Routing entanglement in the quantum internet

npj | Quantum Information

www.nature.com/npjqi

ARTICLE OPEN

Routing entanglement in the quantum internet

Mihir Pant[1,2], Hari Krovi[2], Don Towsley[3], Leandros Tassiulas[4], Liang Jiang[5,6], Prithwish Basu[7], Dirk Englund[1] and Saikat Guha[2,8]

Remote quantum entanglement can enable numerous applications including distributed quantum computation, secure communication, and precision sensing. We consider how a quantum network—nodes equipped with limited quantum processing capabilities connected via lossy optical links—can distribute high-rate entanglement simultaneously between multiple pairs of users. We develop protocols for such quantum "repeater" nodes, which enable a pair of users to achieve large gains in entanglement rates over using a linear chain of quantum repeaters, by exploiting the diversity of multiple paths in the network. Additionally, we develop repeater protocols that enable multiple user pairs to generate entanglement simultaneously at rates that can far exceed what is possible with repeaters time sharing among assisting individual entanglement flows. Our results suggest that the early-stage development of quantum memories with short coherence times and implementations of probabilistic Bell-state measurements can have a much more profound impact on quantum networks than may be apparent from analyzing linear repeater chains. This framework should spur the development of a general quantum network theory, bringing together quantum memory physics, quantum information theory, quantum error correction, and computer network theory.

npj Quantum Information (2019)5:25 ; https://doi.org/10.1038/s41534-019-0139-x

INTRODUCTION

A *quantum network* can generate, distribute, and process quantum information in addition to classical data.[1,2] The most important function of a quantum network is to generate long distance quantum entanglement, which serves a number of tasks including the generation of multiparty shared secrets whose security relies only on the laws of physics,[3,4] distributed quantum computing,[5] improved sensing,[6,7] blind quantum computing (quantum computing on encrypted data),[8] and secure private-bid auctions.[9]

Recent experiments have demonstrated *entanglement links*, viz., entanglement established between quantum memories separated by a few kilometers using a point-to-point optical link,[10] and longer range entanglement with satellites.[11,12] Further near-term demonstrations of long-range terrestrial entanglement are expected.[13]

The conceptually simplest measurement module at a quantum network node is the two-qubit Bell state measurement (BSM), also known as *entanglement swapping*. BSMs have been experimentally demonstrated in many physical systems.[14–19] As illustrated in Fig. 1a, BSMs performed at nodes of a quantum network can glue together small entanglement links into longer-distance entangled clusters. The quantum processing at a quantum network node in the near-term to medium-term will probably be limited to BSMs that are probabilistic because of losses in optical fiber, and inherent limitations of the quantum processing hardware.

In this paper, we develop and analyze *routing* protocols for generating entanglement simultaneously between multiple pairs of users in a quantum network, where each network (repeater) node is equipped with quantum memories, entanglement sources, the ability to perform a BSM between any pair of locally-held qubits, classical computing resources, and a classical communication interface. Our entanglement routing protocols instruct nodes, at every time slot, on how to dynamically choose which BSMs to perform, based on the current knowledge of the network topology, location of end users and current link state knowledge, so as to maximize the entanglement generation rate for a collection of entanglement flows.

We find that even with this limited quantum processing capability at network nodes—the standard primitive to construct and analyze linear repeater chains—entanglement routing on a network affords some very interesting possibilities. For example, we find that *multi-path routing*, i.e., using multiple paths for routing entanglement between a pair of end users, can enable long distance entanglement generation with a superior rate-vs.-distance scaling than what is possible with a single linear repeater chain routing along the shortest path connecting the users (Pirandola recently showed,[20] for an information-theoretic description of repeaters that are ideal fully-error-corrected universal quantum processors, that the optimal rate attainable for multi-path entanglement routing using such ideal repeaters is superior to the rate of a linear chain of ideal repeaters). While an increased rate would be expected with multi-path routing due to a constant factor increase in the number of disjoint paths, we find that the gap between the performance of our multi-path routing protocol and that of a linear repeater chain grows exponentially with distance. Moreover, if the repeater nodes have 'global' link-state knowledge (knowledge of the state of all links in the network at every time step) *and* the entanglement generation success probability over each link is above a (percolation) threshold, we find that multi-path routing enables long-distance entanglement-

[1]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA; [2]Quantum Information Processing group, Raytheon BBN Technologies, 10 Moulton Street, Cambridge, MA, USA; [3]College of Information and Computer Sciences, University of Massachusetts, Amherst, MA, USA; [4]School of Engineering and Applied Science, Yale University, 17 Hill House Avenue, New Haven, CT, USA; [5]Departments of Applied Physics and Physics, Yale University, New Haven, CT, USA; [6]Yale Quantum Institute, Yale University, New Haven, CT, USA; [7]Advanced Networking Systems, Raytheon BBN Technologies, 10 Moulton Street, Cambridge, MA, USA and [8]College of Optical Sciences, University of Arizona, 1630 East University Boulevard, Tucson, AZ, USA
Correspondence: Mihir Pant (mihir.pant@gmail.com)

Received: 22 April 2018 Accepted: 11 February 2019
Published online: 13 March 2019

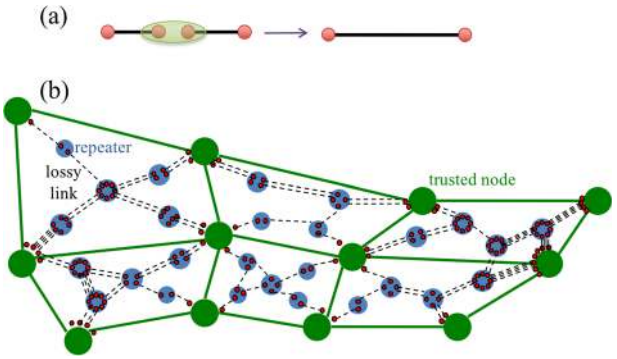

Fig. 1 **a** Example of using Bell measurement (green ovals) to connect two entangled links into a longer entangled link. Red circles represent qubits and black lines represent entanglement. **b** Schematic of a general quantum repeater network. The large (green) circles represent 'trusted' nodes, which are connected via a classical network. The blue circles denote repeater stations, and the red circles inside them represent quantum memories. The dashed lines connecting the red circles are lossy optical channels (e.g., optical fiber). In principle, all network nodes could be equipped with quantum repeaters (i.e., no trusted nodes), in which case depending upon the application need, a node can be a consumer of shared entanglement, or act a router to conduit entanglement flows between other nodes

generation at a rate that depends only linearly on the transmissivity η of a single link in the network, whereas the rate achieved by a linear repeater chain connecting Alice and Bob along the shortest path would be proportional to $\eta^{n_{SP}}$ where n_{SP} is the length of the shortest path. We also present a multi-flow routing protocol that can allow multiple pairs of users to generate shared entanglement simultaneously over the network with simultaneous rates that significantly exceeds with those of achieved when each repeater node makes BSM decisions by simply time-sharing between catering to the individual flows.

