#### **PAPER**

## Quantum repeater architecture with hierarchically optimized memory buffer times

To cite this article: Siddhartha Santra et al 2019 Quantum Sci. Technol. 4 025010

View the article online for updates and enhancements.

#### Recent citations

- Waiting time in quantum repeaters with probabilistic entanglement swapping E. Shchukin *et al*
- Quantum repeaters based on two species trapped ions
   Siddhartha Santra et al



## IOP ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

### **Quantum Science and Technology**



#### RECEIVED

2 November 2018

REVISED
21 February 2019

ACCEPTED FOR PUBLICATION

1 March 2019

PUBLISHED 19 March 2019

#### **PAPER**

# Quantum repeater architecture with hierarchically optimized memory buffer times

Siddhartha Santra<sup>1</sup>, Liang Jiang<sup>2,3</sup> and Vladimir S Malinovsky<sup>1</sup>

- <sup>1</sup> US Army Research Laboratory, Adelphi, MD 20783, United States of America
- <sup>2</sup> Departments of Applied Physics and Physics, Yale University, New Haven, CT 06520, United States of America
- Yale Quantum Institute, Yale University, New Haven, CT 06511, United States of America

E-mail: sidsantra1@@gmail.com

Keywords: quantum repeaters, memory decoherence, optimized architecture

#### Abstract

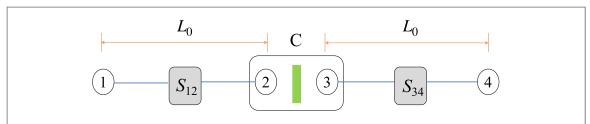
We propose a quantum repeater protocol and architecture that mitigates decoherence of the entangled states by optimizing the quantum memory buffer time. The protocol maximizes the rate of distillable entanglement in the average accessed state at all nesting levels. The achievable rate is higher by orders of magnitude in comparison to a canonical protocol that does not optimize the buffer time. The advantage of the proposed design is observed for all nesting levels of the repeater for technologically feasible memory quality, entanglement generation and swapping success probabilities.

#### 1. Introduction

Spatially distributed entanglement is a valuable resource for quantum communication, computing and sensing [1]. Quantum repeaters (QR) follow a nested divide and conquer strategy to distribute entanglement across large distances [2, 3]. At each nesting level, first, entangled states are generated probabilistically over smaller segments and stored in quantum memories at repeater stations. Second, a swapping operation on the memories doubles the physical range of the entangled state. As states make their way up the levels they spend some time, the memory buffer time, in the decohering quantum memories before being discarded or accessed for use by the next level. A larger buffer time at any nesting level increases the probability to obtain an entanglement length-doubled state but decreases the entanglement quality of the average obtained state due to decoherence. These competing factors determine the entanglement generation rate (EGR) which is the product of the rate of obtaining entanglement length-doubled states and the entanglement of the average obtained state. An optimal buffer time maximizes the EGR. However, most protocols for repeater operation [4] ignore the optimality of buffer time arising due to the interplay of entanglement generation probability and quantum memory decoherence [5].

In practice, it is crucial to include quantum memory decoherence for QR that rely on two-way communication over long distances as shown in [6]. The same reference suggests using decoherence free subspaces or local encoding and repeater operation in blind-mode to suppress memory errors. Other interesting ideas to address this challenge, e.g. addition of more physical resources such as multiplexed quantum memory to reduce memory waiting time [7]; or more complicated operations such as quantum error correction to actively suppress all errors [8–10] are promising in the long term but still very challenging with current experimental capability. Besides asking for more physical resources or complicated operations, it is also important to optimize the parameters of QR protocols. For example, dynamic programming has been introduced to explore the huge parameter space of QR protocols, which can successfully identify efficient protocols with significantly boosted performance in the absence of memory decoherence [11]. So far, there is no efficient method that can include quantum memory decoherence and systematically optimize the design parameters of QR protocols.

Here, we propose an optimized buffer time protocol (OBP) and architecture that maximizes the EGR using hierarchically optimized buffer times for all nesting levels. The optimal buffer time depends on the parameters of quantum memory quality,  $\beta$ , the entanglement generation probability, p, and the swapping success probability,



**Figure 1.** A quantum repeater with two segments at nesting level 1. Quantum memory pairs 1, 2 and 3, 4 store entangled states produced by sources  $S_{12}$  and  $S_{34}$  for entanglement swapping, green rectangle, at repeater station C.

