{X\delta X}(|\psi\rangle) - S_X(|\psi\rangle)$. By the triangle inequality and [55, theorem 11.5.1],

$$\begin{aligned}&|S_{X|\delta X}(U|\psi\rangle) - S_{X|\delta X}(|\psi\rangle)| \\ &\quad\le |S_{X|\delta X}(U|\psi\rangle)| + |S_{X|\delta X}(|\psi\rangle)| \le 2|\delta X|.\end{aligned} \tag{12}$$

By symmetry of the entropy for pure states, we also obtain

$$\begin{aligned}|S_X(U\rho U^\dagger) - S_X(\rho)| &= |S_{\bar{X}|\delta\bar{X}}(U|\psi\rangle) - S_{\bar{X}|\delta\bar{X}}(|\psi\rangle)| \\ &\le 2|\delta\bar{X}|.\end{aligned} \tag{13}$$

The minimum of Eqs. (12) and (13) gives the required bound. ∎

We can saturate this bound in several special cases. A SWAP gate can saturate this bound when the subsystems $\delta X$ and $\delta\bar{X}$ are single qubits that are maximally entangled with the remainder of $\bar{X}$ and $X$, respectively. Furthermore, with sufficient connectivity, we can also saturate this bound in higher dimensions: let

$$|\delta X| = |\delta\bar{X}| \le \min(|X|, |\bar{X}|)/2 \tag{14}$$

and let $\delta X$ and $\delta\bar{X}$ be maximally entangled with the remainder of $\bar{X}$ and $X$, respectively. Then, if we exchange $\delta X$ with $\delta\bar{X}$ through simultaneous SWAPs, the entropy increases by $2|\delta X|$, saturating the bound.

We now prove a lower bound on the time required for *state preparation* of entangled states based on the maximum matching size in the edge boundary of the vertex cut. A *matching* is a set of edges $E' \subseteq E(G)$ such that all vertices in $E'$ are distinct. For any $E' \subseteq E(G)$, we define $m(E') \subseteq E'$ as the maximum(-size) matching in $E'$. State preparation is the task of preparing some target state $\rho$ given an initial state $\rho_0$. A special case of state preparation is routing a particular state. If the change in entanglement between initial state $\rho_0$ and final state $\rho$ is $|S_X(\rho) - S_X(\rho_0)|$, then a simple argument from STE gives a circuit depth lower bound of $|S_X(\rho) - S_X(\rho_0)|/(2|\delta X|)$, and similar arguments have been used with the entanglement capacity [26,56]. However, this does not account for the time required to entangle the boundary subsystem with the bulk subsystem. A careful accounting gives the following, which we later show can be saturated.

**Lemma II.4.** *Given an initial state $\rho_0$ and a target state $\rho$ on the bipartite system consisting of $X$ and $\bar{X}$, define the change in entropy*

$$\Delta S_Z := |S_Z(\rho) - S_Z(\rho_0)| \tag{15}$$

*for any subsystem $Z$. Then any gate-based unitary circuit $\mathcal{C}$ for preparing $\rho$ from $\rho_0$ restricted by an interaction graph $G$ has depth*

$$d \ge \frac{\Delta S_X}{2|m(\partial X)|}, \tag{16}$$

*for any $X \subsetneq V(G)$, and*

$$d \ge \frac{\Delta S_X + \Delta S_Y}{2|m(\partial(\delta X))|}, \tag{17}$$

*for $Y := \bar{X} \setminus \delta X$.*

*Proof.* We can decompose $\mathcal{C}$ into a sequence of disjoint unitaries $U_i$ acting on $X\delta X$ and unitaries $V_i$ acting on $Y\delta X$, where $i \in \mathbb{N}$. To perform operations $U_i$ and $V_i$ simultaneously, they must act on disjoint subsets $X_i, X_i' \subseteq \delta X$, respectively. Between each application of $U_iV_i$, there are local unitary operations within $X$, $\delta X$, and $Y$, labeled as $O_i$, that we allow to be performed instantaneously. The circuit can thus be decomposed as

$$\mathcal{C} = O_d U_d V_d \dots O_1 U_1 V_1 O_0. \tag{18}$$

We lower bound $d$ by considering the change in entropy and applying STE (Lemma II.3). First, we note that the operations $O_i$ cannot change the entropy of the respective subsystems. By STE, $U_i$ can change the entropy of $X$ by at most $2|X_i|$ and $V_i$ can change the entropy of $Y$ by at most $2|X_i'|$. Therefore, we have two inequalities that must be satisfied:

$$\Delta S_X \le 2\sum_{i=1}^{d} |X_i| \tag{19}$$

$$\text{and} \quad \Delta S_Y \le 2\sum_{i=1}^{d} |X_i'|. \tag{20}$$

Noting that $|X_i| \le |m(\partial X)|$, we obtain $\Delta S_X \le 2d|m(\partial X)|$, thus proving Eq. (16). Additionally, we note that $|X_i| + |X_i'| \le |m(\partial(\delta X))|$ so that

$$\Delta S_X + \Delta S_Y \le 2\sum_{i=1}^{d} \left(|X_i| + |X_i'|\right) \le 2d|m(\partial(\delta X))|, \tag{21}$$

which implies Eq. (17). ∎

Entanglement capacity-based state-preparation lower bounds that are proportional to $|\delta X|$ [49] can be weaker than Lemma II.4 by a factor $\Omega(n)$ for some partitions $X$. To see this, consider the graph $L_{2n}$ that consists of two complete graphs $G_1 = K_n$ and $G_2 = K_n$ with additional edges

$$(\{x_1\} \times V(G_2)) \cup (V(G_1) \times \{x_2\}), \tag{22}$$

where we pick arbitrary vertices $x_1 \in V(G_1)$ and $x_2 \in V(G_2)$. For the partition with $X = V(G_1)$, we have $|m(\partial X)| = 2$ whereas $|\delta X| = n$.

Even so, we can obtain a simpler lower bound on the circuit depth as a corollary by relating the change in entropy $\Delta S_{\bar{X}}$ to that of the bulk system $\bar{X} \setminus \delta X$.

**Corollary II.5.** *Given an initial state $\rho_0$ and target state $\rho$, the depth of a gate-based state preparation circuit restricted by interaction graph $G$ with partition $X \subsetneq V(G)$ is lower bounded by*

$$d \geq \frac{\Delta S_X + \Delta S_{\bar{X}} - 2|\delta X|}{2|m(\partial(\delta X))|} \geq \frac{\Delta S_X + \Delta S_{\bar{X}}}{2|\delta X|} - 1. \tag{23}$$

*Proof.* Let $Y := \bar{X} \setminus \delta X$. The entropy of the target state can be upper bounded using subadditivity and $S_{\delta X}(\sigma) \leq |\delta X|$ for any state $\sigma$ as

$$\Delta S_{\bar{X}} = S_{\bar{X}}(\rho) - S_{\bar{X}}(\rho_0) \leq S_Y(\rho) - S_{\bar{X}}(\rho_0) + |\delta X|. \tag{24}$$

The Araki-Lieb triangle inequality,

$$S_{\bar{X}}(\rho_0) \geq |S_Y(\rho_0) - S_{\delta X}(\rho_0)|, \tag{25}$$

then gives

$$\Delta S_{\bar{X}} \leq S_Y(\rho) - S_{\bar{X}}(\rho_0) + |\delta X| \leq \Delta S_Y + 2|\delta X|. \tag{26}$$

We now apply Lemma II.4, giving

$$d \geq \frac{\Delta S_X + \Delta S_{\bar{X}} - 2|\delta X|}{2|m(\partial(\delta X))|} \geq \frac{\Delta S_X + \Delta S_{\bar{X}}}{2|\delta X|} - 1 \tag{27}$$

as claimed, where the second inequality follows from $|m(\partial(\delta X))| \leq |\delta X|$. ■

Corollary II.5 can be saturated by Algorithm II.1 when $|X| = 2k|\delta X|$ for integer $k > 0$, so that a set of $2|\delta X|$ ends of Bell pairs can be exchanged between $X$ and $\bar{X}$ every odd time step. The algorithm makes the additional assumptions that $|X| \leq |\bar{X}|$ to allow for $\Delta S_X = \Delta S_{\bar{X}} = |X|$ and that $\delta X$ has high connectivity with the rest of the graph so that ends of Bell pairs can easily be routed to and from $\delta X$. The algorithm saturates Corollary II.5 after every odd time step up to and including depth $d = 2k - 1$.

We now show that a lower bound on the gate-based quantum routing number follows from Lemma II.4 by preparing an appropriate initial state. See Fig. 3 for an illustration of the proof concept.

**Theorem II.6.** *For any simple graph $G$ and partition $X \subseteq V(G)$ with $|X| \leq |V(G)|/2$,*

$$\mathrm{qrt}(G) \geq \frac{|X|}{|m(\partial X)|} \tag{28}$$

*and*

$$\mathrm{qrt}(G) \geq \max\left(\frac{2|X| - |\delta X|}{|m(\partial(\delta X))|}, \frac{2|X| - |\delta \bar{X}|}{|m(\partial(\delta \bar{X}))|}\right). \tag{29}$$

*Proof.* We augment the subsystems $X$ and $\bar{X}$ with ancilla spaces $x$ and $x'$, respectively, with one ancilla qubit for each vertex in $X$ and $\bar{X}$. Since these ancillas are not connected with the main graph, they cannot help with routing. Each qubit and ancilla pair forms a Bell pair in the initial state $\rho_0$. Then the entropy $S_{Xx}(\rho_0) = S_{\bar{X}x'}(\rho_0) = 0$ since the reduced state is pure.

The gate-based quantum routing number $\mathrm{qrt}(G)$ considers the worst-case permutation of the vertices. So, to show a lower bound, it suffices to pick a permutation $\pi$ that routes all vertices $v \in X$ to $\bar{X}$ arbitrarily and routes $|X|$ vertices $u \in \bar{X}$ to $X$ arbitrarily. Let the resulting state be our target state $\rho$. This gives $S_{Xx}(\rho) = S_{\bar{X}x'}(\rho) = 2|X|$. By

---

**Input:** Initial state of $|X|/2$ Bell pairs on each of $X$ and $\bar{X}$ where each $v \in \delta X$ is part of a distinct Bell pair.

simultaneously SWAP ends of unrouted Bell pairs in $X$ and $\delta X$
**while** $X$ has unrouted ends of Bell pairs **:**
  simultaneously SWAP ends of unrouted Bell pairs in $Y$ and $\delta X$
  simultaneously SWAP ends of unrouted Bell pairs in $X$ and $\delta X$

---

Algorithm II.1: Saturating Corollary II.5 when $|X| = 2k|\delta X| \leq |\bar{X}|$ for integer $k > 0$, assuming every vertex in $\delta X$ is connected to all vertices in $X$ and $Y := \bar{X} \setminus \delta X$. We prepare Bell pairs on $X$ and $\bar{X}$ and route one end of each Bell pair in $X$ to $\bar{X}$ and vice versa.

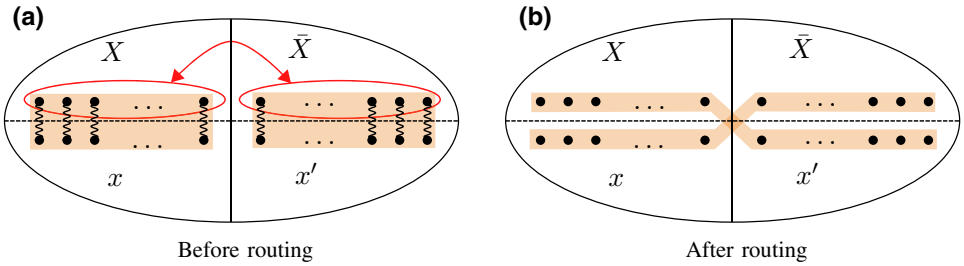


FIG. 3. For the proof of Theorem II.6, we consider a bipartite system consisting of $X$ and $\bar{X}$ with $|X| \le |\bar{X}|$. The subsystems $X$ and $\bar{X}$ consist of qubits represented by vertices and are augmented with ancilla spaces $x$ and $x'$, respectively. To each qubit in $X$ and $\bar{X}$ we associate one ancilla in $x$ and $x'$, respectively. We initialize each qubit-ancilla pair in a Bell state (wavy line). The entropy of subsystem $Xx$ is 0. We then perform routing to exchange $X$ with a subset of $\bar{X}$ (in red). This increases the entropy of $Xx$ to $2|X|$. We bound the entanglement increase for each layer of gates by twice the maximum matching size in $\partial(\delta X)$, thereby lower bounding the circuit depth and $\mathrm{qrt}(G)$.

