HAL
archives-ouvertes.fr

Simulation of quantum walks and fast mixing with classical processes

Simon Apers, Alain Sarlette, Francesco Ticozzi

▶ To cite this version:

Simon Apers, Alain Sarlette, Francesco Ticozzi. Simulation of quantum walks and fast mixing with classical processes. Physical Review A, American Physical Society, 2018, 98 (3), pp.032115. hal-01946763

HAL Id: hal-01946763
https://hal.inria.fr/hal-01946763

Submitted on 6 Dec 2018

Simulation of Quantum Walks and Fast Mixing with Classical Processes

Simon Apers [*1], Alain Sarlette [†1,2], and Francesco Ticozzi [‡3]

[1]*Department of Electronics and Information Systems, Ghent University, Belgium*
[2]*QUANTIC lab, INRIA Paris, France*
[3]*Dipartimento di Ingegneria dell'Informazione, Università di Padova, Italy, and the Department of Physics and Astronomy, Dartmouth College, NH 03755, USA.*

December 6, 2017

Abstract

We compare discrete-time quantum walks on graphs to their natural classical equivalents, which we argue are lifted Markov chains, that is, classical Markov chains with added memory. We show that these can simulate quantum walks, allowing us to answer an open question on how the graph topology ultimately bounds their mixing performance, and that of any stochastic local evolution. The results highlight that speedups in mixing and transport phenomena are not necessarily diagnostic of quantum effects, although superdiffusive spreading is more prominent with quantum walks.

Random walks are both ubiquitous models for natural processes and a powerful, versatile algorithmic tool to explore networks and extract information about their structure. In recent years their quantum analogue, named quantum walks (QWs), was shown to hold similar promises. QWs describe the evolution of the position probability distribution of a "walking" quantum particle on a graph, possibly entangled with other quantum degrees of freedom (the so-called coin). The joint dynamics can be either discrete-time or continuous and must respect the graph locality [1, 2, 3, 4]. Following the realization that QWs on a line can beat the diffusive behavior typical of classical stochastic processes [5, 1], they have been invoked to explain improved transport phenomena in biological systems [6, 7], linked to thermodynamic theories, breakdown models and topological states of matter [8, 9, 10], and simulated in various experiments [11, 12, 13, 14, 15]. Furthermore, they have been intensely studied as a paradigm for quantum computing [16, 17] and to speed up algorithmic tasks [18], in particular, those related to the celebrated Grover search algorithm [19, 20, 21].

Despite impressive advances in their analysis, elucidating the source and extent of quantum advantages from the perspective of QWs, as well as providing general design principles to ensure a quantum speedup, remain ongoing lines of research. A general quadratic speedup by QWs has been established for the hitting time [22, 23, 24, 21, 25, 26], thus searching for a marked node in a graph. The complementary problem of mixing, that is, converging to a particular probability distribution over the nodes, has so far resisted a general QW speedup analysis, although it is closer to the original observation on the line [5]. There is further evidence for a quadratic speedup with respect to classical Markov chains on specific graphs including the cycle [27], the hypercube [28], and the torus [29]. A general characterization of QW mixing would be a fundamental step for investigating quantum *vs.* classical differences in statistical mechanics (thermodynamic equilibrium, transport phenomena, localization defects), and its algorithmic complexity is of key relevance for applications like sampling and Monte-Carlo simulations [30].

In this paper, we characterize mixing performance of QWs by showing that they belong to a class of processes which can be simulated by *classical* Markov chains with additional finite memory, called "lifted Markov chains" (LMCs) [31]. For general graphs, our constructive proof reminds a classical version of the "Feynman clock Hamiltonians" used to prove universality of adiabatic computing [32, 33, 34], in combination with "stochastic bridges" generalizing [35] and [36, 37] to simulate quantum channels for fixed initial conditions. This allows us to derive a tight bound on potential QW mixing speedup, improving the known bounds from [27, 38]. Furthermore, for lattices, on which most QW mixing speedups have been demonstrated, we relate the QWs to fast mixing LMCs that have not only the same mixing performance, but also the same structure [39, 40], making them their natural classical analogue.

---

[*]Corresponding author: simon.apers@ugent.be
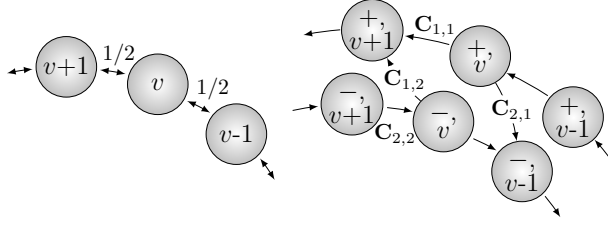[†]alain.sarlette@inria.fr
[‡]ticozzi@dei.unipd.it

Figure 1: (left) random walk $\mathbf{P}_0$ on the N-cycle; (right) quantum walk unitary with coin toss $\mathbf{C}$, lifted Markov chain with a stochastic coin toss $\overline{\mathbf{C}}$, suggesting their comparison.

These results provide several insights. First, an observed speedup in mixing is *not* fundamentally diagnostic of a quantum effect, as it may always be explained by a purely classical memory. Second, QWs are essentially subject to the same bound on their mixing performance as other local processes, induced solely by the topology of the graph. Third, the search for a quantum advantage should focus on identifying efficient designs, in terms of the amount of memory or the graph knowledge required. For lattices, beating efficient classical algorithms is possible only for tasks beyond pure mixing. Whether for statistical mechanics, evolutionarily selected biological systems, or design of faster Monte Carlo algorithms, our results significantly narrow the context in which quantum effects may provide an intrinsic advantage.

QWs and their classical counterparts: a paradigmatic example. – Usually, QWs are presented as the quantum analogues of, and compared to, classical random walks. We next argue that different classical models should be considered towards establishing an intrinsic quantum advantage in mixing, as QWs exhibit genuine *memory effects*. Standard discrete-time QWs [41, 27] describe the evolution of the position distribution p_t of a quantum particle ("walker") over a discrete set of graph nodes $\mathcal{V}$. The quantum evolution of position is conditioned on additional degrees of freedom $\mathcal{C}$, the coin of the walker. The walker state is thus defined on the joint Hilbert space $\mathcal{H} = \mathcal{H}_C \otimes \mathcal{H}_V = \mathrm{span}\{|c\rangle \otimes |v\rangle | (c, v) \in \mathcal{C} \times \mathcal{V}\}$. The cycle graph is a simple example where QWs provide a mixing speedup with respect to a classical walk, see Fig. 1. To the nodes $\mathcal{V} = \{1, 2, \ldots, N\}$ of the cycle, the QW adds a binary coin $\mathcal{C} = \{+, -\}$, see [5, 27]. Denoting $\mathbf{P}^{\pm}$ the cyclic permutation of position, that is, $\mathbf{P}^{\pm}|v\rangle = |(v \pm 1)\mathrm{mod}N\rangle$ for $v \in \mathcal{V}$, the unitary QW primitive reads

$$\mathbf{U} = \mathbf{S} \left(\mathbf{C} \otimes \mathbf{I}_N\right), \quad \mathbf{C} = \begin{bmatrix} e^{-i\phi}\sqrt{1-\alpha} & e^{i\theta}\sqrt{\alpha} \\ -e^{-i\theta}\sqrt{\alpha} & e^{i\phi}\sqrt{1-\alpha} \end{bmatrix},$$

where $\mathbf{S} = |+\rangle \langle+| \otimes \mathbf{P}^+ + |-\rangle \langle-| \otimes \mathbf{P}^-$ expresses a conditional shift, while $\mathbf{C}$ is a general unitary "coin toss" on $\mathcal{H}_C$. The conditional motion can also be viewed as spin-orbit coupling. To actually mix, some decoherence or measurement rule must be added to this unitary evolution, see [42] for a survey. For instance, after every application of $\mathbf{U}$, one can perform with probability q a projective measurement in the canonical basis, after which the unitary evolution is resumed:

$$|\psi_{t+1}\rangle = \begin{cases} \mathbf{U}|\psi_t\rangle & \text{with probability } 1 - q, \\ |c, v\rangle & \text{with probability } q\,|\langle c, v|\mathbf{U}|\psi_t\rangle|^2. \end{cases} \tag{1}$$

A purely unitary QW is obtained with $q = 0$, while $q = 1$ projects the state on the reference basis at each step. The position distribution p_t is obtained by tracing over the coin and considering the probabilities induced in the node basis at time t. The QW of Eq. (1) with, e.g., parameters $\alpha = 1/2, \phi = \theta = 0$ and $q = O(1/N)$, converges towards a uniform p_t in $t = O(N)$ steps, from any initial distribution [5, 27]. In contrast, a classical random walk over $\mathcal{V}$ with transition matrix $\mathbf{P}_0 = (\mathbf{P}^+ + \mathbf{P}^-)/2$ reaches the same distribution only after $O(N^2)$ steps.

Compared to a classical random walk, the QW above clearly adds memory via the coin degrees of freedom. Yet, QWs can exhibit memory effects even without coin. Consider the two-node graph without coin, $\mathcal{H} = \mathrm{span}\{|1\rangle, |2\rangle\}$, equivalent to a qubit, and take the Hadamard gate $U_H = (\sigma_x + \sigma_z)/\sqrt{2}$ as QW primitive. Starting on a given node, after one step, the distribution p_1 over $|1\rangle, |2\rangle$ is uniform, yet at the second step the initial state is perfectly recovered since U_H^2 is the identity operator. This behavior, impossible for any classical Markov process on $\{1, 2\}$, is due to the quantum state storing information in its relative phases, or coherences. Hence, to establish if there is an intrinsic quantum advantage, QWs should be compared to classical local processes with at least a certain amount of additional memory.

Remarkably, a classical walker with memory that mixes fast on the cycle has already been proposed independently of the QW literature [39, 40], and it shares striking similarities. This walker moves among *classical* states

in $\mathcal{C} \times \mathcal{V}$. Its probability distribution $\hat{p}$ over $\mathcal{C} \times \mathcal{V}$ evolves as $\hat{p}_{t+1} = \mathbf{P}\,\hat{p}_t$, with stochastic transition matrix $\mathbf{P}$ having the same structure as $\mathbf{U}$, yet with $\mathbf{C}$ now replaced by a *stochastic* coin toss:

$$\mathbf{P} = \mathbf{S}\left(\bar{\mathbf{C}} \otimes \mathbf{I}_N\right), \quad \bar{\mathbf{C}} = \begin{bmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{bmatrix}. \tag{2}$$

This can be seen as the mixture of two reversible evolutions: with probability $1-\alpha$, the state follows the conditional shift $\mathbf{S}$; or, with probability α, the coin is switched before applying $\mathbf{S}$. The coin allows the classical walker to retain and use information about its previous motion direction, in physical terms its momentum. The similarity between $\mathbf{U}$ and $\mathbf{P}$ carries a deeper connection, as $\mathbf{P}$ in Eq. (2) exactly describes the probabilistic evolution induced by Eq. (1) when starting with $|\psi\rangle = |c, v\rangle$, for some $(c, v) \in \mathcal{C} \times \mathcal{V}$, and with $q = 1$.

This $\mathbf{P}$ mixes over the cycle in $O(N)$ steps [39, 40], provided $\alpha = O(1/N)$. This speedup, only due to classical memory, matches the one provided by the QW in Eq.(1) with $q = O(1/N)$. In both cases, an $O(1/N)$ nonunitarity provides a good tradeoff between fast (deterministic) motion along the graph and losing correlation with the initial condition. From these observations, it appears most natural to compare QWs like Eq.(1) to classical evolutions with memory like Eq.(2), which are formalized as LMCs [31].

QWs and LMCs as local processes with equivalent mixing performance. – Consider a graph with node set $\mathcal{V}$ and edges $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$. The nodes could represent energy levels and the edges allowed transitions. The QW and LMC constructions both start by building a lifted graph, where each node of the initial graph is split into "lifted nodes" or "sublevels". This is done without loss of generality by introducing a coin set $\mathcal{C}$, defining the lifted nodes $\mathcal{C} \times \mathcal{V} = \{(c, v)\}$ and selecting lifted edges in $\{((c, v), (c', v')) \mid (v, v') \in \mathcal{E}\}$, thus without introducing transitions that were not allowed before lifting.