Let us consider a quantum network with topology described by a graph $G(V, E)$. Each of the $N = |V|$ nodes is equipped with a quantum repeater, and each of the $M = |E|$ edges is a lossy optical channel of range L_i (km) and power transmissivity $\eta_i \propto e^{-\alpha L_i}$, $i \in E$. Consider K source-destination (Alice-Bob) pairs (A_j, B_j), $1 \le j \le K$, situated at (not necessarily distinct) nodes in V, each of which would like to generate maximally entangled qubits (i.e., ebits) between themselves (and thus by definition not entangled with any other party, due to the monogamy property of entanglement), at the maximum rates possible R_j (ebits per channel use). The high-level objective is: *Given a class of quantum and classical operations at each of the repeater nodes of the underlying network, what operations should be performed at the repeater stations to maximize the rate region* $(R_1, R_2, \ldots, R_K)$ *simultaneously achievable by the entanglement flows?* More importantly, one would like to address networking questions such as: (a) what is the maximum rate-region attainable, (b) what is the tradeoff between sum throughput and latency of the K entanglement flows, and (c) where should repeater nodes be placed, subject to constraints on device metrics (e.g., memories, sources, and detectors), to maximize the attainable rate region. Ultimately one would like to develop explicit and efficient practical quantum routing protocols that employ quantum operations implemented via lossy and noisy devices, while only requiring local link-state knowledge and limited knowledge of the global network topology, analogous to the classical internet. Ideally, one would also want to look into what benefits are afforded by network nodes being equipped with more complex quantum measurement modules, e.g., a measurement that projects n locally held qubits into one of the 2^n n-qubit GHZ states, with $n > 2$ (BSM corresponds to $n = 2$).

The entanglement-generation rate across a link of transmissivity η, in the absence of any repeater mediation, is limited to:

$$C(\eta) = -\log_2(1-\eta) \text{ ebits per mode}, \quad (1)$$

which $\approx 1.44\eta$ ebits per mode when $\eta \ll 1$.[21] (The achievability of $-\log_2(1-\eta)$ ebits per mode of secret communication rate over the lossy channel with two way authenticated public classical communication was first proven in 2009 by Pirandola.[22] In 2014, Takeoka et al. proved an upper bound to the secret-key agreement capacity, $\log_2[(1+\eta)/(1-\eta)]$ ebits per mode,[23] which equals $\approx 2.88\eta$ ebits per mode when $\eta \ll 1$, thereby establishing that the rate attained by *any* protocol must decay linearly with the channel's transmissivity and hence exponentially with distance L in optical fiber since $\eta \sim e^{\alpha L}$. In 2015, Pirandola et al. proved an improved weak converse upper bound of $-\log_2(1-\eta)$ ebits per mode, which established that as the secret key agreement capacity of the pure loss bosonic channel.[21] Subsequently, Wilde et al.[24] proved $-\log_2(1-\eta)$ ebits per mode as a strong converse upper bound to the secret-key agreement capacity). The number of modes per second is a device-technology-dependent constant, upper bounded by the maximum of the optical bandwidth of the source and the electrical bandwidth of the detector. Since $\eta \sim e^{-\alpha L}$ where L is the length of optical fiber, the ebits-per-mode rate also falls off exponentially with range L. Most analyses of repeater networks have been limited to linear chains, with the objective of outperforming the repeater-less bound.[25–32] Pirandola analyzed entanglement-generation capacities of repeater networks assuming ideal repeater nodes, i.e., those equipped with fully-error-corrected quantum processors and argued that for a single flow ($K = 1$), the maximum entanglement-generation rate R_1 reduces to the classical max-flow min-cut problem with edge e being associated with capacity $C(e) = -\log_2(1-\eta(e))$ ebits per channel use,[20] where $\eta(e)$ is the transmissivity of edge e. Pirandola subsequently argued that classical cut-set bounds with the above link capacity give outer bounds to the K-flow capacity region, but again, for ideal repeater nodes. Azuma et al. independently established an upper bound[33] to the entanglement generation bound to the rate which has the same asymptotic scaling but is not tight. Azuma has also looked at an "aggregated" protocol in which the repeater protocols run in parallel.[34] Schoute and co-authors[35] developed routing protocols on specific network topologies and found scaling laws as functions of N, the number of qubits in the memories at nodes, and the time and space consumed by the routing algorithms, under the assumption that each link generates a perfect, lossless EPR pair in every time-slot, and that the nodes' actions are limited to (perfect) Bell-state measurements (BSMs). Acín and co-authors[36] have considered the problem of entanglement percolation where neighboring nodes share a perfect, lossless pure state. Further, van Meter and co-authors developed explicit networking protocols also restricted to pair-wise EPR pair generation and BSMs, but accounting for imperfect fidelities of the EPR pairs (and thus requiring purification over multiple imperfect pairs), and finite coherence times of the qubit memories.[37] There has also been previous work on quantum network coding[38–41] and linear-optic quantum routers.[42]

RESULTS

Problem statement and notation

Consider a graph $G(V, E)$ that denotes the topology of the repeater network. See Fig. 1b for an illustration. Each node $v \in V$ is a repeater (blue circles), and each edge $e \in E$ is a physical link connecting two repeater nodes. $S(e) \in \mathbb{Z}^+$ is an integer edge weight, which corresponds to the number of parallel (spatial, spectral, and/or polarization mode) channels across the edge e (shown using dashed lines). The number of memories at node v is $\sum_{e \in \mathcal{N}(v)} S(e)$ (see Fig. 1b), where the sum is over $\mathcal{N}(v)$, the set of

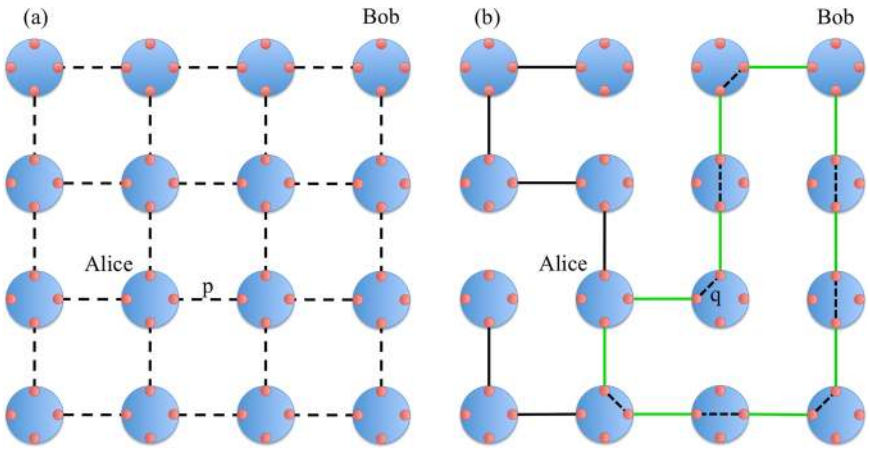

Fig. 2 Schematic of a square-grid topology. The blue circles represent repeater stations and the red circles represent quantum memories. Every cycle (time slot) of the protocol consists of two phases. **a** In the first (external) phase, entanglement is attempted between neighboring repeaters along all edges, each of which succeed with probability p (dashed lines). **b** In the second (internal) phase, entanglement swaps are attempted within each repeater node based on the successes and failures of the neighboring links in the first phase—with the objective of creating an unbroken end-to-end connection between Alice and Bob. Each of these internal connections succeed with probability q. Memories can hold qubits for $T \geq 1$ time slots

nearest neighbor edges of v, with $d(v) = |\mathcal{N}(v)|$ being the degree of node v.