Lemma II.4, the depth of any circuit performing this state preparation and routing task is lower bounded as

$$\mathrm{qrt}(G,\pi) \ge \frac{\Delta S_{Xx}}{2|m(\partial X)|} = \frac{|X|}{|m(\partial X)|}, \tag{30}$$

proving (28). Similarly, Corollary II.5 implies

$$\mathrm{qrt}(G,\pi) \ge \frac{\Delta S_{Xx} + \Delta S_{\bar{X}x'} - 2|\delta X|}{2|m(\partial(\delta X))|} = \frac{2|X| - |\delta X|}{|m(\partial(\delta X))|}. \tag{31}$$

By exchanging the roles of $X$ and $\bar{X}$, Corollary II.5 also gives the lower bound

$$\mathrm{qrt}(G,\pi) \ge \frac{2|X| - |\delta \bar{X}|}{|m(\partial(\delta \bar{X}))|}. \tag{32}$$

Taking the maximum of Eqs. (31) and (32), we obtain Eq. (29) as required. ■

We now show that Theorem II.6 lower bounds the gate-based quantum routing number in terms of the *vertex expansion* (or vertex isoperimetric number)

$$c(G) := \min_{X \subseteq V(G): |X| \le |V(G)|/2} \frac{|\delta X|}{|X|}, \tag{33}$$

which is a well-studied property of graphs [57]. Intuitively, the vertex expansion lower bounds how many vertices neighbor any small enough set $X$. Therefore, the number of vertices in the induced subgraph $G[X \cup N(X)]$, for $N(X)$ the neighborhood of $X$, grows (or "expands") by at least a factor of $1 + c(G)$.

**Corollary II.7.** *For any simple graph $G$,*

$$\mathrm{qrt}(G) \ge \max_{X \subseteq V(G): |X| \le |V(G)|/2} \frac{2|X|}{|\delta X|} - 1 = \frac{2}{c(G)} - 1. \tag{34}$$

*Proof.* By maximizing over all allowed partitions $X$ in Theorem II.6, choosing one branch of Eq. (29), and noting $|m(\partial(\delta Y))| \le |\delta Y|$, for any $Y \subseteq V(G)$, we have

$$\mathrm{qrt}(G) \ge \max_{X \subseteq V(G): |X| \le |V(G)|/2} \frac{2|X| - |\delta X|}{|m(\partial(\delta X))|} \tag{35}$$

$$\ge \max_{X \subseteq V(G): |X| \le |V(G)|/2} \frac{2|X| - |\delta X|}{|\delta X|} \tag{36}$$

$$= \max_{X \subseteq V(G): |X| \le |V(G)|/2} \frac{2|X|}{|\delta X|} - 1 \tag{37}$$

as required. ■

In Appendix A, we show that the vertex expansion and a similarly defined *matching expansion*,

$$m(G) := \min_{X \subseteq V(G): |X| \le |V(G)|/2} \frac{|m(\partial X)|}{|X|}, \tag{38}$$

are asymptotically equivalent, i.e., $c(G) = \Theta(m(G))$.

A simple consequence of Corollary II.7 is that gate-based quantum routing on the star graph, $S_n := K_{1,n}$ [the complete bipartite graph with parts of size 1 and $n$, as shown in Fig. 4(a)], is no faster than classical routing up to a constant factor. A trivial classical routing strategy has a depth upper bounded by $3n/2$, whereas we have $c(S_n) \le 2/n$ so that $\mathrm{qrt}(S_n) \ge n - 1$. This is a consequence of the small vertex cut in the star graph.

### B. Hamiltonian routing

In this section, we consider a stronger model for quantum routing, namely using two-qubit Hamiltonians with fast local operations. The *Hamiltonian routing time*,

$$\mathrm{hqrt}(G) := \max_{\pi} \mathrm{hqrt}(G,\pi), \tag{39}$$

where $\pi$ is a permutation of qubits and $\mathrm{hqrt}(G,\pi)$, is the minimum evolution time, normalized so that a SWAP gate takes time 1 (discussed below), of some time-dependent Hamiltonian $H(t)$ that respects the architecture constraints given by $G$ [i.e., it is 2-local and only has interactions along the edges $E(G)$] and implements $\pi$. Note that here we consider minimizing the time as opposed to circuit depth.

A time scale follows from a normalization condition on the two-qubit interaction strength of the Hamiltonian $H(t)$ at all times. We can write any two-qubit local Hamiltonian in the *canonical form* [58]

$$K := \sum_{j\in\{x,y,z\}} \mu_j \sigma_j \otimes \sigma_j \tag{40}$$

up to local unitaries, where $\mu_x \geq \mu_y \geq |\mu_z| \geq 0$, and $\sigma_x, \sigma_y, \sigma_z$ are the Pauli matrices. We impose the condition that $\|K\| = \sum_j |\mu_j| \leq 3\pi/4$ for all interactions in $H(t)$ at all times $t$ [59], where $\|\cdot\|$ is the spectral norm. Recall that we consider a model in which local operations can be performed arbitrarily quickly. The shortest CNOT time in this model is $1/3$ and the shortest SWAP time is 1 [60]. Furthermore, any two-qubit unitary takes at most time 1 since any such gate can be decomposed into at most three CNOT gates and single-qubit rotations [61]. Therefore, this normalization guarantees $\mathrm{hqrt}(G,\pi) \leq \mathrm{qrt}(G,\pi)$ for any permutation $\pi$, and in particular, $\mathrm{hqrt}(G) \leq \mathrm{qrt}(G)$. We now show that the Hamiltonian routing time is lower bounded by the diameter of the graph over the maximum degree.

**Theorem II.8.** *For any simple graph $G$,*

$$\mathrm{hqrt}(G) = \Omega\left(\frac{\mathrm{diam}(G)}{\max_v d_v}\right), \tag{41}$$

*where $d_v$ is the degree of $v \in V(G)$.*

*Proof.* Pick two vertices $u, v \in V(G)$ at a distance $\mathrm{diam}(G)$. In the Heisenberg evolution picture, routing must be able to map an $X$ operator on $u$ at time 0, $X_u(0)$, to $X_u(T)$ supported on $v$ after some time $T$. This means a $Z$ operator on $v$ at time 0, $Z_v$, has $\|[Z_v, X_u(T)]\| = 2$. Reference [62, Eq. 7] bound this unequal time commutator after time $t$ by

$$\|[Z_v, X_u(t)]\| \leq 2e^{C|t| - \mathrm{diam}(G)}, \tag{42}$$

where

$$C = 2^6 e \max_{w\in V(G)} \sum_{e=(w,w')\in E(G)} \|H^{(e)}\| \leq 3\pi 2^4 e \max_w d_w \tag{43}$$

and $H^{(e)}$ is a two-qubit Hamiltonian term acting only on the ends of the edge $e$. Therefore, the time is lower bounded by $t = \Omega\left(\mathrm{diam}(G)/\max_w d_w\right)$. ∎

The dependence on the maximum degree is necessary when we consider a multigraph with two vertices connected by $k$ edges. We can then speed up any normalized interaction between the two vertices linearly in the degree, $k$. In particular, it is possible to implement a SWAP in time $1/k$. We use a similar idea to show separations between strengthened gate-based and Hamiltonian routing models in Sec. IV. It is an open question whether a Hamiltonian routing protocol on a simple graph can have a routing time that is upper bounded by $o(\mathrm{diam}(G))$.

We show that the Hamiltonian routing time can also be lower bounded by an edge cut in the graph $G$. An edge cut partitions $G$ into two vertex subsets $X \subseteq V(G)$ and $\bar{X}$. The edges leaving $X$ form the *edge boundary* of $X$,

$$\partial X := \{(x,\bar{x}) \in E \mid x \in X, \bar{x} \in \delta X\} = \partial \bar{X}, \tag{44}$$

and are an edge cut. We define the *edge expansion* (or edge isoperimetric number or Cheeger constant) as

$$h(G) := \min_{X\subseteq V(G):|X|\leq |V(G)|/2} \frac{|\partial X|}{|X|}. \tag{45}$$

Intuitively, this corresponds to a lower bound on how many edges leave any small enough set $X$. Therefore, the number of edges in the induced subgraph $G[X \cup N(X)]$ grows (or "expands") by at least $1 + h(G)$.

In the following, we show a lower bound of $\mathrm{hqrt}(G) = \Omega\left(1/h(G)\right)$. Because $|\partial X| \geq |\delta X|$, the edge expansion is always at least as large as the vertex expansion, i.e., $h(G) \geq c(G)$, so this is a weaker lower bound than we showed above on gate-based quantum routing. In particular, the star graph has $h(S_n) = \Theta(1)$ so our lower bound gives $\mathrm{hqrt}(S_n) = \Omega\left(1/h(S_n)\right) = \Omega(1)$. Since $\mathrm{qrt}(S_n) = \Omega(n)$, this does not rule out the possibility of a large separation between Hamiltonian and gate-based quantum routing.

To prove the lower bound on Hamiltonian routing, we use the continuous analog of STE, the *small incremental entangling* (SIE) theorem, adapted to our setting. SIE was conjectured by Kitaev [63] and first proven in Ref. [48].

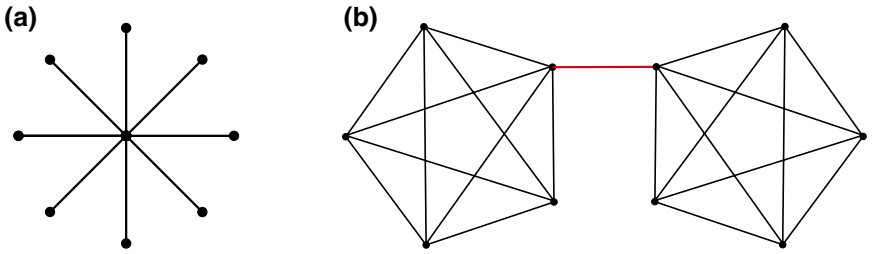


FIG. 4. Gate-based routing models cannot be separated up to a constant additive factor from classical routing on the star graph, but our lower bound on Hamiltonian routing is trivial in this case because of its $\Omega(1)$ edge expansion. Classical and Hamiltonian routing cannot be separated on the barbell graph because of its $O(1/n)$ edge expansion.

**Lemma II.9.** [Small incremental entangling (SIE)]. *Given a finite joint system $X\bar{X}$, any Hamiltonian $H$ with support only on $\delta X \delta\bar{X}$ and any initial pure state $\rho$, the entanglement capacity $\Gamma(H,\rho)$ is bounded as*

$$\Gamma(H,\rho) := \frac{dS_X(\rho(t))}{dt} \leq \alpha \|H\| \log d, \tag{46}$$

*where $\rho(t) = U(t)\rho U(t)^\dagger$ for $U(t) = e^{-iHt}$, $0 < \alpha \leq 4$ is a constant, and $d = \min(|\delta X|, |\delta\bar{X}|)$.*

It is conjectured that $\alpha = 2$ [63] but the best-known bound gives $\alpha = 4$ [64]. No generality is lost by assuming pure states since we can add an ancillary purification system $C$ to $X$ without loss of generality. The resulting state on the joint system $X\bar{X}C$ is pure and constrained by SIE. Since including $C$ as an ancilla can only increase the entanglement capacity (we can always ignore it), we see that the entanglement capacity is also bounded for mixed states on $X\bar{X}$.