A general QW is then described by a quantum channel over the space generated by viewing coin and node as quantum numbers, i.e.,

$$\rho_{t+1} = \sum_k \mathbf{M_k} \rho_t \mathbf{M_k}^\dagger, \tag{3}$$

where ρ_t is a density operator on $\mathcal{H} = \mathrm{span}\{|c, v\rangle \mid (c, v) \in \mathcal{C} \times \mathcal{V}\}$ and the $\mathbf{M_k}$ satisfy $\sum_k \mathbf{M_k}^\dagger \mathbf{M_k} = \mathbf{I}_{\mathcal{C} \times \mathcal{V}}$, with $\mathbf{I}$ denoting the identity [1]. The graph locality is imposed by $\langle c', v' | \mathbf{M_k} | c, v \rangle = 0$ if $(v, v') \notin \mathcal{E}$. To complete the setup, an initial distribution p_0 over $\mathcal{V}$ is mapped onto the lifted nodes (or sublevels) by $F : p_0 \mapsto \rho = \sum_{v \in \mathcal{V}} p_0(v) |c_v, v\rangle \langle c_v, v|$, thus associating some fixed initial coin state c_v to each v. The object of interest is the distribution p_t over $\mathcal{V}$, the main nodes or levels [2], obtained with the partial trace as $p_t = \mathrm{diag}(\mathrm{trace}_\mathcal{C}(\rho_t))$.

Similarly, a LMC follows the dynamics $\hat{p}_{t+1} = \mathbf{P}\,\hat{p}_t$, where $\hat{p}_t$ is a vector representing the probability distribution over $\mathcal{C} \times \mathcal{V}$, and $\mathbf{P}$ is a stochastic matrix expressing the jump probabilities among sublevels. Namely, denoting by $p = e_v$ and $\hat{p} = e_{(c,v)}$ the distributions with probability 1 of being on v and on (c, v), respectively, $e_{(c',v')}^\dagger \mathbf{P}\, e_{(c,v)}$ is the transition probability from (c, v) to (c', v'). Graph locality imposes $e_{(c',v')}^\dagger \mathbf{P}\, e_{(c,v)} = 0$ if $(v, v') \notin \mathcal{E}$. Initial lifted nodes are assigned by $F : p_0 \mapsto \hat{p}_0 = \sum_v p_0(v) e_{(c_v,v)}$. The distribution of interest is obtained by marginalizing over $\mathcal{C}$, thus $p_t(v) = \sum_{c \in \mathcal{C}} \hat{p}_t(c, v)$ for all $v \in \mathcal{V}$.

Clearly, a LMC is a particular QW where populations evolve without coherences, i.e., where ρ_t remains diagonal at all times and $\mathbf{M_k} = \sqrt{e_{(c',v')}^\dagger \mathbf{P}\, e_{(c,v)}} \,|c', v'\rangle \langle c, v|$, with index k running over all nonzero elements of $\mathbf{P}$. The key to our main result will be to observe how, conversely, any QW can be simulated by some LMC (with possibly higher-dimensional coin). In other words, the non-Markovian evolution of p_t under a QW can be described as a classical Markovian evolution of sublevel populations.

We focus on comparing the mixing behavior induced by QWs and LMCs. A QW or LMC mixes to some distribution $\bar{p}$ over $\mathcal{V}$ if for any initial state $F(p_0)$ the induced distribution p_t converges to $\bar{p}$. The mixing time $\tau(\epsilon)$, for any $0 < \epsilon < 1$, is the time required to get ϵ-close to the limit distribution in total variation distance, i.e., the smallest time such that $\frac{1}{2} \sum_{v \in \mathcal{V}} |p_t(v) - \bar{p}(v)| \leq \epsilon$ for all $t \geq \tau(\epsilon)$ and all p_0. A standard "stabilizing" requirement for a process that converges to $\bar{p}$ is that $p_0 = \bar{p}$ should imply $p_t = \bar{p}$ at all times. This holds automatically for the time-invariant $\mathbf{P}$ considered by the LMC framework. The QW framework allows the $\mathbf{M_k}$ to depend on time, but in standard constructions only through the measurement mechanism, like making q time-dependent in the cycle example (see [42] for a review). Such QWs too preserve $p_t = \bar{p}$ at all times when $p_0 = \bar{p}$. We call this property $\bar{p}$-invariance [3] and we will come back to its significance. Our first result shows that the mixing performance of such QW can be closely matched by a LMC.

---

[1] Some authors add a so-called Cesaro averaging routine on top of this QW model [27, 43]. Our results can explicitly capture this and similar extensions via local stochastic maps, see Supplemental Material in appendix A.

[2] Standard literature like [31] defines LMCs with *joint* distribution over $\mathcal{C} \times \mathcal{V}$ as object of interest, without initialization map F. This does not affect our QW results, and in fact it implies no significant difference in general, see [44].

[3] Note that this condition involves both the channel $\mathbf{M_k}$ and the initialization F.
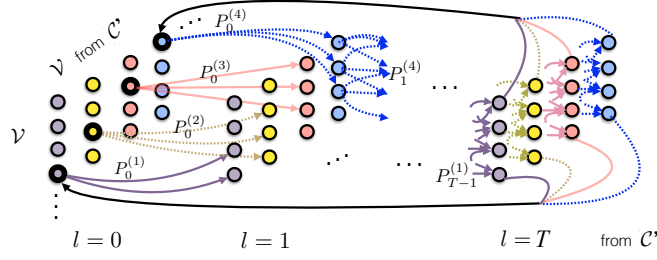
Figure 2: Sketch of the LMC construction proving Thm.1, and implying that QWs feature the same conductance bound on fast mixing as LMCs. A given graph with nodes $\mathcal{V}$ (vertical axis) is lifted with coin space $\mathcal{C}'$ comprising both an initial node index (depth) and a time index (horizontal). LMC transitions are constructed between nodes of this lifted graph. (edges shown partially to avoid clutter)

Theorem 1. *Given a $\bar{p}$-invariant QW with mixing time $\bar{\tau}(\epsilon_0)$ for some $\epsilon_0 \leq 1/4$, we can construct an LMC that has mixing time $\tau(\epsilon) / \bar{\tau}(\epsilon_0) \leq \lceil \log(1/\epsilon) / \log(1/(2\epsilon_0)) \rceil$ for all $\epsilon > 0$.*

Mathematical details for all our results are available in the Supplemental Material, appendix A. The main idea in proving Thm.1 is to simulate the QW over the time interval $[0, \bar{\tau}(\epsilon_0)]$ using a LMC. Indeed, as shown for unitary evolutions in [35], the probability distribution in the *fixed measurement basis* associated to the nodes is not subject to the no-go results for general local hidden variables theories. We extend this result to p_t induced by an arbitrary QW that starts from a given node $v \in \mathcal{V}$. Following p_t step by step, one builds a sequence of stochastic matrices $\mathbf{P}_1^{(v)}, \mathbf{P}_2^{(v)}, ..., \mathbf{P}_t^{(v)}$ acting on $\mathcal{V}$ only, satisfying graph locality, and such that $p_t = \mathbf{P}_t^{(v)} p_{t-1}$ when starting on v. The max-flow min-cut theorem from graph theory ensures that such construction always exists. It can be traced back to a property that holds for LMCs, QWs, and more general local stochastic processes independently of the underlying physical mechanism: a node cannot contain more population at time $t + 1$, than the population at time t on itself and on its neighbors [27]. We thus simulate the QW with a classical process whose jump probabilities depend on time and on the starting node v. To obtain a simulation with a (time-independent) LMC, at least for finite time horizon $t \leq \bar{\tau}(\epsilon_0)$, we encode these dependencies into the coin. This follows the same spirit as adding registers in the clock Hamiltonians by Feynman and Kitaev [32, 33]. Explicitly, we let current time l and initial node v act as a coin degree of freedom $c' = (v, l)$, which conditionally selects the proper transition matrix $\mathbf{P}_l^{(v)}$, see Fig. 2. The resulting LMC describes a distribution over $\mathcal{C}' \times \mathcal{V} \equiv \big(\mathcal{V} \times \{0, 1, ..., T = \bar{\tau}(\epsilon_0)\}\big) \times \mathcal{V}$, with associated stochastic transition matrix

$$\mathbf{P} \equiv \sum_{v \in \mathcal{V}} e_v e_v^\dagger \otimes \left(\sum_{t=0}^{T-1} e_{t+1} e_t^\dagger \otimes \mathbf{P}_t^{(v)} + e_T e_T^\dagger \otimes \mathbf{I}_\mathcal{V} \right)$$

and initial assignment $F : e_v \mapsto (e_v \otimes e_0) \otimes e_v$. Finally, we apply an amplification technique that is exploited in randomized algorithms: the action of $\mathbf{P}$ on e_T is modified to have $\mathbf{P} \, e_{(v_0, T, v)} = e_{(v, 0, v)} = F(v)$, so that the T first steps are repeated iteratively. Thanks to $\bar{p}$-invariance of the QW that was used to generate $\mathbf{P}$, the resulting LMC will contract towards $\bar{p}$ at the announced exponential rate for all $t \geq T$.

Beyond the comparison with LMCs, this construction implies a general bound on the mixing performance of $\bar{p}$-invariant QWs. This tightens and generalizes the bounds of [27, 38], which are restricted to generating uniform $\bar{p}$ with unital quantum channels. The bound involves a function of graph topology and target distribution only, meant to capture the bottlenecks that slow down mixing, called the graph conductance $\Phi_{\bar{p}}$. Specifically, partitioning $\mathcal{V}$ into two subsets $\mathcal{X}$ and $\mathcal{X}^c$, consider that all the stationary population on $\mathcal{X}^c$ is lost; the conductance counts which fraction of the remaining population $\bar{p}(\mathcal{X}) = \sum_{v \in \mathcal{X}} \bar{p}(v)$ jumps back to $\mathcal{X}^c$ in one step. More precisely, if $\mathbf{P}$ on $\mathcal{V}$ has a stationary distribution $\bar{p}$, then

$$\Phi_{\bar{p}}(\mathbf{P}) = \min_{\mathcal{X} : 0 < \bar{p}(\mathcal{X}) \leq \frac{1}{2}} \left(\sum_{v \in \mathcal{X}, v' \in \mathcal{X}^c} (e_{v'}^\dagger \mathbf{P} e_v) \bar{p}(v) \right) / \bar{p}(\mathcal{X}) .$$

The maximal $\Phi_{\bar{p}}(\mathbf{P})$ over all Markov chains that keep $\bar{p}$ invariant on a given graph, is the graph conductance $\Phi_{\bar{p}}$.

The estimate $1/\Phi_{\bar{p}}$ is a well-known lower bound on the mixing time of any classical Markov chain, and it carries over to the convergence of $\hat{p}_t$ in associated LMCs [31]. Conversely, [31] establishes a construction of LMCs that essentially saturate this bound; it however requires to solve a hard multi-commodity flow problem over the entire graph. A novel observation, obtained essentially by fully exploiting the triangle inequality while computing the

4

marginal probabilities, is that the bound keeps holding when taking the marginal p_t over sublevels (i.e., over coin values) of a $\bar{p}$-invariant LMC. Combining this with Theorem 1 provides a tight bound for the ultimately achievable mixing time of QWs.

Theorem 2. *Any $\bar{p}$-invariant QW has a mixing time $\tau(1/4) \geq 1/(4\Phi_{\bar{p}})$, and there exists such a QW that has a mixing time $\tau(\epsilon) \leq O(\log(1/\min_k \bar{p}_k)\log(1/\epsilon) / \Phi_{\bar{p}})$ for all $\epsilon > 0$.*

Besides mixing, the LMC construction has relevance for other tasks, enabling for instance to effectively simulate quantum transport with finite classical resources.
On efficient design of fast mixing QWs and LMCs. – Fast mixing LMCs can often be built significantly more simply than with the general construction of Thm.1, by mirroring the structure of corresponding QWs. Accelerated mixing with QWs has been mostly demonstrated for graphs with strong symmetries, more specifically lattices [5, 27, 28, 29, 43, 45]. Similarly to the QW on the circle above, these examples use coin values to encode the lattice generators among which the walker can select its next move.

