communications physics

A Nature Portfolio journal

Article

https://doi.org/10.1038/s42005-025-02029-w

Reducing classical communication costs in multiplexed quantum repeaters using hardware-aware quasi-local policies

Check for updates

Stav Haldar[1] ✉, Pratik J. Barge[1], Xiang Cheng[2], Kai-Chi Chang[2], Brian T. Kirby[3,4], Sumeet Khatri[5], Chee Wei Wong[2] & Hwang Lee[1]

Future quantum networks will have nodes equipped with multiple quantum memories, allowing for multiplexing and entanglement distillation strategies for long-distance entanglement distribution. In this work, we focus on *quasi-local* policies for multiplexed quantum repeater chains. In fully-local policies, nodes use the knowledge of only their own states, whereas more efficient global policies use knowledge of the entire network state. The classical communication costs of using this knowledge have not been explored in existing literature. We show that quasi-local policies not only obtain improved performance over local policies, but also reduce classical communication costs considerably. Our policies also outperform the widely studied nested purification and doubling policy in practical parameter regimes. We identify parameter regimes where distillation is useful and address the question: "Should we distill before swapping, or vice versa?" Finally, we propose an implementation scheme for a multiplexed repeater chain, experimentally demonstrate the key element, a high-dimensional biphoton frequency comb, and evaluate its anticipated performance using our multiplexing-based policies.

The quantum internet[1–3] has the potential to revolutionize current computation, communication, and sensing technologies by enabling exchange of quantum data. Numerous significant applications have already been recognized, encompassing areas such as secure quantum communication[4–6], distributed quantum computation[7,8], and distributed quantum sensing[9–11], to name a few. However, enormous hardware improvements are needed before a practical quantum internet is realized[12,13]. Currently, a significant challenge lies in effectively distributing entanglement amongst network nodes using quantum repeaters, with the goal of attaining high fidelities and low waiting times, particularly over significant distances. Recent, state-of-the-art experiments have been performed for only a handful of nodes over short distances[14,15]. This challenge stems from the delicate nature of quantum systems, leading to issues like photon losses, imperfect measurements, and quantum memories with short coherence times. Since comprehensive, fault-tolerant quantum error correction has yet to be realized to overcome these obstacles, it becomes important to explore how much we can overcome these limitations through other means. This problem of designing effective repeater protocols for entanglement distribution using noisy, imperfect quantum hardware has been addressed in multiple analytical[16–35] and numerical[36–45] studies, among others; see also refs. 46–52 for reviews.

In order to make the best use of today's noisy, imperfect quantum hardware, protocols/policies for entanglement distribution should be *hardware-aware*. This means tailoring the elementary link generation and entanglement swapping policies to the hardware, instead of employing widely-studied but hardware-agnostic policies such as the doubling[16,17] and the "swap-as-soon-as-possible"[30,31] policies (Another consideration is tailoring the number and locations of repeater stations and entanglement sources to the available hardware, which are architectural questions that have been deeply studied; see, e.g., refs. 3,24,53–56.). Determining optimal hardware-aware policies for elementary link generation and entanglement swapping has received heightened attention only recently[29,43–45,57–60].

At the same time, for the first-generation quantum repeaters that we consider here, it is well known that the two-way end-to-end classical communication involved makes them rather slow as the end-to-end distance increases, when compared to second- and third-generation repeaters, which use quantum error-correction[55,61]. Nevertheless, first-generation quantum repeaters are arguably the most amenable to implementation in

[1]Hearne Institute for Theoretical Physics, Department of Physics and Astronomy, Louisiana State University, Baton Rouge, LA, 70803, USA. [2]Fang Lu Mesoscopic Optics and Quantum Electronics Laboratory, Department of Electrical and Computer Engineering, University of California, Los Angeles, CA, 90095, USA. [3]DEVCOM Army Research Laboratory, Adelphi, MD, 20783, USA. [4]Tulane University, New Orleans, LA, 70118, USA. [5]Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, 14195 Berlin, Germany. ✉e-mail: stavhaldar@gmail.com

the near future. Therefore, it is of interest to develop policies for elementary link generation and entanglement swapping that are not only hardware-aware, but also minimize the amount of classical communication between nodes in the network.

The amount of classical communication in a network is tied to the amount of knowledge nodes have of each other's states when making their decisions regarding elementary link generation and entanglement swapping. Indeed, the more knowledge nodes need to make use of, the more classical communication is required to attain that knowledge. The decisions that nodes make, and the amount of knowledge needed to make the decisions, is dictated by the policy being used. In particular, if a policy calls for the nodes to make use of full, global knowledge of the network—meaning that every node should know the state of every other node at all times—then classical communication over the entire span of the network is required in every time step of the protocol. On the other extreme, for fully local policies, in which nodes need to have knowledge only of their own states, only one round of end-to-end classical communication is required at some prescribed final time. In the middle lie *quasi-local* policies, which we consider in this work. These policies use to their advantage the realization that nodes only need access to knowledge of the connected portion of the chain they are part of. Disconnected regions of the network can determine their actions independently. Thus, barring only a few time steps when long links are present in the network, spanning lengths on the order of the size of the network, full end-to-end classical communication is superfluous and wasteful. Previous works[62–64] have considered quasi-local policies in the context of continuous entanglement distribution, while we consider on-demand entanglement distribution. To the best of our knowledge, a thorough analysis of end-to-end waiting times and fidelities, as a function of the amount of knowledge nodes have of each other's states, and taking classical communication costs into account, has not been conducted. Recent works[44,45] have highlighted the importance of global knowledge in improving waiting times and fidelities, but without fully accounting for the cost of classical communication. We are therefore motivated to ask our first question:

(1) How are waiting times and fidelities affected by the amount of classical communication being performed in a network? Are the benefits of additional network knowledge, beyond that of a fully local policy, negated by the increase in classical communication?In this work, we consider this question in the context of entanglement distribution in linear networks (also called "repeater chains") with multi-memory quantum repeaters and multiplexing[16–18,34,54,65–82]; see Fig. 1. In the multiplexing approach, simultaneous elementary link generation attempts can be made, and when performing entanglement swapping[5,83], it is possible to go beyond the trivial parallel approach and perform cross-channel entanglement swapping.

Many of the aforementioned works on multi-memory quantum repeaters with multiplexing have made use of the hardware-agnostic "nested purification and doubling swapping" protocol devised in the seminal works[16,17] on quantum repeaters. Keeping in mind our goal of developing hardware-aware policies, we therefore ask:

(2) Is the nested purification and doubling swapping policy optimal, particularly in the practically-relevant parameter regimes of low coherence times and low elementary link success probabilities?In the multiplexing approach, there is also the possibility to perform entanglement purification/distillation[84–87]. Distillation introduces additional classical communication overhead. We are therefore motivated to ponder whether the potential benefits of distillation on end-to-end fidelities are worth the additional classical communication overhead. Specifically, we ask:

(3) In what parameter regimes is it beneficial to perform entanglement distillation? What is the best way to combine entanglement swapping with entanglement distillation—should we distill first before swapping, or vice versa?

In this paper, we address the three questions above by extending the framework from ref. 45 for modeling quantum repeater chains in terms of Markov decision processes (MDPs) to the case of multi-memory quantum repeater chains. We define quasi-local multiplexing-based policies, which are generalizations of the so-called "swap-as-soon-as-possible" (SWAP-ASAP) policy[30,31], to the multiplexing setting. For near-term quantum networks with imperfections such as memories with limited coherence time, probabilistic entanglement swapping, etc., classical communication (CC)

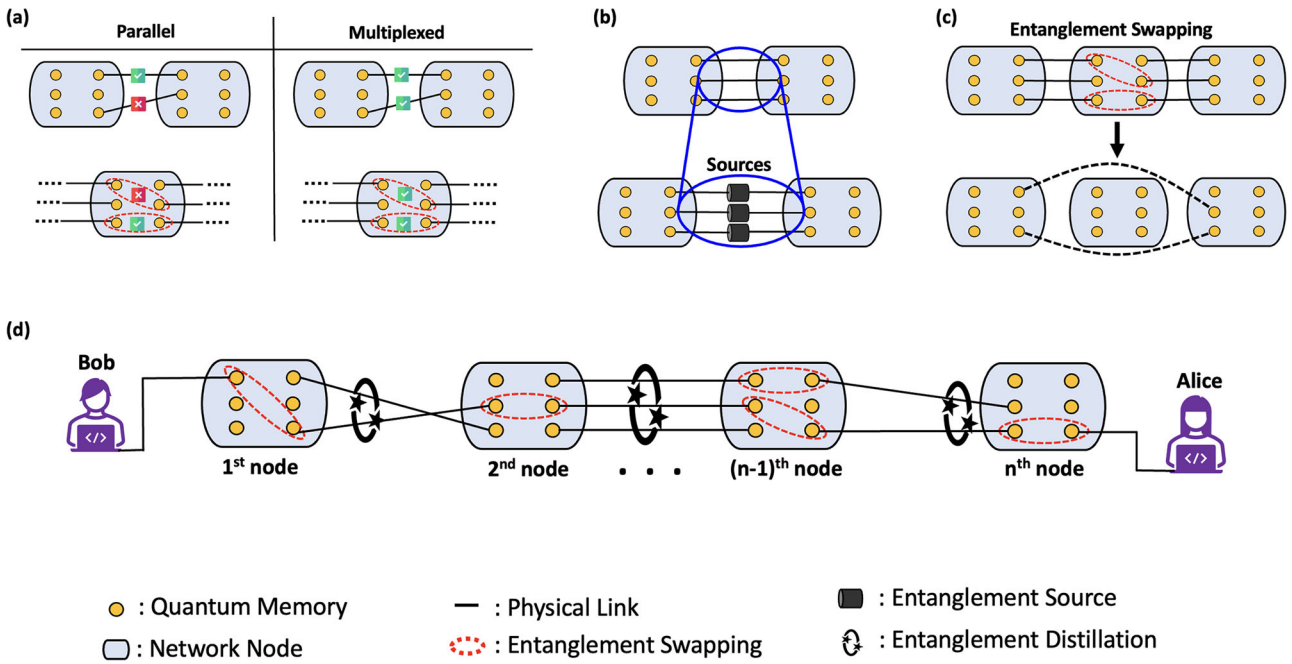

Fig. 1 | Multiplexed entanglement distribution. a In the multiplexing approach to entanglement distribution, it is possible to perform cross-channel elementary link generation and entanglement swapping. **b** Elementary links are created by multiple entanglement generation sources associated with every channel that connects different pairs of quantum memories. **c** Non-neighboring nodes are connected by virtual links generated by performing entanglement swapping operations between neighboring active quantum memories. **d** A linear chain of quantum repeaters with multiple quantum memories capable of establishing multiple physical links, entanglement distillation and entanglement swapping.

costs are considerable, but usually ignored in existing literature. We incorporate the CC costs associated with our policies, and show that these quasi-local policies outperform fully-local policies that do not require any CC. We also show that our policies can outperform the doubling policy in practically relevant parameter regimes. Next, we also add entanglement distillation to our model, and address several policy questions. Finally, we consider a proof-of-principle experimental realization of a multiplexed quantum network equipped using a high-dimensional bi-photon frequency comb (BFC) and assess its performance.

Results and discussion

Overview of results

Our main results and contributions comprise four parts.

1. *Quasi-local policies for multiplexed quantum repeaters*. We define our quasi-local multiplexing-based policies. These policies are generalizations of the so-called "swap-as-soon-as-possible" (SWAP-ASAP) policy[30,31] to the multiplexing setting. We refer to them as *strongest neighbor* (SN) SWAP-ASAP, *farthest neighbor* (FN) SWAP-ASAP, and *random* SWAP-ASAP. These policies are distinguished by how they pair memories within nodes for the purposes of entanglement swapping. Intuitively, the SN SWAP-ASAP policy prioritizes the creation of high-fidelity end-to-end links, and FN SWAP-ASAP prioritizes lowering the waiting time for the creation of end-to-end links. These policies are also "quasi-local", requiring some knowledge of the network state but not full, global knowledge.
2. *The impact of classical communication costs*. In a realistic entanglement distribution setting with imperfect memories, considering classical communication (CC) costs is imperative in order to provide a realistic performance analysis. In *Performance evaluation*, we incorporate the CC costs associated with our policies, and thereby address (Q1). In particular, we show that our multiplexing-based FN and SN SWAP-ASAP policies outperform fully-local policies that do not require any CC. We thus show that the benefit gained by acquiring information about the network's state is retained even when the communication costs of acquiring such information is accounted for. We also address (Q2) by showing that the FN and SN SWAP-ASAP policies can outperform the doubling policy in parameter regimes corresponding to high channel losses, small coherence times of quantum memories, and a small number of multiplexing channels and/or a large number of nodes.
3. *Policies with entanglement distillation*. In *Distillation-based policies*, we also add entanglement distillation to our FN and SN SWAP-ASAP policies, and then address (Q3). We find that if the fresh elementary links have high fidelity, using entanglement distillation turns out to be detrimental. On the other hand, when fresh elementary links have low fidelity, and more generally in resource-constrained scenarios, entanglement distillation can be more useful. Nonetheless, even in such cases, we observe examples in which distillation can in fact be detrimental. In terms of the relative ordering of distillation and swapping, if links have low generation probability or low initial fidelity, then distilling first is more advantageous. When the elementary link generation probability is high, it is better to swap first and then distill.
4. *Design and analysis of real-world repeater chains*. In *Proposed experimental implementation*, we outline a proof-of-principle experimental realization of a quantum network equipped with multiplexing policies using a high-dimensional bi-photon frequency comb (BFC), and we map the relevant experimental parameters to their counterpart parameters within our modeling framework. Then we assess the performance of our multiplexing-based policies within these parameter regimes for two quantum memory platforms, diamond vacancy and rare-earth metal ions. In particular, we consider the number of repeater nodes needed to achieve a desired end-to-end waiting time and fidelity over a distance of 100 km.

Linear chain quantum network with multiple channels

Throughout this work, we consider entanglement distribution in a linear chain network of quantum repeaters. In this section, we present our theoretical model for entanglement distribution in such networks; see also Fig. 1.

- *Nodes:* The linear chain is made up of n nodes. Every node contains multiple quantum memories. Specifically, every node has $2n_{ch}$ quantum memories. Given the linear nature of our networks, $n_{ch} \in \{1, 2, \ldots\}$ corresponds to the number of channels connecting the nearest-neighbor nodes. (In Fig. 1, $n_{ch} = 3$.) Every memory has a finite coherence time of $m^\star \in \{0, 1, \ldots\}$ time steps, which specifies the maximum number of (discrete) times steps that qubits can be stored in the memory.
- *Elementary and virtual links:* Entanglement sources establish physical/elementary links between nearest-neighbor quantum memories with probability $p_\ell \in [0, 1]$. A probabilistic Bell state measurement creates a virtual link between distant quantum memories with success probability $p_{sw} \in [0, 1]$. These values are fundamentally limited by the technology used to create the system, e.g., linear optics makes $p_{sw} \leq 0.5$, and p_ℓ is determined by properties of the source, loss characteristics of the sources, etc. These considerations are relevant for our proposed experimental implementation, and we also provide details on how to determine p_ℓ and $m^\star$ from physical properties of the system.

We keep track of the fidelity of both elementary and virtual links by assigning every link an age. The age m of a link starts from $m = m_0$, when it is first created, with m_0 being a measure of its initial fidelity with respect to a perfect Bell state. The age then increases in discrete steps, such that $m \in \{0, 1, 2, 3\ldots\}$. Once the age of the link reaches $m^\star$, the link is discarded. Throughout this work, we consider a particular Pauli noise model for decoherence of the memories (see "Methods"), such that the fidelity $f(m)$ of a link that has age m is given by

$$f(m) = \frac{1}{4}(1 + 3e^{-\frac{m}{m^\star}}). \tag{1}$$

The age of a link formed by a successful entanglement swapping operation is determined by adding the ages of the two corresponding links: if m_1 and m_2 are the ages of the two links, then if the entanglement swapping is successful the age of the new link is $m' = m_1 + m_2$. This "addition rule" for the age of virtual links is a direct consequence of the Pauli noise model that we use, and a proof can be found in [ref. 45, Appendix C].

- *Entanglement distillation:* Multiple low-fidelity links between any two nodes can be distilled with probability $p_{ds} \in [0, 1]$ to form a higher-fidelity link. Throughout this work, we consider the BBPSSW protocol[84], which allows for distillation of two entangled links to one. We provide a summary of the protocol in "Methods". Briefly, if f_1 and f_2 are the fidelities of the two links being distilled, then the success probability $P_{\text{distill}}(f_1, f_2)$ of the protocol and the resulting fidelity $F_{\text{distill}}(f_1, f_2)$ are given by

$$P_{\text{distill}}(f_1, f_2) = \frac{8}{9}f_1 f_2 - \frac{2}{9}(f_1 + f_2) + \frac{5}{9}, \tag{2}$$

$$F_{\text{distill}}(f_1, f_2) = \frac{1 - (f_1 + f_2) + 10 f_1 f_2}{5 - 2(f_1 + f_2) + 8 f_1 f_2}. \tag{3}$$

We show in "Methods" that the age after distillation with the BBPSSW protocol is given by the following formula:

$$m' = \left\lceil m^\star \log\left(\frac{15 - 6(f_1 + f_2) + 24 f_1 f_2}{32 f_1 f_2 - 2(f_1 + f_2) - 1}\right)\right\rceil, \tag{4}$$

where $f_1 \equiv f(m_1)$, $f_2 \equiv f(m_2)$, m_1 and m_2 are the ages of the two links being distilled, the function $m \mapsto f(m)$ is defined in Eq. (1), and $\lceil x \rceil$ is the smallest integer greater than or equal to x.