 $p_S$ . The minimal parametrization chosen in terms of  $(p, \beta, p_S)$  subsumes implementation-specific details such as source-station geometry, coupling and conversion efficiences or the use of multiplexed memories etc (Entanglement generation probability p, for example, can include source-fiber coupling, wavelength conversion and memory read-in efficiency. Memory read-out may be included in p or swapping success probability  $p_S$ .) We compare the OBP to a canonical repeater protocol (CP) that does not optimize the buffer time and show that the OBP improves the EGR by several orders of magnitude in the technologically relevant parameter region. Moreover, we show that the relative improvement due to the OBP increases with the nesting level for technologically feasible swapping success probability. The protocol works for finite-lifetime quantum memories used to store entangled states in QR which utilize two-way classical communication between its nodes to verify entanglement generation before entanglement swapping is performed.

The layout of the paper is as follows. Section 2 first describes the central idea of the optimized memory buffer time protocol in section 2.1 followed by the definition of optimal memory buffer time in section 2.2 and comparison with a canonical protocol in section 2.3. Section 3 presents a QR architecture compatible with hierarchical optimization of buffer times in section 3.1. Section 3.2 then describes an algorithm that can be used for the hierarchical optimization. Further, section 3.3 shows the comparison of the EGRs for the OBP compared to the CP for all nesting levels. Section 4 concludes the paper with a discussion of the typical advantage one may expect using OBP when used with state of art parameters.

#### 2. QR protocol with optimized memory buffer time

In this section we describe the optimized memory buffer time protocol by focusing on the first nesting level, in section 2.1, of a potentially multi-level QR network. While quantum memories may also suffer from decoherence due to depolarization and loss, we consider dephasing as the only mode of memory decoherence. This highlights the central physical idea while keeping the discussion mathematically simple. The optimal buffer time is described in section 2.2. A comparison of EGRs of the optimized and canonical protocols is presented in section 2.3.

We term a repeater protocol that does not optimize its quantum memory buffer time as a canonical protocol, for example, those in [5] and [6]. As with the optimized protocol, in canonical protocols entanglement generation and swapping occur probabilistically. For comparison with the optimized protocol, the distinctive feature of canonical protocols is that the quantum memories can wait for arbitrarily long times for successful entanglement generation. The rate of entanglement generation in such protocols is inversely proportional to the expected number of entanglement generation attempts needed for success, as shown in appendix A. Subsequent to entanglement generation, for both the canonical and optimized protocols, purification of the generated entangled pairs may be performed if multiple quantum memories at a given nesting level are available at the repeater nodes. In this paper, we compare the protocols without considering purification of entangled states on a finite number of quantum memories. Thus, we compare the two protocols on a single-copy basis and use the distillable entanglement of the average state in the respective protocols as a measure of the entanglement quality.

#### 2.1. Optimized memory buffer time protocol at the first nesting level

Operationally, the optimized memory buffer time protocol can be understood by considering the entanglement swapping of states across two elementary segments at the first nesting level in a QR, figure 1. Sources  $S_{12}$  and  $S_{34}$  supply entangled states with probability, p, to the decohering quantum memory pairs (1, 2) and (3, 4). The memory lifetime is denoted by  $\tau_M$ . Entanglement swapping at the repeater station, C, is performed via informed Bell-state measurements. C checks the two pairs of memories verifying whether they are charged which requires waiting for one unit of one-way classical communication time,  $\tau_C = L_0/c$ , with c the speed of light in the fiber and c0 the length of the segment. If both pairs are charged, c0 performs a swapping operation on the memories 2 and 3 with success probability, c0, producing an entangled state across the remote memories 1 and 4.

Information about the success or failure of the swapping operation is then communicated to the remote memories taking an additional time  $\tau_C$ . In case C finds the memory pairs uncharged (one or both) it classically communicates the need to continue entanglement generation attempt in the segment(s) to the remote memories which also requires  $\tau_C$  amount of time. The OBP limits the number of such entanglement generation attempts to a number  $n_{\rm opt}(p,\beta,p_S)$ , determined by the operating parameters, after which the state from the remote memories is accessed. Subsequently all four memories are refreshed and the entanglement generation process starts over. The CP on the other hand places no limit on the number of entanglement generation attempts which continue till a state is obtained in both segments.