We can derive another expression for $\Gamma(H,\rho)$ by writing

$$\Gamma(H,\rho) = -\frac{d}{dt}\operatorname{Tr}(\rho_X(t)\log\rho_X(t)) \tag{47}$$

$$= -\operatorname{Tr}(\frac{d\rho_X(t)}{dt}\log\rho_X(t)) \tag{48}$$

$$= i\operatorname{Tr}(\operatorname{Tr}_{\bar{X}}([H,\rho])\log\rho_X(t)), \tag{49}$$

where we use the Schrödinger equation $i(d\rho/dt) = [H,\rho]$ (setting $\hbar = 1$). We see that the entanglement capacity is linear in $H$.

The evolution of a system with interaction graph $G$, for any $X \subseteq V(G)$, can be described by a Hamiltonian $H = H_X + H_{\bar{X}} + H_{\delta X \delta\bar{X}}$, where $H_Y$ only has support on the subsystem of vertices $Y \subseteq V(G)$. Operations local to $X$ or $\bar{X}$ do not generate entanglement, so

$$\Gamma(H,\rho) = \Gamma\left(H_{\delta X\delta\bar{X}},\rho\right). \tag{50}$$

We can verify this by first explicitly computing

$$\operatorname{Tr}_{\bar{X}}([H_{\bar{X}},\rho]) = 0 \tag{51}$$

because the partial trace is cyclic on the $\bar{X}$ subsystem. Second,

$$\Gamma(H_X,\rho) = i\operatorname{Tr}([H_X,\rho_X(t)]\log\rho_X(t)) = 0 \tag{52}$$

because $\log\rho_X(t)$ commutes with $\rho_X(t)$ and the trace is cyclic. By linearity, Eq. (50) holds, and we can restrict ourselves to consider only Hamiltonians of the form $H_{\delta X\delta\bar{X}}$.

Now we can bound the entanglement capacity of any edge cut in the graph as specified by the edge boundary of a vertex subset $X$. A slightly weaker result up to constant factors was proved in Ref. [47] by using bounds on the entanglement capacity of bipartite product Hamiltonians [25] instead of SIE.

**Theorem II.10.** *Given any $X \subseteq V(G)$ and any pure state $\rho$, the entanglement capacity of a Hamiltonian $H$ with support only on the joint subsystem $\delta X\delta\bar{X}$ satisfies*

$$\Gamma(H,\rho) = \frac{dS_X(\rho(t))}{dt} \leq \frac{3\pi\alpha}{4}|\partial X|, \tag{53}$$

*for $\alpha$ the constant of SIE.*

*Proof.* We decompose the Hamiltonian into a sum of local terms $H = \sum_{e\in\partial X} H^{(e)}$ where each $H^{(e)}$ is a two-qubit Hamiltonian acting only on the ends of the edge $e$.

By linearity,

$$\Gamma(H,\rho) = \Gamma\left(\sum_{e\in\partial X} H^{(e)},\rho\right) = \sum_{e\in\partial X}\Gamma\left(H^{(e)},\rho\right). \quad (54)$$

We bound each term by SIE (Lemma II.9):

$$\sum_{e\in\partial X}\Gamma\left(H^{(e)},\rho\right) \le \alpha\sum_{e\in\partial X}\|H^{(e)}\|. \quad (55)$$

By unitary similarity (which the norm is invariant under), we can rewrite each term in canonical form (40) and apply our normalization condition such that $\sum_{e\in\partial X}\|H^{(e)}\| \le (3\pi/4)|\partial X|$. ■

Using this relation of entanglement capacity to edge cuts in the graph, we show a lower bound on the time to perform state preparation in the Hamiltonian model dependent on the edge cut.

**Corollary II.11.** *Given an initial pure state $\rho_0$ and target pure state $\rho$ on a bipartite system $X\bar{X}$, define the change in entanglement entropy $\Delta S_X := |S_X(\rho) - S_X(\rho_0)|$. Then any Hamiltonian unitary evolution from $\rho_0$ to $\rho$ restricted by interaction graph $G$ must have evolution time*

$$t \ge \frac{4}{3\pi\alpha}\frac{\Delta S_X}{|\partial X|}. \quad (56)$$

*Proof.* The claim follows directly from Theorem II.10. ■

A lower bound on Hamiltonian routing follows since routing a particular state is a special case of state preparation.

**Theorem II.12.** *For any simple graph $G$,*

$$\mathrm{hqrt}(G) \ge \frac{8}{3\pi\alpha}\frac{1}{h(G)}. \quad (57)$$

*Proof.* We prepare the same initial state as in Theorem II.6, where we have one half of a Bell pair at each vertex $v \in V(G)$ that is entangled with an ancilla. To show a lower bound, we pick some $X \subseteq V(G)$ with $|X| \le |V(G)|/2$ and an associated ancilla space $x$, and pick a permutation $\pi$ that routes all vertices $v \in X$ to $\bar{X}$ arbitrarily and routes $|X|$ vertices $u \in \bar{X}$ to $X$ arbitrarily. Let the resulting state be our target state $\rho$. This gives $\Delta S_{Xx} = S_{Xx}(\rho) = 2|X|$. Corollary II.11 implies that the time to implement this state preparation and routing task is lower bounded as

$$\mathrm{hqrt}(G,\pi) \ge \frac{4}{3\pi\alpha}\frac{\Delta S_{Xx}}{|\partial X|} = \frac{8}{3\pi\alpha}\frac{|X|}{|\partial X|}. \quad (58)$$

We now maximize over all $X$ to lower bound the Hamiltonian routing time

$$\begin{aligned}\mathrm{hqrt}(G) &= \max_{\pi}\mathrm{hqrt}(G,\pi)\\ &\ge \max_{X:|X|\le|V(G)|/2}\frac{8}{3\pi\alpha}\frac{|X|}{|\partial X|} = \frac{8}{3\pi\alpha}\frac{1}{h(G)},\end{aligned} \quad (59)$$

as claimed. ■

One simple example where this rules out a separation between classical and Hamiltonian routing is the *barbell graph*, $C_{2n}$ [65]. The barbell graph consists of two complete graphs, $K_n$, connected by a single edge at some vertex in each complete graph, as shown in Fig. 4(b). Since $h(C_{2n}) \le 2/n$, Theorem II.12 implies the Hamiltonian routing time on this graph is lower bounded as

$$\mathrm{hqrt}(C_{2n}) \ge \frac{4n}{3\pi\alpha}. \quad (60)$$

By routing on its spanning tree, $\mathrm{rt}(C_{2n}) = O(n)$, so classical routing is tight up to a constant factor.

An entanglement capacity bound of $O(|\partial X|)$, as given by Theorem II.10, matches previous results on entanglement area laws for dynamics [49, theorem 1] on lattices of constant dimension. For graphs of superconstant degree, the distinction between bounds on the entanglement capacity proportional to edge cuts (for Hamiltonian routing) and vertex cuts (for gate-based quantum routing) are significant. In general, $|\partial X| \le |\delta X|\max_v d_v$. It remains an open question whether Hamiltonian routing can be separated by a superconstant factor from gate-based quantum routing, and in particular, if the Hamiltonian routing time can also be lower bounded by the vertex expansion $\Omega(1/c(G))$. However, we show in Sec. IV that a stronger model of Hamiltonian routing can be separated from gate-based quantum routing and its routing time cannot be lower bounded by $\Omega(1/c(G))$.

Another case that has been well studied is the path graph, $P_n$. Here, the *odd-even sort* [66] gives a simple classical routing algorithm that upper bounds the circuit depth by $n$. A simple bound on the vertex expansion of the path graph is $c(P_n) \le 2/n$, so $\mathrm{qrt}(P_n) \ge n-1$, matching the diameter lower bound (Theorem II.1) up to an additive constant. Thus, a constant-factor improvement over classical routing on the path is only possible in the Hamiltonian routing model. In that case, we have $h(P_n) \le 2/n$, giving $\mathrm{hqrt}(P_n) \ge 4n/(3\pi\alpha)$. This is slightly weaker (even if $\alpha = 2$) than a specialized bound of $4n/(3\pi\alpha_0) \approx 0.222n$, for $\alpha_0 \approx 1.912$, based on

the entanglement capacity [59]. Indeed, Ref. [67] shows that $\mathrm{hqrt}(P_n) \le (1-\varepsilon)n + O\left(\log^2 n\right)$ for a constant $\varepsilon \approx 0.034$, so, for large enough $n$, $\mathrm{hqrt}(P_n) < \mathrm{qrt}(P_n)$ with a constant-factor speedup.

## III. COMPARISON WITH CLASSICAL ROUTING

Fast classical routing algorithms are already known for some graph families [31,45]. An example is the family of grid graphs, which are Cartesian products of path graphs $P_{L_1}\square P_{L_2}$ with dimensions $L_1, L_2 \in \mathbb{N}$, where we know $\mathrm{rt}(P_{L_1}\square P_{L_2}) \le 2L_1 + L_2$. We can exclude a superconstant quantum advantage simply by the diameter lower bound (Theorem II.8).

In this section, we compare our quantum routing results with general bounds on classical routing. In particular, this gives more conditions for a superpolynomial separation. Our results and proofs are generalizations of results in Ref. [45] from regular graphs to irregular graphs.

In classical routing, we route a permutation $\pi$ in multiple time steps. We first assign to each vertex $v$ a *token* labeled $\pi(v)$. Then, in each time step, we perform SWAP gates on neighboring vertices to exchange their tokens with the constraint that each vertex participates in at most one SWAP. Routing terminates when all tokens have been moved to their destination vertices. The difficulty of classical routing on $G$ is characterized by the *routing number* [45]

$$\mathrm{rt}(G) := \max_{\pi} \mathrm{rt}(G,\pi), \tag{61}$$

where $\mathrm{rt}(G,\pi)$ is defined as the minimal number of time steps needed to implement the permutation $\pi$. Since gate-based routing generalizes SWAP-based routing, $\mathrm{qrt}(G,\pi) \le \mathrm{rt}(G,\pi)$ for any permutation $\pi$, and in particular, $\mathrm{qrt}(G) \le \mathrm{rt}(G)$.

### A. General classical routing

We now describe a classical routing algorithm that performs SWAPs along a set of walks (connecting each token with its destination) that are close to random. The number of SWAPs that act on the same vertices at the same time is bounded from above by the inverse spectral gap of the (normalized) graph Laplacian, leading to high parallelism in graphs with large spectral gap.

The set of vertices is isomorphic to an integer labeling, $V(G) \cong [n]$, so we identify each $v \in V(G)$ with a unique integer index. Let the adjacency matrix $A$ have entries

$$A_{uv} = \begin{cases} 1 & \text{if } (v,u) \in E(G), \\ 0 & \text{otherwise,} \end{cases} \tag{62}$$

for $v, u \in V(G)$, and let the diagonal matrix $T$ have entries $T_{vv} = d_v$ and 0 otherwise, for $d_v = (A\mathbf{1})_v$ the degree of $v$ and $\mathbf{1}$ the all-ones vector. Then the (normalized) graph Laplacian is $\mathcal{L} := \mathbb{1} - T^{-1/2}AT^{-1/2}$. The Laplacian is symmetric and positive semidefinite [57] and has a 0 eigenvalue for the eigenvector $T^{1/2}\mathbf{1}$. Let the *spectral gap*, $\lambda(G)$, be the smallest nonzero eigenvalue of $\mathcal{L}$.

In this section, we assume $n \ge 2$ and show a general bound on the routing number without attempting to minimize the constants. Let $v_1 v_2 \ldots$ denote a walk on the vertices $v_i \in V(G)$ that passes through $v_i$ at time step $i$. We consider memoryless random walks with transition probabilities denoted by $P_{vu} = \mathrm{P}[x_{i+1} = v \mid x_i = u]$. These probabilities form the transition matrix $P$ of the random walk on $G$. We choose the *lazy random walk* $P = (\mathbb{1} + AT^{-1})/2$, i.e.,

$$P_{vu} = \begin{cases} 1/2 & \text{if } u = v, \\ 1/(2d_u) & \text{if } (u,v) \in E(G), \\ 0 & \text{otherwise.} \end{cases} \tag{63}$$

In the following, we refer to lazy random walks simply as random walks. Note that we default to right multiplication with the transition matrix so our probability distributions can be interpreted as column vectors. Therefore, the probability that a random walk starting at $u$ is at $v$ after $i \in \mathbb{N}$ steps is given by

$$\mathrm{P}[x_i = v \mid x_0 = u] = \boldsymbol{e}(v)^{\mathsf{T}} P^i \boldsymbol{e}(u), \tag{64}$$

where $\boldsymbol{e}(v)$ is the column vector with a 1 in position $v$ and 0 otherwise. The stationary distribution of the walk $P$ is $\boldsymbol{\pi} := T\mathbf{1}$ since $PT\mathbf{1} = T\mathbf{1}$.