Entanglement distribution protocols progress in a series of discrete time steps, based on the Markov decision process model developed and used in ref. 45, which we refer the reader to for further details on the model. In every time step, the following events occur.

1. Check the number of active links to the right of every node (except for the right-most node). If there are inactive links, request the corresponding elementary link.
2. Check the number of active links to the left and right of every node, except for the end nodes. If more than one link on either side is active, rank them based on the policy, as described in detail in *Quasi-local multiplexing policies* below. If the doubling policy is being used, pair the links for entanglement swapping only if they are the same length. Perform the entanglement swapping operations in the ranked order. Entanglement swapping is attempted only if the sum of the ages of the two links is less than $m^\star$.
3. If entanglement distillation is part of the policy, attempt to distill all links between the same two nodes until either one or no link survives (also see Supplementary Note 2). If the policy calls for distilling first, then swapping, interchange steps 2 and 3.
4. Increase the age of all active links by one time unit, accounting for decoherence. If classical communication (CC) overheads are to be accounted for, then add the CC time to the age of every active link; see below for details. Discard any link which has age greater than $m^\star$.

The time steps that we consider can be converted into physical times, and also the ages can be converted to fidelity values, by incorporating the network hardware and design parameters. Examples of such a translation are given in *Proposed experimental implementation* and further details are provided in "Methods".

Classical communication. Since entanglement distribution at the elementary link level is not deterministic, it must be heralded via classical signals. This heralding sets a natural time scale for entanglement distribution through a network. The heralding time is equal to the classical communication time between two adjacent nodes. For the purposes of our simulation, this heralding time becomes the duration of one time step. Since entanglement swapping is also often probabilistic, and entanglement distillation requires two-way classical communication[85] between nodes, two distinct classes of entanglement distribution approaches can be envisaged.

- A *local* approach, where all elementary link generation and entanglement swapping attempts are made agnostic to the success or failure of prior swaps. This would then mean that all elementary links (once generated successfully) are retained up to the cutoff time. When an entanglement swap has failed, resulting in both participating links to become inactive, the surrounding nodes must act as if the link is still active until it has reached its expected cutoff. All classical communication about successful swaps is then done at the end of the protocol. See "Methods" for details about how the end of the protocol is established. Such an approach is fully local and therefore all policy decisions are taken by individual nodes, without any collaboration or exchange of information. The benefit is that no classical communication overheads exist except for heralding of elementary links, and therefore every time step has duration equal to the heralding time only. The cost is instead paid by having to wait for a link to reach its cutoff time, even though it might have become inactive long before that. Distillation can be only included for freshly generated elementary links in a fully local scheme, since as links get older their lengths cannot be predicted locally (since results of swaps are not known).
- A *global* approach, on the other hand, allows end-to-end classical communication amongst all nodes in every time step, such that the duration of every time step is equal to $(n-1)\times\Delta t$, where Δt is the CC time across an elementary link. The waiting time is subsequently largely dominated by the CC overheads, putting a huge burden on the coherence time requirements of the memories. The benefit, of course, is that all nodes can now communicate with each other and take decisions collaboratively and in an adaptive fashion.

In previous works such as refs. 44,45, global policies were shown to optimize the average waiting time and fidelity; however, classical communication costs were not included. Furthermore, in ref. 45, a "quasi-local" approach to entanglement distribution policies was proposed, in which there is knowledge of the network state up to some length scales only, and not globally. In this case, an advantage over fully local policies could still be obtained. We now investigate in this work whether the advantages in waiting time and fidelity gained by global and quasi-local knowledge and collaboration can be retained when CC overheads are taken into account.

Below is a summary of how we have taken classical communication into account in our model (Let us make a brief remark. In a full simulation, such as those which can be done via quantum network simulators such as NetSquid[37], QuISP[40], SeQUeNCe[38], etc., time stamps are assigned to each quantum operation and tasks are queued, such that classical communication costs are intrinsically kept track of. Our simulations focus on optimization of policies, similar in spirit to *rule sets* used by many of the above mentioned simulators; see also refs. 88,89).

- We approximate the CC time by adding t_{cc} time steps to every network evolution step, where t_{cc} is equal to the length of the longest link (number of nodes) involved in any entanglement swap in that step times the CC time between two adjacent nodes of the chain. This allows enough time for the classical communication for all entanglement swaps attempted in that MDP step, and at the end of such a CC-accounted MDP step, all nodes are aware of which nodes they are connected to and the ages of those links (which can also be classically communicated). In most cases this estimate acts as an upper bound to the CC cost, since in practice nodes could perform the next required action as soon as they receive the CC tagged for or relevant to them. See "Methods" for details.
- CC costs associated with entanglement distillation are added based on the length of the longest links that are distilled in a time step. For local policies, distillation is only allowed for freshly prepared elementary links (if they are not perfect), and thus the CC cost is always equal to one time step, if distillation is performed, between any two adjacent nodes after heralded entanglement generation.

Quasi-local multiplexing policies

We now introduce our two multiplexing policies for entanglement distribution in a linear chain quantum network. In linear quantum networks without multiplexing (i.e., $n_{ch}=1$ in Fig. 1), one of the simplest and best studied swapping protocols is swap-as-soon-as-possible (SWAP-ASAP). As the name suggests, the SWAP-ASAP policy dictates that entanglement swapping should be attempted as soon as two memories in a given node become active, i.e., as soon as an entanglement swapping policy becomes possible. The SWAP-ASAP policy has been shown to be better than fixed nesting or doubling policies for a large set of parameter regimes in linear networks consisting of a single channel between nodes[26,31,59].

The direct application of SWAP-ASAP in multiplexed linear chains is unspecified due to the increased number of degrees of freedom. Indeed, when multiple links are available between nodes of the network, there are many possibilities for how to perform entanglement swapping. There is also the possibility to perform entanglement distillation between nodes that share several links. We therefore define our policies based on the following considerations.

1. If entanglement swapping of multiple pairs of links is possible, how should we group the pairs? We could, for example, perform entanglement swapping on link pairs that result in virtual links between *farthest* nodes, or the ones that result in *strongest*—highest fidelity—virtual links. Based on this consideration, we define the following two generalizations of SWAP-ASAP for multiplexed linear chains.
   (a) The *Farthest Neighbor (FN) SWAP-ASAP policy:* This policy prioritizes the creation of virtual links between faraway nodes.

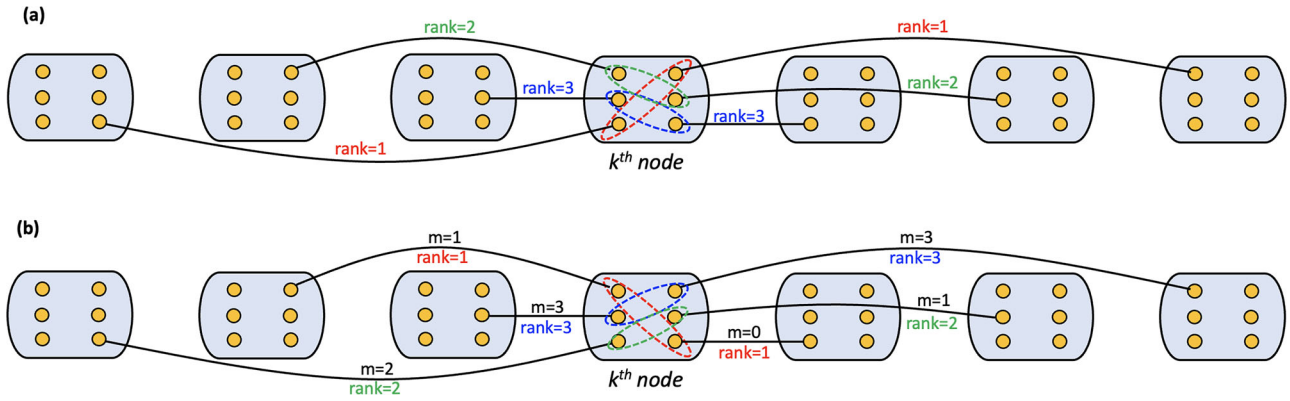

Fig. 2 | Our proposed entanglement swapping policies for multiplexed repeater chains. a Illustrative example of the FN SWAP-ASAP policy. Links on both sides of the k^{th} node are ranked separately on the basis of their lengths. The longest link is given a rank of 1, while shortest is ranked 3. For entanglement swapping, we pair links of rank 1, which would result in the longest possible link. Next links of rank 2 are connected, and so on. (**b**) In the case of the SN SWAP-ASAP policy, link rankings are based on their ages.

Accordingly, we rank the links based on their lengths (farthest links are ranked first), and then entanglement swapping is performed by pairing the links with the same rank. This policy is illustrated with an example in Fig. 2(a). Intuitively, this policy prioritizes minimizing the average waiting time.

(b) The *Strongest Neighbor (SN) SWAP-ASAP policy:* This policy prioritizes the creation of virtual links with high fidelity. Accordingly, we rank the links based on their ages (links with the lowest age are ranked first), and then entanglement swapping is performed by pairing links with same rank; see Fig. 2b. Intuitively, this policy aims to create the strongest or highest-fidelity links between the end nodes of the network. For comparison, we also define the *random SWAP-ASAP* policy. In the random SWAP-ASAP policy, links are randomly paired to perform entanglement swapping operations, without any consideration of their length and/or age. We also define the *parallel SWAP-ASAP* to be the non-multiplexed version of the usual SWAP-ASAP policy.

2. How should we distill multiple links to one? For this, we introduce the *distill-as-soon-as-possible* (DISTILL-ASAP) policy. This is a sequential policy, which can be thought of as a hybrid of the banded and pumping policies[17] (see Supplementary Note 2 for details), and very similar in spirit to the greedy policy defined in ref. 53. We sort the links in increasing order of their ages, and then pair them in increasing order. After one round of pairwise distillation, successfully distilled links are paired again and distillation is attempted again, until only one distilled link or no active link remains. This approach therefore uses the benefits of the banded policy, by pairing links of very similar ages, and at the same time, by realizing the restrictions introduced by decoherence and finite number of channels, by pumping fresh links into the queue of available links as soon as possible.
3. What should be the preference when it comes to the order of entanglement distillation and swapping—should we distill first and then swap, or the other way around? We call these two distinct options the DISTILL-SWAP and SWAP-DISTILL policies. Both of these may be used in conjunction with the SN and FN entanglement swapping policies defined above, leading to four distinct policy combinations.

The FN and SN SWAP-ASAP policies that we have introduced have classical communication (CC) overheads. Indeed, whenever an entanglement swapping decision is made, all the memories at the node are assumed to have knowledge of which nodes/memories they are connected to. This information is needed to decide the order of entanglement swaps based on age (SN) or length (FN) of the links. Thus, every node has some "quasi-local" knowledge of the network state but not the "global" network state.

We demonstrate via Monte Carlo simulations of the underlying Markov decision process that our quasi-local SN and FN SWAP-ASAP multiplexing policies can outperform (in terms of waiting time and end-to-end fidelity) the non-multiplexed, parallel SWAP-ASAP policy (as we would expect), and also the random SWAP-ASAP policy. We present these results in Supplementary Note 1 (also see Supplementary Note 4 for some alternative quasi-local multiplexing policies). In the next section, we present results that demonstrate, more importantly, that our policies can outperform fully local multiplexing policies, as well as the well-known and widely-used nested-purification-and-doubling protocol.

Before proceeding, we remark that very recently in ref. 90, the idea of using fidelity-based ranks to match links for entanglement swapping (similar in spirit to our SN SWAP-ASAP policy) was proposed, but the focus was on second-generation repeaters with error-correction capabilities. Our focus here is on first-generation repeaters. See refs. 55,61 for the definitions of first- and second-generation quantum repeaters.

Performance evaluation

We now evaluate the performance of our policies described above with respect to the average waiting time and average age of the end-to-end link. All of our results were obtained using Monte Carlo simulations, and details about the simulations can be found in "Methods".

In order to benchmark our quasi-local policies fairly against fully-local policies, some modifications need to be made to the way the network evolves under the fully-local policies. These modifications are necessary, because when entanglement swaps are probabilistic, if links need to be restarted after a failed Bell state measurement, but before their actual memory cutoff, some CC will be needed. Therefore, if a policy has to be fully-local and free of any CC overheads, nodes must be agnostic to entanglement swapping failures. All nodes, once they have a heralded entangled link, must retain them until their anticipated memory cutoff age $m^{\star}$. Once this restriction is made, the only available information to nodes that can be used to decide the rank of states is the perceived or local ages of its memories. In other words, every link in the network now has two ages: one is its real age, as determined by probabilistic swaps in the evolution "in real-time"; the second is its *perceived* age, which is the age that the nodes that hold the link assume without any knowledge of the outcomes of entanglement swaps. When the perceived ages are used to make decisions in the network, there is no CC overhead during the evolution. We call such a policy *local age-based* (ranking) multiplexed SWAP-ASAP policy. For such a policy, the CC cost is given by only one round of end-to-end communication time, in the last time step.

From Fig. 3, it is clear that substantial advantage for both the waiting time (around two orders of magnitude reduction) and fidelity is obtained by

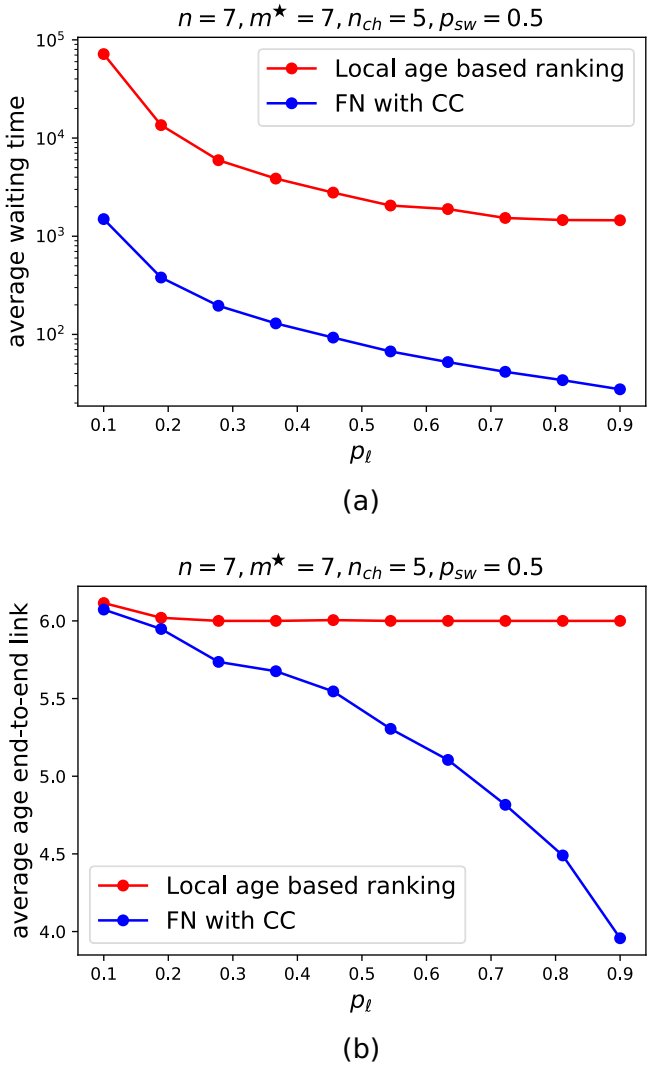

Fig. 3 | Quasi-local policies outperform fully-local policies. Average waiting time (**a**) and average age of an end-to-end link (**b**) as a function of the elementary link success probability when classical communication times are included. We compare the quasi-local FN SWAP-ASAP policy with the fully-local age-based SWAP-ASAP policy. The FN SWAP-ASAP policy provides smaller values for both figures of merit. Similar trends can also be seen for the SN SWAP-ASAP policy.