To simplify our analysis, we assume that time is slotted and that each memory can hold a qubit perfectly for $T \geq 1$ time slots, after which the stored qubit completely decoheres (T should be taken to be much smaller than the memory's coherence time). Each time slot t, $t = 1, 2, \ldots$, is divided into two phases: the "external" phase and the "internal" phase, which occur in that order. During the external phase, each of the $S(e)$ pairs of memories across an edge e attempts to establish a shared entangled (EPR) pair. An entanglement attempt across any one of the $S(e)$ parallel links across edge e succeeds with probability $p_0(e) \sim \eta(e)$,[21,27] where $\eta(e) \sim e^{-\alpha L(e)}$ is the transmissivity of a lossy optical channel of length $L(e)$. Using two-way classical communication over edge $e(u, v)$, neighboring repeater nodes u, v learn which of the $S(e)$ parallel links (if any) succeeded in the external phase, in a given time slot.

Let us assume that neighboring repeaters pick upto one successfully created ebit (i.e., ignore multiple successes if any) as in refs. [29,31], in which case the probability that one ebit is established successfully across the edge e during the external phase is given by: $p(e) = 1 - (1 - p_0)^{S(e)}$. Let us also assume $S(e) = S$, $\forall e \in E$, which in turn gives us $p(e) = p$, $\forall e \in E$. While our results in this paper can be adapted to any network topology, we will henceforth use the 2D regular square grid topology (Fig. 2) to illustrate the performance of our routing algorithms.

One instance of the resulting external links created between repeater nodes after the external phase is shown in Fig. 2b using solid lines. In the internal phase of the time slot, entanglement swaps (BSMs) are attempted locally at each repeater node between pairs of qubit memories (red circles in Fig. 2). We associate these BSM attempts as *internal links*, i.e., links between memories internal to a repeater node, shown using dashed lines inside repeaters in Fig. 2b. If $T > 1$, a repeater node can attempt a BSM between qubits held in two memories that were entangled with their respective neighboring node's qubits in two different time-slots. For minimizing the demands on memory coherence time,[29,31] and to avoid requiring temporal switching, we will assume $T = 1$. So, BSMs will always be attempted between two qubits in distinct memories that were entangled with their respective counterparts at their respective neighboring nodes in the *same* time-slot. Each of these internal-link attempts succeed with probability q. Therefore, after the conclusion of one time-slot, along a path comprising k edges (and thus $k - 1$ repeater nodes), one ebit is successfully shared between the end points of the path with probability $p^k q^{k-1}$. The maximum number of ebits that can be shared between Alice (say, node a) and Bob (say, node b) after one time-slot is $\min\{d(a), d(b)\}$, assuming S is the same over all edges. For the square-grid topology shown, the maximum number of ebits that can be generated between Alice and Bob in each time-slot is 4.

Alice and Bob only learn whether entanglement was generated after a communication lag proportional to the length of the end-to-end channel. For many applications including quantum key distribution,[3,4] secure private-bid auctions,[9] quantum digital fingerprinting[43] and quantum-enhanced sensing,[7] this does not affect the performance of the protocol, except for a latency. However, for some other applications like distributed quantum computing[44] and quantum private queries[45] that require knowledge of whether entanglement was generated in a particular time-slot, we will need an additional memory at Alice and Bob capable of holding entanglement during this latency period. A conventional linear repeater chain would also require these additional memories for these applications.

The remainder of the paper is dedicated to finding the optimal strategy for each repeater node in order to decide which locally held qubits to attempt BSM(s) on during the internal phase of a time slot, based ideally only on knowledge of the outcomes (success or failure) of the nearest neighbor links, i.e., local link-state knowledge, during the respective preceding external phases. We will assume that each repeater node is aware of the overall network topology, as well as the locations of the K Alice-Bob pairs. The goal of the optimal repeater strategy will be to attain the maximum entanglement-generation rate (if there is a single Alice and Bob, i.e., $K = 1$) or the maximum rate-region for multiple flows (i.e., $K > 1$).

Multipath routing of a single entanglement flow

Entanglement routing with global link-state information. We begin with the assumption that global link-state knowledge is available at each repeater node, i.e., the state of every external link in the network after the external phase is known to every repeater in the network and can be used to determine the choice of which internal links to attempt within the nodes. Each memory can only be part of one entanglement swap, i.e., each red node can only be part of one internal edge. Consider the following greedy algorithm to choose the internal links: consider the subgraph induced by the

successful external links and the repeater nodes (at the end of the external phase), and find in it the shortest path connecting Alice and Bob. If no connected path between Alice and Bob exists, no shared ebits are generated in that time slot. If a shortest path of length k_1 is found, all internal links along the nodes of that path are attempted, and the (conditional) probability of a shared ebit is generated by this path is the probability that all $k_1 - 1$ internal link attempts were successful, i.e., q^{k_1-1}. We then remove all the (external and internal) links of the above path from the subgraph, and find a shortest path connecting Alice and Bob in the pruned subgraph. Note that instead of removing the links of the first path from the subgraph, we could simply search for a shortest path in the original subgraph but one that is edge disjoint from the previous path. If such a path exists, we again attempt all internal links at the nodes of this path, so the probability in the path contributes to the generation of an ebit (distinct from the ebit that may have been generated by the first path) is q^{k_2-1} where k_2 is the length of the second path; and so on.

The entanglement generation rate achieved using this greedy algorithm R_g is the sum of expected rates (in ebits per time-slot) from these paths. Given the degree-4 nodes in a square grid topology, there can be a maximum of four edge disjoint paths between Alice and Bob. Figure 2b illustrates our greedy algorithm. Given the set of external links created, the shortest path has length $k_1 = 4$, the next path has length $k_2 = 6$, and no further paths can be found. The two edge-disjoint paths are highlighted in green. Hence, the internal links depicted with the dashed lines in Fig. 2b are attempted and the expected number of shared ebits generated in this time cycle is: $q^{k_1-1} + q^{k_2-1}$. The net entanglement generation rate is the expectation of sums like the above (with up to four terms) over many random instantiations of the (p, $1 - p$) external-link creations during the external phase of many time-slots. Evaluating this expected rate $R_g(p, q)$ achieved by the above routing strategy analytically as a function of the Alice-Bob distance (X_1, X_2) is difficult, even for a square-grid topology.