By limiting the buffer time, the OBP provides an average remote entangled state with a high measure of entanglement since the time for memory decoherence is limited. The entanglement generation in the two segments can succeed at step numbers,  $1 \le k_1$ ,  $k_2 \le n$ , where n is the maximum number of attempts and the probability distribution of successful entanglement generation is  $\mathbb{P}(k_1, k_2) = (1 - p)^{k_1 + k_2 - 2} p^2$ . Without loss of generality, we assume that the memory pairs in the two segments of figure 1 are supplied with the state,  $\rho^- = |\psi^-\rangle \langle \psi^-|$ , where,  $|\psi^\pm\rangle = (|01\rangle \pm |10\rangle)/\sqrt{2}$ . Storing the  $\rho^-$  state in a pair of quantum memories with lifetime  $\tau_M$  for time t results in the state,  $\rho(t) = \rho^-(1 + e^{-2t/\tau_M})/2 + \rho^+(1 - e^{-2t/\tau_M})/2$ , where  $\rho^+ = |\psi^+\rangle \langle \psi^+|$ . The remote state obtained after a successful swap and communication to the remote memories 1 and 4 is

$$\rho^{S}(k_{1}, k_{2}) = \frac{1}{2} (1 + \beta^{|\Delta k|+2}) \rho^{-} + \frac{1}{2} (1 - \beta^{|\Delta k|+2}) \rho^{+}, \tag{1}$$

where  $\beta = \mathrm{e}^{-2\tau_C/\tau_M}$  is the memory quality parameter that quantifies the decoherence in a pair of quantum memories during one round of one-way classical communication, and  $\Delta k = (k_2 - k_1)$ .  $\rho^S(k_1, k_2)$  approaches the totally mixed state exponentially fast with  $\Delta k$ . Thus, the states for large  $\Delta k$  contribute little to the entanglement of the average state. The state in equation (1) further decoheres in the remote memories for a time,  $t = 2\tau_C(n - \max(k_1, k_2))$ , before being accessed, and leads to the state  $\rho_n^S(k_1, k_2) = \rho^-(1 + \beta^{|\Delta k| + 2 + n - \max(k_1, k_2)})/2 + \rho^+(1 - \beta^{|\Delta k| + 2 + n - \max(k_1, k_2)})/2$ . The average remote entangled state is the probabilistically weighted sum of such states,  $\rho^O = \mathcal{N}^{-1}\sum_{k_1=1}^{n,n} k_2 = 1 \mathbb{P}(k_1, k_2) \rho_n^S(k_1, k_2)$ , where  $\mathcal{N} = (1 - (1 - p)^n)^2$  is the total probability of obtaining a remote entangled state across the two segments in n attempts. The average remote entangled state can be expressed in a compact manner as (see appendix A),

$$\rho^{O} = \frac{1}{2} (1 + \gamma^{O}(p, \beta, n)) \rho^{-} + \frac{1}{2} (1 - \gamma^{O}(p, \beta, n)) \rho^{+}, \tag{2}$$

which has a fidelity of  $F^O(p, \beta, n) = \operatorname{tr}\{\rho^-\rho^O\} = \frac{1}{2}(1 + \gamma^O(p, \beta, n))$  and is obtained once every  $n2\tau_C$  period of time. The function  $\gamma^O(p, \beta, n) \in [0, 1]$  can be physically interpreted as the degradation in fidelity of the average state due to memory decoherence during the buffer time. For perfect quantum memories,  $\beta = 1$ , and  $\gamma^O(p, \beta, n) = 1$ .

#### 2.2. Optimal memory buffer time

As an entanglement measure for the mixed state,  $\rho^O$ , we use the upper bound on its distillable entanglement,  $E[F^O] = H\left[\frac{1}{2} + (F^O(1-F^O))^{.5}\right]$  for  $1 \ge F^O > 0.5$  and  $E[F^O] = 0$  for  $.5 \ge F^O \ge 0$ , where  $H[x] = -x\log(x) - (1-x)\log(1-x)$  is the binary entropy function.  $E[F^O]$  expresses the number of pure Bell states that the best distillation protocol can achieve in the limit of asymptotic number of copies. For  $\rho^O$  the bound can be achieved using the hashing protocol [12]. The EGR is thus given by the rate of distillable entanglement

$$R_{\rm DE}^{O}(p,\,\beta,\,n) = \frac{p_{\rm S}(1-(1-p)^n)^2}{n(2\tau_{\rm C})} E[F^{O}(p,\,\beta,\,n)]. \tag{3}$$