Remarkably, the same structure is found in a proposal for designing fast mixing LMCs [40]. For a d-dimensional square lattice of size M, the coin features $2d$ values of type $\pm_k$, with $k \in \{1, 2, ..., d\}$ indicating the axis and $\pm$ the direction of conditional motion among the nodes. At each step, the coin has a probability $\alpha = 1/(2dM)$ to switch to each of the other coin values, thus retaining a high probability $1 - (2d - 1)/(2dM)$ to stay with the same generator. This dynamics precisely corresponds to a QW with diagonally dominant coin update **C** that is projectively measured at each step, as in Eq.(1) with $q = 1$. For fixed dimension d, it also provides the same order of speedup as a QW with $q \ll 1$ [29], and as the best possible QW according to Theorem 2, namely linear in M. Indeed, by counting the probabilities of applying, to each lattice dimension consecutively, the sequence of steps that lead to fast mixing on the cycle, one obtains the following (possibly loose) bound [4].

Theorem 3. *The just described LMC on $\mathbb{Z}_N^d$ has a mixing time $\tau(\epsilon) \leq O(M\,d^2\log(d)\log(1/\epsilon))$.*

Thus, QRW and LMC have the same order of mixing time; the same structure; and they require the same graph knowledge for tuning (α and/or q), namely the time $O(M)$ at which mixing will be considered accomplished.

In summary, we clarify that QWs on a graph induce non-Markovian local processes whose mixing behavior can be simulated by LMCs (Thm.1), and that this has several implications for searching a quantum speedup in mixing processes. The construction of Thm.1 can in fact be extended to abstract local stochastic dynamics (see Supplemental Material in appendix A) beyond the QW model. As a consequence, the hierarchy LMCs $\subseteq$ QWs $\subseteq$ {general local processes} collapses regarding mixing speed, not only in terms of ultimate speedup achievable on general graphs (Thm.2), but also in terms of paradigmatic cases for which efficient mixing designs are known (lattices, Thm.3). In this light, a mixing speedup with respect to Markov chains is not diagnostic of underlying quantum dynamics, but potentially just of a memory effect. This prompts the question whether there is room for a "quantum advantage" at all in QW mixing.

Besides establishing that there is no advantage in terms of best achievable mixing time, our analysis also suggests why this is not the end of the story. While the property of $\bar{p}$-invariance holds and stabilizes the system in typical QW proposals, it does not hold in some mixing-related applications, like simulated annealing. This distinction may be important as, without $\bar{p}$-invariance, the conductance bound of Thm.2 could be broken significantly [44]. As another memory-related aspect, in Eq. (1) on the cycle, taking $\alpha = 1/2$ leads to fast QWs, while the corresponding "projectively measured" LMC boils down to the quadratically slower standard random walk. This shows that coherences can play a beneficial role, and could guide future research towards designing simple yet fast mixing QWs on graphs for which, unlike on lattices, LMCs of simple design do not meet the conductance bound yet. Furthermore, the QW of Eq. (1), taking $\alpha = 1/2$ and $q = 1/N$, turns out to efficiently mix over the t nodes closest to its starting node, *for any number of iterations $t < N$* [5]. Such multiscale mixing cannot be achieved with the LMC of Eq. (2), where tuning $\alpha = 1/N$ to have good mixing at $t = N$ implies almost deterministic motion for $t \ll N$. This feature could point to efficient quantum algorithms addressing tasks related to mixing, yet not directly reducible to it.

The authors want to thank Giuseppe Vallone and Lorenza Viola for valuable suggestions and comments on earlier versions of the manuscript.

References

[1] Y. Aharonov, L. Davidovich, and N. Zagury, "Quantum random walks," *Physical Review A*, vol. 48, no. 2, p. 1687, 1993.

---

[4]This bound was conjectured to hold more generally for all Abelian Cayley graphs in [40], and the authors provide a concrete proof only for the case of the circle.

[2] E. Farhi and S. Gutmann, "Quantum computation and decision trees," *Physical Review A*, vol. 58, no. 2, p. 915, 1998.

[3] J. Watrous, "Quantum simulations of classical random walks and undirected graph connectivity," in *Computational Complexity, 1999. Proceedings. Fourteenth Annual IEEE Conference on*, pp. 180–187, IEEE, 1999.

[4] J. Kempe, "Quantum random walks: an introductory overview," *Contemporary Physics*, vol. 44, no. 4, pp. 307–327, 2003.

[5] A. Ambainis, E. Bach, A. Nayak, A. Vishwanath, and J. Watrous, "One-dimensional quantum walks," in *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, pp. 37–49, ACM, 2001.

[6] G. S. Engel, T. R. Calhoun, E. L. Read, T.-K. Ahn, T. Mančal, Y.-C. Cheng, R. E. Blankenship, and G. R. Fleming, "Evidence for wavelike energy transfer through quantum coherence in photosynthetic systems," *Nature*, vol. 446, no. 7137, pp. 782–786, 2007.

[7] M. Mohseni, P. Rebentrost, S. Lloyd, and A. Aspuru-Guzik, "Environment-assisted quantum walks in photosynthetic energy transfer," *The Journal of chemical physics*, vol. 129, no. 17, p. 11B603, 2008.

[8] A. Romanelli, R. Donangelo, R. Portugal, and F. de Lima Marquezino, "Thermodynamics of n-dimensional quantum walks," *Physical Review A*, vol. 90, no. 2, p. 022329, 2014.

[9] T. Oka, N. Konno, R. Arita, and H. Aoki, "Breakdown of an electric-field driven system: a mapping to a quantum walk," *Physical review letters*, vol. 94, no. 10, p. 100602, 2005.

[10] T. Kitagawa, M. A. Broome, A. Fedrizzi, M. S. Rudner, E. Berg, I. Kassal, A. Aspuru-Guzik, E. Demler, and A. G. White, "Observation of topologically protected bound states in photonic quantum walks," *Nature communications*, vol. 3, p. ncomms1872, 2012.

[11] M. Karski, L. Förster, J.-M. Choi, A. Steffen, W. Alt, D. Meschede, and A. Widera, "Quantum walk in position space with single optically trapped atoms," *Science*, vol. 325, no. 5937, pp. 174–177, 2009.

[12] A. Peruzzo, M. Lobino, J. C. Matthews, N. Matsuda, A. Politi, K. Poulios, X.-Q. Zhou, Y. Lahini, N. Ismail, K. Wörhoff, *et al.*, "Quantum walks of correlated photons," *Science*, vol. 329, no. 5998, pp. 1500–1503, 2010.

[13] M. Genske, W. Alt, A. Steffen, A. H. Werner, R. F. Werner, D. Meschede, and A. Alberti, "Electric quantum walks with individual atoms," *Physical review letters*, vol. 110, no. 19, p. 190601, 2013.

[14] P. M. Preiss, R. Ma, M. E. Tai, A. Lukin, M. Rispoli, P. Zupancic, Y. Lahini, R. Islam, and M. Greiner, "Strongly correlated quantum walks in optical lattices," *Science*, vol. 347, no. 6227, pp. 1229–1233, 2015.

[15] E. Flurin, V. Ramasesh, S. Hacohen-Gourgy, L. Martin, N. Yao, and I. Siddiqi, "Observing topological invariants using quantum walks in superconducting circuits," *Physical Review X*, vol. 7, no. 3, p. 031023, 2017.

[16] A. M. Childs, "Universal computation by quantum walk," *Physical review letters*, vol. 102, no. 18, p. 180501, 2009.

[17] N. B. Lovett, S. Cooper, M. Everitt, M. Trevers, and V. Kendon, "Universal quantum computation using the discrete-time quantum walk," *Physical Review A*, vol. 81, no. 4, p. 042330, 2010.

[18] A. Ambainis, "Quantum walks and their algorithmic applications," *International Journal of Quantum Information*, vol. 1, no. 04, pp. 507–518, 2003.

[19] N. Shenvi, J. Kempe, and K. B. Whaley, "Quantum random-walk search algorithm," *Physical Review A*, vol. 67, no. 5, p. 052307, 2003.

[20] A. M. Childs and J. Goldstone, "Spatial search by quantum walk," *Physical Review A*, vol. 70, no. 2, p. 022314, 2004.

[21] F. Magniez, A. Nayak, J. Roland, and M. Santha, "Search via quantum walk," *SIAM Journal on Computing*, vol. 40, no. 1, pp. 142–164, 2011.

[22] A. M. Childs, R. Cleve, E. Deotto, E. Farhi, S. Gutmann, and D. A. Spielman, "Exponential algorithmic speedup by a quantum walk," in *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pp. 59–68, ACM, 2003.

[23] M. Szegedy, "Quantum speed-up of markov chain based algorithms," in *Foundations of Computer Science, 2004. Proceedings. 45th Annual IEEE Symposium on*, pp. 32–41, IEEE, 2004.

[24] J. Kempe, "Discrete quantum walks hit exponentially faster," *Probability theory and related fields*, vol. 133, no. 2, pp. 215–235, 2005.

[25] H. Krovi, F. Magniez, M. Ozols, and J. Roland, "Quantum walks can find a marked element on any graph," *Algorithmica*, vol. 74, no. 2, pp. 851–907, 2016.

[26] P. Hoyer and M. Komeili, "Efficient quantum walk on the grid with multiple marked elements," *arXiv preprint arXiv:1612.08958*, 2016.

[27] D. Aharonov, A. Ambainis, J. Kempe, and U. Vazirani, "Quantum walks on graphs," in *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, pp. 50–59, ACM, 2001.

[28] C. Moore and A. Russell, "Quantum walks on the hypercube," in *International Workshop on Randomization and Approximation Techniques in Computer Science*, pp. 164–178, Springer, 2002.

[29] P. C. Richter, "Quantum speedup of classical mixing processes," *Physical Review A*, vol. 76, no. 4, p. 042306, 2007.

[30] A. Sinclair, *Algorithms for random generation and counting: a Markov chain approach*. Springer Science & Business Media, 2012.

[31] F. Chen, L. Lovász, and I. Pak, "Lifting markov chains to speed up mixing," in *Proceedings of the thirty-first annual ACM symposium on Theory of computing*, pp. 275–281, ACM, 1999.

[32] R. P. Feynman, "Simulating physics with computers," *International journal of theoretical physics*, vol. 21, no. 6, pp. 467–488, 1982.

[33] A. Y. Kitaev, A. Shen, and M. N. Vyalyi, *Classical and quantum computation*, vol. 47. American Mathematical Society Providence, 2002.

[34] D. Aharonov, W. Van Dam, J. Kempe, Z. Landau, S. Lloyd, and O. Regev, "Adiabatic quantum computation is equivalent to standard quantum computation," *SIAM review*, vol. 50, no. 4, pp. 755–787, 2008.

[35] S. Aaronson, "Quantum computing and hidden variables," *Physical Review A*, vol. 71, no. 3, p. 032325, 2005.

[36] M. Pavon and F. Ticozzi, "Discrete-time classical and quantum markovian evolutions: Maximum entropy problems on path space," *Journal of Mathematical Physics*, vol. 51, no. 4, p. 042104, 2010.

[37] T. T. Georgiou and M. Pavon, "Positive contraction mappings for classical and quantum schrödinger systems," *Journal of Mathematical Physics*, vol. 56, no. 3, p. 033301, 2015.

[38] K. Temme, M. J. Kastoryano, M. Ruskai, M. M. Wolf, and F. Verstraete, "The χ 2-divergence and mixing times of quantum markov processes," *Journal of Mathematical Physics*, vol. 51, no. 12, p. 122201, 2010.

[39] P. Diaconis, S. Holmes, and R. M. Neal, "Analysis of a nonreversible markov chain sampler," *Annals of Applied Probability*, pp. 726–752, 2000.

[40] P. Diaconis and L. Miclo, "On the spectral analysis of second-order markov chains," *Annales de la Faculté des Sciences de Toulouse. Mathématiques. Série 6*, vol. 22, no. 3, pp. 573–621, 2013.

[41] D. A. Meyer, "From quantum cellular automata to quantum lattice gases," *Journal of Statistical Physics*, vol. 85, no. 5-6, pp. 551–574, 1996.

[42] V. Kendon, "Decoherence in quantum walks–a review," *Mathematical Structures in Computer Science*, vol. 17, no. 6, pp. 1169–1220, 2007.

[43] F. Marquezino, R. Portugal, G. Abal, and R. Donangelo, "Mixing times in quantum walks on the hypercube," *Physical Review A*, vol. 77, no. 4, p. 042312, 2008.