The intuition behind this simple greedy algorithm is that the entanglement generation rate along a path of length k decays exponentially as q^{k-1}, suggesting that attempting internal links to facilitate connections along the shortest path first would optimize the expected rate. However, it is possible to draw random instances of successes of external links, where either one of the two possible options—(1) picking the shortest path (which disrupts all other paths) and (2) picking two edge disjoint (but longer) paths—could yield either a larger or a smaller expected rate than the other, depending upon the value of q. If q is larger than a threshold, option (2) will have a larger expected rate and vice versa. Finding the global optimal rule remains an open problem. It is easy, however, to prove that the greedy algorithm achieves a rate within a factor of 4 of the optimum algorithm employing global link-state knowledge, $R_{\text{opt}}(p, q)$. Let us denote the length of the shortest path between Alice and Bob with Manhattan distance (X, Y) in the induced subgraph after the external phase, as $n_{\text{SP}}(p)$. This quantity is of interest in percolation theory, and is not known analytically, even for simple graph topologies. It undergoes a sharp transition (i.e., starts out large and suddenly jumps to a much smaller value) as p crosses the bond-percolation threshold p_c of the graph G, from below to above. Clearly, $R_{\text{g}}(p,q) \geq \mathbb{E}[q^{n_{\text{SP}}(p)-1}]$ since using the shortest path is the first step of the greedy algorithm. Furthermore, since the optimal rule can create entanglement over a maximum of four edge-disjoint paths in each time-step, each of which must have a length no less than the length of the shortest path, $R_{\text{opt}}(p,q) \leq \mathbb{E}[4q^{n_{\text{SP}}(p)-1}] \stackrel{\Delta}{=} R^{(\text{UB})}_{\text{opt}}(p,q)$. Therefore,

$$R_{\text{opt}}(p,q) \geq R_{\text{g}}(p,q) \geq \frac{R_{\text{opt}}(p,q)}{4}, \tag{2}$$

i.e., the greedy rule achieves the same rate-vs.-distance scaling as the optimal algorithm that employs global link-state knowledge, and at worst is lower that the optimal rate only by a constant factor equaling the node degree.

In Fig. 3a we plot $R_{\text{g}}(p, q)$ as a function of the Alice-Bob X separation (measured in number of hops). We choose the $X = Y$ direction here (45° with respect to the grid axes) but other directions show similar behavior, and a 3D plot with all directions is in Supplementary Fig. 1 in Supplementary Note 1. When $q = 1$ and $p > p_c$ ($p_c = 0.5$ for the square lattice), a *giant connected component* is formed by the external links alone at the end of the first (external) phase of a time slot. Recall that the rate along a length k path is $p^k q^{k-1}$, where $p \sim \eta$ is the transmissivity of each link. In the network case, when $p > p_c$ and $q = 1$, we find that the p^k portion of the rate expression becomes immaterial for scaling with Alice-Bob distance. This behavior can be explained by percolation theory: in a large lattice in this regime, the probability of a connected path between Alice and Bob along successful external links in each time-slot approaches a non-zero constant as the Alice-Bob separation is increased. Furthermore, we numerically find that finite size effects do not have a significant impact on this behavior, even when the Alice-Bob separation is as small as 5 hops. So, if $q = 1$, $R_{\text{g}}(p, q)$ remains essentially distance invariant. When $p < p_c$, the rate falls off exponentially with distance (even when $q = 1$). It is instructive to note here that the optimal rate (entanglement-generation *capacity*) achievable on a single length

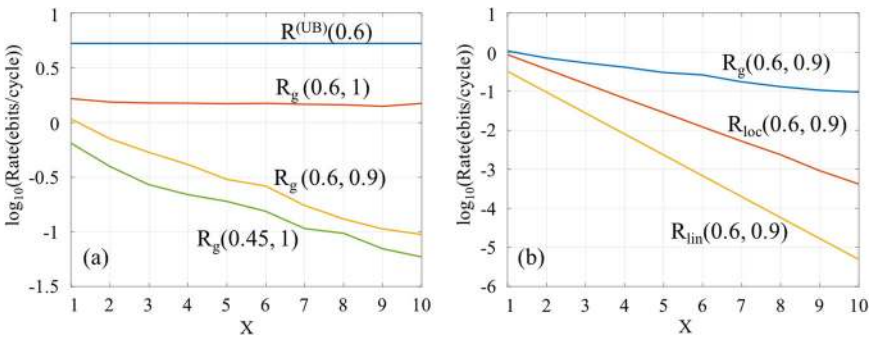

Fig. 3 Entanglement generation rate vs. Alice-Bob X separation for different (p, q). We choose direction $X = Y$ here but other directions show similar behavior, and a 3D plot with all directions is in Supplementary Fig. 1 in Supplementary Note 1; (**a**) $R_{\text{g}}(p, q)$ is the rate attained by the global-knowledge-based protocol. For $q = 1$, R_{g} is distance independent when p is greater than the bond percolation threshold (0.5 for the square lattice). $R^{(\text{UB})}(0.6)$ is the distance-independent rate upper bound for $p = 0.6$, achieving which requires perfect quantum processing at repeater nodes. $R_{\text{g}}(0.6, 1)$ is also distance independent, and within a factor 3.6 of $R^{(\text{UB})}(0.6)$. With $q < 1$, e.g., $R_g(0.6, 0.9)$, the rate decays exponentially with distance. **b** R_{loc} is attained by our local link-state knowledge protocol. The rate-distance scaling exponent of R_{loc} is clearly worse than R_{g}, but is superior to that of a linear repeater chain along the shortest path, R_{lin}, demonstrating multi-path routing advantage even with local link-state knowledge

k path does not depend on k, and only on the transmissivity of the lossiest link in the path, i.e., $C \sim \eta$,[20] but achieving this requires infinite-coherence-time quantum memories and ideal quantum operations at nodes. The multi-path gain in the $p > p_c$ regime lets us achieve a distance-independent rate, but with memories whose coherence times are no more than one time slot, and only using BSMs. The rates were calculated using Monte-Carlo simulations which resulted in some numerical noise as is apparent in Fig. 3.

A general upper bound on the entanglement generation rate is given by the min-cut of the graph,[20] and for a square lattice, is given by:

$$R^{(\mathrm{UB})}(p) = -\log_2[(1-p)^4]. \tag{3}$$

$R^{(\mathrm{UB})}(0.6)$ is plotted in Fig. 3a. The known methods for achieving $R^{(\mathrm{UB})}$ require infinite coherence time memories and error-corrected quantum processors at each node. For our implementation (assuming global link state knowledge), $R_{\mathrm{g}}(0.6, 1)$ is also plotted in Fig. 3a. Although our protocol only requires memories to hold entanglement for one time step, the multi path advantage gives us the same constant rate-distance scaling and within a factor of ~3.6 of $R^{(\mathrm{UB})}(0.6)$. The assumption of perfect BSMs is unrealistic and thus $q < 1$, in which case $R_g(p, q)$ falls off exponentially with distance; even when $p > p_c$, as seen in the plot for $R_{\mathrm{g}}(0.6, 0.9)$.