The optimal buffer time,  $n_{\text{opt}}$ , maximizes this rate for given p and  $\beta$  values, i.e.

$$n_{\text{opt}}(p, \beta) = \text{ArgMax}_n[R_{\text{DE}}^O(p, \beta, n)].$$
 (4)

with the obvious condition that  $n_{\rm opt}(p,\,\beta)\geqslant 1$ . Note that the optimal buffer time found using (4) is obtained in units of the two-way classical communication time  $2\tau_C$ . The behavior of the optimal buffer time,  $n_{\rm opt}(p,\,\beta)$ , in different regions of the  $(p,\,\beta)$  parameter space is described in appendix B. While the optimal buffer time at the first nesting level depends only on the entanglement generation probability p and the memory quality parameter  $\beta$ , for higher nesting levels it depends also on the swapping success probability of the previous level  $p_S$ . In this paper, for simplicity, we assume that the parameters  $p,\,\beta,\,p_S$  remain constant for all nesting levels. However, our analysis outlined in section 3 can be used to address various distributions of the parameters. In case of multiplexed quantum memories [7], expression (4) can be used to determine the optimal buffer time by using the effective entanglement generation probability between the nodes. Also, note that the asymptotic value of the fidelity,  $F^O(p,\,\beta,\,n) = \frac{1}{2}(1\,+\,\gamma^O(p,\,\beta,\,n))$ , is at least 0.5 when dephasing is the only mode of decoherence.

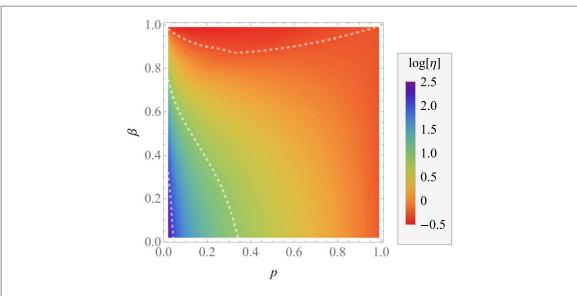


Figure 2. Logarithm (base 10) of the ratio of entanglement generation rate in the optimized memory buffer time protocol,  $R_{DE}^O$ , to that in the canonical protocol,  $R_{DE}^O$ , at the first nesting level. Dotted white lines are contours with  $\log[\eta] = 2$ , 1, 0 from left to right.

Indeed, dephasing is the dominant mode of decoherence for repeater-relevant timescales in quantum memories based on nuclear spins in diamond NV centers and the hyperfine electron levels in ion traps [13]. However, when loss and depolarization are also considered, the asymptotic fidelity can fall below the distillable entanglement threshold of  $F^O > 0.5$ . The optimized memory buffer time protocol works in this general case as well but now the maximum memory buffer time,  $n_{\text{max}}$  is limited by the threshold condition,  $F^O(p, \beta, n_{\text{max}}) > 0.5$ .

#### 2.3. EGR comparison of the optimized and canonical protocols

To compare the EGRs of the optimized and canonical protocols we next obtain the average remote entangled state of the canonical protocol. The average remote entangled state in the canonical protocol has a low measure of entanglement since it is an average over states that have decohered in the memories for arbitrarily long times. The average state,  $\rho^C = \sum_{k=1,k=1}^{\infty,\infty} \mathbb{P}(k_1, k_2) \rho^S(k_1, k_2)$ , again takes a compact form (see appendix A)

$$\rho^{C} = \frac{1}{2} (1 + \gamma^{C}(p, \beta)) \rho^{-} + \frac{1}{2} (1 - \gamma^{C}(p, \beta)) \rho^{+}.$$
 (5)

Such states of fidelity  $F^C(p, \beta) := \operatorname{tr}\{\rho^- \rho^C\} = \frac{1}{2}(1 + \gamma^C(p, \beta))$  are obtained at the rate of the inverse of the waiting time  $\langle k \rangle = (3 - 2p)/p(2 - p)$  [14]. The EGR in the canonical protocol is

$$R_{\rm DE}^{C}(p,\,\beta) = \frac{p_{\rm S}}{\langle k \rangle (2\tau_{C})} E[F^{C}(p,\,\beta)]. \tag{6}$$