[44] S. Apers, F. Ticozzi, and A. Sarlette, "Lifting markov chains to mix faster: Limits and opportunities," *arXiv preprint arXiv:1705.08253*, 2017.

[45] F. L. Marquezino, R. Portugal, and G. Abal, "Mixing times in quantum walks on two-dimensional grids," *Physical Review A*, vol. 82, no. 4, p. 042341, 2010.

[46] L. R. Ford and D. R. Fulkerson, "Maximal flow through a network," *Canadian journal of Mathematics*, vol. 8, no. 3, pp. 399–404, 1956.

[47] D. A. Levin and Y. Peres, *Markov chains and mixing times*, vol. 107. American Mathematical Soc., 2017.

[48] D. Aldous and J. Fill, "Reversible markov chains and random walks on graphs," 2002.

[49] R. Montenegro, P. Tetali, *et al.*, "Mathematical aspects of mixing times in markov chains," *Foundations and Trends® in Theoretical Computer Science*, vol. 1, no. 3, pp. 237–354, 2006.

A   Supplemental Material

The objective of the paper is the comparison of Quantum Walks (QWs) and Lifted Markov Chains (LMCs). However, the main results can be extended to a more general setting that includes both QWs and LMCs, namely, local stochastic processes preserving the target distribution. An example of such a generalized setting would be Cesaro averaging, i.e., to consider as output distribution the uniform time average of the evolution generated by a QW or LMC. Here we shall provide detailed proofs of our results directly in this generalized setting. The main ideas remain the same as for the particular case of QWs.

The supplemental material is organized as follows. We start with some notation and defining the generalized class of processes that will be studied. We then explicitly show how QWs fall under this setting. The next three sections are respectively devoted to a detailed proof, with all mathematical details worked out, of each of the three theorems of the main paper.

Notation – We first recall some notation that will be used throughout these notes. Let $\mathcal{G}$ be a graph with node set $\mathcal{V}$ and edge set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. By convention, we include in the edge set all (v, v), $v \in \mathcal{V}$. We define the in-neighborhood or simply neighborhood of a set $\mathcal{X} \subseteq \mathcal{V}$ as $\mathcal{B}(\mathcal{X}) = \{v \in \mathcal{V} \backslash \mathcal{X} : (v, v') \in \mathcal{E} \text{ for some } v' \in \mathcal{X}\}$. Note that we will keep, throughout the report, this notation such that "probability mass flows from v to v'". We create a *lifted graph* by expanding the node set to $\mathcal{C} \times \mathcal{V}$, for some finite set $\mathcal{C}$, so the nodes for the lifted graph are pairs (c, v), with $c \in \mathcal{C}$, $v \in \mathcal{V}$, and we let the edge set be a subset of $\{((c, v), (c', v')) \mid (v, v') \in \mathcal{E}\}$. We associate the Hilbert space $\mathcal{H}_{\mathcal{C} \times \mathcal{V}} = \{|c, v\rangle \mid c \in \mathcal{C}, v \in \mathcal{V}\}$ to this graph and call $\mathcal{D}(\mathcal{H}_{\mathcal{C} \times \mathcal{V}})$ the set of density operators over $\mathcal{H}_{\mathcal{C} \times \mathcal{V}}$, that is, positive semidefinite Hermitian operators of trace one. For any subset $\mathcal{X} \subseteq \mathcal{V}$ and $\rho \in \mathcal{D}(\mathcal{H}_{\mathcal{C} \times \mathcal{V}})$, we define $\mathbb{P}_\rho(\mathcal{X}) = \text{trace}(\mathbf{\Pi}_\mathcal{X} \rho)$, where $\mathbf{\Pi}_\mathcal{X} = \mathbf{I} \otimes \sum_{v \in \mathcal{X}} |v\rangle \langle v|$ is the projector onto the subspace associated to the subset of nodes $\mathcal{X}$ of the original graph $\mathcal{G}$. More generally, we will use the standard notation $\mathbb{P}_p(E)$ to denote the probability of some event E according to the probability measure p; occasionally we also use the notation $\mathbb{P}_p(E) = p(E)$. We will also denote, as in the main paper, by e_v the probability (column) vector with all weight on the node $v \in \mathcal{V}$, and by $e_v^\dagger$ the dual classical (row) vector. Using the tensor product $e_c \otimes e_v = e_{(c,v)}$ also known as the Kronecker product of vectors, we get the probability vector with all weight on the single event $(c, v) \in \mathcal{C} \times \mathcal{V}$.

A.1   Local stochastic processes

In this section we introduce the concept of local stochastic dynamics, and we show how QWs and LMCs fall under this general framework. A stochastic map $\mathbf{\Psi}$ over $\mathcal{V}$ is function that maps a probability distribution p_0 to another probability distribution p_1; it is linear and preserves both the positivity and the sum of the components of p_0. The general stochastic processes which we consider are a family of stochastic linear maps $\mathbf{\Psi}_t$, indexed by time $t \in \mathbb{N}$, and which map an initial probability distribution p_0 over $\mathcal{V}$ to a probability distribution $p_t = \mathbf{\Psi}_t[p_0]$ over $\mathcal{V}$ at each time t. We say that the family $\mathbf{\Psi}_t$ is *local* with respect to a graph $\mathcal{G}$ with nodes $\mathcal{V}$ if and only if [27]:

$$\text{For all } \mathcal{X} \subseteq \mathcal{V}, p_0, t > 0, \text{ it holds that } \mathbb{P}_{p_{t+1}}(\mathcal{X}) \leq \mathbb{P}_{p_t}(\mathcal{X}) + \mathbb{P}_{p_t}(\mathcal{B}(\mathcal{X})). \tag{4}$$

This formula expresses the intuitive statement from the main paper, that a node $\mathcal{X} = \{v\}$ cannot contain more population at time $t + 1$, than the population at time t on itself and on its neighbors $\mathcal{B}(v)$.

The family $\mathbf{\Psi}_t$ is *invariant* with respect to a distribution $\bar{p}$, or short $\bar{p}$-invariant, if and only if $\mathbf{\Psi}_t[\bar{p}] = \bar{p}$, $\forall t \in \mathbb{N}$. This expresses that $p_t = \bar{p}$ for all $t \geq 0$ when $p_0 = \bar{p}$, and it ensures that the process *stabilizes* $\bar{p}$ at all times.

A.1.1 Quantum channels as $\bar{p}$-invariant local stochastic processes

We will now show how such a family of abstract processes $\boldsymbol{\Psi}_t$ explicitly covers the specific case of p_t induced by QWs. Thereto, let $\boldsymbol{\Gamma} : \mathcal{D}(\mathcal{H}_{\mathcal{C}\times\mathcal{V}}) \to \mathcal{D}(\mathcal{H}_{\mathcal{C}\times\mathcal{V}})$ be a completely positive trace-preserving (CPTP) map representing a QW, defined by

$$\boldsymbol{\Gamma}[\rho] = \sum_{\mathbf{k}} \mathbf{M_k} \rho \mathbf{M_k}^\dagger,$$

together with a linear initialization map $\mathbf{F} : p_0 \mapsto \sum_{v \in \mathcal{V}} p_0(v) |c_v, v\rangle \langle c_v, v|$. For any starting condition $p_0 = e_v$, we can compute the distribution p_t induced by the QW as the diagonal of the partial trace over $\mathcal{C}$ of $\boldsymbol{\Gamma}^\mathbf{t}[\mathbf{F}[\mathbf{p_0}]]$. Thanks to linearity of all these steps, the computation of the resulting evolutions $p_0, p_1, ...$ can be described by a family of linear maps $\boldsymbol{\Psi}_t$ such that $p_t = \boldsymbol{\Psi}_t[p_0]$, for general p_0 too. Of course they preserve positivity and total probability, so they are stochastic; and if a target distribution $\bar{p}$ is invariant under $\boldsymbol{\Gamma}$, then obviously it is invariant too under the family of induced $\boldsymbol{\Psi}_t$. In the following lemma we prove that if $\boldsymbol{\Gamma}$ is local, in the sense that the $\mathbf{M_k}$ have zero entries where nodes are not connected in $\mathcal{G}$, then so is $\boldsymbol{\Psi}_t$ in the sense of Eq.(4).

Lemma 4. *Let $\boldsymbol{\Gamma}$ be a quantum channel. The following statements are equivalent:*

(a) *For all $c, c' \in \mathcal{C}$, $v, v' \in \mathcal{V}$, it holds: if $(v, v') \notin \mathcal{E}$ then $\langle c', v' | \mathbf{M}_l | c, v \rangle = 0 \; \forall l$.*

(b) *For all $\mathcal{X} \subseteq \mathcal{V}$ and $\rho \in \mathcal{D}(\mathcal{H}_{\mathcal{C}\times\mathcal{V}})$, it holds that $\mathbb{P}_{\boldsymbol{\Gamma}[\rho]}(\mathcal{X}) \leq \mathbb{P}_\rho(\mathcal{X}) + \mathbb{P}_\rho(\mathcal{B}(\mathcal{X}))$.*

Proof. "(a) $\Rightarrow$ (b)": We will show that the inequality in *(b)* holds for a one-dimensional projection $\rho = |\psi\rangle \langle\psi|$, $|\psi\rangle \in \mathcal{H}_{\mathcal{C}\times\mathcal{V}}$; due to linearity of the involved operators in ρ, the inequality must then necessarily hold also for all density operators $\rho \in \mathcal{D}(\mathcal{H}_{\mathcal{C}\times\mathcal{V}})$, being convex combinations of projections. We can write

$$
\begin{aligned}
\mathbb{P}_{\boldsymbol{\Gamma}[|\psi\rangle\langle\psi|]}(\mathcal{X}) &= \sum_{c' \in \mathcal{C}, v' \in \mathcal{X}} \langle c', v' | \boldsymbol{\Gamma}[|\psi\rangle \langle\psi|] | \mathbf{c}', \mathbf{v}' \rangle \\
&= \sum_{c' \in \mathcal{C}, v' \in \mathcal{X}} \sum_l \langle c', v' | \mathbf{M}_l | \psi \rangle \langle \psi | \mathbf{M}_l^\dagger | c', v' \rangle \\
&= \sum_{c' \in \mathcal{C}, v' \in \mathcal{X}} \sum_l | \langle c', v' | \mathbf{M}_l | \psi \rangle |^2.
\end{aligned}
$$

Since we assume that *(a)* holds, we have that $\forall l$, $\langle c', v' | \mathbf{M}_l | c, v \rangle = 0$ for $v' \in \mathcal{X}$ and $v \in \mathcal{X}^c \backslash \mathcal{B}(\mathcal{X})$, where $\mathcal{X}^c = \mathcal{V} \backslash \mathcal{X}$. If we now write $|\psi\rangle = |\psi_\mathcal{X}\rangle + |\psi_{\mathcal{B}_\mathcal{X}}\rangle + |\psi_{\mathcal{X}^c \backslash \mathcal{B}_\mathcal{X}}\rangle$, where $|\psi_\mathcal{Y}\rangle \equiv \boldsymbol{\Pi}_\mathcal{Y} |\psi\rangle$ for any $\mathcal{Y} \subseteq \mathcal{V}$, then *(a)*, in particular, implies that $\forall c' \in \mathcal{C}, v' \in \mathcal{X}$ and all l, we have $\langle c', v' | \mathbf{M}_l | \psi_{X^c \backslash B_X} \rangle = 0$. Intuitively, this expresses that $|\psi_{\mathcal{X}^c \backslash \mathcal{B}_\mathcal{X}}\rangle$ does not contribute to the probability of observing Π_X after the action of $\boldsymbol{\Gamma}$. Inserting $\langle c', v' | \mathbf{M}_l | \psi_{X^c \backslash B_X} \rangle = 0$ into the above sum, we thus get:

$$
\begin{aligned}
\mathbb{P}_{\boldsymbol{\Gamma}[|\psi\rangle\langle\psi|]}(\mathcal{X}) &= \mathbb{P}_{\boldsymbol{\Gamma}[(|\psi_\mathcal{X}\rangle + |\psi_{\mathcal{B}(\mathcal{X})}\rangle)(\langle\psi_\mathcal{X}| + \langle\psi_{\mathcal{B}(\mathcal{X})}|)]}(\mathcal{X}) \\
&\quad \text{\footnotesize (going from population on $\mathcal{X} \cup \mathcal{B}(\mathcal{X})$ to population on $\mathcal{X}$ after applying $\boldsymbol{\Gamma}$)} \\
&\leq \text{trace}(\boldsymbol{\Gamma}[(|\psi_\mathcal{X}\rangle + |\psi_{\mathcal{B}(\mathcal{X})}\rangle)(\langle\psi_\mathcal{X}| + \langle\psi_{\mathcal{B}(\mathcal{X})}|)]) \\
&\quad \text{\footnotesize ($\boldsymbol{\Gamma}$ is trace-preserving and $|\psi_\mathcal{X}\rangle$ orthogonal to $|\psi_{\mathcal{B}(\mathcal{X})}\rangle$)} \\
&= \text{trace}((|\psi_\mathcal{X}\rangle + |\psi_{\mathcal{B}(\mathcal{X})}\rangle)(\langle\psi_\mathcal{X}| + \langle\psi_{\mathcal{B}(\mathcal{X})}|)) \\
&= \text{trace}(|\psi_\mathcal{X}\rangle \langle\psi_\mathcal{X}|) + \text{trace}(|\psi_{\mathcal{B}_\mathcal{X}}\rangle \langle\psi_{\mathcal{B}_\mathcal{X}}|) \\
&= \mathbb{P}_{|\psi\rangle\langle\psi|}(\mathcal{X}) + \mathbb{P}_{|\psi\rangle\langle\psi|}(\mathcal{B}(\mathcal{X})).
\end{aligned}
$$