Entanglement routing with local link-state information. $R_{\mathrm{g}}(p, q)$, the rate attained by the protocol described in the previous subsection that employs global link-state knowledge, is re-plotted in Fig. 3b. We also plot:

$$R_{\mathrm{lin}}(p,q) = p^{n_{\mathrm{SP}}(1)} q^{n_{\mathrm{SP}}(1)-1}, \tag{4}$$

the rate attained by a single linear repeater chain, where $n_{\mathrm{SP}}(1)$ is the shortest-path length between Alice and Bob along the edges of the underlying square grid. The assumption of global link-state knowledge in large networks is unrealistic, as it requires memories whose coherence time increases with the network size due to the time required for the traversal of link-state information across the entire network. In this section, we describe a more realistic protocol in which knowledge of success and failure of an external link at each time slot is communicated only to the two repeater nodes connected by the link, as is the case in the analysis of many 'second-generation' linear repeater chains.[26,27,29] Repeater nodes need to decide on which pair(s) of memories BSMs should be attempted (i.e., which internal links to attempt), based only on information about the states of external links adjacent to them. We assume that network topology and positions of Alice and Bob are known to each repeater station, and communicated classically beforehand.

Every repeater, except Alice and Bob which do not attempt any internal links, uses the same local rule, which is illustrated in Fig. 4 using the example of repeater u inside the dotted box. A repeater decides which internal edges to attempt based on the information of (1) which of the four neighboring external edges have been successfully created in the external phase and (2) The distance of it's four neighbors to Alice and Bob. The distance of a repeater to Alice and Bob is denoted, respectively, by d_A and d_B. We use the $\mathbb{L}^2$ norm for both these distances (other distance metrics are discussed in Supplementary Note 3). The rules used to determine the internal links to be attempted at a repeater are:

- If less than one of the neighboring external links is successful: no internal links are attempted, since this repeater node cannot be part of a path from Alice to Bob in that time slot.
- If two or more neighboring external links are successful: of all the nearest neighbor nodes of u whose links to u were successful in that time slot, we label the one that has the minimum d_A as v. Similarly, the neighbor with a successful external link with u and the minimum d_B is labeled w. If two neighbors have the same values of d_A and d_B, an unbiased coin is tossed to determine the choice of v and w, to preserve symmetry in the protocol. If v and w are the same node, v (or w) is replaced by node u's nearest-neighbor node with the next smallest value of d_A (or d_B). The choice of whether to replace v or w is made in a manner that minimizes the sum of d_A and d_B from the eventually chosen two neighbors to connect. An internal link is attempted between the memories connected to v and w respectively, as shown in Fig. 4a.
- If all four neighboring external links are successful: in addition to the internal link attempted in the previous point, an additional internal link is attempted between the remaining two memories as shown in Fig. 4b, since the addition of this internal link can only increase the entanglement generation rate.

The entanglement generation rate $R_{\mathrm{loc}}(p, q)$ achieved by the above described local rule is plotted in Fig. 3b and compared to $R_{\mathrm{g}}(p, q)$ and $R_{\mathrm{lin}}(p, q)$. We use $p = 0.6$ and $q = 0.9$, the same values used for the global-information rate plots in Fig. 3a. As one expects, the rate-distance scaling of R_{loc} is worse than that of R_{g}. However, the rate-distance scaling exponent achieved by the local rule is superior to that of a linear chain, even though the physical elements employed to build the repeaters are identical. In other words,

$$R_g(p,q) > R_{\mathrm{loc}}(p,q) > R_{\mathrm{lin}}(p,q), \tag{5}$$

where each of the above three rates can be expressed as an exponential decay with the distance L, i.e., $R \sim e^{-\alpha L}$, where the exponents satisfy:

$$\alpha_g(p,q) < \alpha_{\mathrm{loc}}(p,q) < \alpha_{\mathrm{lin}}(p,q), \tag{6}$$

where it is known that $\alpha_{\mathrm{lin}}(p, q) < \alpha_{\mathrm{fiber}}$, the loss coefficient of a fiber, which is the rate scaling exponent when no repeaters are used.[29] This is proven analytically in Supplementary Note 2A.

The scaling advantage of R_{loc} over R_{lin} arises because the local rule allows the entanglement-generation flow between Alice and Bob to find different (and potentially simultaneously multiple) paths in different time slots, and does not have to rely on all links along a linear chain to be successful. This is analogous to multi-path routing in a classical computer network. The contour plot in Supplementary Fig. 1d in Supplementary Note 1 further illustrates this point: there is a noticeable enhancement of R_{loc} along the $X = Y$ line because the diagonal direction contains the largest spatial density of possible paths between Alice and Bob. The scaling advantage over R_{lin} persists in any direction, i.e., along $Y = 0$ as well.

Sweeping over different values of p and q, we find that the multi-path advantage relative to a linear repeater chain increases as p decreases from unity, but there is little relative improvement as q is varied (see Supplementary Note 2B).

Clearly, other distance metrics (e.g., $\mathbb{L}^p$ norm for $p \geq 1$) can be used in lieu of the $\mathbb{L}^2$ norm in the algorithm described above. In Supplementary Note 3, we present a recursive numerical evaluation technique to find the rate-optimal distance metric, which can be applied to any network topology. For planar network topologies, we find that the $\mathbb{L}^2$ norm is near-optimal for our local routing algorithm.

An analytical enumeration of the expected number of edge-disjoint paths as a function of p between Alice and Bob separated by a given distance (X, Y) in a bond-percolation instance (i.e., with $p > p_c$) of a network is an open question, the solution of which will enable a firmer quantitative understanding of the multi-path advantage in entanglement generation in a repeater network.

Simultaneous entanglement flows

In this section, we consider simultaneous entanglement-generation flows between two Alice-Bob pairs, using local link state knowledge at all repeater nodes. Consider two pairs Alice 1 - Bob 1 (red nodes) and Alice 2 - Bob 2 (green nodes) as shown in the two scenarios in Fig. 5. In Fig. 5a, the shortest paths