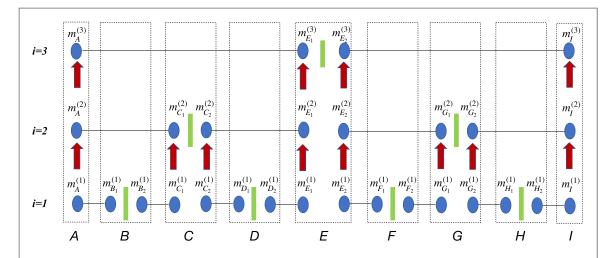
The optimized buffer time protocol provides manifold increase of EGRs in most of the  $(p,\beta)$ -parameter space, even at the first nesting level, as shown in figure 2. In particular, for the low  $p,\beta$ -region the ratio,  $\eta(p,\beta,n_{\rm opt})=R_{\rm DE}^O(p,\beta,n_{\rm opt})/R_{\rm DE}^C(p,\beta)\sim 1/p$  (see appendix C). Only for  $\beta\sim 1$ , i.e. for near-perfect quantum memories, does the canonical protocol provide better rates. The optimal buffer time,  $n_{\rm opt}$ , depends on the operating point in parameter space. For short-lived quantum memories,  $\beta\ll 1$ , it is numerically found that  $n_{\rm opt}=1$ . For long-lived quantum memories,  $\beta\to 1$ , and low entanglement generation probability,  $p\to 0$ , the optimal buffer time scales as,  $n_{\rm opt}\sim p^{-1}\log(1/\beta)^{-1}=(1/p)(\tau_M/2\tau_C)$ .

#### 3. Hierarchical buffer time optimization-compliant repeater architecture

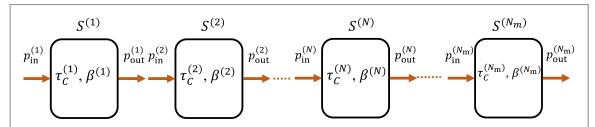
We now present a repeater architecture which can operate all its nesting levels based on the optimized memory buffer time protocol in section 3.1. This is followed by a description of the algorithm to hierarchically optimize the buffer time in section 3.2. The section ends by presenting a comparison of the EGRs of the optimized and canonical protocols for all nesting levels in section 3.3.

#### 3.1. Optimization-compliant architecture for all nesting levels

A QR architecture capable of supporting hierarchical optimization of buffer times requires a new set of quantum memories at each nesting level as shown in figure 3. The average remote entangled state output by nesting level i is transferred to the new quantum memories at level (i + 1) for  $i = 1, 2, ..., (N_m - 1)$ , where  $N_m$  is the



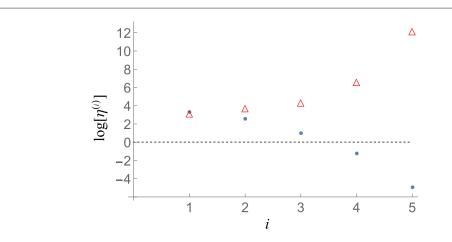
**Figure 3.** Quantum repeater architecture based on the optimized buffer time protocol. Shown are three nesting levels (i = 1, 2, 3) and 9 repeater nodes (A - I). A new set of quantum memories (blue ovals) are required at each nesting level. Green rectangles represent entanglement swapping operations. Entanglement length-doubled states obtained at any level are transferred to the quantum memories of the next level using coherent operations (red arrows).



**Figure 4.** Organization of nesting levels in a quantum repeater architecture as a sequence of self-similar systems. Each system  $S^{(i)}$  is characterized by the one-way classical communication time  $\tau_C^{(i)}$  and its memory quality parameter  $\beta^{(i)}$ .  $S^{(i)}$  receives a state  $\rho_{\text{in}}^{(i)}$  with probability  $p_{\text{in}}^{(i)}$  every  $n_{\text{in}}^{(i)} \tau_C^{(i)}$  amount of time from the preceding system  $S^{(i-1)}$ .  $S^{(i)}$  outputs a state  $\rho_{\text{out}}^{(i)}$  with probability  $p_{\text{out}}^{(i)}$  every  $n_{\text{out}}^{(i)}$  of its input cycles.