"(b) $\Rightarrow$ (a)": Assume that *(a)* does not hold. Thus, there exists some l, some $c, c' \in \mathcal{C}$, some $v' \in \mathcal{X}$ and $v \in \mathcal{X}^c \backslash \mathcal{B}_\mathcal{X}$ such that $\langle c', v' | \mathbf{M}_l | c, v \rangle \neq 0$. If we now consider $|\psi\rangle = |c, v\rangle$, then $\mathbb{P}_{|c,v\rangle\langle c,v|}(\mathcal{X}) + \mathbb{P}_{|c,v\rangle\langle c,v|}(\mathcal{B}(\mathcal{X})) = 0$, whereas $\mathbb{P}_{\boldsymbol{\Gamma}[|\mathbf{c},\mathbf{v}\rangle\langle\mathbf{c},\mathbf{v}|]}(\mathcal{X}) \geq \mathbb{P}_{\boldsymbol{\Gamma}[|\mathbf{c},\mathbf{v}\rangle\langle\mathbf{c},\mathbf{v}|]}((c', v')) = |\langle c', v' | \mathbf{M}_l | c, v \rangle|^2 > 0$. So *(b)* does not hold when *(a)* does not; thus conversely, if *(b)* holds then *(a)* must hold too. $\qquad\square$

In the light of the above lemma, a quantum channel is said to be *local* with respect to a reference lifted graph if (a), or equivalently (b), holds; and from (b) thus, the associated $\boldsymbol{\Psi}_t$ will be local in the sense of Eq.(4) too.

A.2 Proof of Theorem 1

In the main paper, Theorem 1 essentially states that QW mixing can be simulated by an LMC, and the main steps of its proof are described for this setting. Here we provide a formal proof for a more general statement: the mixing performance of any stochastic process that is local and invariant can be simulated using a suitably constructed local LMC.

A.2.1 Simulability of stochastic linear maps

In the following, we first show that the p_t generated by a local stochastic map, starting *from any given initial distribution*, can always be simulated by a sequence of *stochastic transition matrices* which *each satisfy the graph locality*. This sequence will be in general dependent on the initial distribution. The lemma and proof are a generalization of the result by [35] from unitary evolution to abstract stochastic linear maps.

Lemma 5 (Local simulability). *If $\mathbf{\Psi}_t$ is local, then for every pair (p_0, t) with $t > 0$ there exists a local stochastic matrix $\mathbf{P}_t^{(p_0)}$ such that $p_t = \mathbf{P}_t^{(p_0)} p_{t-1}$, where $p_t = \mathbf{\Psi}_t[p_0]$.*

Proof. Call $y = p_{t-1}$ and $z = p_t$. In order to prove the above statement, it is convenient to resort to results concerning *flows over capacitated networks* [46], and, in particular, consider the graph shown in Figure 3, where each edge is assigned a corresponding weight, or capacity. The network consists of a source node s, a sink node r, and two copies $\mathcal{W}$ and $\mathcal{W}'$ of the set of node states $\mathcal{V}$. Node s is connected with capacity $y(v)$ to any node $v \in \mathcal{W}$; any node $v \in \mathcal{W}$ is connected with capacity 1 to any node $v' \in \mathcal{W}'$ iff $(v, v') \in \mathcal{E}$, else the nodes are not connected; and any node $v' \in \mathcal{W}'$ is connected with capacity $z(v')$ to node r. The capacities $y(v)$ and $z(v')$, respectively from s and to r, thus reflect the probability distributions to be mapped. The key observation is the following: if this network can route a steady flow of value 1 from node s to node r, then the fraction from $v \in \mathcal{W}$ that is routed towards $v' \in \mathcal{W}'$ directly defines the entry $e_{v'}^{\dagger} \mathbf{P}_t^{(p_0)} e_v$ that we need, and also denoted $\mathbf{P}_t^{(p_0)}(v', v)$. Indeed, to route a flow of value 1 the edges from r to $\mathcal{W}$ will have to be used to their full capacities $y(v)$, such that the flow through the edges from $\mathcal{W}'$ to s becomes $z(v') = \sum_{v \in \mathcal{V}} \mathbf{P}_t^{(p_0)}(v', v) \, y(v)$; so we would have $\mathbf{P}_t^{(p_0)} y = z$ as claimed.
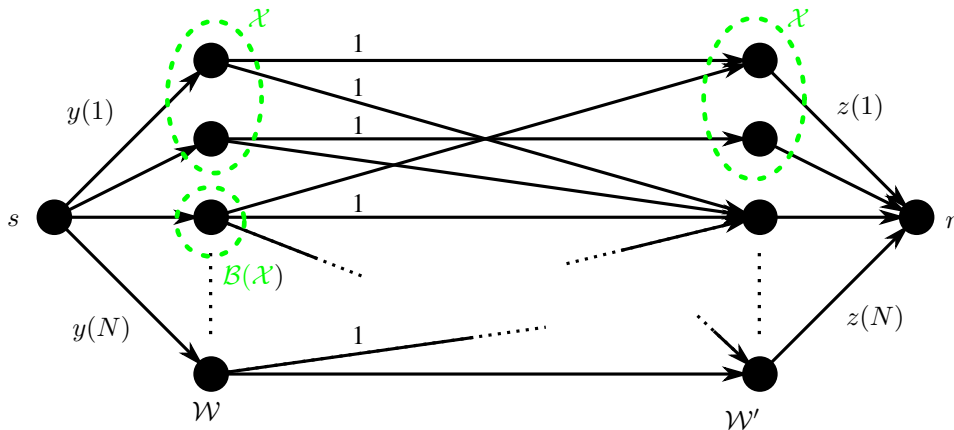

Figure 3: Capacitated network construction used in Lemma 5.

The max-flow min-cut theorem [46] states that the maximum steady flow which can be routed from node s to node r is equal to the minimum cut value of the graph, where a cut value is the sum of the capacities of a set of edges that disconnects s from r. It is clear that cutting all edges arriving at r disconnects the graph, with a cut value of 1, whereas cutting any middle edge between $\mathcal{W}$ and $\mathcal{W}'$ gives a cut value ≥ 1. So the minimum cut should not include any of these "middle" edges, and it must be some combination of edges starting on s or arriving at r. Assume that we know the optimal cut, and let $\mathcal{X} \subseteq \mathcal{W}'$ such that the cut involves the edges from the *complement* of $\mathcal{X} \subseteq \mathcal{W}'$ to r. To block any flow from s to r while keeping all middle edges, we must then cut the edges from s to all the $l \in \mathcal{W}$ which have an edge to $\mathcal{X}$. This corresponds to all $l \in \mathcal{X} \cup \mathcal{B}(\mathcal{X})$. The value of this cut is thus

$$1 - \mathbb{P}_z(\mathcal{X}) + \mathbb{P}_y(\mathcal{X}) + \mathbb{P}_y(\mathcal{B}(\mathcal{X})).$$

Recalling that $y = p_t$ and $z = p_{t+1}$, locality imposes

$$\mathbb{P}_z(\mathcal{X}) \leq \mathbb{P}_y(\mathcal{X}) + \mathbb{P}_y(\mathcal{B}(\mathcal{X})),$$

from which it follows that the minimum value of the cut is ≥ 1. This minimum is attained (among others) with cutting all edges arriving at r, i.e., with $\mathcal{X}$ the empty set. Hence, the minimum cut value is 1 and a $\mathbf{P}_t^{(p_0)}$ solution to our problem exists. $\qquad \square$

A.2.2 Amplification lemma

Lemma 5 is instrumental in proving Thm.1 of the main paper for a finite time frame, by simulating the QW up to some given time. The following will be instrumental to prove the theorem for arbitrary time, showing that a

finite-memory process is sufficient to extend this mixing performance to arbitrarily small $\epsilon > 0$. In particular, we now show that, given an evolution map that mixes up to a certain total variation distance, we can iterate this map in order to mix to arbitrarily small distance, a process informally known as *amplification*.

Lemma 6 (Amplification lemma). *Assume that $\boldsymbol{\Psi}_t$ is a family of stochastic linear maps that mixes to an invariant $\bar{p}$, and admits a mixing time $\overline{\tau}(\epsilon)$ for all $\epsilon > 0$. Then for any $\epsilon_0 < 1/2$, its amplified version defined as*

$$\widetilde{\boldsymbol{\Psi}}_t = \boldsymbol{\Psi}_{t \bmod T} \left(\boldsymbol{\Psi}_T \right)^{\lfloor t/T \rfloor},$$

with $T = \overline{\tau}(\epsilon_0)$, has a mixing time $\tau(\epsilon) \leq \overline{\tau}(\epsilon_0) \cdot \lceil \log(1/\epsilon) \, / \, \log(1/(2\epsilon_0)) \rceil$ for all $\epsilon > 0$.

Proof. We will thus check that at any time $t \geq T \cdot \lceil \log(1/\epsilon) \, / \, \log(1/(2\epsilon_0)) \rceil = \kappa \cdot T$, $\kappa \in \mathbb{N}$, the total variation distance to $\bar{p}$ is lower than ϵ. The proof uses invariance of $\bar{p}$ under $\boldsymbol{\Psi}_t$ to transform $\boldsymbol{\Psi}_t[p] - \bar{p}$ into $\boldsymbol{\Psi}_t[p - \bar{p}]$.

For $t = \kappa \cdot T$, we get

$$\max_p \| \widetilde{\boldsymbol{\Psi}}_{\kappa \cdot T}[p] - \bar{p} \|_{TV} \quad = \quad \max_p \| \left(\boldsymbol{\Psi}_T \right)^{\kappa} [p] - \bar{p} \|_{TV} \quad \leq \quad \max_{p,p'} \| \left(\boldsymbol{\Psi}_T \right)^{\kappa} [p] - \left(\boldsymbol{\Psi}_T \right)^{\kappa} [p'] \|_{TV}$$

$$\text{(see justification below)}$$

$$\leq \quad \left(\max_{p,p'} \| \boldsymbol{\Psi}_T[p] - \boldsymbol{\Psi}_T[p'] \|_{TV} \right)^{\kappa}$$

$$\leq \quad \left(2 \max_p \| \boldsymbol{\Psi}_T[p] - \bar{p} \|_{TV} \right)^{\kappa} \leq (2\epsilon_0)^{\kappa} \leq \epsilon \; .$$

On the last line we have used that $\max_p \| \boldsymbol{\Psi}_T[p] - \bar{p} \|_{TV} \leq \epsilon_0$, and that $(2\epsilon_0)^{\kappa} \leq \epsilon$ as soon as $\kappa \geq \log(1/\epsilon)/\log(1/(2\epsilon_0))$; from first to second line we have used the submultiplicativity of the total variation norm under any linear map in the form stated in [47].