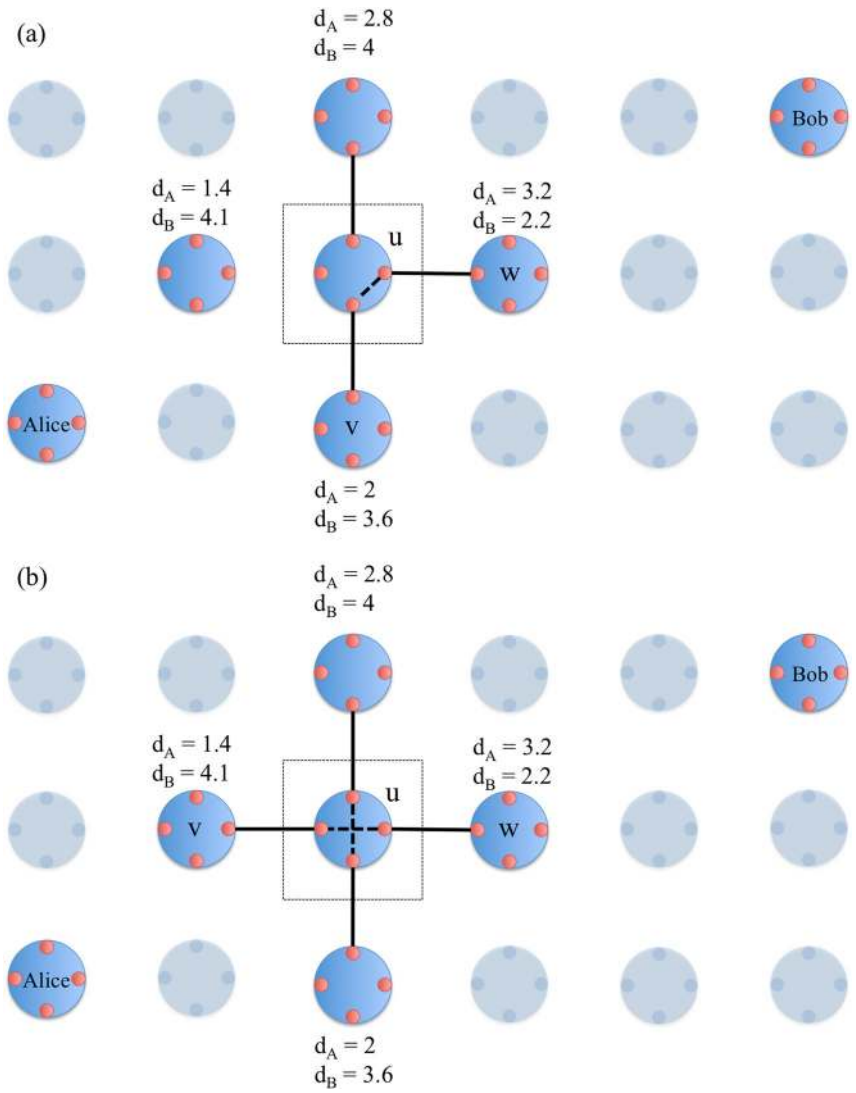

Fig. 4 The entanglement swap rule used at the repeater C in the dotted box in the case of local link-state knowledge. A and B are the repeaters closest to Alice and Bob, respectively, with a direct edge to C. **a** If two or three links are up, the memories linked to A and B undergo an entanglement swap. **b** If four links are up, the remaining two memories also undergo an entanglement swap

connecting the two Alice-Bob pairs do not cross, but they do in Fig. 5b. In both cases, they are placed at the four corners of a 6×6 square grid, embedded within a large square grid network. Denote by R_1 and R_2 the entanglement generation rates achieved by the two Alice-Bob pairs respectively. We first consider the case of non-intersecting flows shown in Fig. 5a. A simple strategy is for every single repeater node (including the nodes labeled as the two Alices and Bobs) to use the local rule described in the previous section tailored to support the Alice 1-Bob 1 flow for a fraction, λ, of the time slots and to support the Alice 2-Bob 2 flow for the remaining $1-\lambda$ fraction. For $p = q = 0.9$, the rate region attained by varying $\lambda \in [0, 1]$ is depicted with the blue line in Fig. 5b, which we refer to as *single-flow time-share*. However, if every repeater with the exception of the Alices and Bobs carry out the above time-sharing strategy, even when all repeater nodes support flow 1, there is still some 'left-over' non-zero R_2 that is attained. This *multi-flow time-share* rate region is shown using the red line in Fig. 5b.

In Fig. 6, for the case that a single Alice and a single Bob are separated by 6 hops on the square grid, we plot a color map of p_{usage}, the probability a given repeater node is involved in a successful creation of a shared ebit generated between Alice and Bob when our local rule (for multi-path entanglement routing described in the previous section) is employed. We observe that only the repeaters lying in a small spatial region surrounding the straight line joining Alice and Bob are used significantly.

This observation motivates a *multi-flow spatial-division* rule for routing multi-flow entanglement, in which we divide the network between two spatial regions corresponding to the two flows, as shown in Fig. 5a. Any repeater in the red shaded region follows the local rule tied to the Alice 1 - Bob 1 flow while repeaters in the green region operate with the local rule tied to the Alice 2 - Bob 2 flow. The placement of the boundary determines the rates R_1 and R_2. The rate region attained is plotted with the yellow line in Fig. 5b. This significantly outperforms time sharing. The two flows can co-exist and operate with a very small reduction from their individual best rates, because the repeaters they respectively benefit the most from, form (almost) disjoint sets.

In the other extreme, we consider two Alice-Bob pairs, still separated by six hops, but with their shortest paths crossing as shown in Fig. 5c. The rate region attained by multi-flow time sharing, shown by the line segment BC, still provides an improvement over single-flow time-sharing, shown by the line segment AD, as shown in Fig. 5d. It is interesting to note that the

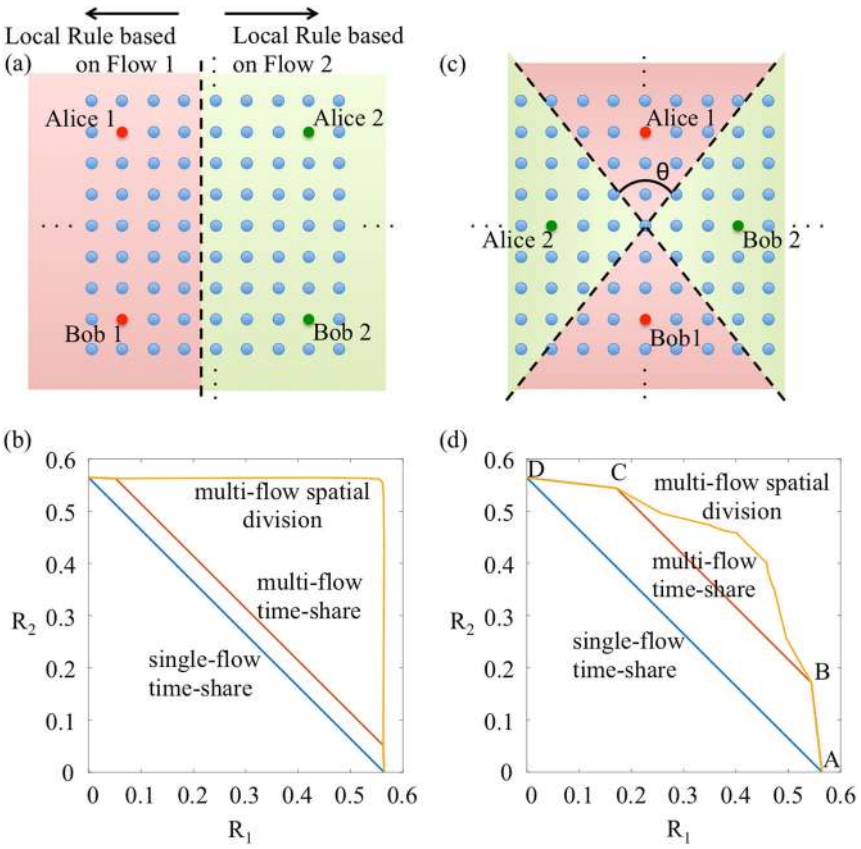

Fig. 5 **a** Multi-flow routing for two Alice-Bob pairs that lie along the sides of a 6×6 square, embedded in a 100×100 grid; (**b**) rate region (R_1, R_2) with different rules at repeater nodes, each employing local link-state knowledge, for $p = q = 0.9$. **c** Multi flow routing when the Alice-Bob paths cross (**d**) multiflow rate region for two local-knowledge rules

maximum R_1 under multi-flow time sharing (point B) is slightly lower than maximum R_1 with the single-flow time-share rule (point A). This happens because unlike in single-flow time-share, the nodes at Alice 2 and Bob 2 do not contribute to R_1 under multi-flow time-share. A point along AB represents time sharing between the strategies at points A and B. To further increase the rate, we adopt a multi-flow spatial division strategy in which nodes in the red region are configured to assist flow 1 and nodes in the green region are configured to assist flow 2. Varying the angle θ demarcating those regions results in the rate region shown by the yellow line in Fig. 5d. This time, the improvement due to the spatial-division rule is not as pronounced, since the spatial regions corresponding to 'useful' repeater nodes for the two flows are not disjoint.