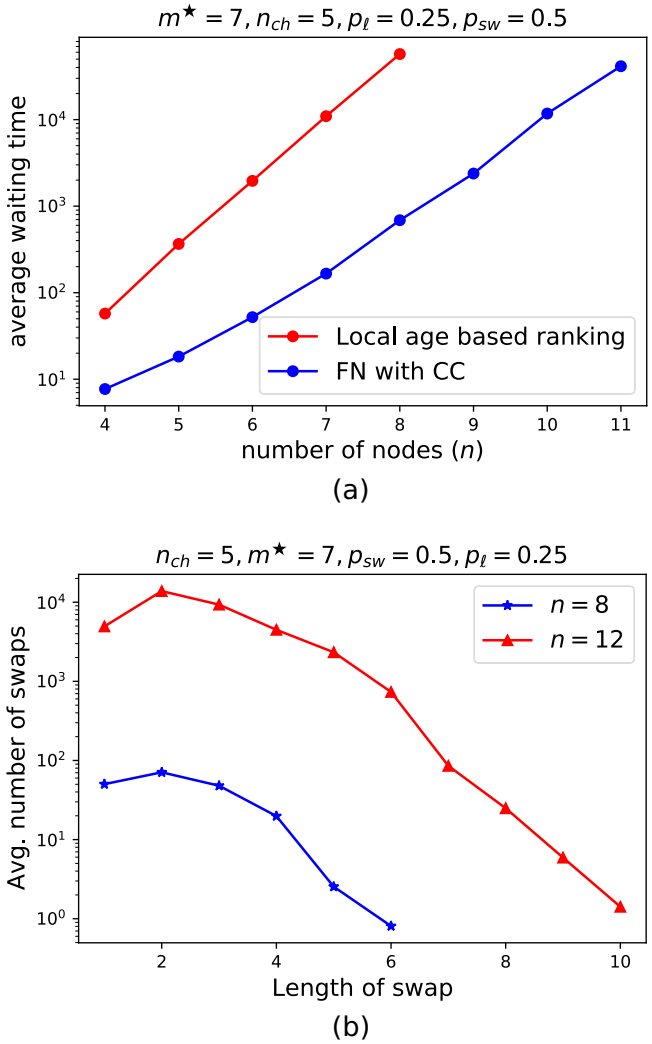

Fig. 4 | Advantage of quasi-local policies is scalable and grows with increasing number of nodes. a Average waiting time as a function of the number of nodes when classical communication times are included. We compare the quasi-local FN SWAP-ASAP policy with the fully-local age-based SWAP-ASAP policy, and we find that the FN SWAP-ASAP policy provides smaller waiting times. More surprisingly, the advantage grows with an increasing number of nodes. This provides evidence of the scalability of the advantage gained by using knowledge of the network state. **b** Number distribution of lengths of entanglement swaps in the FN SWAP-ASAP policy. We observe an exponential suppression of entanglement swaps of the order of the length of the chain, which is an important reason for the retained advantage of this policy, even with classical communication included.

using our policies compared to the local age-based SWAP-ASAP policy, even when CC times are taken into consideration.

A more surprising observation is the fact that even as the number of nodes increases, the advantage of FN SWAP-ASAP over the local age-based ranking policy remains strong—in fact, the advantage increases with increasing number of nodes. In Fig. 4a, we show the average waiting times for the local age-based ranking and the FN SWAP-ASAP policies as a function of the number of nodes. Intuitively, we might have guessed that as the number of nodes increases, so too would the CC cost, because longer and longer swaps would be required to distribute entanglement between the end points of the chain, making quasi-local policies worse off in terms of average waiting times. It is therefore surprising that not only a large improvement using the FN and SN SWAP-ASAP policies is obtained for long chains, but further that the improvement increases with the growing scale (number of nodes) of the network.

Nonetheless, we can explain this non-intuitive result as follows. First, we remind the reader that although fully-local policies have the advantage of no CC overheads, they suffer in terms of average waiting time due to the agnostic entanglement swapping. Indeed, waiting to restart links that are already dead is detrimental to the average waiting time. Second, because we use quasi-local policies, the CC costs in a given time step are upper-bounded by the length of the longest entanglement swapping operation, rather than the length of the entire repeater chain. (Recall that we define the length of an entanglement swapping operation as the length, in terms of number of nodes, of the longer of two links involved in the swap.) In a typical time-evolution of a network, from completely disconnected to an end-to-end connected link, for the FN SWAP-ASAP policy the number of entanglement swapping operations of length equal to the number of nodes is roughly $O(1)$, i.e., roughly constant with respect to the number of nodes. This is seen in Fig. 4b. The average number distribution of the length of entanglement swaps shows an exponentially suppressed tail. Therefore, there are very few time steps in which the CC cost is given by the entire length of the repeater chain. This justifies and supports the use of quasi-local policies over global policies for large networks, because in the latter, the CC cost is equal for

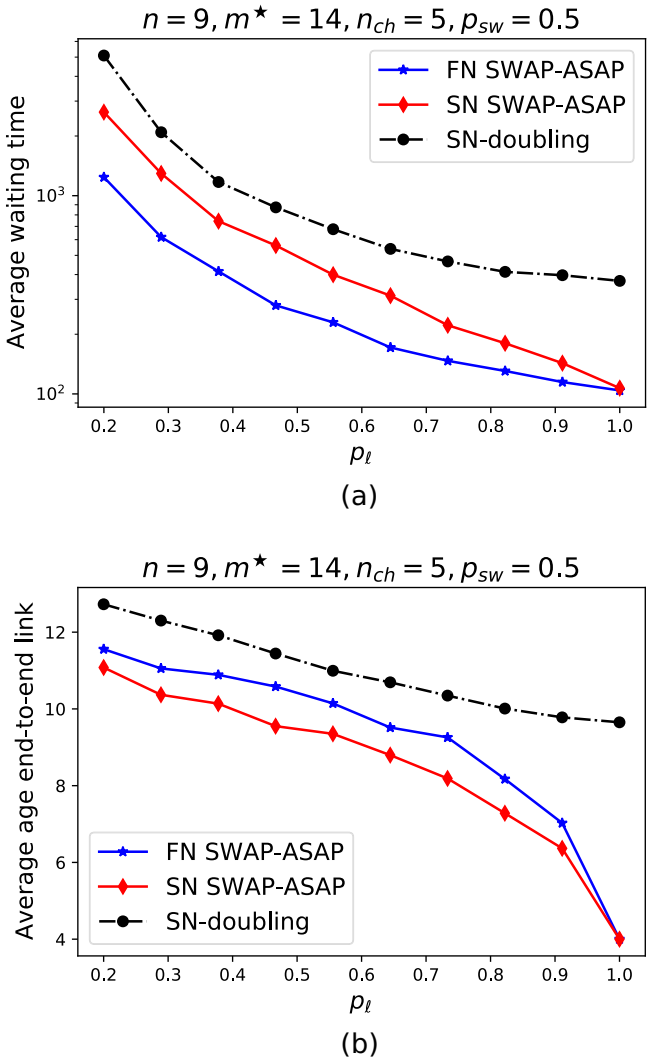

Fig. 5 | Strongest neighbor and farthest neighbor SWAP-ASAP policies outperform the doubling policy. Comparison of (**a**) average waiting times and (**b**) average age of the end-to-end link for the FN and SN SWAP-ASAP policies and the SN DOUBLING policy. FN SWAP-ASAP outperforms both SN SWAP-ASAP and SN DOUBLING in terms of average waiting time, and SN SWAP-ASAP performs the best in terms of average age of the end-to-end link, for all elementary link success probabilities (p_ℓ).

each time step and given by the entire length of the repeater chain, while for quasi-local policies the CC cost can be considerably less.

Comparison with doubling. Let us now compare our policies with those in existing literature. In particular, we compare our policies with one widely studied in previous works, namely, the *doubling* (or nesting) policy; see, e.g., refs. 16,17,53,54,66,71. In this policy, which only applies to repeater chains with 2^N elementary links, the lengths of links are always doubled. As an example, consider the following situation. Suppose two adjacent elementary links are formed, an entanglement swap is performed successfully, and a link of length two is obtained. Now, in contrast to the SWAP-ASAP policy, this length-two link cannot be extended by swapping it with links of arbitrary lengths—links of length two must be swapped with links of length two only—which means that we must wait for another link of length two adjacent to the original link to be created before an entanglement swap can be attempted. In the meantime, the first produced virtual link keeps aging. We can thus see that this doubling policy is more restricted compared to SWAP-ASAP. Accordingly, it was shown in ref. 59 that for low elementary link success probabilities and moderate swapping success probabilities (such as 50%, as in the case of linear optics), the SWAP-ASAP policy outperforms the doubling policy, although infinite coherence times for memories was assumed in that work.

Now, within the multiplexing scenario, the doubling policy can be implemented by ranking links based on age when more than one option of entanglement swapping is available. Thus, we call this policy *Strongest-Neighbor doubling*, or SN DOUBLING. This doubling policy is also quasi-local, in the same sense as SN and FN SWAP-ASAP, because a node only needs to know the length of its links, i.e., it needs to know which nodes it is connected to.

Our aim now is to see if, within the multiplexing setting, and with finite coherence times for the quantum memories, our FN and SN SWAP-ASAP policies can outperform the SN DOUBLING policy. Our results are shown in Fig. 5. We see that FN SWAP-ASAP outperforms both SN SWAP-ASAP and SN DOUBLING in terms of average waiting time. At the same time, SN SWAP-ASAP performs better than SN DOUBLING, thus further strengthening our intuition that SWAP-ASAP policies outperform DOUBLING policies in practical resource-constrained scenarios. We also see that SN SWAP-ASAP performs the best in terms of average age of the end-to-end link, for all elementary link success probabilities p_ℓ. This is expected, because SN SWAP-ASAP prioritizes the creation of younger links.

Distillation-based policies. Now we present results for distillation-based policies when CC overheads are included. As mentioned earlier in *Classical communication*, the CC cost of distillation for any quasi-local policy is given by the length of longest links that are used for distillation. For local policies, distillation is only allowed at the elementary link level, and only when the links are freshly prepared. We continue using the DISTILL-ASAP policy described in *Quasi-local multiplexing policies*. The CC overheads make the already probabilistic BBPSSW protocol even more costly. Furthermore, the CC costs of distilling longer (virtual) links is higher than distilling elementary links, and thus the question of whether to distill or not and whether to swap first or to distill first depends crucially on the CC overheads. We show the average waiting times for FN SWAP-ASAP policy without distillation and both distillation-based policies (SWAP-DISTILL and DISTILL-SWAP) in Fig. 6a, b.

Finally, in Fig. 6c, d, we again compare the different distillation-based versions of quasi-local policies amongst themselves, viz., FN SWAP-ASAP, SN SWAP-ASAP and the SN DOUBLING policies. We see that, with distillation (DISTILL-SWAP version) and the relevant CC costs added, the trends remain similar to Fig. 5, i.e., FN SWAP-ASAP is still the best policy for reducing the average waiting time and SN SWAP-ASAP for average age of the youngest end-to-end link.

The results of the numerical simulations presented in this section show that including entanglement distillation in multiplexing policies introduces many new considerations, such as the fact that, in some parameter regimes, it is better to distill first and in others the other way around. In the next section, we take a deeper dive into these factors surrounding the use of entanglement distillation.

Unraveling the role of entanglement distillation policies for multiplexing

The BBPSSW entanglement distillation protocol considered in this work non-deterministically converts two links with low fidelity into one link with a higher fidelity. This immediately raises the following question: "Is distillation useful for all network parameter regimes?" Due to its non-deterministic nature, in certain parameter regimes, the slight increase in fidelity obtained by distillation might not be worth the potential loss of links, in case it fails. Furthermore, even without the consideration of CC costs, there exists an underlying asymmetry between the way entanglement swapping and distillation protocols work. In the former, the success probability is independent of the ages of the links, and thus it does not discriminate between longer (generally older) and shorter links, whereas the latter's success crucially depends on the ages of the input links. This

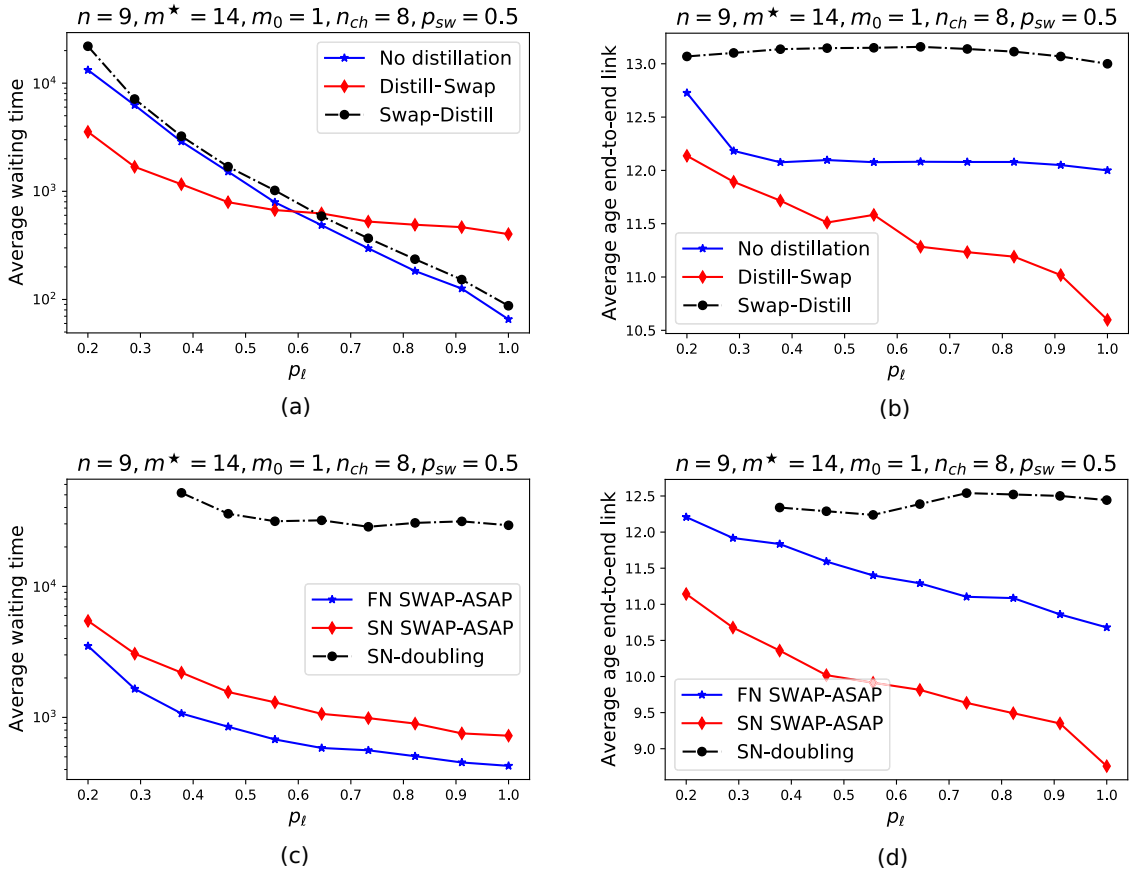

Fig. 6 | Performance of various quasi-local policies with distillation. Comparison of average waiting times (**a**) and average age of the end-to-end link (**b**) for the FN SWAP-ASAP policy with and without distillation. For the average waiting time, DISTILL-SWAP outperforms the policy without distillation for low elementary link success probabilities, but it is not useful to distill when the success probability is high. At the same time, we see that it is never useful to SWAP-DISTILL, both in terms of the average waiting time and the age of the end-to-end link. The CC costs of distillation are much higher when long virtual links are distilled, which provides intuition for the inefficacy of the SWAP-DISTILL policy. Average waiting times (**c**) and average age of end-to-end link (**d**) for the DISTILL-SWAP versions of FN SWAP-ASAP, SN SWAP-ASAP and the SN DOUBLING policies when classical communication costs are included. FN SWAP-ASAP is the best policy for reducing the average waiting time and SN SWAP-ASAP for average age of the youngest end-to-end link.

asymmetry leads to the following question: "Should we distill and then swap or the other way around?" This consideration leads to two distinct policies, namely DISTILL-SWAP and SWAP-DISTILL, with different behaviors in different parameter regimes, as we show in Fig. 6. The aim of this section is to pursue the mechanisms behind these different behaviors and unravel them layer by layer. We make one simplifying assumption in this section, which is that classical communication costs associated with entanglement swapping and distillation are ignored. This is motivated by two reasons: firstly, for short chains, CC costs are usually ignored in existing literature on entanglement distribution policies[31,43,44,91,92], thus this assumption allows us to fairly compare the performance of our policies with those in the existing literature. Secondly, ignoring CC simplifies our model and makes some analyses more transparent. At the same time, the conclusions we come to in this section hold qualitatively even when CC is accounted for.

We begin by noting that entanglement distillation is useful and worthwhile only when it increases the fidelity of the distilled link as compared to the fidelities of the original two links. In particular, the fidelity after distillation should be greater than the highest-fidelity link of the two being distilled, i.e., $F_{\text{distill}}(f_1, f_2) > \max\{f_1, f_2\}$. In terms of ages, this translates to the following condition for entanglement distillation with the BBPSSW protocol to be useful: $m' < \min\{m_1, m_2\}$, where m' is the age after distillation, as given by Eq. (4). If $m' \geq \min\{m_1, m_2\}$, then we might as well use the youngest of the two links rather than attempt entanglement distillation and risk its failure. In particular, if one of the ages is 0, corresponding to a perfect Bell pair, then distillation should not be performed.