We first define a useful notion of interference between walks.

**Definition III.1:** (Interfering walks, Fig. 5). Two walks $W$ and $W'$ are said to *interfere* if $W_i = W'_j$ for some $i,j \in \mathbb{N}$ with $|i-j| \le 1$.

The condition that $|i-j| > 1$ ensures that tokens can be swapped along $W$ in parallel with token swaps along $W'$, namely a token being swapped along $W$ has SWAPs that overlap at a location for two time steps.

Now, let us perform a simple random walk of a given length starting at each vertex $u$ and call this walk $W(u)$. We show that, with high probability and for sufficiently long walks, the number of walks that interfere with a given walk can be bounded from above. This is a generalization of Ref. [45, lemma 2] to irregular graphs, where we explicitly analyze the dependence on the degree. In particular, the entries of $T/\min_v d_v$ are bounded from above by the *degree ratio*

$$d_* := \frac{\max_v d_v}{\min_v d_v}. \tag{65}$$

**Lemma III.2.** *Let $G$ be a connected simple graph on $n$ vertices and suppose $l \ge \ln(n)/\lambda(G)$. For every $v \in V(G)$, let*

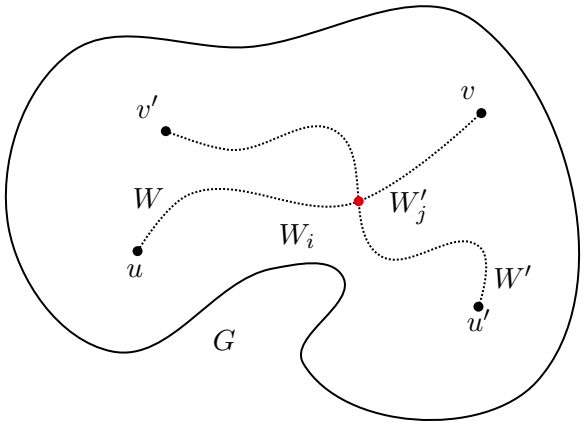


FIG. 5. Shown are walks $W$ from $u$ to $v$ and $W'$ from $u'$ to $v'$. Walks may intersect at a vertex (red) such that the $i$th location of $W$ and the $j$th location of $W'$ are the same, i.e., $W_i = W'_j$. We say that $W$ and $W'$ *interfere* if there exist $i,j$ with $|i-j| \le 1$ such that $W_i = W'_j$. We show that, with high probability, there exists a set of walks for a permutation $\sigma$ of order two on $V(G)$ that connect $v$ to $\sigma(v)$ such that the number of interfering walks can be bounded. SWAPs along sets of walks that do not interfere with each other significantly parallelize the routing process.

*$W(v)$ denote a random walk of length $l$ starting at vertex $v$. Let $I(v)$ denote the total number of other walks $W(u)$ that interfere with $W(v)$. Then with probability at most $n^{-20}$ there is a vertex $v \in V(G)$ with $I(v) > 30ld_*$.*

*Proof.* We wish to bound $I(v)$ for any $v \in V(G)$. We introduce an indicator random variable depending on the random walks $W(v)$:

$$X_{uv} := \begin{cases} 1 & \text{if } W(u) \text{ and } W(v) \text{ interfere,} \\ 0 & \text{otherwise.} \end{cases} \tag{66}$$

We include the random walk starting at $v$ in the total which only increases the expectation of $I(v)$. By summing over $u \in V(G)$ and including $v$, the expected value of $I(v)$ over random walks is bounded by

$$\mathbf{E}[I(v)] \le \mathbf{E}\left[\sum_u X_{uv}\right] = \sum_{u \ne v} \mathrm{P}[X_{uv} = 1] \tag{67}$$

$$= \sum_u \mathrm{P}\left[\bigvee_{i\in[l]} \bigvee_{j:|i-j|\le 1} W(v)_i = W(u)_j\right] \tag{68}$$

$$\le \sum_u \sum_{i\in[l]} \sum_{j:|i-j|\le 1} \mathrm{P}[W(v)_i = W(u)_j]. \tag{69}$$

Using Eq. (64), we have

$$\sum_u \sum_{\substack{i\in[l] \\ j:|i-j|\le 1}} \mathrm{P}[W(v)_i = W(u)_j]$$

$$= \sum_u \sum_{i\in[l]} \boldsymbol{e}(W(v)_i)^{\mathsf{T}} \sum_{j:|i-j|\le 1} P^j \boldsymbol{e}(u) \tag{70}$$

$$= \sum_{i\in[l]} \boldsymbol{e}(W(v)_i)^{\mathsf{T}} \sum_{j:|i-j|\le 1} P^j \mathbf{1}. \tag{71}$$

The transposed vector $e(W(v)_i)^{\mathsf{T}} P^j$ on the right-hand side has non-negative entries for all $i,j$, therefore an upper bound follows from substituting $\mathbf{1}$ by the entrywise larger vector $\boldsymbol{\pi} / \min_v d_v$ as

$$\sum_{i\in[l]} \boldsymbol{e}(W(v)_i)^{\mathsf{T}} \sum_{j:|i-j|\le 1} P^j \mathbf{1}$$

$$\le \sum_{i\in[l]} \boldsymbol{e}(W(v)_i)^{\mathsf{T}} \sum_{j:|i-j|\le 1} P^j \frac{\boldsymbol{\pi}}{\min_v d_v}. \tag{72}$$

The distribution $\boldsymbol{\pi}$ is stationary under the walk $P$, so

$$\sum_{i\in[l]} \boldsymbol{e}(W(v)_i)^{\mathsf{T}} \sum_{j:|i-j|\le 1} P^j \frac{\boldsymbol{\pi}}{\min_v d_v}$$

$$\le 3 \sum_{i\in[l]} \boldsymbol{e}(W(v)_i)^{\mathsf{T}} \frac{\boldsymbol{\pi}}{\min_v d_v} \tag{73}$$

$$\le 3ld_*, \tag{74}$$

since $T_{uu} \le \max_v d_v$ for any $u \in V(G)$. Therefore, $\mathbf{E}[I(v)] \le 3ld^*$.

We now bound the tail probability of $I(v)$. We use the multiplicative Chernoff bound, which states that for a random variable $Y = \sum_i Y_i$ with mean $\mu$ where the $Y_i$ are independent random variables, $\mathrm{P}[X > (1+\delta)\mu] \le \left(e^\delta/(1+\delta)^{1+\delta}\right)^\mu$ for any $\delta > 0$. We see that the Chernoff bound applies to $I(v) \le \sum_u X_{uv}$ since the walks $W(u)$ are independent (note that they may depend on $v$). Applying the Chernoff bound with $\delta = 9$, we have

$$\mathrm{P}[I(v) > 30ld_*] \le \mathrm{P}[I(v) > 10\,\mathbf{E}[I(v)]] < e^{3\ln(e^9/10^{10})ld_*}. \tag{75}$$

Given that $3\ln(e^9/10^{10}) < -42$ and using the lemma's assumption that $l \ge \ln(n)/\lambda(G)$, we obtain

$$e^{3\ln(e^9/10^{10})ld_*} < e^{-42ld_*} \le n^{-42d_*/\lambda(G)}. \tag{76}$$

We lower bound $d_* \ge 1$ and $1/\lambda(G) \ge 1 - 1/n \ge 1/2$ [57, lemma 1.7] (we assume $n \ge 2$) to obtain

$$n^{-42d_*/\lambda(G)} \le n^{-21}. \tag{77}$$

Since there are $n$ vertices, the probability that there exists a vertex $v$ with $I(v) > 30ld_*$ is at most $n^{-20}$. The lemma follows from the contrapositive. ■

We can "glue" together pairs of random walks starting at the $k/2$ pairs of vertices that are mapped to each other in a permutation of order two to obtain a set of $k$ glued walks. We show that, with high probability, no glued walk in this set will have many interfering other glued walks. This is an adaptation of Ref. [45, lemma 3] to irregular graph.

**Lemma III.3.** *Let $G$ be a simple connected graph on $n$ vertices, let $\sigma$ be a permutation of order two on $V(G)$ with $k$ vertices $v \in V(G)$ such that $\sigma(v) \neq v$, and let $l = 20/\lambda(G)\ln n$. Then there is a set of $k$ walks $W(v)$ of length $2l$, where both $W(v)$ and $W(\sigma(v))$ have endpoints $v$ and $\sigma(v)$ and traverse the same edges (in opposite directions), satisfying the following: if $I(v)$ denotes the total number of other walks $W(u)$ that interfere with either $W(v)$ or $W(\sigma(v))$, then $I(v) < 120ld_*$ for all $v$ with probability at least $1 - O\left(n^{-7}\right)$.*

*Proof.* We first show the existence of $k$ conditioned random walks (defined below) of length $l$, one for each vertex $v \in V(G)$ with $\sigma(v) \neq v$, that are close to random walks.

Define the probability of an *open random walk* starting at $v$ and ending at a random (not *a priori* specified) vertex $w \in V(G)$, after $t \in \mathbb{N}$ steps, as

$$P_v^{(t)}(w) := \mathrm{P}[W(v)_t = w] \tag{78}$$

$$= \sum_{v_2,\ldots,v_{t-1} \in V(G)} \mathrm{P}[W(v) = (v, v_2, \ldots, v_{t-1}, w)]. \tag{79}$$

We now define the *relative pointwise distance*, $\Delta : V(G) \to \mathbb{R}$, of $P_v^{(t)}(w)$ to the stationary distribution $\boldsymbol{\pi}$ as [57]

$$\Delta(t) := \max_{v,w} \frac{|P_v^{(t)}(w) - \boldsymbol{\pi}(w)|}{\boldsymbol{\pi}(w)}. \tag{80}$$

All random walks of length $l$ are close to stationary with respect to the relative pointwise distance since [57,68, theorem 1.16]

$$\Delta(l) \leq 2e^{-\frac{l\lambda(G)}{2}} \frac{|E(G)|}{\min_x d_x} = 2n^{-10} \frac{|E(G)|}{\min_x d_x} < 2n^{-8}, \tag{81}$$

where we use $|E(G)| < n^2$, and $\min_x d_x \geq 1$.

Now we compare the statistics of an open random walk $W(v)$ of length $l$ to a *conditioned random walk* where we condition on the last vertex being $w \in V(G)$, which is sampled according to the stationary distribution. The probability of a particular open random walk $W(v)$ can be related to that of the conditioned random walk by

$$\mathrm{P}[W(v)] = \sum_w \mathrm{P}[W(v)|W(v)_l = w] P_v^{(l)}(w) \tag{82}$$

$$\leq \sum_w \mathrm{P}[W(v)|W(v)_l = w](1 + 2n^{-8})\boldsymbol{\pi}(w) \tag{83}$$

$$= 2n^{-8} + \sum_w \mathrm{P}[W(v)|W(v)_l = w]\boldsymbol{\pi}(w), \tag{84}$$

and a corresponding lower bound can be derived similarly. Therefore, for large $n$ a conditioned random walk has vanishing deviation from an open random walk.

For all vertices $v$ with $\sigma(v) \neq v$, we now condition $W(v)$ and $W(\sigma(v))$ to have the same terminal vertex $w$, which is sampled once from $\boldsymbol{\pi}$. We call the combined walk of $W(v)$ followed by the reverse of $W(\sigma(v))$ the *glued walk* for $v$. There are $k$ glued walks that connect $v$ to $\sigma(v)$ in pairs.