DISCUSSION

We proposed and analyzed quantum repeater protocols for entanglement generation in a quantum network in an architecture that uses the same elements as in many theoretical proposals and analyses of linear repeater chains. We accounted for channel losses between repeater nodes and the probabilistic nature of entanglement swaps at each repeater stemming from device inefficiencies, as well as the probabilistic nature of Bell-state measurements (e.g., due to inherent limitations of using linear optics and lossy detectors). The rate attained for a single entanglement-generation flow can far outperform that is attainable over a linear repeater chain along the shortest path, even when the nodes only have local link-state knowledge, due to a multi-path routing advantage. We also proposed a modified version of our routing protocol for supporting simultaneous entanglement generation flows between multiple Alice-Bob pairs. We found multi-flow entanglement routing strategies that far outperform the rate region attained when each repeater node's local action simply time shares among assisting each flow. Our results suggest that building and connecting quantum repeaters in non-trivial network topologies could provide a substantial benefit over building linear repeater chains. Seen another way, given constraints on the number and quality of quantum memories, link losses between nodes, and limited and imperfect processing capabilities at repeater nodes, a 2D network topology can outperform the repeater-less rate-vs.-distance upper limits[21,23] more easily than a linear repeater chain connecting the communicating parties.

Our work has also opened up a number of new questions. Even in our simplified model—an abstraction that applies when the only source of imperfection for each component (including the quantum memories) is *pure loss*—the rate-optimal (single-flow and multi-flow) entanglement routing protocol remains open. Rate-distance performance of the class of protocols we studied when allowing for temporal switching at the repeater nodes—BSM between locally-held qubits that were entangled with their respective neighboring-node counterpart qubits in different time slots—remains open. Since our protocol only requires a quantum memory to hold a qubit for one entanglement attempt between neighboring stations, photon loss would indeed be the major source of imperfection in many implementations of the protocol. However, accounting for more general errors (such as detector excess noise, qubit decoherence models in the memory, multi-pair generation probabilities in the entanglement sources, etc.) will require purification of entanglement,[25,46,47] i.e., converting several poorer-quality EPR pairs into a few good ones using local quantum operations and classical communication, accounting which will

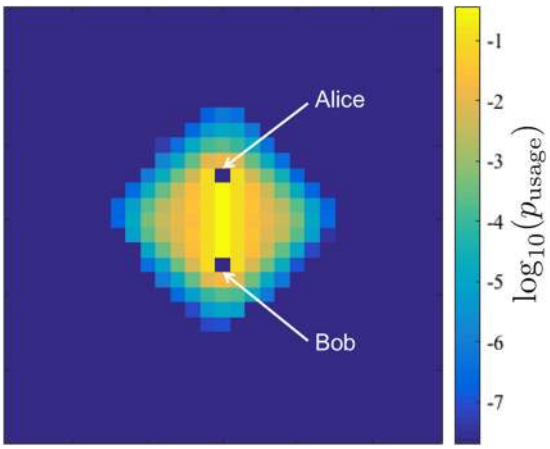

Fig. 6 A heat map plotting p_{usage}, the probability that a given repeater node is involved in a successful creation of a shared ebit generated between Alice and Bob, separated by 6 hops in an underlying square grid topology, when our local rule is employed. We assume $p = 0.9$ and $q = 0.9$

require us to introduce the Fidelity of shared entanglement at intermediate steps of the protocol. Furthermore, we restricted our analysis to nodes making 2-qubit measurements. Being able to perform multi-qubit unitary operations and multi-qubit measurements at repeater nodes (e.g., a 3-qubit GHZ projection across three locally held qubits) may improve the achievable rate regions. The idea of using a distance metric to choose the measurements at the repeater station could be used in protocols that use measurements of more than two qubits as well. Finally, it will be interesting to consider repeater protocols for the distillation of multi-partite entanglement shared between more than two parties, and a repeater network that can support multiple simultaneous flows of generation of multi-partite entanglement.

Code availability

All the numerical data presented in this paper are results of C simulations conducted by MP on idealized grid network topologies. The code used to generate this data will be made available to the interested reader upon reasonable request.

DATA AVAILABILITY

The data sets generated during and or analyzed during the current study are available from the corresponding author on reasonable request.

ACKNOWLEDGEMENTS

S.G. would like to thank Stefano Pirandola, Zachary Dutton, and Dongning Guo for valuable discussions. M.P., D.E., and L.J. acknowledge support from the Air Force Office of Scientific Research MURI (FA9550-14-1-0052) and the Army Research Laboratory (ARL) Center for Distributed Quantum Information (CDQI). S.G., H.K., P.B., M.P., and D.E. would like to acknowledge the Office of Naval Research program *Communications and Networking with Quantum Operationally-Secure Technology for Maritime Deployment* (CONQUEST), awarded under prime contract number N00014-16-C-2069. S.G., D.T., P. B., and L.T. acknowledge the ARL DAIS-ITA program; this work would not have occurred without the collaborations and ideas seeded by this program.

AUTHOR CONTRIBUTIONS

This work took seed in 2015, when S.G., H.K., P.B., and D.T. started looking into network extensions of routing entanglement using quantum repeaters that are based on BSMs and multiplexing–based on two papers on analyzing line repeaters that S.G. and H.K. collaborated on–and they found rate-region bounds for multi-flow routing on a tree topology network. S.G., H.K., and D.T. found example protocols for two-flow entanglement that outperformed simple time sharing at all repeater nodes. New insights on multi flow routing were subsequently developed in a brainstorming on quantum networks during when S.G. and M.P. visited L.J. and L.T. in Yale. The majority of the work thereafter, including the numerical simulations and theoretical performance bounds were done by M.P., under the supervision of D.T., D.E., and S.G.

ADDITIONAL INFORMATION

Supplementary Information accompanies the paper on the *npj Quantum Information* website (https://doi.org/10.1038/s41534-019-0139-x).