Another important consideration is that both links being distilled should be entangled. If one of the links is not entangled, then it is better to simply discard the unentangled link and keep the entangled one. We elaborate on this point in "Methods" when discussing the BBPSSW distillation protocol. With this in mind, let us note that under the Pauli noise model that we consider, the fidelity according to Eq. (1) is bounded from below by 0.25 (in the limit $m \to \infty$), and it is equal to approximately 0.5259 when $m = m^\star$. The entanglement threshold for the states that we consider is 0.5, meaning that the link is entangled if and only if its fidelity strictly exceeds 0.5, and thus the age at which the link is no longer entangled is $m = \lceil m^\star \log(3) \rceil \approx \lceil 1.09 m^\star \rceil$ (also see "Methods"). This means that, depending on the value of $m^\star$, it can be useful to distill links that are older than $m^\star$ time steps. However, the probability of successful distillation falls with increasing ages of the links, and hence using a very old link in conjunction with a very young link increases the likelihood of wasting the high-quality young link. Thus, while distillation might be useful in this scenario, it is not necessarily worthwhile. Also, for the moderate values of $m^\star$ considered in our simulations, we choose to discard links at $m^\star$ time steps, instead of the strict limit of $\lceil m^\star \log(3) \rceil$, even when distillation is considered.

We now present simulation results for a seven-node chain with seven channels. We also provide analytical results for some simpler cases in Supplementary Note 3. A somewhat large value of the cutoff is chosen ($m^\star = 24$) to show some important scaling behavior with increasing age m_0 of the fresh (newly created) links. Figure 7 shows the average waiting time and average age of end-to-end links as a function of the ages of the

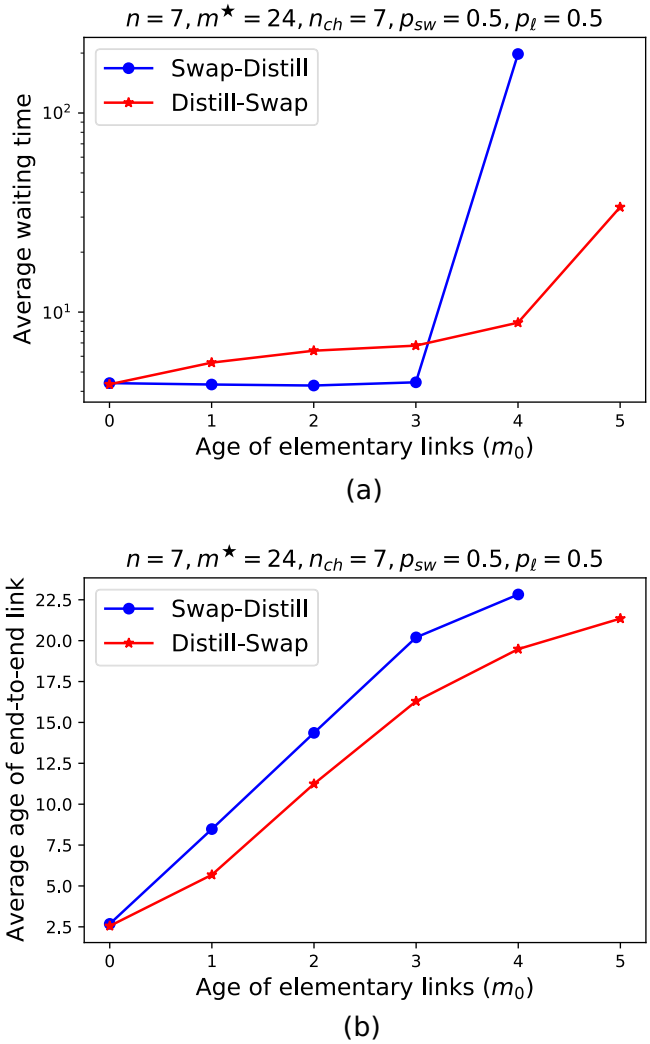

Fig. 7 | Should we swap-then-distill or distill-then-swap? Average waiting time (**a**) and average age of end-to-end links (**b**) as a function of the age m_0 of newly-created elementary links. The entanglement swapping policy is fixed to FN SWAP-ASAP. It is clear that SWAP-DISTILL is the policy of choice in terms of reducing the waiting time when fresh elementary links have high fidelity (low age); otherwise, DISTILL-SWAP performs better.

elementary links m_0 when they are fresh, i.e., newly created. For both SWAP-DISTILL and DISTILL-SWAP, we have fixed the entanglement swapping policy to FN SWAP-ASAP.

It is clear that SWAP-DISTILL is the policy of choice in terms of reducing the waiting time when fresh elementary links have high fidelity (low age); otherwise, DISTILL-SWAP performs better. At the same time, DISTILL-SWAP always performs better in terms of the age of the end-to-end link. We also verify that the waiting time rapidly increases beyond a threshold value of m_0 for both SWAP-DISTILL and DISTILL-SWAP, the threshold being smaller for SWAP-DISTILL. Beyond the m_0 values reported in Fig. 7, end-to-end entanglement generation time increases so much in the simulations that we are unable to find any reasonable estimate of an average value. Thus, for all practical purposes, the waiting time is essentially infinite. This is also indicated by the average ages approaching $m^\star$.

Some estimates of the m_0-thresholds up to which the two distillation policies are effective can be made using some rather simple arguments. If the fresh elementary links have age m_0, the age of the distilled link saturates to a minimum value $m^d_{\min}(m_0)$ which is a function of m_0 (see, e.g., the pumping policy in Supplementary Note 2). In order to get an end-to-end link via entanglement swapping of such distilled links, the following condition must be satisfied:

$$(n-1)\cdot m^d_{\min}(m_0) \le m^\star. \tag{5}$$

For the configuration in Fig. 7, the above condition translates to $m^d_{\min} \lessapprox 4$, which in terms of m_0 is $m_0 \le 5$, and this is the value around which we see that the average waiting time diverges.

In the case of SWAP-DISTILL, a typical trajectory to get an end-to-end link would be to first perform $(n-2)$ swaps to get end-to-end links of age $(n-1)m_0$, which are then distilled to links of age $m^d_{\min}((n-1)\cdot m_0)$. The threshold condition thus takes the following form:

$$m^d_{\min}\big((n-1)\cdot m_0\big) \le m^\star, \tag{6}$$

but since distillation is ineffective when both links involved are above the cutoff age, this condition is satisfied whenever

$$(n-1)\cdot m_0 \le m^\star. \tag{7}$$

This also explains why the threshold is lower for SWAP-DISTILL, since $m^d_{\min}(m_0) \le m_0$. Again, for the configuration in Fig. 7, this translates to $m_0 \lessapprox 4$, which agrees well with the simulation results. Furthermore, the threshold in the case of the SWAP-DISTILL policy is identical to a policy without distillation.

We further elaborate on our answer to the question of whether to distill first or to swap first, and whether or not distillation is useful in Fig. 8. To do this, we plot the waiting time *improvement factor* for distillation, which we define to be the ratio of the average waiting time without distillation to the average waiting time with distillation. Apart from the quality of fresh links (the parameter m_0), as shown in Fig. 7, the choice between the two distillation orderings also depends on the other network parameters. As an illustration, we show that for the parameters $(n, n_{ch}, m^\star, p_\ell)$ chosen in Fig. 8a, when p_ℓ is low, DISTILL-SWAP performs better, and when p_ℓ is high SWAP-DISTILL performs better. This observation can be understood in light of the analytical results in Supplementary Note 3 and Fig. 7. Indeed, we observe that for low p_ℓ, links are typically older when they are ready to be swapped, and we have seen in the previous discussions in this section that DISTILL-SWAP performs better in this case, because long entanglement swaps can only be possible once the ages of the participating links have been reasonably reduced. Also, distilling after entanglement swapping is highly improbable to be successful because the ages are going to only increase due to the swapping. On the other hand, when links have a high p_ℓ, SWAP-DISTILL performs better, because we observe that such links are typically younger when they are ready to be swapped, and they are much more conducive to successful swapping. Furthermore, in this case, the non-deterministic swaps become the rate determining step, and hence having more opportunities to perform entanglement swapping is more important. Distilling links first reduces the number of such opportunities. Furthermore, we see that for large p_ℓ, distillation is either not very useful or even detrimental. Note that these trends were also qualitatively seen in Fig. 6, when CC costs were accounted for.

The discussion above leads us to further explore the question pertaining to the usefulness of distillation-based policies. Distillation is a probabilistic process, and even when it succeeds it leads to a reduction in the number of active links. Hence, it is not obvious that a policy without distillation, such SN or FN SWAP-ASAP alone, cannot perform better than its distillation-based counterpart. Let us look at two cases: when fresh elementary links are perfect, i.e., $m_0 = 0$, and when they are imperfect, say $m_0 = 1$ for concreteness. In Fig. 8b–d we look at a seven-node chain with $n_{ch} = 12$, $p_{sw} = 0.5$, and $m^\star = 8$ ($m_0 = 1$ corresponds to roughly 90% fidelity in this case). We choose FN SWAP-ASAP as our swapping policy. We show the average waiting time improvement factor for distillation-based policies, defined as the ratio of average waiting time without distillation and with distillation as a function of the elementary link success probability p_ℓ. An improvement factor of greater than 1

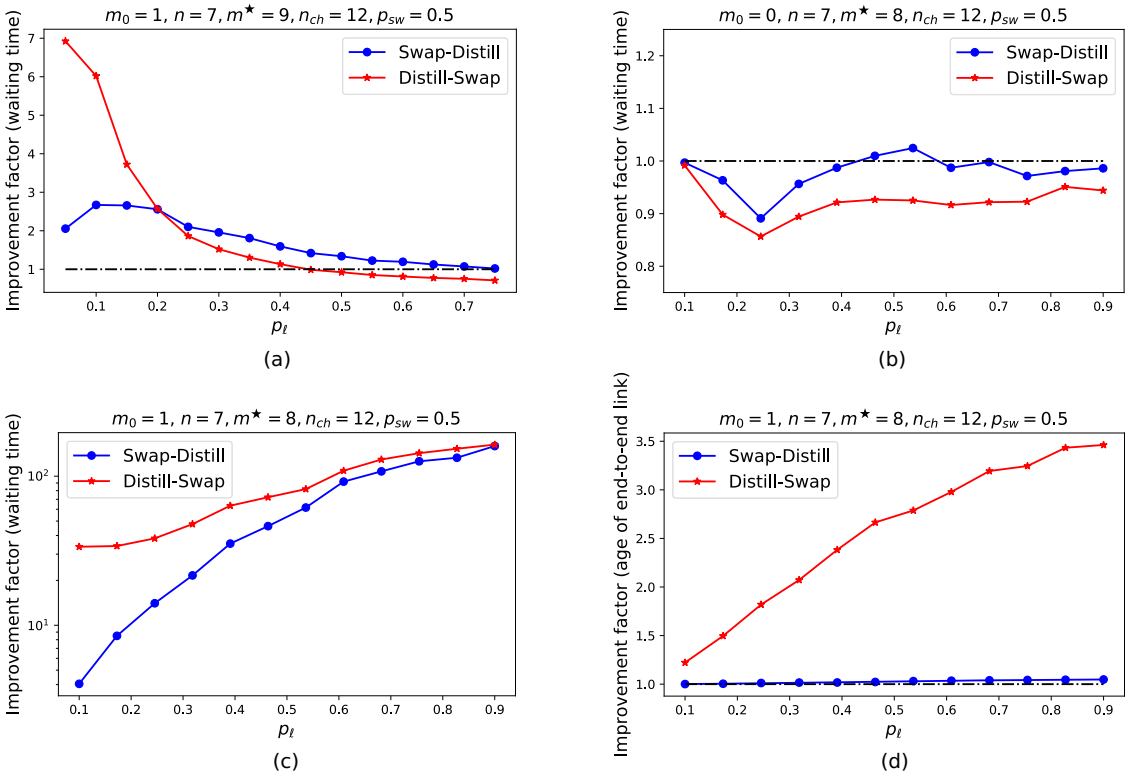

Fig. 8 | Is distillation useful? Average waiting time improvement factor for distillation with the SWAP-DISTILL and DISTILL-SWAP policies (with FN SWAP-ASAP entanglement swapping). **a** When p_ℓ is low, DISTILL-SWAP performs better, and when p_ℓ is high SWAP-DISTILL performs better. These are general trends and in concurrence with (**c**), where for low p_ℓ DISTILL-SWAP is better but as p_ℓ increases SWAP-DISTILL catches up. Furthermore, for large p_ℓ, distillation is either not very useful, and can even be detrimental. Distillation is not very useful, in fact mostly detrimental, when (**b**) fresh elementary links are perfect ($m_0 = 0$, 100% fidelity), but when fresh elementary links are imperfect (**c**) ($m_0 = 1$, approx. 90% fidelity), a dramatic reduction in waiting time can be obtained by performing distillation. **d** Improvement factor for average age of youngest end-to-end link using distillation over non-distillation based (FN) policy. SWAP-DISTILL provides negligible advantage but DISTILL-SWAP provides substantial advantage.

indicates an advantage in using distillation and vice versa. We see in Fig. 8b that distillation is mostly detrimental when freshly produced elementary links are perfect (100% fidelity). Both SWAP-DISTILL and DISTILL-SWAP increase the average waiting time, for all values of the elementary link probability. Further, the disadvantage is higher in the case of DISTILL-SWAP. This can understood by the following observation. The primary role of distillation in the case when elementary links are perfect is to reduce the age of long or older links obtained as a consequence of waiting for other links and/or via entanglement swaps. The longest links tend to be the oldest ones in the network, due to the $m_1 + m_2$ age-update rule for entanglement swapping. If we distill first, then initially, since most links are perfect, distillation is never invoked and after entanglement swaps, when distillation is needed, we have to wait an extra time step before distillation is attempted. This extra time step reduces the success probability of distillation of already old links, following Eq. (2). This effect is even more magnified when CC costs are added, since swaps will now lead to even older links. This is also the reason why SWAP-DISTILL has some positive improvement in some parameter regimes. Although not shown in Fig. 8b, there exists a small parameter regime of low $m^\star$, p_ℓ and n_{ch} in which SWAP-DISTILL leads to a small positive improvement (3–5%). The choice of parameters in the figure is motivated by the intent to show the most prominent and overarching trends. On the other hand, when elementary links are imperfect, distillation is essential in increasing the scale, fidelity and throughput of the network, indicated by the significant reduction in waiting time seen in Fig. 8c. When the elementary link probability is low, DISTILL-SWAP performs much better, and when it is is high, SWAP-DISTILL catches up in terms of improvement. This is in agreement with the results shown in Fig. 7 and 8a. In terms of the end-to-end link fidelity SWAP-DISTILL provides negligible advantage but DISTILL-SWAP provides substantial advantage as shown in Fig. 8d. It is noteworthy that the trends of improvement factor variation with increasing elementary link probability change between Fig. 8a, c substantially with a small change in $m^\star$ from 8 to 9. This happens because $m^\star = 9$ lies at the threshold near which distillation becomes beneficial, hence the high sensitivity to $m^\star$. As an aid to intuition, recall that if no distillation is used, then to establish an end-to-end link, when newly-created links have age $m_0 = 1$ in a chain of seven nodes, the smallest cutoff time that ensures an end-to-end entangled link is $m^\star = 6$, and note that $m^\star = 8, 9$ are very close to this limit.

Main messages for distillation. In this section, we have addressed the following questions.

- *To distill or not to distill?*

Yes (generally), when fresh elementary links are not perfect. In this case, distillation leads to a significant improvement in both the average waiting time and the average age of the end-to-end link. Nonetheless, the decision to distill links is not straightforward, since in extremely resource constrained scenarios, such as very low p_ℓ or p_{sw} it might again become disadvantageous to distill (see *Figures of merit and repeater chain design for existing memory platforms*), even when fresh links are imperfect.

No, when fresh elementary links are perfect. In this case, distillation offers either negligible advantage or is in fact detrimental for most parameter regimes.