maximum nesting level. This transfer can be achieved by using a two-qubit quantum SWAP gate [15]. In figure 3 the quantum memories are labeled by the nesting level as superscript and the node label as subscript, for example,  $m_{E_1}^{(1)}$  denotes quantum memory number 1 at the first nesting level of node E. The memories at any level follow the OBP with an optimal buffer time that is determined by the effective probability with which it receives entangled states and memory quality parameter relative to its classical communication time. All levels follow the informed Bell-state measurement procedure followed by communication to the remote memories just as in the first nesting level described earlier. The state output by a nesting level is therefore the probabilistically weighted sum of the average state received from the previous level. Periodic SWAP operations between two quantum memories at a node, for example, between  $m_{E_1}^{(1)} \rightarrow m_{E_1}^{(2)} \rightarrow m_{E_1}^{(3)}$  in figure 3, are used to feed forward the average state to the quantum memories of the higher nesting level.

The nesting levels in such an architecture can be modeled as a sequence of self-similar input-output systems,  $\{S^{(i)}\}_i$ ,  $i \in \{1, 2, ..., N_m\}$ , as shown in figure 4. Each system is characterized by its classical communication time  $\tau_C^{(i)} = 2^{i-1}\tau_C$  and memory quality parameter  $\beta^{(i)} = \beta^{2^{i-1}}$ . Further, for each system  $S^{(i)}$  the input-cycle time  $n_{\text{in}}^{(i)}$  specifies the number of two-way classical communication cycles over which it receives one average state from the previous system with probability  $p_{\text{in}}^{(i)}$ . This takes  $n_{\text{in}}^{(i)} 2\tau_C^{(i)}$  amount of time. The output-cycle time  $n_{\text{out}}^{(i)}$  is the buffer time for system  $S^{(i)}$  in terms of the number of input-cycles for system  $S^{(i)}$ , i.e. one average state is output by the system in  $n_{\text{out}}^{(i)} n_{\text{in}}^{(i)} 2\tau_C^{(i)}$  amount of time with probability  $p_{\text{out}}^{(i)}$ . Any system  $S^{(i)}$  is able to receive or output an average state only at the end of a time period that is a multiple of its two-way classical communication cycle time. Two adjacent systems  $S^{(i)}$  and  $S^{(i+1)}$  are synchronized if the physical times at which the ith system outputs its average state corresponds to the physical times at which the (i+1)th system can receive the state. Therefore, successive input-cycle times and output-cycle times have to obey the condition for synchronization of the



**Figure 5.** Logarithm (base 10) of the ratio of entanglement generation rate for the optimized buffer time protocol and the canonical protocol at different nesting levels for p=0.02,  $\beta=0.2$ ,  $p_T=1$  and two different values of  $p_S=0.75$  (blue dots) and  $p_S=0.5$  (red triangles). The buffer times of the nesting levels were approximately optimal in both cases.

systems,  $n_{\text{out}}^{(i)} n_{\text{in}}^{(i)} 2\tau_C^{(i)} = n_{\text{in}}^{(i+1)} 2\tau_C^{(i+1)}$ , which implies

$$n_{\rm in}^{(i+1)} = n_{\rm out}^{(i)} n_{\rm in}^{(i)} / 2,$$
 (7)

for  $i=1,\ldots,N_m-1$ , so that  $n_{\rm in,out}^{(i)}$  are positive integers, and  $n_{\rm in}^{(1)}=1$ . The output probability of system  $S^{(i)}$  is related to its input probability as  $p_{\rm out}^{(i)}=p_s[1-(1-p_{\rm in}^{(i)})^{\eta_{\rm out}^{(i)}}]^2$  with  $p_{\rm in}^{(1)}=p$ . While the input probability of system  $S^{(i)}$  is related to the output probability of system  $S^{(i-1)}$  as  $p_{\rm in}^{(i)}=p_Tp_{\rm out}^{(i-1)}$ , where  $p_T$  is the probability to successfully transfer states from the memories of one nesting level to the next, and  $p_{\rm out}^{(0)}=p$  to maintain consistency.