For $t = t' + \kappa \cdot T$ with any $t' > 0$, we know that

$$\| \boldsymbol{\Psi}_t[p] - \bar{p} \|_{TV} = \| \boldsymbol{\Psi}_t[p - \bar{p}] \|_{TV} = \| \boldsymbol{\Psi}_{t'}[\boldsymbol{\Psi}_{\kappa \cdot T}[p - \bar{p}]] \|_{TV} \leq \| \boldsymbol{\Psi}_{\kappa \cdot T}[p - \bar{p}] \|_{TV}, \quad \text{for all } p,$$

thanks to contractivity of the 1-norm under stochastic maps. So finally we find that, for arbitrary $t \geq 0$,

$$\max_p \| \widetilde{\boldsymbol{\Psi}}_t[p] - \bar{p} \|_{TV} \leq \max_p \| \widetilde{\boldsymbol{\Psi}}_{\lfloor t/T \rfloor \cdot T}[p] - \bar{p} \|_{TV} \leq \epsilon,$$

if $t \geq \overline{\tau}(\epsilon_0) \cdot \lceil \log(1/\epsilon) \, / \, \log(1/(2\epsilon_0)) \rceil$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \square$

A.2.3 Proof of main theorem

We can now finalize the proof of a generalized form of Theorem 1 of the main paper for a general stochastic process associated to a family of linear maps $\boldsymbol{\Psi}_t$ over the node set $\mathcal{V}$ that are local, and that leave the target distribution $\bar{p}$ invariant.

Theorem 7 (main paper Thm.1, generalized version). *Let $\boldsymbol{\Psi}_t$ be a stochastic linear map that mixes to some distribution $\bar{p}$ with mixing time $\overline{\tau}(\epsilon)$, satisfying some locality constraint and leaving $\bar{p}$ invariant. Then for any $\epsilon_0 < 1/2$ we can construct an LMC that satisfies the same locality constraint and that mixes to $\bar{p}$ with a mixing time $\tau(\epsilon) = \overline{\tau}(\epsilon)$ for all $\epsilon \geq \epsilon_0$, and a mixing time $\; \tau(\epsilon) \leq \overline{\tau}(\epsilon_0) \cdot \lceil \log(1/\epsilon) \, / \, \log(1/(2\epsilon_0)) \rceil$ for all $\epsilon > 0$.*

Proof. The proof essentially combines the two previous lemmas, and for the rest it follows the construction from the main text. We shall first use Lemma 5 to build a lifted Markov chain that simulates the dynamics of this channel up to time $t = \tau(\epsilon_0)$, and next apply the amplification lemma 6 to prove exponential convergence for $\epsilon < \epsilon_0$ and thus $t > \tau(\epsilon_0)$.

● *[First part: construction for $t \leq \tau(\epsilon_0)$]* Lemma 5 tells us that for every initial state p_0 and given time bound T, there exists a local *stochastic bridge* $\{ \mathbf{P}_t^{p_0} \}_{t=1,2,\ldots,T}$ such that, for all $t \in (0, T]$, we have

$$\boldsymbol{\Psi}_t[p_0] = \mathbf{P}_t^{p_0} \ldots \mathbf{P}_1^{p_0} p_0.$$

This allows us to construct the operator sets $\{ \mathbf{P}_t^{e_v} \}_{t=1,2,\ldots,T}$ for $v \in \mathcal{V}$, where we recall that e_v is the elementary vector corresponding to node v, representing the classical probability vector whose entire weight is on node v. We will combine these bridges into a *single and time-invariant lifted Markov chain* $\mathbf{P}$, mapping probability distributions over the extended set $\hat{\mathcal{V}} = \mathcal{C}' \times \mathcal{V} = (\mathcal{V} \times \{1, 2, \ldots, T\}) \times \mathcal{V}$, where $T = \overline{\tau}(\epsilon_0)$. Let $e_{v_0} \otimes e_l \otimes e_v$ denote

11

the probability (column) vector over $\hat{\mathcal{V}}$ whose weight is centered on element $(v_0, l, v) \in \hat{\mathcal{V}}$, and by $e_{v_0}^\dagger \otimes e_l^\dagger \otimes e_v^\dagger$ we denote the dual or adjoint (row) vector. Now we can define

$$\mathbf{P} \equiv \sum_{v \in \mathcal{V}} e_v e_v^\dagger \otimes \left(\sum_{t=0}^{T-1} e_{t+1} e_t^\dagger \otimes P_t^{e_v} + e_T e_T^\dagger \otimes \mathbf{I}_{\mathcal{V}} \right). \tag{5}$$

To complete the construction, the above evolution should be locally initialized according to the map $\mathbf{F}$, which maps any probability vector p over $\mathcal{V}$ to a probability vector $\mathbf{F}[p]$ over $\hat{\mathcal{V}}$, defined as

$$\mathbf{F}[p] = \sum_{v \in \mathcal{V}} p(v) e_v \otimes e_0 \otimes e_v. \tag{6}$$

The probability vector $\mathbf{F}[p]$ is such that $\mathbf{P}^t \mathbf{F}[p]$ induces the same marginal distribution on $\mathcal{V}$ as $\mathbf{\Psi}_t[p]$ for a fixed time frame:

$$f(\mathbf{P}^t \mathbf{F}[p]) = \mathbf{\Psi}_t[p] \quad \text{for all } t \in [0, \overline{\tau}(\epsilon_0)] ,$$

where f computes the marginal probability distribution induced by the lift on $\mathcal{V}$, i.e.,

$$f \left(\sum_{v_0, l, v} p(v_0, l, v) e_{v_0} \otimes e_l \otimes e_v \right) = \sum_v \left(\sum_{v_0, l} p(v_0, l, v) \right) e_v.$$

As a consequence, initial states $\mathbf{F}[p]$, with an arbitrary p over $\mathcal{V}$, will mix on $\mathcal{V}$ with the same mixing time $\overline{\tau}(\epsilon)$ as $\mathbf{\Psi}_t$ for any $\tau(\epsilon) \le \tau(\epsilon_0)$, i.e., for any $\epsilon \ge \epsilon_0$. This proves the first part of the theorem.

• *[Second part: modifying the construction towards $t > \tau(\epsilon_0)$]* As the size of the lift transition matrix $\mathbf{P}$ scales linearly with $\overline{\tau}(\epsilon_0)$, which in general is unbounded for $\epsilon_0 \to 0$, the above construction only makes sense for fixed ϵ_0. Towards building a lift for arbitrary $\epsilon > 0$, and thus prove the second part of the theorem, we invoke the amplification lemma 6. The lemma shows that, instead of the full process $\mathbf{\Psi}_t$, we can simulate the simpler one $\widetilde{\mathbf{\Psi}}_t$, defined as

$$\widetilde{\mathbf{\Psi}}_t = \mathbf{\Psi}_{t \bmod T} \left(\mathbf{\Psi}_T \right)^{\lfloor t/T \rfloor} ,$$

and ensure a mixing time $\tau(\epsilon) \le \tau(\epsilon_0) \cdot \lceil \log(1/\epsilon) / \log(1/(2\epsilon_0)) \rceil$. It is not difficult to show that the evolution induced by $\widetilde{\mathbf{\Psi}}_t$ can in fact be simulated for an arbitrary number of steps, with a lift of fixed size. To this aim, we modify the lift construction of the first part, namely replacing the unit probability of staying at (v_0, T, v) by a unit probability to jump from (v_0, T, v) to $(v, 0, v)$. Explicitly, we thus adapt the lift as follows:

$$\mathbf{P} \equiv \sum_{v \in \mathcal{V}} \sum_{t=0}^{T-1} e_v e_v^\dagger \otimes e_{t+1} e_t^\dagger \otimes P_t^{e_v} + \sum_{v, v_0 \in \mathcal{V}} e_v e_{v_0}^\dagger \otimes e_0 e_T^\dagger \otimes e_v e_v^\dagger. \tag{7}$$

When associated to the same initialization map $\mathbf{F}$ and marginalization f, this transition matrix gives exactly the same output distributions p_t over $\mathcal{V}$ as the LMC of the first part, for all $t \le T$. At $t = T$, in fact (7) takes the output $p_T = f(\mathbf{P}^T \mathbf{F}[p]) = \mathbf{\Psi}_T[p]$ of the LMC constructed in the first part, and reinitiates the walk with $F(p_T)$ for the next steps. It follows that with (7) we have:

$$f(\mathbf{P}^t \mathbf{F}[p]) = \widetilde{\mathbf{\Psi}}_t[p], \quad \forall t \ge 0.$$

Lemma 6 implies the conclusion about mixing time for all $\epsilon > 0$. $\qquad \square$

The version reported in the main text is the second part of the above theorem for the special case of a stochastic process generated from a QW.

A.3 Proof of conductance bound

We have just shown that quantum channels, and stochastic linear maps in general, can be simulated by lifted Markov chains under appropriate conditions. Accordingly, we can prove a conductance bound for quantum channels and stochastic linear maps by building on a similar bound for LMCs that we provide in Lemma 9 below. It is a generalization of the bound formulated in for instance [31], to the setting where the Markov chain is initialized on a lifted space with some local map $\mathbf{F}$ and where the convergence to a limit distribution only involves the *marginal* over $\mathcal{V}$.

Before going into the statement of the lemmas, let us formalize some concepts more rigorously with the notation of this Supplemental Material. We say that an LMC $\mathbf{P}$ with initialization map $\mathbf{F} : \mathcal{V} \to \mathcal{C} \times \mathcal{V}$ mixes to $\overline{p}$

with a *mixing time* $\tau(\epsilon)$ for all $\epsilon > 0$ if, for any p over $\mathcal{V}$, the induced distribution of $\mathbf{P}^t \mathbf{F}[p]$ over $\mathcal{V}$ is ϵ-close in total variation distance to $\bar{p}$ for all $t \geq \tau(\epsilon)$. We will bound this mixing time using the *conductance*, a quantity that we can associate to a general irreducible Markov chain $\mathbf{P}$ on $\mathcal{V}$. We recall that if $\mathbf{P}$ has a stationary distribution $\bar{p}$, then its conductance $\Phi(\mathbf{P})$ is defined as

$$\Phi(\mathbf{P}) = \min_{\mathcal{X} \subseteq \mathcal{V}: 0 < \bar{p}(\mathcal{X}) \leq \frac{1}{2}} \Phi_{\mathcal{X}}(\mathbf{P}), \quad \text{with} \ \Phi_{\mathcal{X}}(\mathbf{P}) = \frac{\mathbf{Q}_{\mathbf{P}}(\mathcal{X}^c, \mathcal{X})}{\pi(\mathcal{X})},$$

where $\bar{p}(\mathcal{X}) = \mathbb{P}_{\bar{p}}(\mathcal{X})$ and $\mathbf{Q}_{\mathbf{P}}(\mathcal{X}^c, \mathcal{X}) = \sum_{v \in \mathcal{X}, v' \in \mathcal{X}^c} \mathbf{P}(v', v)\bar{p}(v)$ is the ergodic flow from $\mathcal{X}$ to its complement. Here we use as earlier the notation $\mathbf{P}(v', v) = e_{v'}^{\dagger} \mathbf{P} e_v$. We can also associate a conductance to a graph and distribution, without specifying an associated Markov chain. The *graph conductance* $\Phi_{\bar{p}}$ with respect to some distribution $\bar{p}$ is defined as $\Phi_{\bar{p}} = \max_{\mathbf{P}'} \Phi(\mathbf{P}')$, where the maximization runs over all $\mathbf{P}'$ satisfying the locality of the graph and $\mathbf{P}' \bar{p} = \bar{p}$.