Finally, we bound the number of interfering glued walks with high probability by applying Lemma III.2. We first arbitrarily partition the graph into $X \subseteq V(G)$ and $\bar{X}$ such that the vertices in each pair $(v, \sigma(v))$ lie in different parts. For any conditioned random walk $W(v)$, we can write the number of walk interferences $I(v)$ as a sum of two random variables, $I(v)_X$ and $I(v)_{\bar{X}}$, defined as the number of interferences with conditioned random walks $W(u)$ with $u$ from $X$ and $\bar{X}$, respectively, excluding $W(\sigma(v))$. Note that since the terminal vertices of each conditioned random walk in $X$ (or $\bar{X}$) are sampled independently, we can apply Lemma III.2 individually to $I(v)_X$ and $I(v)_{\bar{X}}$ [taking advantage of Eq. (84)]. We can then similarly bound $I(\sigma(v))_X$ and $I(\sigma(v))_{\bar{X}}$. The number of glued walks that interfere with a given $(v, \sigma(v))$ glued walk is a random variable bounded above by the sum

$$I(v) + I(\sigma(v)) = I(v)_X + I(v)_{\bar{X}} + I(\sigma(v))_X + I(\sigma(v))_{\bar{X}} \tag{85}$$

that can now be bounded from above by Lemma III.2. This gives $I(v) + I(\sigma(v)) < 120ld_*$ with probability at least $1 - O\left(n^{-7}\right)$. ■

The existence of walks between opposite ends of an order-two permutation with few intersections leads to a classical routing algorithm that divides the walks into disjoint sets that do not intersect. This adapts Ref. [57, theorem 4.10] to the irregular graph setting using our previous lemmas.

**Theorem III.4.** *Let $\sigma$ denote a permutation of order two on the vertex set of a connected graph $G$. Then, for $l = 20\ln(n)/\lambda(G)$,*

$$\mathrm{rt}(G, \sigma) \leq 2l(120ld_* + 1) = O\left(\frac{d_*}{\lambda(G)^2}\log^2 n\right). \tag{86}$$

*Proof.* Let $W(v)$ be a system of walks of length $2l$ satisfying Lemma III.3. Let $H$ be the graph whose vertices are the walks $W(v)$ and in which $W(v)$ and $W(u)$ are adjacent if there exist two indices $0 \le i,j \le l, |i-j| \le 1$, so $W(v)_i = W(u)_j$ or $W(v)_i = W(u)_{2l-j}$. By Lemma III.3, the maximum degree of $H$ is at most $120ld_*$ with high probability, hence by Brooks' theorem it is vertex colorable with at most $120ld_* + 1$ colors. We can therefore divide the walks $W(v)$ into at most $120ld_* + 1$ sets of independent walks of length $2l$.

We now present the routing algorithm. For each set of independent walks we sequentially do the following. For step $i$ with $1 \le i \le l$, we flip tokens along the edges numbered $i$ and $2l-1-i$ in each of the walks. After $l$ steps, the tokens at either end of the walk have been exchanged and the tokens not involved in any walk have not moved. After repeating this for all independent sets, all tokens have reached their destinations.

Since this routing algorithm succeeds with positive probability, there exists an algorithm achieving the claimed routing number. ■

Now we generalize to all permutations.

**Corollary III.5.** *For every connected simple graph $G$ and $l = 20\ln(n)/\lambda(G)$,*

$$\mathrm{rt}(G) \le 4l\,(120ld_* + 1) = O\left(\frac{d_*}{\lambda(G)^2}\log^2 n\right). \tag{87}$$

*Proof.* Any permutation of $V(G)$ can be written as a product of two permutations of order two. Use Theorem III.4 to route each sequentially to obtain the result. ■

To the best of our knowledge, Corollary III.5 provides novel upper bounds for certain irregular graphs. Of particular interest are irregular graphs where $d_*/\lambda(G)^2 = o(n)$. One such example is an Erdös-Rényi graph $G_{n,p}$, which is an $n$-vertex graph where each edge is independently present with some probability $p$. Hoffman *et al.* [69] showed that for $p \ge (1+\delta)\log n/n$, for constant $\delta > 0$, there is a constant $C(\delta)$ such that

$$|1-\lambda(G)| < \frac{C(\delta)}{\sqrt{p(n-1)}} = O\left(\frac{1}{\sqrt{n}}\right) \tag{88}$$

with high probability. Thus, we have that $\lambda(G) = \Omega(1)$ with high probability for such $p$ and large enough $n$. Moreover, the degree ratio $d_* \to 1$ for $n \to \infty$ with high probability, though it does not exactly equal 1 for finite $n$, giving some irregularity. Under these conditions, Corollary III.5 shows that $\mathrm{rt}(G_{n,p}) = O\left(\log^2 n\right)$ with high probability.

### B. Conditions for a superpolynomial separation

To compare our upper bound on the routing number and the Hamiltonian routing time lower bound, we bound the Hamiltonian routing time in terms of the spectral gap. We use the *Cheeger inequality* [57,70] that we state here without proof.

**Lemma III.6.** (Cheeger inequality). *For any connected graph $G$,*

$$2h_G \ge \lambda(G) > \frac{h_G^2}{2}, \tag{89}$$

*where the Cheeger constant is*

$$h_G := \min_{X \subseteq V(G):|X| \le |V(G)|/2} \frac{|\partial X|}{\sum_{x\in X} d_x}. \tag{90}$$

The edge expansion $h(G)$ relates to $h_G$ as

$$h(G) = \min_X \frac{|\partial X|}{|X|} \le \min_X \frac{|\partial X|}{\sum_{x\in X} d_x} \max_v d_v = h_G \max_v d_v, \tag{91}$$

where $X \subseteq V(G)$ and $|X| \le |V(G)|/2$. We now rewrite the Hamiltonian routing time lower bound, Theorem II.12, in terms of the spectral gap.

**Lemma III.7.** *For a connected graph $G$,*

$$\mathrm{hqrt}(G) \ge \frac{8}{3\pi\alpha \max_v d_v}\sqrt{\frac{1}{2\lambda(G)}}. \tag{92}$$

*Proof.* Using Theorem II.12, Eq. (91) and Lemma III.6, we have

$$\mathrm{hqrt}(G) \ge \frac{8}{3\pi\alpha \cdot h(G)} \tag{93}$$

$$\ge \frac{8}{3\pi\alpha \cdot h_G \max_v d_v} \tag{94}$$

$$> \frac{8}{3\pi\alpha \max_v d_v}\sqrt{\frac{1}{2\lambda(G)}} \tag{95}$$

as claimed. ■

A simple way to bound the slowdown when a classical routing algorithm is used instead of a Hamiltonian routing algorithm is the ratio of the routing times. By Corollary III.5 and Lemma III.7, we have

$$\frac{\mathrm{rt}(G)}{\mathrm{hqrt}(G)} = O\left(\frac{d_* \max_v d_v}{\lambda(G)^{3/2}}\log^2 n\right). \tag{96}$$

By routing on a spanning tree of $G$, we have $\mathrm{rt}(G) = O(n)$ [45], and, trivially, $\mathrm{hqrt}(G) = \Omega(1)$. Therefore, Eq. (96) is

nontrivial if

$$d_* \max_v d_v = o\left(\frac{n}{\log^2 n}\lambda(G)^{3/2}\right). \tag{97}$$

Moreover, it is possible to bound the routing number by a polynomial in the Hamiltonian routing time when $\lambda(G)$ is sufficiently small.

**Corollary III.8.** *For a simple connected graph $G$,* $\mathrm{rt}(G) = O(\mathrm{poly}(\mathrm{hqrt}(G)))$*, where* $\mathrm{poly}(x)$ *is a polynomial in $x$, if*

$$\frac{1}{h(G)} = \Omega\left(\mathrm{poly}\left(d_*, \frac{1}{\lambda(G)}, \log n\right)\right). \tag{98}$$

*Proof.* We wish to understand when $\mathrm{rt}(G)$ is some polynomial of $\mathrm{qrt}(G)$, i.e., $\mathrm{rt}(G) = O\left(\mathrm{hqrt}(G)^k\right)$ for constant $k \geq 1$. By Corollary III.5 and Theorem II.12, this is the case if

$$\frac{\log \mathrm{rt}(G)}{\log \mathrm{hqrt}(G)} = O\left(\frac{\log\left(\frac{d_*}{\lambda(G)}\log n\right)}{\log \frac{1}{h(G)}}\right) \tag{99}$$

is bounded by a constant. Equation (98) is a sufficient condition for this to hold. ∎

Similarly, we can use Cheeger's inequality and the diameter lower bound to obtain conditions for polynomially relating the routing number and the Hamiltonian routing time.

**Theorem III.9.** *For a simple connected graph $G$,* $\mathrm{rt}(G) = O(\mathrm{poly}(\mathrm{hqrt}(G)))$ *if*

$$\max\left(\frac{1}{\lambda(G)\max_v d_v}, \mathrm{diam}(G)\right) = \Omega\left(\mathrm{poly}\left(d_*, \frac{1}{\lambda(G)}, \log n\right)\right). \tag{100}$$

*Proof.* By Corollary III.5, Lemma III.7, and the diameter lower bound, this happens when

$$\frac{\log \mathrm{rt}(G)}{\log \mathrm{hqrt}(G)} = O\left(\frac{\log\left(\frac{d_*}{\lambda(G)}\log n\right)}{\log \max\left(\frac{1}{\lambda(G)\max_v d_v}, \mathrm{diam}(G)\right)}\right) \tag{101}$$

can be upper bounded by a constant. Equation (100) is a sufficient condition. ∎

We define a *separation* between the routing number and the Hamiltonian routing time as a function $f : \mathbb{R} \to \mathbb{R}$ such that

$$\mathrm{rt}(G) = \Omega\left(f(\mathrm{hqrt}(G))\right). \tag{102}$$

For example, a quadratic separation corresponds to $f(x) = x^2$. Theorem III.9 bounds the separation to polynomial for trivial cases such as graphs with $\mathrm{diam}(G) = \Omega(n^c)$ for $c > 0$ since $\mathrm{rt}(G) = O(n)$ from routing on a spanning tree [46]. Furthermore, there is no superpolynomial separation for bounded-degree graphs $G$, since $\mathrm{diam}(G) = \Omega(\log n)$ such that Eq. (100) simplifies to

$$\max\left(\frac{1}{\lambda(G)}, \log n\right) = \Omega\left(\mathrm{poly}\left(\frac{1}{\lambda(G)}, \log n\right)\right), \tag{103}$$

which is always satisfied. In particular, the separation is quadratic in the case $\lambda(G) = \Omega(1)$.

There are families of graphs where Theorem III.9 limits the separation to polynomial that cannot be obtained from the diameter lower bound on Hamiltonian routing, Theorem II.8, and results for classical routing on regular graphs [45]. An example is given by a family of irregular bounded-degree graphs constructed by Ref. [71] with arbitrary $h(G)$. The diameter of this graph family is $\Theta(1/h(G))$. Thus, when we pick a subpolynomial $1/h(G)$, i.e., $1/h(G) = o(n^c)$ for all constant $c > 0$, Theorem III.9 implies a polynomial limit on the separation that does not follow from the diameter lower bound on Hamiltonian routing.

However, there are graphs with large spectral gap but unbounded degree that are not restricted to a polynomial separation by Theorem III.9. The star graph $S_n$ has $\lambda(S_n) = 1$ [57] but is a poor vertex expander since $c(S_n) = O\left(n^{-1}\right)$, giving $\mathrm{qrt}(S_n) = \Theta(n)$. We cannot exclude the possibility that Hamiltonian quantum routing could exhibit a superpolynomial separation in this case, since our lower bound on $\mathrm{hqrt}(S_n)$ from Theorem II.12 is trivial. We take a first step toward exhibiting separations in the next section.

## IV. TOWARD A SEPARATION

We have given necessary conditions for a superpolynomial separation between Hamiltonian and classical routing, but we are not even aware of any superconstant separation. In this section, we describe separations in stronger routing models.

### A. A quadratic separation with ancillas

First we show that such a separation is possible in a variant of the Hamiltonian routing model that allows local ancilla qubits. The main idea of our construction is to consider a vertex bottleneck. This argument also shows that the Hamiltonian routing with ancillas cannot be lower bounded by $\Omega(1/c(G))$.