More generally, the more resource constrained the scenario, i.e., in terms of network's hardware parameters, the more useful distillation is. The

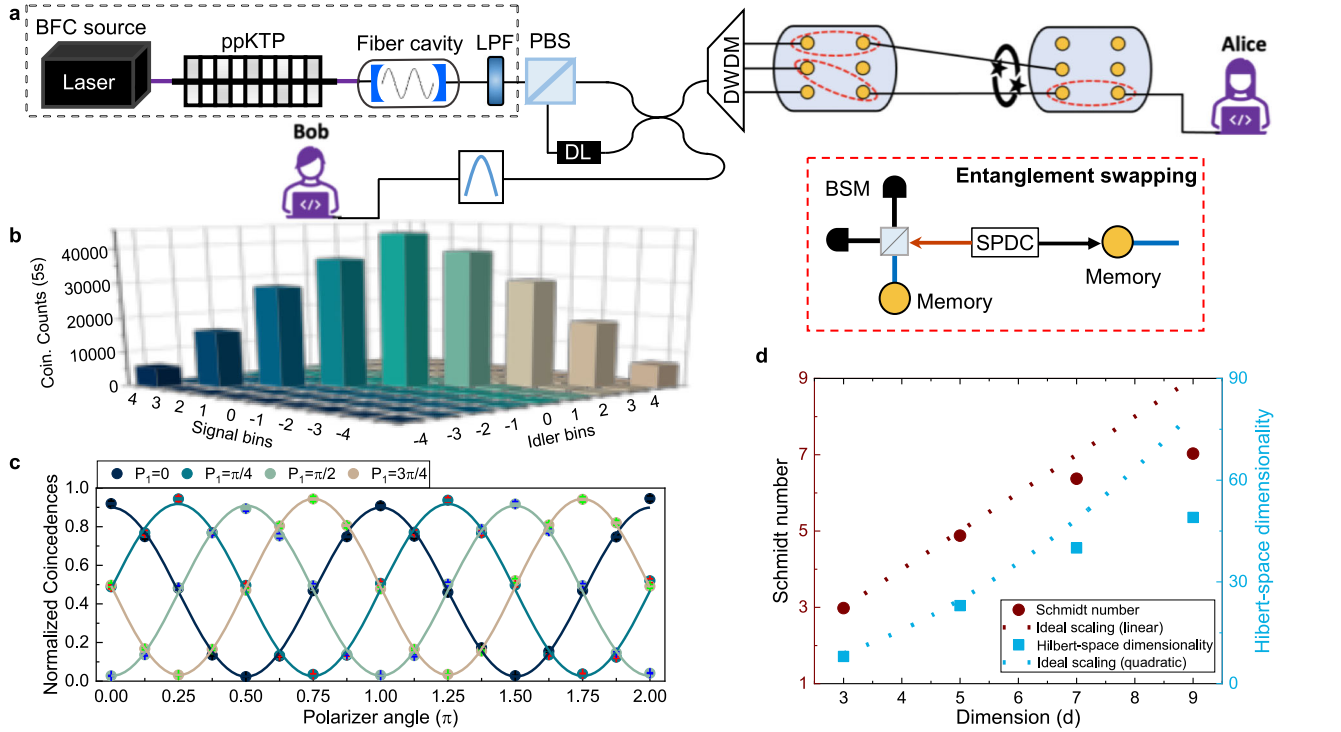

Fig. 9 | Proposed experimental implementation of a multiplexed repeater chain. **a** Proposed implementation of a multiplexed repeater chain using a high-dimensional biphoton frequency comb (BFC). **b** Measured frequency correlation matrix of the BFC for the multiplexed elementary link that we implemented. **c** Polarization entanglement measurements of the implemented BFC. **d** Schmidt number and Hilbert-space dimensionality scaling for different dimensions in the implemented BFC.

observations made in this section show that detailed simulations are indispensable for analysis of the role of distillation in repeater chains.

- *To distill first or to swap first?*

If links have low p_ℓ or low initial fidelity, distilling first is more advantageous, in the opposite case it is better to swap first.

More generally, in resource constrained scenarios, i.e., in terms of network's hardware parameters, it is more useful to distill first and then swap. Further, classical communication overheads only lead to quantitative changes in these trends.

Proposed experimental implementation

Based on the multiplexed model introduced in Fig. 1, in this section we propose a proof-of-principle experimental proposal of a quantum network with multiplexing using a high-dimensional biphoton frequency comb (BFC) in Fig. 9.

As part of this proposal, we present results of an implementation of a multiplexed elementary link using a high-dimensional BFC. First, the BFC state is prepared by passing the spontaneous parametric down-conversion (SPDC) photons generated from a type-II periodically-poled KTiOPO4 (ppKTP) waveguide through a fiber Fabry-Pérot cavity (FFPC)[93–95]. Due to the type-II phase-matching condition, the signal and idler photons can be efficiently separated by a polarization beam-splitter (PBS), enabling deterministic BFC generation without post-selection. The signal and idler photons are then temporally overlapped at a fiber beamsplitter with orthogonal polarizations, by using a tunable delay line (DL) and polarization controllers[93]. This configuration generates polarization entangled BFC state

$$|\psi\rangle = \sum_{m=-N}^{N} |\omega_p/2 + m\Delta\Omega\rangle_1 |\omega_p/2 - m\Delta\Omega\rangle_2 \otimes (|H\rangle_1|V\rangle_2 + |V\rangle_1|H\rangle_2), \quad (8)$$

where $2N+1$ is the number of cavity lines passed by an overall bandwidth-limiting filter (the number of frequency modes), $\Delta\Omega$ is the free spectral range of the FFPC, and ω_p is the pump frequency[95]. This scalable frequency-polarization hyperentangled BFC state[93] facilitates the multiplexing network policy by transmitting polarization qubits via different frequency channels to quantum network nodes, where entanglement swapping is performed in polarization basis. Specifically, within the context of our theoretical model, $n_{ch} = 2N + 1$. As shown in Fig. 9a, one output channel of the beamsplitter is directed to Bob, and the other is connected to a commercial dense wavelength-division multiplexing (DWDM) module for de-multiplexing and sent to quantum network nodes, consisting of multiple quantum memories. Each frequency channel is assigned to couple to one quantum memory at the quantum network node. The entanglement swapping inside a node is assisted by an auxiliary SPDC source, where the signal photons are sent to Bell-state measurement and idler photons are sent to another memory to establish a fiber link to the next node.

Figure 9b shows the measured frequency correlation matrix of a BFC, using a FFPC of 50 GHz free spectral range (FSR) and finesse of 10. We measured 9 correlated frequency-bin pairs between signal and idler photons with a type-II SPDC source of ≈250 GHz phase-matching bandwidth. Figure 9c shows the measured high-quality polarization fringes for the BFC state before de-multiplexing, which shows a raw fringe visibility up to $94.1 \pm 0.5\%$ and violates the Clauser-Horne-Shimony-Holt (CHSH) inequality by 19 standard deviations with score $S = 2.666 \pm 0.034$. This polarization state yields a state fidelity of $93.3 \pm 0.1\%$ with respect to a Bell state $\frac{1}{\sqrt{2}}(|H\rangle_1|V\rangle_2 + |V\rangle_1|H\rangle_2)$. We note that, ideally, the polarization entanglement fidelity remains the same after multiplexing. However, due to the falloff of the frequency correlation matrix, there will be deviation for the frequency-bin pairs away from the central bin[93,96]. Such high-fidelity polarization entanglement facilitates the entanglement swapping between quantum memories.

We perform the Schmidt mode decomposition of the measured joint spectral intensity[93], and summarize the extracted Schmidt mode number (K), and the corresponding Hilbert space dimension (K^2). In particular, K represents the effective orthogonal modes in the system. When reconstructing the density matrix, we lower bound the Schmidt mode number, which scales linearly with the dimension d of the biphoton frequency comb.

The Hilbert-space dimensionality can then be estimated by K^2. In Fig. 9d, we plot the Schmidt mode number and the Hilbert-space dimension as a function of the dimension d, and we compare it with the ideal, theoretical scaling. We observe that the Schmidt mode number is $K = 2.98$ with $d = 3$, and $K = 7.03$ when we increase the dimension to 9. Such BFC state provides up to seven effective frequency modes for multiplexing. The deviation from ideal scaling in Fig. 9d mainly comes from the falloff from the sinc function of the SPDC spectrum, which can be circumvented by using a flat-top broadband SPDC source. The Hilbert-space dimension can be further increased by using a SPDC source with broader spectrum bandwidth and FFPC with smaller FSR. We also note that by passing only the signal photons through a FFPC, the idler photons will still exhibit a comb-like behavior[96,97]. Such a scheme can provide higher photon flux due to less filtering loss, and with potentially higher secure key rates for quantum key distribution[98–101]. Although we have already experimentally implemented the multiplexed source required for the proposed experimental implementation, integration of the scheme with quantum memories, implementing entanglement swapping, and distillation of links has not been done yet. At the same time, there has been great progress in each of these key steps for the construction of a long-distance quantum network.

Calculation of model parameters. In this section, we show how to translate physical parameters into the values of p_ℓ, $m^\star$, and Δt within our theoretical framework.

Elementary link success probability p_ℓ. In general, we have that

$$p_\ell = \eta_r \eta_\ell \eta^2, \tag{9}$$

where $\eta_\ell = \exp(-\frac{\ell}{12\,\mathrm{km}})$ is the photon-loss contribution, with ℓ being the length of the elementary link. For $\ell = 40$km, we have $\eta_\ell \approx 0.036$. The factor of $\frac{1}{12\,\mathrm{km}}$ in the exponent corresponds to the 3.63 dB loss in the 10 km channel reported in the experiment in ref. 102. In addition, $\eta = 0.69$ is the Debye–Waller factor, the ideal efficiency of the memory (fraction of photons captured in the memory out of all the incident photons)[103], for a rare-Earth metal based quantum memory[104], and $\eta_r = 0.79$ (approximately 1 dB) is the loss from other residual factors, which could include detector efficiencies, optical component loss, etc.[93]. Altogether, we have that $p_\ell = \eta_r \eta_\ell \eta^2 \approx 0.0134$. In the case of emissive memories like those based on Diamond vacancies, an effective Debye–Waller factor of 0.5 is assumed, since a linear-optical Bell state measurement is needed at each node to teleport the state of the incoming SPDC photon into the memory.

Simulation time step Δt. This is given by the elementary link distance ℓ and the source rate per channel as

$$\Delta t = \max\left\{\frac{1}{R}, \frac{n\ell}{c}\right\} \text{ seconds}, \tag{10}$$

where n is the refractive index of the communication medium, ℓ is length of the elementary link, and c is the speed of light in vacuum and R is the rate of ebit pairs generated by the dimmest channel of the multiplexed SPDC source (see Fig. 9b).

Memory cutoff $m^\star$. Using the memory coherence time of $T_2 \approx 1$ ms, as for a rare-Earth metal based quantum memory[104], we can find the memory cutoff $m^\star$ as

$$m^\star = \frac{T_2}{\Delta t}. \tag{11}$$

For $\ell = 40$ km, we have $m^\star = 5$.

Fresh elementary link age m_0. This is given by the fidelity f_s of the polarization entangled state produced by the SPDC source (which we assume to be the same for all channels, for simplicity) and the fidelity f_0 of the entangled state when it is freshly absorbed (emitted) by the quantum memory. The fidelity of the elementary link is thus given by $f_e := f_s f_0^2$. This can be converted to the initial age of the elementary links using Eq. (26). Note that this also depends on $m^\star$.

For example, if the elementary link distance ℓ is varied from 5 to 25 km, p_ℓ varies from 0.25 to 0.05, and Δt goes from 0.025 to 0.125 ms. The latter in turn implies that $m^\star$ varies from 40 to 8, assuming a T_2 of approximately 1 ms.

Figures of merit and repeater chain design for existing memory platforms. Let us now address the following question. Given a particular number n_{ch} of channels for multiplexing, and given other physical parameters corresponding to channel losses, quantum memory coherence times, etc., what is the optimal number of repeater nodes for a linear quantum repeater chain spanning a certain distance, and what would be the rate and fidelity of end-to-end entanglement distribution for this optimal setting? Intuitively, we do expect such an optimal number of nodes to exist, because as the number of nodes increases, even though the individual links become smaller and thus p_ℓ and $m^\star$ both increase, the non-deterministic nature of entanglement swapping and/or the "age addition" rule, will at some point begin to adversely affect the waiting time as the number of nodes increases. Furthermore, in a (source-)rate limited setting, $m^\star$ also saturates to a maximum value with decreasing elementary link length.

To answer the question posed above, let us set the end-to-end distance to be 100 km. We also use the FN SWAP-ASAP policy for illustrative purposes, and we also include the classical communication overheads, but we leave out distillation-based policies. The physical parameters are chosen to reflect the proposed experimental implementation above. They are as follows:

- Number of multiplexing channels n_{ch}: We choose $n_{ch} = 5$, since although the SPDC spectrum in Fig. 9(b) can be divided into nine channels, both the ebit rates of polarization entangled states and their fidelity with respect to the maximally entangled Bell states fall as one moves away from the center of the spectrum.
- Source rate R: We choose $R = 5000$ ebits/s. Since the original SPDC source is divided into different frequency channels, the number of ebits per channel is at least lower by a factor of n_{ch}. Furthermore, as shown in Fig. 9(b) the sinc function of the SPDC spectrum makes the ebit rates fall as one moves away from the central channel. Thus, to make a conservative estimate, we choose R to be rate corresponding to the dimmest channel. This also allows us to choose the same rate for all channels (consistent with our simulation assumptions that all channels are identical).
- Fidelity of state at source f_s: We choose $f_s = 0.93$, as obtained in the experiment mentioned above. Recall that the fresh elementary link fidelity is given by $f_e = f_s f_0^2$.

Since quantum memories and entanglement swapping have not yet been implemented in the experiment, we choose two paradigmatic values for the swapping success probability, which are $p_{sw} = 0.5$ (linear optics) and $p_{sw} = 1.0$ (perfect gates on solid-state memories). For the quantum memories, we choose two quantum memory platforms for our analysis, a rare-Earth metal based memory (Pr^{3+} ions) and a diamond vacancy based memory. The state-of-the-art physical parameters for both the Rare-Earth and Diamond-based quantum memories are taken from [ref. 104, Table 1]. Given the above SPDC source and memory parameters, we can calculate all the relevant simulation parameters following the prescription given above. We plot the average rate and fidelity of end-to-end entanglement distribution both for near-term and state-of-the-art parameter values in Fig. 10. We tabulate the values for the fidelity and entanglement distribution rate for the two memory platforms in the optimal setting, along with the hardware parameters, in Table 1. The optimal number of nodes for the near-term (state-of-the-art) settings in terms of the waiting time seems to be around

Table 1 | Optimized figures of merit for a repeater chain using frequency multiplexing and quantum memories

State-of-the-art parameters								
Memory	T_2	η	f_e	p_ℓ	$m^\star$	m_0	R	f
Rare-Earth	1.2 ms	0.69	0.89	0.023	7	1	53 Hz	0.55
Diamond	13 ms	0.5	0.74	0.003	52	22	10 Hz	0.54
Near-term parameters								
Rare-Earth	1.2 ms	0.69	0.91	0.046	7	0	190 Hz	0.59
Diamond	13 ms	0.5	0.91	0.04	78	9	342 Hz	0.59

Hardware parameters for two quantum memory platforms (rare-earth ion and diamond vacancy) [ref. 104, Table 1][111,112], and their respective optimal entanglement distribution rates (R) and end-to-end fidelities (f), for a 100 km repeater chain, based on the data in Fig. 10. T_2 is the coherence time of the memory, η is the absorption efficiency, i.e., Debye–Waller factor (DWF). in the case of Rare-Earth memory, for emissive memories like the Diamond vacancy-based memory an effective DWF of 0.5 is assumed. f_e is the fidelity of a fresh elementary link (which includes the fidelity of the SPDC source state f_s and the initial entanglement generated or absorbed by the quantum memories f_0). For the near-term case we consider $f_0 = 0.99$. Since, solid-state memories are considered, we assume swapping to be deterministic. Also, shown are the parameters p_ℓ, $m^\star$, and m_0 of our theoretical model.

five to seven (two to four) for both memory platforms, which have T_2 times on the order of a few milliseconds. In the rate optimal setting (number of nodes), we anticipate entanglement distribution rates for the near-term (state-of-the-art) parameters ranging from a few Hertz to tens of Hertz, with an average end-to-end link fidelity of 60-70% (53-55%). We also see that as the swapping success probability is increased, especially close to the optimal choice of number of repeaters, there is a substantial increase in the rates, although the end-to-end fidelity does not substantially improve. This is also an artifact of choosing the farthest neighbor policy, which prioritizes forming long links. Also, the results of both platforms indicate that the end-to-end link fidelity decreases with an increasing number of nodes.

Finally, we look at the performance of SN DOUBLING and distillation-based multiplexing policies for the linear repeater chain considered above. Significant differences between SWAP-ASAP and DOUBLING policies appear for $n > 5$ (four links or more). For example, for a nine-node chain using a rare-Earth metal-based memory and 50% swapping success probability, the SN DOUBLING policy leads to around 60% higher waiting time compared to FN SWAP-ASAP. Similar improvement using the SWAP-ASAP policy can also be seen for the Diamond vacancy-based memory.

Since the elementary links have (nearly) unit fidelity for the rare-Earth metal-based memory, distillation is detrimental as shown by the results in Fig. 8. For the diamond vacancy-based memory with state-of-the-art initial state fidelity of 89%, we also find that for the network parameters under consideration, distillation is in fact detrimental. This observation can be attributed to the higher CC overheads compared to the small increase in fidelity that distillation offers in this case. For example, the maximum rates available (for optimal number of repeaters) with deterministic entanglement swapping and using FN-based DISTILL-SWAP policy is around 1 Hz, which is at least five times smaller than the rates available without distillation.

Conclusions

Near-term, resource-constrained quantum networks require the use of *hardware-aware* and *network state-aware* policies, in order to achieve high performance in terms of end-to-end waiting times and fidelities. These policies should be quasi-local, in order to reduce the impact of classical communication (CC) costs on performance. A careful assessment of the trade-off between CC costs and performance is crucial for the identification of optimal entanglement distribution policies for different hardware parameter regimes, particularly for first-generation quantum repeaters.