To any lifted Markov chain on a lifted graph, we can associate an *induced chain* on the original graph, as introduced in [48]. Thereto, let $\mathbf{P}$ be an irreducible lifted Markov chain on the nodes of a lifted graph $\mathcal{C} \times \mathcal{V}$, having stationary distribution $\hat{\bar{p}}$. The induced chain $\mathbf{P}_{\mathcal{V}}$ over $\mathcal{V}$ is defined by

$$\mathbf{P}_{\mathcal{V}}(v', v) = \sum_{c, c' \in \mathcal{C}} \frac{\hat{\bar{p}}(c, v)}{\bar{p}(v)} \mathbf{P}((c', v'), (c, v)),$$

where $\bar{p}$ represents the stationary distribution of $\mathbf{P}_{\mathcal{V}}$, defined by $\bar{p}(v) = \hat{\bar{p}}(\mathcal{C} \times v)$. This definition is motivated by obtaining matching ergodic flows $\mathbf{Q}_{\mathbf{P}_{\mathcal{V}}}(v', v) = \mathbf{Q}_{\mathbf{P}}(\mathcal{C} \times v', \mathcal{C} \times v)$ and so for any subset $\mathcal{X} \subseteq \mathcal{V}$:

$$\Phi_{\mathcal{X}}(\mathbf{P}_{\mathcal{V}}) = \Phi_{\mathcal{C} \times \mathcal{X}}(\mathbf{P}). \tag{8}$$

This readily implies that $\Phi(\mathbf{P}_{\mathcal{V}}) \geq \Phi(\mathbf{P})$. We also have that $\Phi_{\bar{p}} \geq \Phi(\mathbf{P}_{\mathcal{V}})$, with $\Phi_{\bar{p}}$ the graph conductance associated to $\bar{p}$ on the non-lifted graph. Indeed by definition $\mathbf{P}_{\mathcal{V}}$ obeys the graph locality and $\mathbf{P}_{\mathcal{V}} \bar{p} = \bar{p}$, i.e., it is an element of the set over which the graph conductance $\Phi_{\bar{p}}$ is maximized.

We next borrow standard techniques, as presented in for instance [49] and [47], to prove two instrumental lemmas.

Lemma 8. *Consider an irreducible Markov chain $\mathbf{P}$ over a set $\mathcal{V}$, with unique stationary distribution $\bar{p}$. Then*

$$\mathbb{P}_{\mathbf{P}^t \bar{p}_{\mathcal{X}}}(\mathcal{X}^c) \leq t \, \Phi_{\mathcal{X}}(\mathbf{P}) \quad \text{for all } \mathcal{X} \subseteq \mathcal{V}, t \geq 0 \,,$$

where $\bar{p}_{\mathcal{X}}(v) = \bar{p}(v)/\bar{p}(\mathcal{X})$, for $v \in \mathcal{X}$, and zero elsewhere.

Proof. Note that

$$\mathbb{P}_{\mathbf{P} \bar{p}_{\mathcal{X}}}(X^c) = \sum_{v \in \mathcal{X}, v' \in \mathcal{X}^c} \frac{\mathbf{P}(v', v) \cdot \bar{p}(v)}{\bar{p}(\mathcal{X})} = \Phi_{\mathcal{X}}(\mathbf{P}).$$

We will first prove the following inequalities:

$$\mathbb{P}_{\mathbf{P} \bar{p}_{\mathcal{X}}}(\mathcal{X}^c) = \|\mathbf{P} \bar{p}_{\mathcal{X}} - \bar{p}_{\mathcal{X}}\|_{TV}, \quad \text{and} \quad \mathbb{P}_{\mathbf{P}^t \bar{p}_{\mathcal{X}}}(X^c) \leq \|\mathbf{P}^t \bar{p}_{\mathcal{X}} - \bar{p}_{\mathcal{X}}\|_{TV}, \ \forall t \geq 0. \tag{9}$$

To obtain the first inequality, we can use the equivalent definition of the total variation distance:

$$\|\mathbf{P} \bar{p}_{\mathcal{X}} - \bar{p}_{\mathcal{X}}\|_{TV} = \sum_{v \in V: (\mathbf{P} \bar{p}_{\mathcal{X}})(v) \geq \bar{p}_{\mathcal{X}}(v)} (\mathbf{P} \bar{p}_{\mathcal{X}})(v) - \bar{p}_{\mathcal{X}}(v).$$

We then observe that on the set $\mathcal{X}^c$, $\mathbf{P} \bar{p}_{\mathcal{X}}$ will be elementwise larger than or equal to $\bar{p}_{\mathcal{X}}$, since the latter is zero on $\mathcal{X}^c$; whereas on the set $\mathcal{X}$, $\mathbf{P} \bar{p}_{\mathcal{X}}$ will be elementwise smaller than $\bar{p}_{\mathcal{X}}$:

$$(\mathbf{P} \bar{p}_{\mathcal{X}})(v') = \frac{\sum_{v \in \mathcal{X}} \mathbf{P}(v', v) \bar{p}(v)}{\bar{p}(\mathcal{X})} \leq \frac{\sum_{v \in \mathcal{V}} \mathbf{P}(v', v) \bar{p}(v)}{\bar{p}(\mathcal{X})} = \frac{\bar{p}(v')}{\bar{p}(\mathcal{X})} = \bar{p}_{\mathcal{X}}(v') \quad \text{since } P\bar{p} = \bar{p}.$$

Now we can rewrite $\sum_{v \in \mathcal{V}: (\mathbf{P} \bar{p}_{\mathcal{X}})(v) \geq \bar{p}_{\mathcal{X}}(v)} (\mathbf{P} \bar{p}_{\mathcal{X}})(v) - \bar{p}_{\mathcal{X}}(v) = \sum_{v \in \mathcal{X}^c} (\mathbf{P} \bar{p}_{\mathcal{X}})(v) = \mathbb{P}_{\mathbf{P} \bar{p}_{\mathcal{X}}}(\mathcal{X}^c)$. To obtain the inequality in (9) we expand the total variation norm:

$$\|\mathbf{P}^t \bar{p}_{\mathcal{X}} - \bar{p}_{\mathcal{X}}\|_{TV} = \frac{1}{2} \sum_{v \in \mathcal{V}} |(\mathbf{P}^t \bar{p}_{\mathcal{X}})(v) - \bar{p}_{\mathcal{X}}(v)| = \frac{1}{2} \mathbb{P}_{\mathbf{P}^t \bar{p}_{\mathcal{X}}}(X^c) + \frac{1}{2} \sum_{v \in \mathcal{X}} |(\mathbf{P}^t \bar{p}_{\mathcal{X}})(v) - \bar{p}_{\mathcal{X}}(v)|$$

$$\geq \frac{1}{2} \mathbb{P}_{\mathbf{P}^t \bar{p}_{\mathcal{X}}}(\mathcal{X}^c) + \frac{1}{2} \left| \sum_{v \in \mathcal{X}} (\mathbf{P}^t \bar{p}_{\mathcal{X}})(v) - \bar{p}_{\mathcal{X}}(v) \right|$$

$$= \frac{1}{2} \mathbb{P}_{\mathbf{P}^t \bar{p}_{\mathcal{X}}}(\mathcal{X}^c) + \frac{1}{2} |1 - \mathbb{P}_{\mathbf{P}^t \bar{p}_{\mathcal{X}}}(X^c) - 1| = \mathbb{P}_{\mathbf{P}^t \bar{p}_{\mathcal{X}}}(X^c).$$

Starting with this inequality, we apply the triangle inequality on $\mathbf{P}^t p - p = (\mathbf{P}^t p - \mathbf{P}^{t-1} p) + (\mathbf{P}^{t-1} p - \mathbf{P}^{t-2} p) + \cdots + (\mathbf{P} p - p)$; next we bound each term by $\|\mathbf{P}\bar{p}_{\mathcal{X}} - \bar{p}_{\mathcal{X}}\|_{TV}$ thanks to contractivity, i.e., as at the end of the proof of Lemma 6, the fact that for arbitrary distributions p, p' and a stochastic matrix $\mathbf{P}$, we have $\|\mathbf{P}p - \mathbf{P}p'\|_{TV} \leq \|p - p'\|_{TV}$; and finally we apply the equality from (9). This yields:

$$\mathbb{P}_{\mathbf{P}^t \bar{p}_{\mathcal{X}}}(X^c) \leq \left\|\mathbf{P}^t \bar{p}_{\mathcal{X}} - \pi_{\mathcal{X}}\right\|_{TV} \leq t \left\|\mathbf{P}\bar{p}_{\mathcal{X}} - \bar{p}_{\mathcal{X}}\right\|_{TV} = t\,\Phi_{\mathcal{X}}(\mathbf{P}).$$

$\square$

Lemma 9. *Consider an irreducible lifted Markov chain $\mathbf{P}$ on a lifted state space $\mathcal{C} \times \mathcal{V}$, and call $\mathbf{P}_{\mathcal{V}}$ its induced chain on $\mathcal{V}$. Then its mixing time on $\mathcal{V}$ satisfies*

$$\tau(1/4) \geq \frac{1}{4\Phi(\mathbf{P}_{\mathcal{V}})} \geq \frac{1}{4\Phi_{\bar{p}}},$$

where $\Phi_{\bar{p}}$ is the graph conductance associated to the corresponding limit distribution on $\mathcal{V}$.

Proof. Let f be the function computing the marginal distribution over $\mathcal{V}$ from a distribution over $\mathcal{C} \times \mathcal{V}$. By applying the reverse triangle inequality, it is easily seen that

$$\left\|\mathbf{P}^t p - \hat{\bar{p}}\right\|_{TV} = \frac{1}{2} \sum_{(c,v) \in \mathcal{C} \times \mathcal{V}} \left|(\mathbf{P}^t p)(c,v) - \hat{\bar{p}}(c,v)\right|$$

$$\geq \frac{1}{2} \sum_{v \in V} \left|\sum_{c \in \mathcal{C}} (\mathbf{P}^t p)(c,v) - \hat{\bar{p}}(c,v)\right| = \left\|f(\mathbf{P}^t p) - f(\hat{\bar{p}})\right\|_{TV}.$$

Now take a subset $\mathcal{X} \subseteq \mathcal{V}$ such that $\hat{\bar{p}}(\mathcal{C} \times \mathcal{X}) \leq 1/2$. Using this subset we define a second marginalization $g_{\mathcal{X}}$ mapping distributions over the nodes of $\mathcal{V}$ to the binary property $\{v \in \mathcal{X}\}$ or $\{v \notin \mathcal{X}\}$, i.e., $g_{\mathcal{X}}(p)$ can be represented as a vector $[\sum_{v \in \mathcal{X}} p(v) \,;\, \sum_{v \notin \mathcal{X}} p(v)]$. By a similar reasoning we get

$$\left\|f(\mathbf{P}^t p) - f(\hat{\bar{p}})\right\|_{TV} \geq \left\|g_{\mathcal{X}}(f(\mathbf{P}^t p)) - g_{\mathcal{X}}(f(\hat{\bar{p}}))\right\|_{TV}.$$

If we take $p = \hat{\bar{p}}_{\mathcal{C} \times \mathcal{X}}$ as defined in Lemma 8 and we use the triangle inequality again, it follows that

$$\left\|g_{\mathcal{X}}(f(\mathbf{P}^t \hat{\bar{p}}_{\mathcal{C} \times \mathcal{X}})) - g_{\mathcal{X}}(f(\hat{\bar{p}}))\right\|_{TV}$$

$$\geq \left\|g_{\mathcal{X}}(f(\hat{\bar{p}}_{\mathcal{C} \times \mathcal{X}})) - g_{\mathcal{X}}(f(\hat{\bar{p}}))\right\|_{TV} - \left\|g_{\mathcal{X}}(f(\hat{\bar{p}}_{\mathcal{C} \times \mathcal{X}})) - g_{\mathcal{X}}(f(\hat{\mathbf{P}}^t \hat{\bar{p}}_{\mathcal{C} \times \mathcal{X}}))\right\|_{TV}$$

$$\geq \frac{1}{2} - \mathbb{P}_{\mathbf{P}^t \hat{\bar{p}}_{\mathcal{C} \times \mathcal{X}}}(\mathcal{C} \times \mathcal{X}^c) \geq \frac{1}{2} - t\Phi_{\mathcal{C} \times \mathcal{X}}(\mathbf{P}) = \frac{1}{2} - t\Phi_{\mathcal{X}}(\mathbf{P}_{\mathcal{V}}).$$

From first to second line, we have used that $\mathbb{P}_{g_{\mathcal{X}}(f(\hat{\bar{p}}_{\mathcal{C} \times \mathcal{X}}))}(x) = 1$, while $\mathbb{P}_{g_{\mathcal{X}}(f(\hat{\bar{p}}))}(x) \leq 1/2$; this ensures that on the right hand side of the first line, the first term is $\geq 1/2$, while the second term boils down exactly to the probability to be in $\mathcal{X}^c$. The last inequalities follow from Lemma 8 and equation (8). We thus find altogether that

$$\left\|f(\mathbf{P}^t p) - f(\hat{\bar{p}})\right\|_{TV} \geq \frac{1}{2} - t\Phi_{\mathcal{X}}(\mathbf{P}_{\mathcal{V}}).$$