Competing interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Kimble, H. J. The quantum internet. *Nature* **453**, 1023–1030 (2008).
2. Wehner, S., Elkouss, D. & Hanson, R. Quantum internet: a vision for the road ahead. *Science* **362**, eaam9288 (2018).
3. Ekert, A. K. Quantum cryptography based on Bell's theorem. *Phys. Rev. Lett.* **67**, 661–663 (1991).
4. Bennett, C. H. & Brassard, G. Quantum cryptography: public key distribution and coin tossing. *Theor. Comput. Sci.* **560**, 7–11 (2014).
5. Cirac, J., Ekert, A., Huelga, S. & Macchiavello, C. Distributed quantum computation over noisy channels. *Phys. Rev. A.* **59**, 4249–4254 (1999).
6. Gottesman, D., Jennewein, T. & Croke, S. Longer-baseline telescopes using quantum repeaters. *Phys. Rev. Lett.* **109**, 070503 (2012).
7. Kómár, P. et al. A quantum network of clocks. *Nat. Phys.* **10**, 582–587 (2014).
8. Broadbent, A., Fitzsimons, J. & Kashefi, E. Universal blind quantum computation. In *2009 50th Annual IEEE Symposium on Foundations of Computer Science*, 517–526 (IEEE, Atlanta, GA, USA, 2009).
9. Guha, S., Hogg, T., Fattal, D., Spiller, T. & Beausoleil, R. G. Quantum auctions using adiabatic evolution: the corrupt auctioneer and circuit. *Int. J. Quantum Inf.* **06**, 815–839 (2008).
10. Hensen, B. et al. Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres. *Nature* **526**, 682–686 (2015).
11. Yin, J. et al. Satellite-based entanglement distribution over 1200 kilometers. *Science* **356**, 1140–1144 (2017).
12. Ren, J.-G. et al. Ground-to-satellite quantum teleportation. *Nature* **549**, 70–73 (2017).
13. Courtland, R. China's 2,000-km quantum link is almost complete. *IEEE Spectr.* **53**, 11–12 (2016).
14. Yuan, Z.-S. et al. Experimental demonstration of a BDCZ quantum repeater node. *Nature* **454**, 1098–1101 (2008).
15. Bernien, H. et al. Heralded entanglement between solid-state qubits separated by three metres. *Nature* **497**, 86–90 (2013).
16. Olmschenk, S. et al. Quantum teleportation between distant matter qubits. *Science* **323**, 486–489 (2009).
17. Pan, J.-W., Bouwmeester, D., Weinfurter, H. & Zeilinger, A. Experimental entanglement swapping: entangling photons that never interacted. *Phys. Rev. Lett.* **80**, 3891–3894 (1998).
18. Chou, C. W. et al. Measurement-induced entanglement for excitation stored in remote atomic ensembles. *Nature* **438**, 828–832 (2005).
19. Moehring, D. L. et al. Entanglement of single-atom quantum bits at a distance. *Nature* **449**, 68–71 (2007).
20. Pirandola, S. Capacities of repeater-assisted quantum communications. Preprint at arXiv: http://arxiv.org/abs/1601.00966 (2016).
21. Pirandola, S., Laurenza, R., Ottaviani, C. & Banchi, L. Fundamental limits of repeaterless quantum communications. *Nat. Commun.* **8**, 15043 (2017).
22. Pirandola, S., Garca-Patron, R., Braunstein, S. L. & Lloyd, S. Direct and reverse secret-key capacities of a quantum channel. *Phys. Rev. Lett.* **102**, 050503 (2009).
23. Takeoka, M., Guha, S. & Wilde, M. M. Fundamental rate-loss tradeoff for optical quantum key distribution. *Nat. Commun.* **5**, 5235 (2014).
24. Wilde M. M., Tomamichel M. & Berta M. Converse bounds for private communication over quantum channels. *IEEE Transactions on Information Theory* **63**, 1792–1817 (2017).
25. Briegel, H.-J., Dür, W., Cirac, J. & Zoller, P. Quantum repeaters: the role of imperfect local operations in quantum communication. *Phys. Rev. Lett.* **81**, 5932–5935 (1998).
26. Jiang, L. et al. Quantum repeater with encoding. *Phys. Rev. A.* **79**, 032325 (2009).
27. Muralidharan, S. et al. Optimal architectures for long distance quantum communication. *Sci. Rep.* **6**, 20463 (2016).

28. Sinclair, N. et al. Spectral multiplexing for scalable quantum photonics using an atomic frequency comb quantum memory and feed-forward control. *Phys. Rev. Lett.* **113**, 053603 (2014).
29. Guha, S. et al. Rate-loss analysis of an efficient quantum repeater architecture. *Phys. Rev. A.* **92**, 022357 (2015).
30. Azuma, K., Tamaki, K. & Lo, H.-K. All-photonic quantum repeaters. *Nat. Commun.* **6**, 6787 (2015).
31. Pant, M., Krovi, H., Englund, D. & Guha, S. Rate-distance tradeoff and resource costs for all-optical quantum repeaters. *Phys. Rev. A.* **95**, 012304 (2017).
32. Ewert, F., Bergmann, M. & van Loock, P. Ultrafast long-distance quantum communication with static linear optics. *Phys. Rev. Lett.* **117**, 210501 (2016).
33. Azuma, K., Mizutani, A. & Lo, H. K. Fundamental rate-loss trade-off for the quantum internet. *Nat. Commun.* **7**, ncomms13523 (2016).
34. Azuma, K. & Kato, G. Aggregating quantum repeaters for the quantum internet. *Phys. Rev. A.* **96**, 032332 (2017).
35. Schoute, E., Mancinska, L., Islam, T., Kerenidis, I. & Wehner, S. Shortcuts to quantum network routing. *arXiv*: 1610.05238 1–45 (2016).
36. Acín, A., Cirac, J. I. & Lewenstein, M. Entanglement percolation in quantum networks. *Nat. Phys.* **3**, 256–259 (2007).
37. Van Meter, R. *Quantum Networking* (Wiley, NJ, USA, 2014).
38. Hayashi, M, Iwama, K, Nishimura, H, Raymond, R. & Yamashita, S. in *Quantum network coding STACS 2007*. 610–621 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2007).
39. Kobayashi, H., Le Gall, F., Nishimura, H. & Rötteler, M. General scheme for perfect quantum network coding with free classical communication, Vol. 5555. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 622–633 (LNCS, Berlin, Heidelberg, 2009).
40. Satoh, T., Le Gall, F. & Imai, H. Quantum network coding for quantum repeaters. *Phys. Rev. A* **86**, 032331 (2012).
41. Satoh, T., Ishizaki, K., Nagayama, S. & Van Meter, R. Analysis of quantum network coding for realistic repeater networks. *Phys. Rev. A* **93**, 032302 (2016).
42. Lemr, K., Bartkiewicz, K., Černoch, A. & Soubusta, J. Resource-efficient linear-optical quantum router. *Phys. Rev. A* **87**, 062333 (2013).
43. Buhrman, H., Cleve, R., Watrous, J. & de Wolf, R. Quantum fingerprinting. *Phys. Rev. Lett.* **87**, 167902 (2001).
44. Zhang, Z., Mouradian, S., Wong, F. N. C. & Shapiro, J. H. Entanglement-enhanced sensing in a lossy and noisy environment. *Phys. Rev. Lett.* **114**, 110506 (2015).
45. Giovannetti, V., Lloyd, S. & Maccone, L. Quantum private queries. *Phys. Rev. Lett.* **100**, 230502 (2008).
46. Deutsch, D. et al. Quantum privacy amplification and the security of quantum cryptography over noisy channels. *Phys. Rev. Lett.* **77**, 2818–2821 (1996).
47. Bennett, C. H. et al. Purification of noisy entanglement and faithful teleportation via noisy channels. *Phys. Rev. Lett.* **76**, 722–725 (1996).