In this paper, we have taken steps in these directions. We have presented practical, quasi-local multiplexing-based policies for long-distance entanglement distribution using quantum repeaters with multiple memories. We call these policies farthest neighbor (FN) SWAP-ASAP and strongest neighbor (SN) SWAP-ASAP, adapting swap-as-soon-as-possible (SWAP-ASAP) policies to multiplexing-based linear networks. These policies go beyond fully local policies, such that the nodes have knowledge of states of the other nodes in the chain, but not necessarily full, global knowledge.

We have shown that not only do our quasi-local multiplexed policies retain their advantage over fully-local policies when CC costs are included, but this advantage can also increase with an increasing number of nodes. This is a surprising and counter-intuitive result, and shows that policies that use some knowledge of the network state (but not necessarily full, global knowledge) can enhance network performance, even when the CC costs associated with such knowledge are accounted for. The advantage attained by our quasi-local policies is rooted in the fact that they only require CC between connected regions of the chain; therefore, for most time steps, end-to-end CC is not required. This is an important conclusion from the point of view determining useful policies beyond the fully-local ones, especially for large quantum networks. We also benchmarked our policies against the widely studied doubling policy. We have shown via simulations that the FN and SN SWAP-ASAP policies can yield a considerable advantage over the doubling policy, both in terms of reducing the average end-to-end waiting time and increasing the fidelity of the end-to-end link. These advantages occur in the most relevant parameter regimes for near-term quantum networks, which correspond to the most resource-constrained settings.

We then considered policies with entanglement distillation, and we proposed a new policy that we call DISTILL-ASAP. This policy combines the benefits of existing distillation policies, like the banded, greedy, and pumping approaches, and thus is able to outperform the doubling policy nested with distillation. We also provide answers to two important policy questions related to distillation: When is distillation useful, and when it is useful, should we distill then swap, or the other way around? In this direction, the next question of immediate interest for future work is: How much to distill? Indeed, in this work, we only considered entanglement distillation policies that take N links and distill them to one, but one could instead distill N links to some K links greater than one, and such policies have been the study of recent work[105].

Finally, we performed an experimental demonstration of multiplexing using a high-dimensional biphoton frequency comb (BFC), which would form the backbone of our proposed experimental implementation of a linear-chain quantum network with multiplexing capabilities. We then assessed the anticipated performance of such a network over 100 km for two concrete memory platforms, namely rare-earth ion and diamond vacancy based quantum memories, when using our multiplexing-based policies.

Moving forward, it would be interesting to assess the optimality of the policies presented in this work. Optimal policies for multiplexed quantum repeater chains can be obtained using reinforcement learning (RL), using the methods developed in refs. 43–45. By adding the appropriate classical communication costs to such policies, we could assess not only the optimality of the FN and SN SWAP-ASAP policies presented in this work, but also we could determine whether quasi-local policies in general can outperform fully-global ones, when CC costs are accounted for.

Throughout this work, we have also considered average values of the waiting time and fidelity of the end-to-end link. The behavior of the higher moments of these quantities, and in particular the distributions of these quantities, is an important consideration that has an impact on how well the average estimates the real behavior. We anticipate that this will involve new techniques, and is an interesting direction for future work.

Methods

Noise model and decoherence

We consider the following Pauli channel noise model for qubit decoherence[106,107]:

$$\mathcal{N}_{m_1^\star,m_2^\star}(\rho) = p_I\rho + p_X X\rho X + p_Y Y\rho Y + p_Z Z\rho Z, \tag{12}$$

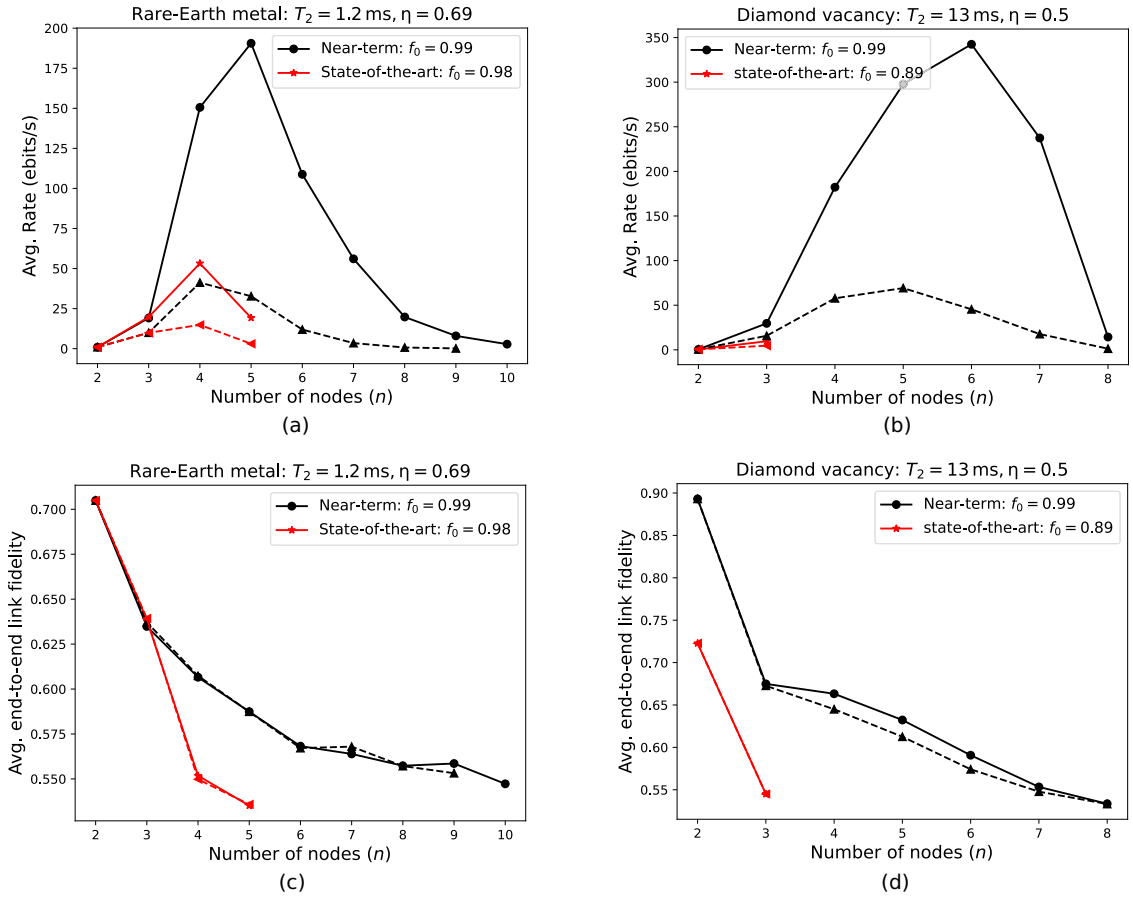

Fig. 10 | Performance of the proposed experimental implementation using our multiplexing policies. Average rate of end-to-end entanglement distribution (**a**, **b**), including classical communication overheads, and average fidelity of the end-to-end link (**c**, **d**) for entanglement distribution along a 100 km repeater chain with five channels with rare-Earth (Pr^{3+} ion) (**a**, **c**) and diamond vacancy (**b**, **d**) based memories using the FN SWAP-ASAP policy. The dotted (solid) lines represent $p_{sw} = 0.5 (p_{sw} = 1)$. Both state-of-the-art and near-term values for the fidelity f_0 of fresh entangled states generated (absorbed) by memory platforms are shown. Using three to five repeaters between the end nodes seems to be optimal for both memory platforms in terms of waiting times.

where X, Y, Z are the single-qubit Pauli operators, defined as

$$X := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y := \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \quad Z := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{13}$$

and the probabilities p_I, p_X, p_Y, p_Z are defined as

$$p_I = \frac{1 + \mathrm{e}^{-\frac{1}{m_2^\star}}}{2} - \frac{1 - \mathrm{e}^{-\frac{1}{m_1^\star}}}{4}, \tag{14}$$

$$p_X = \frac{1 - \mathrm{e}^{-\frac{1}{m_1^\star}}}{4}, \tag{15}$$

$$p_Y = \frac{1 - \mathrm{e}^{-\frac{1}{m_1^\star}}}{4}, \tag{16}$$

$$p_Z = \frac{1 - \mathrm{e}^{-\frac{1}{m_2^\star}}}{2} - \frac{1 - \mathrm{e}^{-\frac{1}{m_1^\star}}}{4}, \tag{17}$$

where $m_1^\star, m_2^\star \in \{1, 2, \ldots\}$. This channel is the Pauli-twirled version of the concatenated amplitude damping and dephasing channels.

Let us also define the two-qubit Bell states as follows:

$$|\Phi^\pm\rangle := \frac{1}{\sqrt{2}}(|0,0\rangle \pm |1,1\rangle), \quad \Phi^\pm := |\Phi^\pm\rangle\langle\Phi^\pm|, \tag{18}$$

$$|\Psi^\pm\rangle := \frac{1}{\sqrt{2}}(|0,1\rangle \pm |1,0\rangle), \quad \Psi^\pm := |\Psi^\pm\rangle\langle\Psi^\pm|. \tag{19}$$

Decoherence of an entangled qubit pair. Now, let us suppose that the initial state of an entangled qubit pair is the perfect maximally-entangled Bell state Φ^+. Then, after $m \in \{1, 2, 3, \ldots\}$ time steps, it is straightforward to show that the decohered entangled state is equal to

$$\begin{aligned} &(\mathcal{N}_{m_1^\star, m_2^\star} \otimes \mathcal{N}_{m_1^\star, m_2^\star})^{\circ m}(\Phi^+) \\ &= \frac{1}{4}\Big((1 + \mathrm{e}^{-2m/m_1^\star} + 2\mathrm{e}^{-2m/m_2^\star})\Phi^+ \\ &\quad + (1 + \mathrm{e}^{-2m/m_1^\star} - 2\mathrm{e}^{-2m/m_2^\star})\Phi^- \\ &\quad + (1 - \mathrm{e}^{-2m/m_1^\star})\Psi^+ \\ &\quad + (1 - \mathrm{e}^{-2m/m_1^\star})\Psi^-\Big). \end{aligned} \tag{20}$$

For a proof, we refer to [ref. 45, Appendix E]. This implies that the fidelity of the state after m time steps is $\frac{1}{4}(1 + \mathrm{e}^{-2m/m_1^\star} + 2\mathrm{e}^{-2m/m_2^\star})$.

In order to connect with the results presented in "Results", for a given value of $m^\star$ as presented there, let us take

$$m_1^\star = 2m^\star, \quad m_2^\star = 2m^\star. \tag{21}$$

The state in (20) is then

$$\sigma(m) := (\mathcal{N}_{2m^\star,2m^\star} \otimes \mathcal{N}_{2m^\star,2m^\star})^{\circ m}(\Phi^+) \tag{22}$$

$$= \frac{1}{4}(1 + 3e^{-\frac{m}{m^\star}})\Phi^+ + \frac{1}{4}(1 - e^{-\frac{m}{m^\star}})(\Phi^- + \Psi^+ + \Psi^-) \tag{23}$$

$$= f(m)\Phi^+ + \frac{1 - f(m)}{3}(\Phi^- + \Psi^+ + \Psi^-), \tag{24}$$

where the fidelity $f(m)$ after m time steps is equal to

$$f(m) := \frac{1}{4}(1 + 3e^{-\frac{m}{m^\star}}). \tag{25}$$

Note that we have chosen the value of $m_2^\star$ in Eq. (21) such that, at the cutoff time $m^\star$, the fidelity of the entangled qubit pair is $f(m^\star) = \frac{1}{4}(1 + 3/e) \approx 0.5259$. We emphasize that our results can be applied to other choices of $m_1^\star$ and $m_2^\star$—in particular, choices that could take dephasing as the dominant source of noise[108], in order to make connections to prior works[21,24,31,43,66].

We can invert the fidelity function in Eq. (25), such that for a given fidelity F, the corresponding age of the qubits is given by

$$f^{-1}(F) = \lceil -m^\star \log((4F - 1)/3) \rceil, \tag{26}$$

for all $F \in (f(m^\star), 1)$, where we take the ceiling $\lceil \cdot \rceil$ because we want an integer for the age.

We also remark that the decohered entangled state in Eq. (20) is a Bell-diagonal state of the form

$$\rho_{(a,b,c)} := (a + b)\Phi^+ + (a - b)\Phi^- + c\Psi^+ + c\Psi^-, \tag{27}$$

with $a = \frac{1}{4}(1 + e^{-2m/m_1^\star})$, $b = \frac{1}{2}e^{-2m/m_2^\star}$, and $c = \frac{1}{4}(1 - e^{-2m/m_1^\star})$. As such, its entanglement can be characterized entirely by the quantity $a + b$, namely, the fidelity of the state with respect to Φ^+. In particular, the state is entangled if and only if this fidelity is strictly greater than $\frac{1}{2}$; see, e.g., [ref. 58, Chapter 2]. Therefore, under the assumptions in Eq. (21), such that the fidelity is given by Eq. (25), we find that the state in Eq. (20) is entangled for all time steps $m \in \{0, 1, 2, \ldots, m^\star\}$. Furthermore, inserting $F = \frac{1}{2}$ in Eq. (26), we find that the age at which entanglement is lost is $\lceil m^\star \log(3) \rceil$.

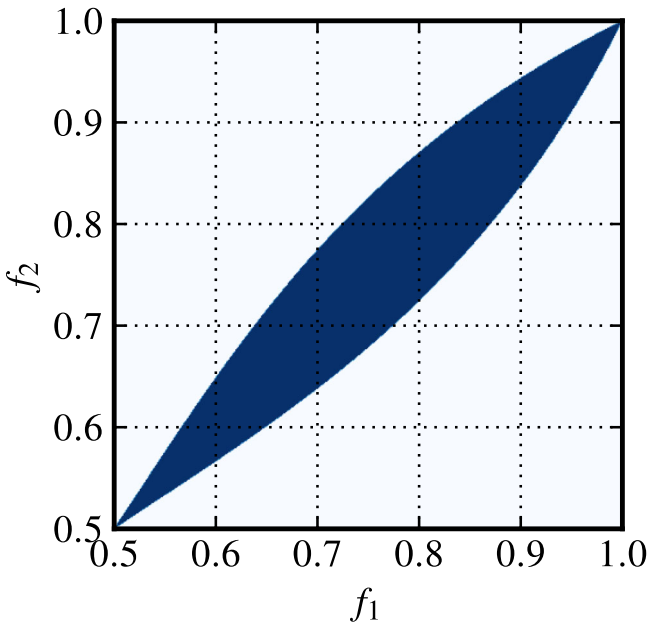

Fig. 11 | Usefulness of distillation with the BBPSSW protocol. The dark blue shaded region indicates the values of the initial fidelities, f_1 and f_2, for which the fidelity after distillation with the BBPSSW protocol[84] exceeds the fidelity of the best of the two links, i.e., when $F_{\text{distill}}(f_1, f_2) \geq \max\{f_1, f_2\}$.

Entanglement distillation with the BBPSSW protocol

In this work, the entanglement distillation protocol we consider is the BBPSSW protocol introduced in ref. 84. This protocol, involving the distillation of two entangled pairs to one, consists of a CNOT gate applied locally to each pair of qubits, followed by measuring the local target qubits of the CNOTs in the Pauli-Z basis. If f_1 and f_2 are the fidelities of the two entangled qubit pairs, then the success probability of the distillation procedure is

$$P_{\text{distill}}(f_1, f_2) = \frac{8}{9}f_1 f_2 - \frac{2}{9}(f_1 + f_2) + \frac{5}{9}, \tag{28}$$

and the fidelity of the resulting state is

$$F_{\text{distill}}(f_1, f_2) = \frac{1}{P_{\text{distill}}(f_1, f_2)}\left(\frac{10}{9}f_1 f_2 - \frac{1}{9}(f_1 + f_2) + \frac{1}{9}\right) \tag{29}$$

$$= \frac{1 - (f_1 + f_2) + 10 f_1 f_2}{5 - 2(f_1 + f_2) + 8 f_1 f_2}. \tag{30}$$

Note that these formulas apply only when the input states to the distillation protocol are so-called *isotropic states*, given by $f\Phi^+ + (\frac{1-f}{3})(\Phi^- + \Psi^+ + \Psi^-)$, with $f \in [0, 1]$ being the fidelity. Also, distillation should be considered useful only when it can improve upon the fidelity of the best of the two links. In other words, in order for distillation to be useful, we should have $F_{\text{distill}}(f_1, f_2) > \max\{f_1, f_2\}$. In Fig. 11, we plot the (f_1, f_2) region defined by this inequality. General restrictions on useful entanglement distillation protocols can be found in ref. 109.

From ref. 84, we have the following relations, indicating how the Bell states $\Phi^\pm$ and $\Psi^\pm$ are transformed by the BBPSSW protocol.