For $t = \tau(1/4)$, the left hand side must be smaller than $1/4$ and rearranging terms yields $\tau(1/4) \geq 1/(4\Phi_{\mathcal{X}}(\mathbf{P}_{\mathcal{V}}))$. The same obviously holds true when minimizing $\Phi_{\mathcal{X}}$ over $\mathcal{X}$, yielding the statement with $\Phi(\mathbf{P}_{\mathcal{V}})$. The fact that $\Phi_{\bar{p}} \geq \Phi(\mathbf{P}_{\mathcal{V}})$ was already discussed after equation (8). $\square$

Combining Theorem 7 with the bound on the mixing time of LMCs provided by Lemma 9, leads to the following bound for quantum channels:

Theorem 10 (main paper Thm.2, generalized version). *Any local and invariant stochastic linear map has a mixing time $\tau(1/4) \geq 1/(4\Phi_{\bar{p}})$. As a consequence, any $\bar{p}$-invariant QW has a mixing time $\tau(1/4) \geq 1/(4\Phi_{\bar{p}})$. There exists such a QW that has a mixing time $\tau(\epsilon) \leq O(\log(1/\min_k \bar{p}_k) \log(1/\epsilon) / \Phi_{\bar{p}})$.*

Proof. If the stochastic linear map has a $1/4$-mixing time $\tau(1/4)$, then according to Theorem 7 with $\epsilon_0 \leq 1/4$ we can construct a local LMC with a marginal $1/4$-mixing time equal to $\tau(1/4)$. However, from Lemma 9 we can bound the $1/4$-mixing time of any such LMC by the graph conductance $\Phi_{\bar{p}}$. At the beginning of the second section we explain how we can associate a stochastic linear map that is local and invariant, to any local quantum channel that leaves the same target distribution invariant. This readily implies that the lower bound on mixing time holds for such QWs. The existence result follows from the same existence result for lifted Markov chains, proven in [31]. Its validity for QWs follows by recognizing that lifted Markov chains are a special class of quantum channels. $\square$

A.4 Quantum Walks on lattices

Consider a d-dimensional periodic lattice $\mathbb{Z}_M^d$ of side M, encoded in a graph with node set

$$\mathcal{V} = \{(i_1, i_2, \ldots, i_d) | 1 \leq i_k \leq M, \forall k\}.$$

Similar to the QW/LMC construction for the cycle, we lift this graph by adding a set of coin states $\mathcal{C} = \{+_k, -_k | 1 \leq k \leq d\}$. An LMC on this graph thus takes place on the vector space $\mathcal{H}_\mathcal{C} \otimes \mathcal{H}_\mathcal{V} = \{ e_{c,v} | (c,v) \in \mathcal{C} \times \mathcal{V} \}$. With operator $\mathbf{P}_k^\pm$ defined on $\mathcal{H}_\mathcal{V}$ as the cyclic permutation of the k-th dimension, that is, $\mathbf{P}_k^\pm e_{\ldots,i_{k-1},i_k,i_{k+1}\ldots} = e_{\ldots,i_{k-1},(i_k\pm 1)\bmod M,i_{k+1},\ldots}$ for all $1 \leq k \leq d$, the LMC defined in [40] writes:

$$\mathbf{P} = \left(\sum_k e_{+_k} e_{+_k}^\dagger \otimes \mathbf{P}_k^+ + e_{-_k} e_{-_k}^\dagger \otimes \mathbf{P}_k^- \right) \cdot (\mathbf{S} \otimes \mathbf{I}_\mathcal{V})$$

$$= \begin{bmatrix} \mathbf{P}_1^+ & & & & \\ & \mathbf{P}_1^- & & & \\ & & \mathbf{P}_2^+ & & \\ & & & \ddots & \\ & & & & \mathbf{P}_d^- \end{bmatrix} \cdot (\mathbf{S} \otimes \mathbf{I}_\mathcal{V}), \tag{10}$$

where we now specifically select

$$\mathbf{S} = \begin{bmatrix} 1-(2d-1)\alpha & \alpha & \ldots & \alpha \\ \alpha & 1-(2d-1)\alpha & \ldots & \alpha \\ \vdots & \vdots & \ddots & \vdots \\ \alpha & \alpha & \ldots & 1-(2d-1)\alpha \end{bmatrix},$$

with $\alpha = 1/(2dM)$, and we recall that $\mathbf{I}_\mathcal{V}$ is the identity matrix on $\mathcal{H}_\mathcal{V}$.

We will prove the mixing time for M odd. For M even, the LMC shows a parity problem: starting from a single state, at any given time the walk will be supported only on the even or only on the odd nodes. This is easily remedied by for instance modifying $\mathbf{P}$ to $(\mathbf{P} + \mathbf{I}_{\mathcal{C}\times\mathcal{V}})/2$, which changes the mixing time only by a constant factor, or by randomizing the parity of the initial state. To facilitate its reading, we again structure the proof using two technical lemmas.

Lemma 11. *Assume that the LMC in Eq. (10) starts with any $p_0 \in \{e_{(c,v)}\}$, i.e., with all its weight concentrated on a single vertex and single coin choice $c \in \{k_+, k_-\}$ for some k. Then the resulting distribution after $2M$ time steps has uniformly mixed i_k with a probability $\geq 1/(16d)$, in the sense that $\mathbb{P}_{p_{2M}}(i_k = n) \geq 1/(16d\,M)$ for all $n \in \{1, 2, ..., M\}$.*

Proof. By symmetry, we can consider without loss of generality that the initial distribution is

$$p_0 = e_{+_1, 1, \ldots, 1}.$$

Writing $\mathbf{S} = (1 - 2d\alpha)\mathbf{I}_\mathcal{C} + 2d\alpha \mathbf{1}/2d$, where $\mathbf{1}$ denotes the matrix of all ones, we say that at each time step with probability $2d\alpha = 1/M$ a coin toss takes place. Then the probability of a single coin toss happening over $2M$ steps is given by

$$\mathbb{P}_{2M \text{ steps}}(1 \text{ coin toss}) = \binom{2M}{1} \cdot \frac{1}{M} \cdot \left(1 - \frac{1}{M}\right)^{2M-1}$$

$$= 2\left(1 - \frac{1}{M}\right)^{2M-1} \geq \frac{1}{8} \quad \text{for } M \geq 2,$$

where the inequality follows from the fact that $(1 - 1/M)^M$ is an increasing function of M, going from $1/4$ for $M = 2$ to $1/e$ for M large. From this, the event E_1 that a single coin toss takes place and switches the coin state from $+_1$ to $-_1$, occurs with probability

$$\mathbb{P}_{2M \text{ steps}}(E_1) = \frac{1}{2d} \cdot \mathbb{P}_{2M \text{ steps}}(1 \text{ coin toss}) \geq \frac{1}{16d}.$$

When E_1 holds true with the single coin toss at time $T \in [1, 2M]$, the distribution at time $2M$ equals

$$p_{2M} = e_{-_1,\ 1+(T-1)-(2M-T+1)\bmod M\ ,1,\ldots,1} = e_{-_1,\ (2T-1)\bmod M\ ,1,\ldots,1}\ .$$

15

Yet, conditional on the fact that E_1 holds true, the timing T of the single coin toss is uniformly distributed between 1 and $2M$. As a consequence, $2T - 1$ is uniformly distributed over $1, 3, 5, ..., M, 2, 4, 6, ..., M - 1, 1, 3, 5, ..., M, 2, 4, 6, ..., M - 1$, i.e., effectively over the integers from 1 to M. Thus

$$\mathbb{P}_{p_{2M}}(i_k = n) \geq \mathbb{P}_{E_1}(i_k = n) \cdot \mathbb{P}_{2M \text{ steps}}(E_1) = \frac{1}{16d} \frac{1}{M} \quad \forall n \in \{1, 2, ..., M\},$$

which proves the statement. $\qquad\square$

Our application of Lemma 11 to prove the following result is loose. Its sequential use, coordinate by coordinate, leaves further room for improvement, and this is why we think that it should be possible to win a factor d on the mixing time. However, the resulting estimate is sufficient for the story of the main paper.

Lemma 12. *Consider the LMC defined in Eq.(10) on $\mathbb{Z}_M^d$, with M odd. For any initial distribution p_0, the distribution p_T after $T = 3M \cdot d(d \log(d) + d)$ steps satisfies*

$$p_T = \mathbf{P}^T p_0 = q \cdot \bar{p} + (1 - q) \cdot \left(\frac{p_T - q \cdot \bar{p}}{1 - q} \right), \tag{11}$$

with $\tilde{p}_T = (p_T - q \cdot \bar{p}) / (1 - q)$ a positive distribution; $\bar{p}$ the stationary distribution of $\mathbf{P}$, which is the uniform distribution over $\mathcal{C} \times \mathcal{V}$; and $q = (1 - 1/e)/2$ where $e = \exp(1)$.

Proof. We consider time intervals of $3M$ steps, which we analyze as follows:

- in the first M steps: As in Lemma 11, we say that at each time step with probability $2d\alpha = 1/M$ a coin toss takes place. Now, we use that this completely randomizes the coin state. The probability that no such coin toss has happened after M steps is $(1 - 1/M)^M \leq 1/e$.

- in the next $2M$ steps: By lemma 11 the coordinate corresponding to the randomized coin state is mixed uniformly with a probability $\geq 1/16d$. If that coordinate was already in a more mixed state than in the hypothesis of lemma 11, then the resulting mixing can only be better.

Each interval of $3M$ steps will thus uniformly mix a random coordinate with probability $r \geq 1/(16e\,d)$. Using Cantelli's inequality for a binomial process with success probability r, we find the following bound for the number of successful mixing episodes k:

$$\mathbb{P}_{2l/r \text{ iterations}}(k \geq l) \geq 1/2.$$

So if we go through $(d \log(d) + d) \cdot 32e\,d$ such intervals of $3M$ steps, then with a probability $1/2$ we will have mixed $d \log(d) + d$ coordinates; the latter are chosen randomly according to independent uniform processes with repetition. According to the coupon collector's problem, $d \log(d) + d$ random choices selects all coordinates with a probability $(1 - 1/e)$. This implies that we can bound the state after $T = 3M \cdot (d \log(d) + d) \cdot 32e\,d$ steps as $p_T \geq q \cdot \pi$, with $q = (1 - 1/e)/2$. $\qquad\square$

We now have all the pieces to prove the actual result.

Theorem 13 (main paper Thm.3)**.** *The LMC defined in Eq.(10) on $\mathbb{Z}_M^d$, with M odd, has a mixing time $\tau(\epsilon) \leq O(M\ d^2 \log(d) \log(1/\epsilon))$. (We recall that the assumption that M is odd is a standard technicality, to avoid discussing all possible easy ways to break the parity symmetry.)*

Proof. We have shown in Lemma 12 that for any initial distribution p_0 over $\mathcal{V}$, with a fixed probability q the state will be uniformly mixed after $T = 3M \cdot d(d \log(d) + d)$ steps, i.e., p_T will be of the form (11). Then after $2T$ steps, we get

$$p_{2T} = \mathbf{P}^T p_T = \mathbf{P}^T (q \cdot \bar{p} + (1 - q) \cdot \tilde{p}_T) = q \cdot \bar{p} + q(1 - q) \cdot \bar{p} + (1 - q)^2 \cdot \tilde{p}_{2T}.$$

And after another $(k - 2)T$ steps we find by an iterative argument that

$$p_{kT} = (1 - (1 - q)^k) \cdot \bar{p} + (1 - q)^k \cdot \tilde{p}_{kT}.$$

This shows that

$$\|\mathbf{P}^t p_0 - \bar{p}\|_{TV} \leq (1 - q)^{\lfloor t/T \rfloor} \quad \forall t \geq 0, p_0,$$

and thus $\|\mathbf{P}^t p_0 - \bar{p}\|_{TV} \leq \epsilon$ provided $t \geq T \cdot \left(1 + \frac{\log \epsilon^{-1}}{\log (1-q)^{-1}} \right)$. As q is a fixed constant below 1, and $T \in O(M \cdot d^2 \log(d))$, this proves the claimed mixing time. $\qquad\square$