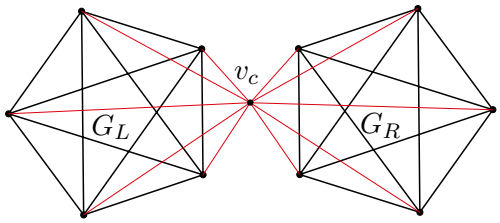


FIG. 6. The vertex barbell graph $B_{2n}$, for $n = 5$, consisting of two complete graphs connected through an additional vertex. We have $\lambda(G) \leq 2/n$ by Lemma III.6.

We show a separation on a graph $B_{2n}$, for $n \in \mathbb{N}$, that we call the *vertex barbell graph* (see Fig. 6). It consists of two complete graphs, $G_L$ and $G_R$, of $n$ vertices each and a central vertex $v_c$ where each complete graph is fully connected with $v_c$, forming two complete graphs of size $n+1$ joined at a vertex. We have $\mathrm{qrt}(B_{2n}) = \Theta(n)$: Corollary II.7 with $1/c(B_{2n}) \geq n$ implies the lower bound and a trivial SWAP routing strategy implies the upper bound. The Hamiltonian routing time is not similarly bounded because Theorem II.12 implies only a trivial lower bound, $\mathrm{hqrt}(B_{2n}) = \Omega(1)$, since $1/h(B_{2n}) = \Omega(1)$, making the vertex barbell graph a potential candidate for a separation.

We are able to show a separation in the stronger model of *Hamiltonian routing with ancillas*. This model is based on Hamiltonian routing with two additional assumptions: (i) each qubit has one associated ancilla qubit available, and (ii) the ancilla can perform a SWAP with its associated qubit in negligible time. We denote the *Hamiltonian routing time with ancilla* as

$$\mathrm{hqrt}_{\mathrm{a}}(G) := \max_{\pi} \mathrm{hqrt}_{\mathrm{a}}(G, \pi), \tag{104}$$

where $\mathrm{hqrt}_{\mathrm{a}}(G, \pi)$ is the routing time in the Hamiltonian routing with ancilla model of $\pi$ on graph $G$. As a point of comparison, we may define a modified gate-based quantum routing number $\mathrm{qrt}_{\mathrm{a}}(G)$ analogously. Due to the vertex bottleneck, we still have $\mathrm{qrt}_{\mathrm{a}}(B_{2n}) = \Theta(n)$.

We can use a protocol for fast state transfer [72] to implement Hamiltonian routing with ancillas for the hard case of routing on $B_{2n}$.

**Theorem IV.1.** *Given a vertex barbell graph $B_{2n}$ and a permutation $\sigma$ that permutes all vertices from $G_L$ to $G_R$ and vice versa, we have*

$$\mathrm{hqrt}_{\mathrm{a}}(B_{2n}, \sigma) = O\left(\sqrt{n}\right). \tag{105}$$

*Proof.* We define a Hamiltonian to construct a $W$ state [72],

$$W(x, \mathcal{S}) := \sum_{v \in \mathcal{S}} c_x^\dagger c_v + \mathrm{h.c.}, \tag{106}$$

where $\mathcal{S} \subseteq V(B_{2n})$, $x \in V(B_{2n}) \setminus \mathcal{S}$, and $c_y = |0\rangle_y \langle 1|_y$ (respectively, $c_y^\dagger$) are annihilation (respectively, creation) operators acting on qubit $y \in V(B_{2n})$. Evolving for time $\pi/(2\sqrt{|\mathcal{S}|})$ with initial state $|\psi\rangle = a_0|0\rangle + a_1|1\rangle$ on $x$, we have

$$e^{-iW(x,\mathcal{S})T}(a_0|0\rangle_x + a_1|1\rangle_x)|0\rangle_{\mathcal{S}} = |0\rangle_x(a_0|0\rangle_{\mathcal{S}} + a_1|W\rangle_{\mathcal{S}}), \tag{107}$$

where $|W\rangle := 1/\sqrt{|\mathcal{S}|}\sum_{v\in\mathcal{S}} c_v^\dagger|0\rangle_{\mathcal{S}}$ is the $W$ state over the qubits $\mathcal{S}$ (an equal superposition over Hamming weight 1 strings).

The protocol is then as follows. We first use (fast) SWAPs between each qubit and its ancilla so all data qubits in the graph are in the state $|0\rangle$. We now pick some vertex $v \in V(G_L)$ and show how to route the state originally at $v$ to $\sigma(v)$. We SWAP the data qubit at $v$ with its ancilla to return $v$ to its initial state. Then we evolve by the Hamiltonian $W(v, V(G_L) \setminus \{v\})$ to encode the state on $v$ on the data qubits associated with $V(G_L) \setminus \{v\}$, creating a state similar to Eq. (107), followed by the inverse operation $W(v_c, V(G_L) \setminus \{v\})^\dagger$. Overall, this sends the state from $v$ to the central vertex $v_c$ in total time $2T = \pi/\sqrt{n-1}$ (see also Fig. 7). We repeat this process to transfer the qubit from $v_c$ to $\sigma(v)$ in time $2T$. Then we SWAP the qubit at $\sigma(v)$ with its ancilla. If the qubit that is now at $\sigma(v)$ needs to be routed, we follow an analogous procedure and send it to $\sigma(\sigma(v))$. If it does not, we pick some other vertex in $V(G_R)$ that still needs to be routed. We iterate in this way, alternately handling a vertex from $G_L$, then $G_R$, until all vertices are routed to their destination ancillas. Finally, we simultaneously SWAP all qubits with their ancillas to finish the routing. The total time is $4T \cdot 2n = O\left(\sqrt{n}\right)$. ■

We can now generalize the algorithm to all permutations on $B_{2n}$.

**Corollary IV.2.** *For the vertex barbell graph $B_{2n}$, we have*

$$\mathrm{hqrt}_{\mathrm{a}}(B_{2n}) = O\left(\sqrt{n}\right). \tag{108}$$

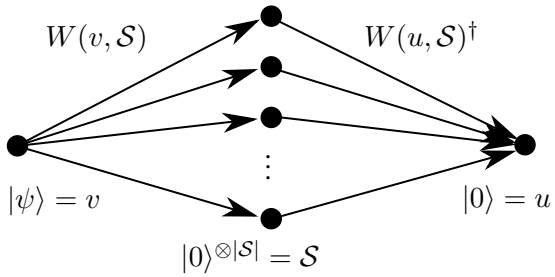


FIG. 7. In our routing protocol for the vertex barbell graph, we transfer the state $|\psi\rangle$ on qubit $v$ to $u$ by using the intermediate qubits $\mathcal{S}$ as ancillas (in the $|0\rangle^{\otimes|\mathcal{S}|}$ state). The operation $W(v, \mathcal{S})$ encodes $|\psi\rangle$ in a subspace spanned by $|0\rangle^{\otimes|\mathcal{S}|}$ and the $W$ state in time $\pi/(2\sqrt{|\mathcal{S}|})$ [72]. Since this procedure is unitary, we can use its inverse to transfer the state to $u$. We repeat this procedure in $G_R$ to transfer the state to its destination.

*Proof.* Let $\sigma$ be any permutation of the vertices $V(G)$. First, we SWAP $\sigma(v_c)$ with its ancilla, SWAP $v_c$ with $\sigma(v_c)$, and finally SWAP $\sigma(v_c)$ with its ancilla again. Then, we route all vertices that are permuted only within $G_L$ or $G_R$ in $O(1)$ time using SWAPs since $\mathrm{rt}(K_n) \leq 2$ [45]. Consider now the vertex barbell subgraph of the remaining vertices that need to move between $G_L$ and $G_R$, together with $v_c$. This routing can be done in time $O\left(\sqrt{n}\right)$ by Theorem IV.1 and starting at $\sigma(v_c)$. Finally, we SWAP $\sigma^{-1}(v_c)$ with its ancilla, and then with $v_c$. ■

Corollary IV.2 shows a quadratic separation

$$\mathrm{qrt_a}(B_{2n}) = \Omega\left(\mathrm{hqrt_a}(B_{2n})^2\right). \tag{109}$$

It also shows $\mathrm{hqrt_a}(B_{2n}) \notin \Omega\left(c(B_{2n})^{-1}\right) = \Omega(n)$, so Corollary II.7 does not generalize to Hamiltonian routing with ancillas.

### B. Optimal routing with fast local interactions

In this section, we show optimal routing for stronger models of classical and Hamiltonian routing that allow arbitrarily fast interactions within partitions of the graph $G$. Given a partition $X \subsetneq V(G)$, let us define the *fast classical routing model* as classical routing with arbitrarily fast SWAP operations within the vertex-induced subgraphs $G[X]$ and $G[\bar{X}]$. Then we can define the *$X$-fast routing number* of $G$, $\mathrm{rt}_f(G,X)$, as the worst-case time to route any permutation in the fast classical routing model for a given graph $G$ and partition $X$. We denote with subscript "$f$" that the interactions within the partitions $X$ and $\bar{X}$ are fast.

We now show that $\mathrm{rt}_f(G,X) = \lceil |X|/|m(\partial X)| \rceil$ for any connected simple graph $G$ and partition $X \subseteq V(G)$ with $|X| \leq |V(G)|/2$. The upper bound is given by the following routing algorithm.

**Theorem IV.3.** *For any connected simple graph $G$ and partition $X \subseteq V(G)$ with $|X| \leq |V(G)|/2$,*

$$rt_f(G,X) \leq \lceil \frac{|X|}{|m(\partial X)|} \rceil. \tag{110}$$

*Proof.* For any permutation $\pi$ to be routed, call the $k$ vertices $x \in X$ such that $\pi(x) \notin X$ *marked*. Similarly, we mark the vertices $x' \in \bar{X}$ such that $\pi(x') \notin \bar{X}$. Fix a maximum matching $m(\partial X) = \{(x_i, x_i')\}_{i=1}^{k}$ with $x_i \in X$ and $x_i' \in \bar{X}$. We repeat the following two steps for $\lceil |X|/|m(\partial X)| \rceil$ times:

1. Route as many marked vertices in $X$ as possible to $x_1, \ldots, x_k$ in order, and route as many marked vertices in $\bar{X}$ as possible to $x_1', \ldots, x_k'$ in order.
2. Perform parallel SWAPs along $(x_i, x_i')$ for all $i$ less than the number of remaining marked vertices.

This routes all marked vertices to their destination partitions. The only contribution to the $X$-fast routing number of $G$ is a unit contribution of Item 2 every iteration. Finally, we route all qubits within $X$ and $\bar{X}$ to their destinations using fast SWAPs. ■

The lower bound (28) on the gate-based quantum routing number applies to the model with fast interactions since STE still upper bounds the change in entropy for any unitary acting on $m(\partial X)$. The lower bound (rounded up, since the routing number cannot be a fraction) is attained by the $X$-fast classical routing algorithm for $G$, and thus the algorithm is optimal for all gate-based models.

Given a partition $X \subsetneq V(G)$, let us define the *fast Hamiltonian routing model* as Hamiltonian routing with ancilla and arbitrary interactions within the vertex-induced subgraphs $G[X]$ and $G[\bar{X}]$. Then we can define the *$X$-fast Hamiltonian routing time* of $G$, $\mathrm{hqrt}_{fa}(G,X)$, as the worst-case time to route any permutation in the fast Hamiltonian routing model for a given graph $G$ and partition $X$. The additional subscript "$a$" indicates the presence of an ancilla at every qubit.

We give an $X$-fast Hamiltonian routing algorithm that attains the lower bound on Hamiltonian routing up to a constant factor $3\pi\alpha/8$ (where $\alpha$ is the constant of SIE, Lemma II.9) for any graph $G$. We first prove that it is possible to perform any two-qubit unitary in the fast Hamiltonian routing model in time $O(1/|\partial X|)$.

**Lemma IV.4.** *For a connected simple graph $G$, any two-qubit unitary $U$ can be performed in the $X$-fast Hamiltonian routing model with partition $X \subsetneq V(G)$ in time at most $1/|\partial X|$.*

*Proof.* We show how to perform a CZ operation between $v, u \in V(G)$ in time $t = 1/3|\partial X|$. By a decomposition of $U$ into at most three CZ operations plus single-qubit operations [61], the result follows. The result is trivial if $v$ and $u$ are both within $X$ or $\bar{X}$.