	Φ^+	Φ^-	Ψ^+	Ψ^-
Φ^+	Φ^+	Φ^-	0	0
Φ^-	Φ^-	Φ^+	0	0
Ψ^+	0	0	Ψ^+	Ψ^-
Ψ^-	0	0	Ψ^-	Ψ^+

Here, the rows correspond to the first link, and the columns correspond to the second link. From this, it follows that the unnormalized state corresponding to success of the BBPSSW protocol, when both input states are of the form (24), is

$$\begin{aligned}\sigma(m_1) \otimes \sigma(m_2) \mapsto &\left(f(m_1)f(m_2) + \frac{1 - f(m_1)}{3}\frac{1 - f(m_2)}{3}\right)\Phi^+ \\ &+ \left(f(m_1)\frac{1 - f(m_2)}{3} + \frac{1 - f(m_2)}{3}f(m_1)\right) \\ &\Phi^- + 2\frac{1 - f(m_1)}{3}\frac{1 - f(m_2)}{3}(\Psi^+ + \Psi^-)\end{aligned} \tag{31}$$

$$\begin{aligned}\mapsto &\left(\frac{10}{9}f(m_1)f(m_2) - \frac{1}{9}(f(m_1) + f(m_2)) + \frac{1}{9}\right)\Phi^+ \\ &+ \left(-\frac{10}{9}f(m_1)f(m_2) + \frac{1}{9}(f(m_1) + f(m_2)) + \frac{8}{9}\right)\frac{1}{3}(\Phi^- + \Psi^+ + \Psi^-),\end{aligned} \tag{32}$$

where in the last line we applied the isotropic twirling map $X \mapsto \text{Tr}[\Phi^+ X]\Phi^+ + \frac{1 - \text{Tr}[\Phi^+ X]}{3}(\Phi^- + \Psi^+ + \Psi^-)$ [ref. 110, Example 7.25], which is required in order to use the BBPSSW protocol recursively.

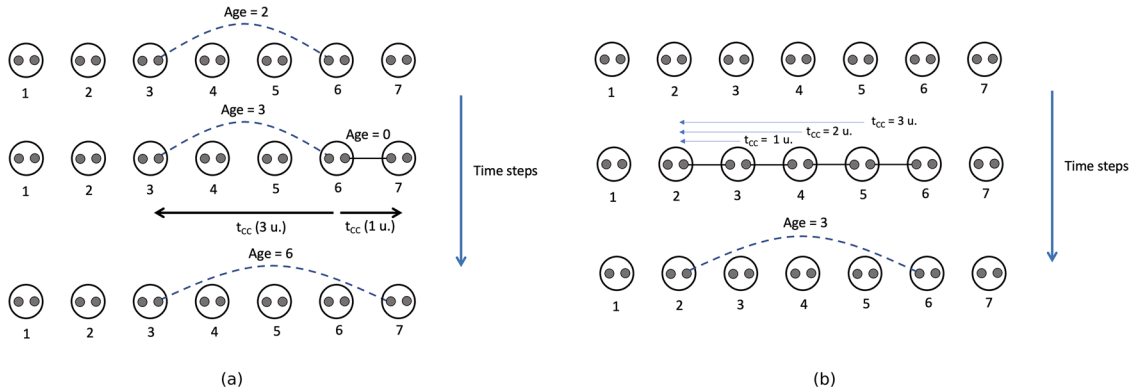

Fig. 12 | Classical communication costs of entanglement swapping. a A generic network evolution trajectory. The CC cost is estimated in our model as the length of the longest link involved in a swap in a given time-step. The swap at node 6 leads to a CC cost of 3 units, where 1 unit is equal to the heralding time for the elementary link. **b** Evolution of a section of the chain with no connections initially. The end nodes stop waiting for CC if there is a gap in the stream of CC packets they receive.

Therefore, after normalization, the state is

$$F_{\text{distill}}(f(m_1),f(m_2))\Phi^+ + \frac{1-F_{\text{distill}}(f(m_1),f(m_2))}{3}(\Phi^- + \Psi^+ + \Psi^-). \tag{33}$$

We observe that this is a state of the form $\sigma(m')$ in (24), for some age m'. We can determine the age m' corresponding to the state after distillation via $m' = f^{-1}(F_{\text{distill}}(f(m_1),f(m_2)))$, with the expression for f^{-1} given in Eq. (26). In particular, the updated age m' is given by:

$$m' = \left\lceil m^* \log\left(\frac{15-6(f(m_1)+f(m_2))+24f(m_1)f(m_2)}{32f(m_1)f(m_2)-2(f(m_1)+f(m_2))-1}\right)\right\rceil. \tag{34}$$

Classical communication costs for local, quasi-local and global policies

In this section, we aim to discuss some of the details related to our model for calculating the classical communication costs for different policies. As was mentioned in "Results", based on the amount of information available to nodes when making policy decisions, policies can be classified as local, quasi-local, and global. We remind the reader that in our model, for all policies, no CC is allowed within one MDP step. Consequently, if at the beginning of an MDP step it is determined that multiple interdependent swaps or distillation attempts are to be made, the failure of one leads to the failure of all dependent swaps or distillation attempts.

We begin by discussing local policies. Distillation, if performed, is always performed at the elementary link level when the links are freshly prepared, and thus no extra CC cost has to be added for distillation. For entanglement swapping, we considered that all the CC cost is paid only at the end of the protocol, since nodes determine their policy decisions or actions based on the locally perceived ages of their links. One concrete implementation through which this can be achieved is the following. Consider a central control or information processing node. All the Bell measurement outcomes (success or failure, and the outcomes obtained in case of success) are continuously transmitted by each node to this central unit. The central node waits until it receives a CC signal indicating a successful swap that leads to end-to-end entanglement generation. When such a signal is received, it relays this information to the end nodes, indicating the end of the protocol, along with the instructions about the list of local operations (Pauli rotations) they must perform to get the final end-to-end virtual link or links. Thus, no node waits for CC from any other node while taking decisions, but the end nodes know when to perform local operations to extract the end-to-end state. End nodes need not know beforehand when to stop, because perceived ages are always lower than the real ages of links, and hence there is no risk of them discarding their links before the CC arrives from the central unit indicating the end of one round of the protocol. Of course, the assumption of an additional central processing node is just an aid to understanding; this role could very well be played by, e.g., the end nodes, or the central node of the chain itself.

For quasi-local policies, we considered that the CC cost for every MDP step is equal to the length (number of nodes) of the longest link involved in a swap multiplied by the elementary link CC time (heralding time). We stress here that this is a simple but useful estimate for the CC cost, and the actual CC cost will depend on the exact time evolution of the network and the concrete protocol used to share and process the classical information emerging from different nodes of the network at different times. Each node sees a constantly evolving picture of the entire network, as new classical communication signals reach it, and thus can decide to make its decisions based on any of these perceived snapshots. A concrete CC policy determination can only be done aided by a quantum network simulator such as those developed in refs. 37,40, etc. The aim of our work is not to replace such a simulator, but to understand the merits and demerits of different classes of policies based on their use of network-state knowledge. Quasi-local policies utilize the fact that each node, using CC, can determine the nodes that it is connected to and the ages of its links. Ranking of links for swaps is done based on this knowledge. Thus, the most important question we must answer is the following: How do links know when to stop anticipating more CC signals and make a decision to perform subsequent swaps?

Consider the example of a generic network state as shown in Fig. 12(a). We ignore distillation in this discussion, the CC cost for which is much easier to estimate and has been explained in "Results". Let us focus on the actions of node 6 and isolate one of the channels between the nodes of the chain. Node 6 knows at the beginning of the MDP step that it is connected to node 3. Now it will try to generate an elementary link with node 7. As soon as such a link is generated, it will perform a swap. Some of the time evolution trajectories that can emerge are as follows.

- Case 1: If the swap at node 6 fails, it will know that either the swap it attempted failed or a swap at node 3 must have been attempted and failed in the meantime. In any case, it is free to restart its memory. As per our model, a swap at node 3 or node 6 will lead to a CC cost of three time steps, because that was the longest link involved in those swaps. Therefore, we overestimate the CC cost in this case. That is, in our policies, we unnecessarily keep node 6 waiting for two extra time steps.
- Case 2: The swap at 6 succeeds, in which case node 6 will know that it must wait for at least three time steps, since this will give it enough time to transmit information to nodes 3 and 7 that they now share a link. In this case, our estimate of CC cost is correct. Of course, node 6 could have started to attempt generating new elementary links and performed subsequent swaps in the meantime, but in our policies we make

the choice to not do so. Node 6 restarts only after the required CC time has elapsed.

- Case 3: In the time that the link between nodes 6 and 7 is produced, if the link between nodes 2 and 3 had also been produced, this information would be available at node 3. Now swaps will be attempted both at node 3 and 6. In this case, the longest link involved in a swap will be between nodes 3 and 7 (or equivalently between 2 and 6), and hence the CC cost is four time steps. This will give enough time for node 3 to communicate to nodes 2 and 7 that they are now connected.

Of course, the question now arises: why should nodes 2 and 7 wait for four time steps and not try to perform swaps as soon as possible? The answer is again similar in spirit to the reasoning above. When the elementary link between 2 and 3 is heralded, 3 can send a simultaneous CC signal to 2 indicating that it already has a connection with node 6 or node 7, as the situation may be at the instant the CC signal is generated. This will inform node 2 that if it attempts a swap and it is successful, this indicates that the swap at node 3 must have also been successful. Thus, it must wait for four time steps to find out whether it connects to node 6 or node 7. Our estimate of CC cost is therefore also correct in this case.

More generally, the cases of network evolution where many interdependent or connected swaps are attempted in the same time step are extremely complex, with many possible time evolution trajectories and different CC costs associated with each. But following the principle that CC can be shared transitively between nodes as soon as a new member joins a connected portion of the chain, many of these complexities can be resolved and a concrete swapping policy can be established. Taking such an approach, and as illustrated via the example above, the CC cost estimated by taking the length of the longest link involved in a swap is an upper bound on the real CC cost.

To elaborate on such a scheme further, we give another example in Fig. 12b. We consider a portion of the network in which there are no links to begin with and show how a connected segment of the chain is established, with each node aware of its connections.

At the beginning of the protocol, each node tries to generate an elementary link. Suppose all of these attempts succeed. Now, each node will attempt swaps simultaneously, because the information about their respective elementary link generation success is available in one time step (heralding time). Now, if one of the swaps fails all of them must fail, and this information will be available at each node, allowing them to restart their memories in the next time step. This will also mark the end of one round of the protocol for this portion of the chain and no extra CC will be required. The other possibility is that all the swaps have succeeded, in which case the internal nodes will be free to restart without the need for any CC. They will just send their measurement outcomes to their next node, that is the node they were connected to at the beginning of the outcome, this CC signal will now be relayed along the chain with more information being added at each node until all this information reaches both the end nodes (see Fig. 12(b)). One of the end nodes can keep utilizing this classical information to perform the local operations necessary to create an end-to-end entangled link. (Which one? We can have a set convention that the node on the left, i.e., one receiving CC signals from the right of it, will do Pauli rotations.) As more and more CC has reached them, they get informed about the success and failure of swaps that occurred farther and farther away, and they get to know that they now share a link with some far-off node that has a certain age. The end nodes will stop waiting for more CC once there is a gap of one time step in the CC packets reaching them. This will indicate that they have received all CC from the connected region of the chain.

Error estimates for Monte-Carlo simulations

All simulation results in this work were produced by performing Monte Carlo simulations. Therefore, all average waiting time and fidelity (end-to-end link age) values have some statistical errors. For most hardware parameter regimes considered in the work, averages of the figures of merit were obtained by performing 20 batches of 1000 runs each. One run is considered as starting from a fully disconnected chain and evolving the network until at least one end-to-end link is generated. The average value reported is the mean of means of all the 20 batches. The standard deviation of this mean (of means) falls with increasing number of batches. The choice of number of batches was made to keep the standard deviation less than 5% of the reported mean values. Therefore, statistical error (due to finite statistics generated by the Monte Carlo simulations) is less than 5% for all values reported in our work. Here we would also like to point out that higher moments of the waiting times and fidelities are also an important figure of merit, e.g., if the variance of the waiting time is very large, having a small average does not guarantee that most runs will take a small time. The full probability distributions can also be obtained using our simulations, but this was not the focus of our work. Existing works have looked at these distributions in detail, for example ref. 26.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data generated through the simulations performed in this work and the experimental data that support the findings of this study are available on request from the corresponding authors.

Code availability

The custom code developed to perform the simulations in this work that support the findings of this study are available on request from the corresponding authors.

Received: 9 May 2024; Accepted: 6 March 2025;
Published online: 01 April 2025

References

1. Wehner, S., Elkouss, D. & Hanson, R. Quantum internet: A vision for the road ahead. *Science* **362** eaam9288 (2018).
2. Dowling, J.*Schrödinger's Web: Race to Build the Quantum Internet* (Taylor & Francis, 2020).
3. Van Meter, R.*Quantum Networking* (John Wiley & Sons, Ltd, 2014).
4. Bennett, C. H. & Brassard, G. Quantum cryptography: Public key distribution and coin tossing. In *International Conference on Computer System and Signal Processing, IEEE*, 175–179 (1984).
5. Bennett, C. H. et al. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.* **70**, 1895–1899 (1993).
6. Xu, F., Ma, X., Zhang, Q., Lo, H.-K. & Pan, J.-W. Secure quantum key distribution with realistic devices. *Rev. Mod. Phys.* **92**, 025002 (2020).
7. Cirac, J. I., Ekert, A. K., Huelga, S. F. & Macchiavello, C. Distributed quantum computation over noisy channels. *Phys. Rev. A* **59**, 4249–4254 (1999).
8. Barz, S. et al. Demonstration of blind quantum computing. *Science* **335**, 303–308 (2012).
9. Ge, W., Jacobs, K., Eldredge, Z., Gorshkov, A. V. & Foss-Feig, M. Distributed quantum metrology with linear networks and separable inputs. *Phys. Rev. Lett.* **121**, 043604 (2018).
10. Komar, P. et al. A quantum network of clocks. *Nat. Phys.* **10**, 582–587 (2014).
11. Proctor, T. J., Knott, P. A. & Dunningham, J. A. Multiparameter estimation in networked quantum sensors. *Phys. Rev. Lett.* **120**, 080501 (2018).
12. Heshami, K. et al. Quantum memories: emerging applications and recent advances. *J. Mod. Opt.* **63**, 2005–2028 (2016).
13. Awschalom, D. et al. Development of quantum interconnects (QuICs) for next-generation information technologies. *PRX Quantum* **2**, 017002 (2021).