Suppose, without loss of generality, $v \in X$ and $u \in \bar{X}$. We use (fast) SWAPs between qubits on the boundary $\delta X \cup \delta\bar{X} \setminus \{u, v\}$ and their ancillas. Suppose the qubit at $v$ is in the state $a_0|0\rangle + a_1|1\rangle$ and $u$ is in the state $a_0'|0\rangle + a_1'|1\rangle$ (by linearity, the protocol also works if $v$ and $u$ are initially entangled with other qubits). We then encode the state of $v$ onto $\delta\bar{X}$ as

$$a_0|0\ldots0\rangle + a_1|1\ldots1\rangle = a_0|\bar{0}\rangle + a_1|\bar{1}\rangle \tag{111}$$

by fast unitaries. Similarly, with some abuse of notation, we encode the state of $u$ onto $\delta X$ as $a_0'|\bar{0}\rangle + a_1'|\bar{1}\rangle$ where we disregard the different register sizes with the overline notation.

Notice that a Hamiltonian on $x, x' \in V(G)$

$$\frac{3\pi}{4}(\mathbb{1} - Z_x - Z_{x'} + Z_x Z_{x'}) = 3\pi|11\rangle\langle 11|_{xx'}, \tag{112}$$

with $Z_x$ a Pauli-$Z$ operator acting on the qubit at $x$, consists of local terms and one normalized $ZZ$ interaction. We can therefore evolve by the Hamiltonian

$$H = 3\pi \sum_{xx' \in \partial X} \mathbb{1}_{\overline{xx'}} \otimes |11\rangle\langle 11|_{xx'}, \tag{113}$$

where $\mathbb{1}_{\overline{xx'}}$ is the identity operator on all subsystems besides $xx'$, using fast local unitaries and $ZZ$ interactions along each edge in $\partial X$. By commutativity of the terms in $H$, we see that

$$\begin{aligned} &e^{-itH}(a_0|\bar{0}\rangle + a_1|\bar{1}\rangle) \otimes (a_0'|\bar{0}\rangle + a_1'|\bar{1}\rangle) \\ &\quad = a_0 a_0'|\bar{0}\bar{0}\rangle + a_0 a_1'|\bar{0}\bar{1}\rangle + a_0' a_1|\bar{1}\bar{0}\rangle + a_1 a_1' e^{-i3\pi t|\partial X|}|\bar{1}\bar{1}\rangle. \end{aligned} \tag{114}$$

After performing our initial SWAP and encoding operations in reverse, we apply a CZ operation between $u$ and $v$ in time $t$. ■

With the ability to quickly perform arbitrary two-qubit operations in the fast Hamiltonian routing model, we give a routing algorithm with the $X$-fast Hamiltonian routing time of $G$ upper bounded by $|X|/|\partial X|$.

**Theorem IV.5.** *For a connected simple graph $G$ and partition $X \subsetneq V(G)$,*

$$hqrt_{fa}(G, X) \leq \frac{|X|}{|\partial X|}. \tag{115}$$

*Proof.* Let $\pi$ be any permutation to be routed, and suppose there are $k$ *marked* vertices $x \in X$ such that $\pi(x) \notin X$. Then there are also $k$ marked vertices $x' \in \bar{X}$ such that $\pi(x') \notin \bar{X}$. We perform $k$ SWAPs between each pair of marked vertices, which, by Lemma IV.4, can be done in time $k/|\partial X|$. Finally, we use fast local SWAPs to route all qubits to their destination. The result follows since $k \leq |X|$. ■

Theorem II.10 also bounds the entanglement capacity of any Hamiltonian acting on the edge boundary $\partial X$ in the fast Hamiltonian model. Therefore, Eq. (58) implies

$$\mathrm{hqrt}_{fa}(G, X) \geq \frac{8}{3\pi\alpha}\frac{|X|}{|\partial X|} \tag{116}$$

for any $X \subsetneq V(G)$ with $|X| \leq |V(G)|/2$. It follows that Theorem IV.5 is tight up to a multiplicative constant $3\pi\alpha/8 \leq 4.713$.

When we compare the $X$-fast Hamiltonian routing time with the $X$-fast classical routing time (even with ancilla), we see that

$$\frac{\mathrm{rt}_f(G, X)}{\mathrm{hqrt}_{fa}(G, X)} = \Theta\left(\frac{|\partial X|}{|m(\partial X)|}\right). \tag{117}$$

The vertex barbell graph with the partition $V(G_L)$ is an example where a speedup of $\Theta(n)$ is realized.

## V. CONCLUSION

In this paper, we have explored the power of gate-based and Hamiltonian models of quantum routing, investigating both lower bounds and separations. We showed conditions on the spectrum of the architecture graph for a superpolynomial separation. In particular, our conditions exclude bounded-degree graphs from exhibiting a superpolynomial separation. We also gave an example graph [71], where diameter-based lower bounds and known classical routing algorithms [45] cannot exclude such a separation.

One natural open question is whether the star graph $S_n$, which has $\lambda(S_n) = 1$, can exhibit a superconstant quantum routing separation. While our results imply that gate-based quantum routing essentially gives no improvement over classical routing since $\mathrm{qrt}(S_n) \geq n - 1$, the same cannot be said for Hamiltonian routing, for which the corresponding lower bound is trivial. In fact, if the Hamiltonian model is strengthened by allowing a constant number of ancillas per qubit, a quadratic separation holds on the vertex barbell graph, which also exhibits a similar vertex bottleneck. By allowing fast interactions within certain regions of the graph, we can give optimal routing algorithms for gate-based and Hamiltonian models and exhibit a speedup from $\Theta(n)$ for gate-based models to $O(1)$ for Hamiltonian models.

Our depth (or time) lower bounds can be strengthened to include computational models with local operations and classical communication (LOCC). LOCC models give a stronger class of quantum routing and would allow, e.g., teleportation to bridge long distances. Trivially, this can exceed the Lieb-Robinson velocity [27] and seemingly invalidates simple lower bounds based on the diameter of the graph. Piroli *et al.* [50] showed LOCC circuit lower bounds on state preparation for lattices and inspired us to show similar state-preparation results for general interaction graphs and to lower bound routing. Since our depth (and time) lower bounds follow from entropic arguments and the entropy is nonincreasing under LOCC, we see that STE, SIE, and our state-preparation bounds (Lemma II.4 and Corollary II.11) generalize to models including LOCC when the entropy is nondecreasing. Thus our quantum routing bounds (Theorems II.6 and II.12) also generalize to models including LOCC. How much stronger models of routing with LOCC can be is studied in Ref. [73].

## ACKNOWLEDGMENTS

We thank Minh Tran for suggesting the possibility of using the state transfer protocol of Ref. [72], Michael Gullans for pointing out properties of Erdös-Rényi graphs, and Dhruv Devulapalli and Andrew Guo for helpful discussions.

A.B. and A.V.G. acknowledge funding by the NSF PFCQC program, ARO MURI, DoE QSA, DoE ASCR Quantum Testbed Pathfinder program (Award No. DE-SC0019040), NSF QLCI (Award No. OMA-2120757), DoE ASCR Accelerated Research in Quantum Computing program (Award No. DE-SC0020312), DARPA SAVaNT ADVENT, AFOSR, AFOSR MURI, and U.S. Department of Energy Award No. DE-SC0019449. E.S. and A.M.C. acknowledge support by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Quantum Testbed Pathfinder program (Award No. DE-SC0019040) and the U.S. Army Research Office (MURI Award No. W911NF-16-1-0349). E.S. acknowledges support from an IBM PhD Fellowship.


## APPENDIX: ASYMPTOTIC EQUIVALENCE OF THE MATCHING EXPANSION AND VERTEX EXPANSION

We show that the matching expansion (38) is equivalent to the vertex expansion, i.e., $c(G) = \Theta(m(G))$. The lower bound $m(G) \leq c(G)$ follows from the trivial bound $|m(\partial X)| \leq |\delta X|$ for any $X$. The following theorem provides the upper bound.

**Theorem A.1.** *For any simple graph G,*

$$c(G) \leq 2m(G) + O\left(m(G)^2\right). \tag{A1}$$

*Proof.* We note that $m(G) \in [0,1]$. If $m(G) = 1$, then the theorem holds since $c(G) \in [0,1]$.

We now consider the case $m(G) \in [0,1)$. Let $X \subseteq V(G)$ be a partition that attains the minimum in the matching expansion, and $Y \subseteq V(G)$ is the set of vertices in $m(\partial X)$. The set $X' := X \setminus Y$ is nonempty because $m(G) < 1$. We show

$$|\delta X'| \leq |Y| = 2|m(\partial X)|. \tag{A2}$$

Suppose, toward a contradiction, that there are adjacent vertices $x' \in X'$ and $x \in \overline{X'} \setminus Y$. Then $m(\partial X)$ is not maximal since $m(\partial X) \cup \{(x,x')\}$ is a larger matching. Therefore, $\delta X'$ must consist only of vertices in $Y$, giving $|\delta X'| \leq |Y|$ as claimed.

It follows from Eq. (A2) that

$$c(G) \leq \frac{|\delta X'|}{|X'|} \leq \frac{2|m(\partial X)|}{|X'|} = \frac{2|m(\partial X)|}{|X| - |m(\partial X)|} = \frac{2m(G)}{1 - m(G)}. \qquad \blacksquare$$


[1] D. Kielpinski, C. Monroe, and D. J. Wineland, Architecture for a large-scale ion-trap quantum computer, Nature **417**, 709 (2002).

[2] P. Murali, D. M. Debroy, K. R. Brown, and M. Martonosi, in: *ISCA '20: Proceedings of the ACM/IEEE 47th Annual International Symposium on Computer Architecture* (2020), ISBN: 9781728146614.

[3] C. Monroe and J. Kim, Scaling the Ion Trap Quantum Processor, Science **339**, 1164 (2013).

[4] C. Monroe, R. Raussendorf, A. Ruthven, K. R. Brown, P. Maunz, L.-M. Duan, and J. Kim, Large-scale modular quantum-computer architecture with atomic memory and photonic interconnects, Phys. Rev. A **89**, 022317 (2014).

[5] T. Brecht, W. Pfaff, C. Wang, Y. Chu, L. Frunzio, M. H. Devoret, and R. J. Schoelkopf, Multilayer microwave integrated quantum circuits for scalable quantum computing, npj Quantum Inf. **2**, 16002 (2016).

[6] N. C. Jones, R. Van Meter, A. G. Fowler, P. L. McMahon, J. Kim, T. D. Ladd, and Y. Yamamoto, Layered Architecture for Quantum Computing, Phys. Rev. X **2**, 031007 (2012).

[7] A. Yu. Kitaev, Quantum computations: Algorithms and error correction, Russ. Math. Surv. **52**, 1191 (1997).

[8] D. Bluvstein, H. Levine, G. Semeghini, T. T. Wang, S. Ebadi, M. Kalinowski, A. Keesling, N. Maskara, H. Pichler, M. Greiner, V. Vuletić, and M. D. Lukin, A quantum processor based on coherent transport of entangled atom arrays, Nature **604**, 451 (2022).

[9] J. M. Pino, J. M. Dreiling, C. Figgatt, J. P. Gaebler, S. A. Moses, M. S. Allman, C. H. Baldwin, M. Foss-Feig, D. Hayes, K. Mayer, C. Ryan-Anderson, and B. Neyenhuis, Demonstration of the trapped-ion quantum CCD computer architecture, Nature **592**, 209 (2021).

[10] H. Bombin, I. H. Kim, D. Litinski, N. Nickerson, M. Pant, F. Pastawski, S. Roberts, and T. Rudolph, Interleaving: Modular architectures for fault-tolerant photonic quantum computing, March 15, (2021).

[11] S. Bose, Quantum Communication through an Unmodulated Spin Chain, Phys. Rev. Lett. **91**, 207901 (2003).

[12] S. Bose, Quantum communication through spin chain dynamics: An introductory overview, Contemp. Phys. **48**, 13 (2007).