14. Pompili, M. et al. Realization of a multi-node quantum network of remote solid-state qubits. *Science* **372**, 259–264 (2021).
15. Hermans, S. L. N. et al. Qubit teleportation between non-neighbouring nodes in a quantum network. *Nature* **605**, 663–668 (2022).
16. Briegel, H.-J., Dür, W., Cirac, J. I. & Zoller, P. Quantum repeaters: The role of imperfect local operations in quantum communication. *Phys. Rev. Lett.* **81**, 5932–5935 (1998).
17. Dür, W., Briegel, H.-J., Cirac, J. I. & Zoller, P. Quantum repeaters based on entanglement purification. *Phys. Rev. A* **59**, 169–181 (1999).
18. Collins, O. A., Jenkins, S. D., Kuzmich, A. & Kennedy, T. A. B. Multiplexed memory-insensitive quantum repeaters. *Phys. Rev. Lett.* **98**, 060502 (2007).
19. Bernardes, N. K., Praxmeyer, L. & van Loock, P. Rate analysis for a hybrid quantum repeater. *Phys. Rev. A* **83**, 012323 (2011).
20. Azuma, K., Tamaki, K. & Lo, H.-K. All-photonic quantum repeaters. *Nat. Commun.* **6** (2015).
21. Rozpędek, F. et al. Parameter regimes for a single sequential quantum repeater. *Quantum Sci. Technol.* **3**, 034002 (2018).
22. Das, S., Khatri, S. & Dowling, J. P. Robust quantum network architectures and topologies for entanglement distribution. *Phys. Rev. A* **97**, 012335 (2018).
23. Khatri, S., Matyas, C. T., Siddiqui, A. U. & Dowling, J. P. Practical figures of merit and thresholds for entanglement distribution in quantum networks. *Phys. Rev. Res.* **1**, 023032 (2019).
24. Rozpędek, F. et al. Near-term quantum-repeater experiments with nitrogen-vacancy centers: Overcoming the limitations of direct transmission. *Phys. Rev. A* **99**, 052330 (2019).
25. Vinay, S. E. & Kok, P. Statistical analysis of quantum-entangled-network generation. *Phys. Rev. A* **99**, 042313 (2019).
26. Shchukin, E., Schmidt, F. & van Loock, P. Waiting time in quantum repeaters with probabilistic entanglement swapping. *Phys. Rev. A* **100**, 032322 (2019).
27. Li, B., Coopmans, T. & Elkouss, D. Efficient optimization of cutoffs in quantum repeater chains. *IEEE Trans. Quantum Eng.* **2**, 1–15 (2021).
28. Vardoyan, G., Guha, S., Nain, P. & Towsley, D. On the exact analysis of an idealized quantum switch. *Perform. Evaluat.* **144**, 102141 (2020).
29. Khatri, S. Policies for elementary links in a quantum network. *Quantum* **5**, 537 (2021).
30. Coopmans, T. *Tools for the design of quantum repeater networks*. Ph.D. thesis, Delft University of Technology. https://repository.tudelft.nl/islandora/object/uuid:90d06f1d-4f23-48cc-8f96-51500258020f?collection=research (2021).
31. Kamin, L., Shchukin, E., Schmidt, F. & van Loock, P. Exact rate analysis for quantum repeaters with imperfect memories and entanglement swapping as soon as possible. *arXiv:2203.10318* (2022).
32. Khatri, S. On the design and analysis of near-term quantum network protocols using Markov decision processes. *AVS Quantum Sci.* **4**, 030501 (2022).
33. Coopmans, T., Brand, S. & Elkouss, D. Improved analytical bounds on delivery times of long-distance entanglement. *Phys. Rev. A* **105**, 012608 (2022).
34. Dai, W. & Towsley, D. Entanglement swapping for repeater chains with finite memory sizes. *arXiv:2111.10994* (2021).
35. Sadhu, A., Somayajula, M. A., Horodecki, K. & Das, S. Practical limitations on robustness and scalability of quantum Internet. *arXiv:2308.12739* (2023).
36. You, X., Li, X., Xu, Y., Feng, H. & Zhao, J. Toward Packet Routing with Fully-distributed Multi-agent Deep Reinforcement Learning. In *2019 International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOPT)*, 1–8 (2019).
37. Coopmans, T. et al. NetSquid, a NETwork Simulator for QUantum Information using Discrete events. *Commun. Phys.* **4**, 164 (2021).
38. Wu, X. et al. SeQUeNCe: a customizable discrete-event simulator of quantum networks. *Quantum Sci. Technol.* **6**, 045027 (2021).
39. Wallnöfer, J., Hahn, F., Wiesner, F., Walk, N. & Eisert, J. ReQuSim: Faithfully simulating near-term quantum repeaters. *arXiv:2212.03896* (2022).
40. Satoh, R. et al. QuISP: a Quantum Internet Simulation Package. In *2022 IEEE International Conference on Quantum Computing and Engineering (QCE)*, 353–364 (2022).
41. da Silva, F. F., Torres-Knoop, A., Coopmans, T., Maier, D. & Wehner, S. Optimizing entanglement generation and distribution using genetic algorithms. *Quantum Sci. Technol.* **6**, 035007 (2021).
42. Bäuml, S., Azuma, K., Kato, G. & Elkouss, D. Linear programs for entanglement and key distribution in the quantum internet. *Commun. Phys.* **3**, 55 (2020).
43. Reiß, S. D. & van Loock, P. Deep reinforcement learning for key distribution based on quantum repeaters. *arXiv:2207.09930* (2022).
44. Álvaro, G. I. ñesta, Vardoyan, G., Scavuzzo, L. & Wehner, S. Optimal entanglement distribution policies in homogeneous repeater chains with cutoffs. *npj Quantum Inf.* **9**, 46 (2023).
45. Haldar, S., Barge, P. J., Khatri, S. & Lee, H. Fast and reliable entanglement distribution with quantum repeaters: Principles for improving protocols using reinforcement learning. *Phys. Rev. Appl.* **21**, 024041 (2024).
46. Sangouard, N., Simon, C., de Riedmatten, H. & Gisin, N. Quantum repeaters based on atomic ensembles and linear optics. *Rev. Mod. Phys.* **83**, 33–80 (2011).
47. Cacciapuoti, A. S. et al. Quantum internet: Networking challenges in distributed quantum computing. *IEEE Netw.* **34**, 137–143 (2020).
48. Cacciapuoti, A. S., Caleffi, M., Van Meter, R. & Hanzo, L. When entanglement meets classical communications: Quantum teleportation for the quantum internet. *IEEE Trans. Commun.* **68**, 3808–3833 (2020).
49. Azuma, K., Bäuml, S., Coopmans, T., Elkouss, D. & Li, B. Tools for quantum network design. *AVS Quantum Sci.* **3**, 014101 (2021).
50. Munro, W. J., Piparo, N. L., Dias, J., Hanks, M. & Nemoto, K. Designing tomorrow's quantum internet. *AVS Quantum Sci.* **4**, 020503 (2022).
51. Illiano, J., Caleffi, M., Manzalini, A. & Cacciapuoti, A. S. Quantum Internet protocol stack: A comprehensive survey. *Comput. Netw.* **213**, 109092 (2022).
52. Azuma, K. et al. Quantum repeaters: From quantum networks to the quantum internet. *Rev. Mod. Phys.* **95**, 045006 (2023).
53. Van Meter, R., Ladd, T. D., Munro, W. J. & Nemoto, K. System design for a long-line quantum repeater. *IEEE/ACM Trans. Netw.* **17**, 1002–1013 (2009).
54. Razavi, M., Thompson, K., Farmanbar, H., Piani, M. & Lütkenhaus, N. Physical and architectural considerations in quantum repeaters. In Arakawa, Y., Sasaki, M. & Sotobayashi, H. (eds.) *Quantum Communications Realized II*, vol. 7236, 18–30. International Society for Optics and Photonics (SPIE, 2009).
55. Muralidharan, S. et al. Optimal architectures for long distance quantum communication. *Sci. Rep.* **6**, 20463 (2016).
56. Jones, C., Kim, D., Rakher, M. T., Kwiat, P. G. & Ladd, T. D. Design and analysis of communication protocols for quantum repeater networks. *N. J. Phys.* **18**, 083015 (2016).
57. Wallnöfer, J., Melnikov, A. A., Dür, W. & Briegel, H. J. Machine learning for long-distance quantum communication. *PRX Quantum* **1**, 010301 (2020).
58. Khatri, S.*Towards a General Framework for Practical Quantum Network Protocols*. Ph.D. thesis, Louisiana State University. https://arxiv.org/abs/2412.20472 (2021).
59. Shchukin, E. & van Loock, P. Optimal entanglement swapping in quantum repeaters. *Phys. Rev. Lett.* **128**, 150502 (2022).

60. Cacciapuoti, A. S., Illiano, J., Viscardi, M. & Caleffi, M. Entanglement Distribution in the Quantum Internet: Knowing when to Stop! *arXiv:2307.05123* (2023).
61. Munro, W. J., Azuma, K., Tamaki, K. & Nemoto, K. Inside quantum repeaters. *IEEE J. Sel. Top. Quantum Electron.* **21**, 78–90 (2015).
62. Chakraborty, K., Rozpędek, F., Dahlberg, A. & Wehner, S. Distributed Routing in a Quantum Internet. *arXiv:1907.11630* (2019).
63. Kolar, A., Zang, A., Chung, J., Suchara, M. & Kettimuthu, R. Adaptive, Continuous Entanglement Generation for Quantum Networks. In *IEEE INFOCOM 2022 - IEEE Conference on Computer Communications Workshops*, 1–6 (2022).
64. Iñesta, A. G. & Wehner, S. Performance metrics for the continuous distribution of entanglement in multiuser quantum networks. *Phys. Rev. A* **108**, 052615 (2023).
65. Ladd, T. D., van Loock, P., Nemoto, K., Munro, W. J. & Yamamoto, Y. Hybrid quantum repeater based on dispersive CQED interactions between matter qubits and bright coherent light. *N. J. Phys.* **8**, 184 (2006).
66. Razavi, M., Piani, M. & Lütkenhaus, N. Quantum repeaters with imperfect memories: Cost and scalability. *Phys. Rev. A* **80**, 032301 (2009).
67. Munro, W. J., Harrison, K. A., Stephens, A. M., Devitt, S. J. & Nemoto, K. From quantum multiplexing to high-performance quantum networking. *Nat. Photonics* **4**, 792–796 (2010).
68. Sheng, Y.-B. & Deng, F.-G. Deterministic entanglement purification and complete nonlocal Bell-state analysis with hyperentanglement. *Phys. Rev. A* **81**, 032307 (2010).
69. Sheng, Y.-B. & Deng, F.-G. One-step deterministic polarization-entanglement purification using spatial entanglement. *Phys. Rev. A* **82**, 044305 (2010).
70. Simon, C., de Riedmatten, H. & Afzelius, M. Temporally multiplexed quantum repeaters with atomic gases. *Phys. Rev. A* **82**, 010304 (2010).
71. Bratzik, S., Abruzzo, S., Kampermann, H. & Bruß, D. Quantum repeaters and quantum key distribution: The impact of entanglement distillation on the secret key rate. *Phys. Rev. A* **87**, 062335 (2013).
72. Sheng, Y.-B., Zhou, L. & Long, G.-L. Hybrid entanglement purification for quantum repeaters. *Phys. Rev. A* **88**, 022302 (2013).
73. 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).
74. Zhou, L., Zhong, W. & Sheng, Y.-B. Purification of the residual entanglement. *Opt. Expr.* **28**, 2291–2301 (2020).
75. Hu, X.-M. et al. Long-distance entanglement purification for quantum communication. *Phys. Rev. Lett.* **126**, 010503 (2021).
76. Dhara, P., Patil, A., Krovi, H. & Guha, S. Subexponential rate versus distance with time-multiplexed quantum repeaters. *Phys. Rev. A* **104**, 052612 (2021).
77. Dhara, P., Linke, N. M., Waks, E., Guha, S. & Seshadreesan, K. P. Multiplexed quantudhara2022multiplexedm repeaters based on dual-species trapped-ion systems. *Phys. Rev. A* **105**, 022623 (2022).
78. Huang, C.-X. et al. Experimental one-step deterministic polarization entanglement purification. *Sci. Bull.* **67**, 593–597 (2022).
79. Ecker, S., Sohr, P., Bulla, L., Ursin, R. & Bohmann, M. Remotely Establishing Polarization Entanglement Over Noisy Polarization Channels. *Phys. Rev. Appl.* **17**, 034009 (2022).
80. Chakraborty, T. et al. Towards a spectrally multiplexed quantum repeater. *arXiv:2205.10028* (2022).
81. Milligen, E. A. V. et al. Entanglement Routing over Networks with Time Multiplexed Repeaters. *arXiv:2308.15028* (2023).
82. Zang, A. et al. Entanglement distribution in quantum repeater with purification and optimized buffer time. In *IEEE INFOCOM 2023 - IEEE Conference on Computer Communications Workshops*, 1–6 (2023).
83. Żukowski, M., Zeilinger, A., Horne, M. A. & Ekert, A. K. 'Event-ready-detectors' Bell experiment via entanglement swapping. *Phys. Rev. Lett.* **71**, 4287–4290 (1993).
84. Bennett, C. H. et al. Purification of noisy entanglement and faithful teleportation via noisy channels. *Phys. Rev. Lett.* **76**, 722–725 (1996).
85. Bennett, C. H., Bernstein, H. J., Popescu, S. & Schumacher, B. Concentrating partial entanglement by local operations. *Phys. Rev. A* **53**, 2046–2052 (1996).
86. Bennett, C. H., DiVincenzo, D. P., Smolin, J. A. & Wootters, W. K. Mixed-state entanglement and quantum error correction. *Phys. Rev. A* **54**, 3824–3851 (1996).
87. Yan, P.-S., Zhou, L., Zhong, W. & Sheng, Y.-B. Advances in quantum entanglement purification. *Sci. China Phys., Mech. Astron.* **66**, 250301 (2023).
88. Matsuo, T., Durand, C. & Van Meter, R. Quantum link bootstrapping using a ruleset-based communication protocol. *Phys. Rev. A* **100**, 052320 (2019).
89. Matsuo, T. Simulation of a dynamic, RuleSet-based quantum network. *arXiv:1908.10758* (2019).
90. Rozpędek, F., Seshadreesan, K. P., Polakos, P., Jiang, L. & Guha, S. All-photonic Gottesman-Kitaev-Preskill–qubit repeater using analog-information-assisted multiplexed entanglement ranking. *Phys. Rev. Res.* **5**, 043056 (2023).
91. Kumar, V., Chandra, N. K., Seshadreesan, K. P., Scheller-Wolf, A. & Tayur, S. Optimal entanglement distillation policies for quantum switches. *2023 IEEE Int. Conf. Quantum Comput. Eng. (QCE)* **01**, 1198–1204 (2023).
92. Azari, M., Polakos, P. & Seshadreesan, K. P. Quantum switches for Gottesman-Kitaev-Preskill qubit-based all-photonic quantum networks. In *IEEE Transactions on Quantum Engineering* vol. 5, pp. 1–15 (2024).
93. Chang, K.-C. et al. 648 Hilbert-space dimensionality in a biphoton frequency comb: entanglement of formation and Schmidt mode decomposition. *npj Quantum Inf.* **7**, 48 (2021).
94. Chang, K.-C., Cheng, X., Sarihan, M. C. & Wong, C. W. Time-reversible and fully time-resolved ultra-narrowband biphoton frequency combs. *APL Quantum* **1**, 016106 (2024).
95. Xie, Z. et al. Harnessing high-dimensional hyperentanglement through a biphoton frequency comb. *Nat. Photonics* **9**, 536–542 (2015).
96. Cheng, X. et al. High-dimensional time-frequency entanglement in a singly-filtered biphoton frequency comb. *Commun. Phys.* **6**, 278 (2023).
97. Chang, K.-C., Cheng, X., Sarihan, M. C. & Wong, C. W. Towards optimum franson interference recurrence in mode-locked singly-filtered biphoton frequency combs. *Photonics Res.* **11**, 1175 (2023).
98. Sarihan, M. C. et al. High dimensional quantum key distribution with biphoton frequency combs through energy-time entanglement. In *Conference on Lasers and Electro-Optics* (OSA, 2019).
99. Cheng, X. et al. Secure high dimensional quantum key distribution based on wavelength-multiplexed time-bin encoding. In *Conference on Lasers and Electro-Optics* (Optica Publishing Group, 2021).
100. Sarihan, M. C., Cheng, X., Chang, K.-C. & Wong, C. W. Wavelength-multiplexed multi-user quantum network based on high-dimensional time-bin encoding. In *CLEO 2023* (Optica Publishing Group, 2023).
101. Chang, K.-C., Sarihan, M. C., Cheng, X., Zhang, Z. & Wong, C. W. Large-alphabet time-bin quantum key distribution and einstein-podolsky-rosen steering via dispersive optics. *Quantum Sci. Technol.* **9**, 015018 (2023).
102. Chang, K.-C. et al. High-dimensional Energy-time Entanglement Distribution via a Biphoton Frequency Comb. In *Conference on Lasers and Electro-Optics* (Optica Publishing Group, 2021).

103. Meng, R.-R. et al. Solid-state quantum nodes based on color centers and rare-earth ions coupled with fiber fabry-pérot microcavities. *Chip* 100081 (2024).
104. Lei, Y. et al. Quantum optical memory for entanglement distribution. *Optica* **10**, 1511 (2023).
105. Goodenough, K. et al. Near-term *n* to *k* distillation protocols using graph codes. *arXiv:2303.11465* (2023).
106. Sarvepalli, P. K., Klappenecker, A. & Rötteler, M. Asymmetric quantum codes: constructions, bounds and performance. *Proc. R. Soc. A: Math. Phys. Eng. Sci.* **465**, 1645–1672 (2009).
107. Ghosh, J., Fowler, A. G. & Geller, M. R. Surface code with decoherence: An analysis of three superconducting architectures. *Phys. Rev. A* **86**, 062318 (2012).
108. Ourari, S. et al. Indistinguishable telecom band photons from a single Er ion in the solid state. *Nature* **620**, 977–981 (2023).
109. Zang, A., Chen, X., Chitambar, E., Suchara, M. & Zhong, T. No-Go Theorems for Universal Entanglement Purification. *arXiv:2407.21760* (2024).
110. Watrous, J.*The Theory of Quantum Information* (Cambridge University Press, 2018).
111. Sukachev, D. D. et al. Silicon-vacancy spin qubit in diamond: A quantum memory exceeding 10 ms with single-shot state readout. *Phys. Rev. Lett.* **119**, 223602 (2017).
112. Hedges, M. P., Longdell, J. J., Li, Y. & Sellars, M. J. Efficient quantum memory for light. *Nature* **465**, 1052–1056 (2010).

Acknowledgements

We thank Thomas Searles and Sanjaya Lohani for helpful discussions. This work was supported by the Army Research Office Multidisciplinary University Research Initiative (ARO MURI) through the grant number W911NF2120214. S.H. also acknowledges support from the RCS program of Louisiana Boards of Regents through the grant LEQSF(2023-25)-RD-A-04. P.B. and H.L. also acknowledge the support the US-Israel Binational Science Foundation. S.K. acknowledges financial support from the German BMBF (Hybrid).

Author contributions

S.H., S.K., H.L., and P.B. conceived the main idea of the work. S.K., P.B., and S.H. developed the theory and performed the computations. S.K. and S.H. performed the analytical calculations. X.C. and K.C. performed the experiments supervised by C.W. B.K. encouraged S.H. to investigate the role of classical communication. S.K., B.K., C.W., and H.L. supervised the findings of this work. All authors discussed the results and contributed to the writing of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42005-025-02029-w.

Correspondence and requests for materials should be addressed to Stav Haldar.

Peer review information *Communications Physics* thanks Naphan Benchasattabuse and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025