[13] M. Christandl, N. Datta, T. C. Dorlas, A. Ekert, A. Kay, and A. J. Landahl, Perfect transfer of arbitrary states in quantum spin networks, Phys. Rev. A **71**, 032312 (2005).

[14] L. Campos Venuti, C. Degli Esposti Boschi, and M. Roncaglia, Qubit Teleportation and Transfer across Antiferromagnetic Spin Chains, Phys. Rev. Lett. **99**, 060401 (2007).

[15] C. Di Franco, M. Paternostro, and M. S. Kim, Perfect State Transfer on a Spin Chain without State Initialization, Phys. Rev. Lett. **101**, 230502 (2008).

[16] L. Banchi, A. Bayat, P. Verrucchi, and S. Bose, Nonperturbative Entangling Gates between Distant Qubits Using Uniform Cold Atom Chains, Phys. Rev. Lett. **106**, 140501 (2011).

[17] N. Y. Yao, L. Jiang, A. V. Gorshkov, Z.-X. Gong, A. Zhai, L.-M. Duan, and M. D. Lukin, Robust Quantum State Transfer in Random Unpolarized Spin Chains, Phys. Rev. Lett. **106**, 040505 (2011).

[18] C. Albanese, M. Christandl, N. Datta, and A. Ekert, Mirror Inversion of Quantum States in Linear Registers, Phys. Rev. Lett. **93**, 230502 (2004).

[19] T. Shi, Y. Li, Z. Song, and C.-P. Sun, Quantum-state transfer via the ferromagnetic chain in a spatially modulated field, Phys. Rev. A **71**, 032309 (2005).

[20] P. Karbach and J. Stolze, Spin chains as perfect quantum state mirrors, Phys. Rev. A **72**, 030301(R) (2005).

[21] R. Raussendorf, Quantum computation via translation-invariant operations on a chain of qubits, Phys. Rev. A **72**, 052301 (2005).

[22] J. Fitzsimons and J. Twamley, Globally Controlled Quantum Wires for Perfect Qubit Transport, Mirroring, and Computing, Phys. Rev. Lett. **97**, 090502 (2006).

[23] W. Dür, G. Vidal, J. I. Cirac, N. Linden, and S. Popescu, Entanglement Capabilities of Nonlocal Hamiltonians, Phys. Rev. Lett. **87**, 137901 (2001).

[24] A. M. Childs, D. W. Leung, F. Verstraete, and G. Vidal, Asymptotic entanglement capacity of the Ising and anisotropic Heisenberg interactions, Quantum Inf. Comput. **3**, 97 (2003).

[25] A. M. Childs, D. W. Leung, and G. Vidal, Reversible simulation of bipartite product hamiltonians, IEEE Trans. Inf. Theory **50**, 1189 (2004).

[26] C. H. Bennett, A. W. Harrow, D. W. Leung, and J. A. Smolin, On the capacities of bipartite hamiltonians and unitary gates, IEEE Trans. Inf. Theory **49**, 1895 (2003).

[27] E. H. Lieb and D. W. Robinson, The finite group velocity of quantum spin systems, Commun. Math. Phys. **28**, 251 (1972).

[28] M. Saeedi, R. Wille, and R. Drechsler, Synthesis of quantum circuits for linear nearest neighbor architectures, Quantum Inf. Process. **10**, 355 (2011). ISSN:1573-1332,.

[29] A. Lye, R. Wille, and R. Drechsler, in *The 20th Asia and South Pacific Design Automation Conference* (Jan. 2015), p. 178.

[30] A. Zulehner and R. Wille, in *Proceedings of the 24th Asia and South Pacific Design Automation Conference on - ASPDAC '19* (2019).

[31] A. M. Childs, E. Schoute, and C. M. Unsal, in *14th conference on the theory of quantum computation, communication and cryptography (tqc 2019)*, Vol. 135, edited by W. van Dam and L. Mancinska, Leibniz International Proceedings in Informatics (LIPIcs) (2019), 3:1–3:24, ISBN: 978-3-95977-112-2.

[32] T. S. Metodi, D. D. Thaker, A. W. Cross, F. T. Chong, and I. L. Chuang, in *Quantum information and computation IV*, edited by E. J. Donkor, A. R. Pirich, and H. E. Brandt (May 12, 2006), p. 6244.

[33] C.-C. Lin, S. Sur-Kolay, and N. K. Jha, PAQCS: Physical design-aware fault-tolerant quantum circuit synthesis, IEEE Trans. Very Large Scale Int. (VLSI) Syst. **23**, 1221 (2015).

[34] R. Wille, O. Keszocze, M. Walter, P. Rohrs, A. Chattopadhyay, and R. Drechsler, in *2016 21st Asia and South Pacific Design Automation Conference (ASP-DAC)* (Jan. 2016).

[35] P. Murali, J. M. Baker, A. J. Abhari, F. T. Chong, and M. Martonosi, in *ASPLOS '19: Proceedings of the twenty-fourth international conference on architectural support for programming languages and operating systems*, edited by I. Bahar, M. Herlihy, E. Witchel, and A. R. Lebeck (2019), p. 1015, ISBN: 978-1-4503-6240-5.

[36] H. Abraham, *et al.*, Qiskit: An Open-source Framework for Quantum Computing, IBM, (2019).

[37] Y. Hirata, M. Nakanishi, S. Yamashita, and Y. Nakashima, in *2009 Third International Conference on Quantum, Nano and Micro Technologies* (February, 2009).

[38] R. Beals, S. Brierley, O. Gray, A. W. Harrow, S. Kutin, N. Linden, D. Shepherd, and M. Stather, Efficient distributed quantum computing, Proc. R. Soc. A: Math., Phys. Eng. Sci. **469**, 20120686 (2013).

[39] A. Shafaei, M. Saeedi, and M. Pedram, in *2014 19th Asia and South Pacific Design Automation Conference (ASP-DAC)* (January, 2014), p. 495.

[40] M. Pedram and A. Shafaei, Layout Optimization for quantum circuits with linear nearest neighbor architectures, IEEE Circuits Syst. Mag. **16**, 62 (2016). ISSN: 1531-636X,.

[41] S. Brierley, Efficient implementation of quantum circuits with limited qubit interactions, Quantum Inf. Comput. **17**, 1096 (2017).

[42] D. Maslov, S. M. Falconer, and M. Mosca, Quantum Circuit Placement, IEEE Trans. Comput.-Aided Des. Int. Circuits Syst. **27**, 752-763 (2008).

[43] I. Banerjee and D. Richards, in *Fundamentals of computation theory* (Springer Berlin Heidelberg, 2017), p. 69.

[44] T. Miltzow, L. Narins, Y. Okamoto, G. Rote, A. Thomas, and T. Uno, in *24th Annual European Symposium on Algorithms (ESA 2016): Algorithms*, Vol. 57, edited by P. Sankowski and C. Zaroliagis, Leibniz International Proceedings in Informatics (LIPIcs) (February, 16, 2016), 66:1–66:15, ISBN: 978-3-95977-015-6.

[45] N. Alon, F. R. K. Chung, and R. L. Graham, Routing permutations on graphs via matchings, SIAM J. Discrete Math. **7**, 513 (1994).

[46] L. Zhang, Optimal bounds for matching routing on trees, SIAM J. Discrete Math. **12**, 64 (1999).

[47] S. Bravyi, M. B. Hastings, and F. Verstraete, Lieb-Robinson Bounds and the Generation of Correlations and Topological Quantum Order, Phys. Rev. Lett. **97**, 050401 (2006).

[48] K. Van Acoleyen, M. Mariën, and F. Verstraete, Entanglement Rates and Area Laws, Phys. Rev. Lett. **111**, 170501 (2013).

[49] Z.-X. Gong, M. Foss-Feig, F. G. S. L. Brandão, and A. V. Gorshkov, Entanglement Area Laws for Long-Range Interacting Systems, Phys. Rev. Lett. **119**, 050501 (2017).

[50] L. Piroli, G. Styliaris, and J. I. Cirac, Quantum Circuits Assisted by Local Operations and Classical Communication: Transformations and Phases of Matter, Phys. Rev. Lett. **127**, 220503 (2021).

[51] P. Nation, H. Paik, A. Cross, and Z. Nazario, *The IBM quantum heavy hex lattice*, (July 7, 2021) https://research.ibm.com/blog/heavy-hexlattice (visited on January 21, 2022).

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

[53] We do not limit routing circuits to a particular gate set. If necessary, any such circuit can be approximated by a universal, inverse-closed gate set with at most polylogarithmic overhead by the Solovay-Kitaev theorem [7].

[54] M. Mariën, K. M. R. Audenaert, K. Van Acoleyen, and F. Verstraete, Entanglement rates and the stability of the area law for the entanglement entropy, Commun. Math. Phys. **346**, 35 (2016).

[55] M. M. Wilde, *Quantum Information Theory*, 2nd ed. (Cambridge University Press, Cambridge, UK New York, Feb. 2017), ISBN: 9781316809976.

[56] Z. Eldredge, L. Zhou, A. Bapat, J. R. Garrison, A. Deshpande, F. T. Chong, and A. V. Gorshkov, Entanglement bounds on the performance of quantum computing architectures, Phys. Rev. Res. **2**, 033316 (2020).

[57] F. Chung, *Spectral graph theory: Revised and improved*, (2006) https://mathweb.ucsd.edu/∼fan/research/revised.html (visited on 01/12/2023); repr. of Spectral graph theory (American Mathematical Society, Dec. 1996), ISBN: 0821803158.

[58] C. H. Bennett, J. I. Cirac, M. S. Leifer, D. W. Leung, N. Linden, S. Popescu, and G. Vidal, Optimal simulation of two-qubit Hamiltonians using general local operations, Phys. Rev. A **66**, 012305 (2002).

[59] A. Bapat, E. Schoute, A. V. Gorshkov, and A. M. Childs, Nearly optimal time-independent reversal of a spin chain, Phys. Rev. Res. **4**, L012023 (2022).

[60] G. Vidal, K. Hammerer, and J. I. Cirac, Interaction Cost of Nonlocal Gates, Phys. Rev. Lett. **88**, 237902 (2002).

[61] F. Vatan and C. Williams, Optimal quantum circuits for general two-qubit gates, Phys. Rev. A **69**, 032315 (2004).

[62] B. Nachtergaele and R. Sims, Lieb-Robinson bounds and the exponential clustering theorem, Commun. Math. Phys. **265**, 119 (2006).

[63] S. Bravyi, Upper bounds on entangling rates of bipartite Hamiltonians, Phys. Rev. A **76**, 052319 (2007).

[64] K. M. R. Audenaert, Quantum skew divergence, J. Math. Phys. **55**, 112202 (2014).

[65] A. Ghosh, S. Boyd, and A. Saberi, Minimizing effective resistance of a graph, SIAM Rev. **50**, 37 (2008).

[66] D. E. Knuth, in *The art of computer programming: volume 3: sorting and searching* (2nd edition) (Addison-Wesley Professional, 1998), p. 219, ISBN: 0201896850.

[67] A. Bapat, A. M. Childs, A. V. Gorshkov, S. King, E. Schoute, and H. Shastri, Quantum routing with fast reversals, Quantum **5**, 533 (2021).

[68] [57, theorem 1.16] shows a slightly stronger version for a modified lazy random walk, which has different transition probability to remain at the same vertex.

[69] C. Hoffman, M. Kahle, and E. Paquette, Spectral Gaps of Random Graphs and Applications. *International Mathematics Research Notices*, (2019),.

[70] J. Cheeger, in *Problems in analysis: A symposium in honor of Salomon Bochner (PMS-31)*, edited by R.C. Gunning (1971).

[71] P. Raghavan and E. Upfal, in *Proceedings of the Twenty-Sixth Annual ACM Symposium on Theory of Computing - STOC '94* (1994).

[72] A. Y. Guo, M. C. Tran, A. M. Childs, A. V. Gorshkov, and Z.-X. Gong, Signaling and scrambling with strongly long-range interactions, Phys. Rev. A **102**, 010401(R) (2020).

[73] D. Devulapalli, E. Schoute, A. Bapat, A. M. Childs, and A. V. Gorshkov, *Quantum routing with teleportation*, April 8, (2022).