The average remote entangled state  $\rho^{O,\,(i)}$  obtained in the OBP based architecture at any nesting level i depends on the parameter values for the preceding levels, i.e. on  $\{\tau_C^{(j)},\,\beta_C^{(j)},\,n_{\rm in}^{(j)}\}$  for  $1\leqslant j\leqslant i$ . In addition, it depends on the initial entanglement generation probability, p, the swapping success probability,  $p_S$ , and the transfer success probability,  $p_T$ . This state

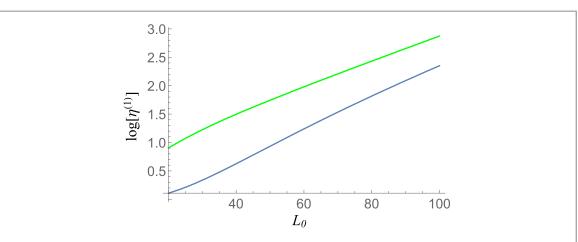
$$\rho^{O,(i)} = \frac{1 + \prod_{j=1}^{i} \left[ \gamma^{O,(j)} \right]^{2^{i-j}}}{2} \rho^{-} + \frac{1 - \prod_{j=1}^{i} \left[ \gamma^{O,(j)} \right]^{2^{i-j}}}{2} \rho^{+}, \tag{8}$$

is obtained once every  $n_{\rm in}^{(i)} n_{\rm out}^{(i)} 2 \tau_C^{(i)}$  amount of time.  $\gamma^{O,\,(\,j)}$  as a function of the relevant parameters can be found in the supplementary material. Physically,  $\gamma^{O,(j)} \leqslant 1$  can be understood as the degradation of the fidelity during the buffer time at nesting level j.  $\rho^{O,\,(i)}$  has a fidelity  $F^{O,(i)} = \operatorname{tr}\{\rho^-\rho^{O,(i)}\} = (1+\prod_{j=1}^i [\gamma^{O,(j)}]^{2^{i-j}})/2$ . This fidelity approaches the distillation threshold of 1/2 as  $(1/2)\prod_{j=1}^i [\gamma^{O,(j)}]^{2^{i-j}}$ , which accounts for the degradation in all nesting levels prior to i. Therefore, the EGR of system  $S^{(i)}$  is

$$R_{\rm DE}^{O,(i)} = \frac{p_{\rm out}^{(i)}}{n_{\rm in}^{(i)} n_{\rm out}^{(i)} 2\tau_C^{(i)}} E \left[ \frac{1 + \prod_{j=1}^i \left[ \gamma^{O,(j)} \right]^{2^{i-j}}}{2} \right]. \tag{9}$$

#### 3.2. Hierarchical optimization of memory buffer time at every nesting level

In a repeater with  $N_m$  levels the final EGR can be maximized by hierarchically optimizing buffer times of all nesting levels. We search over the set of positive integers  $\{n_{\rm in}^{(i)}, n_{\rm out}^{(i)}\}$ , subject to synchronization condition constraints, equation (7). The optimization proceeds by maximizing the EGR sequentially starting from nesting level 1. If at any nesting level, i, the optimal value for  $n_{\rm out,opt}^{(i)} = {\rm ArgMax}_{n_{\rm out}^{(i)}}[R_{\rm DE}^{O,(i)}]$  does not satisfy the synchronization condition one of the neighboring values,  $\tilde{n}_{\rm out,opt}^{(i)} \in \{n_{\rm out,opt}^{(i)} - 1, n_{\rm out,opt}^{(i)} + 1\}$  whichever provides a higher  $R_{\rm DE}^{O,(i)}$ , is chosen and used for calculating  $n_{\rm in}^{(i+1)}$  and  $p_{\rm in}^{(i+1)}$ . This procedure is followed for all nesting levels upto  $i = N_m$  and produces an approximately-optimal synchronized sequence of  $n_{\rm in}^{(i)}$  and  $n_{\rm out}^{(i)}$ . A repeater operating its nesting levels based on the sequence of approximately-optimal buffer times still gives manifold increase of EGR as shown by the logarithm of the ratio of the rates at nesting level i,  $\log[\eta^{(i)}(p,\beta,n)] = \log[R_{\rm DE}^{O,(i)}(p,\beta,n)/R_{\rm DE}^{C,(i)}(p,\beta)]$ , in figure 5.



**Figure 6.** Logarithm (base 10) of the ratio of entanglement generation rate for the optimized buffer time protocol and the canonical protocol versus the distance between the repeater nodes,  $L_0$ , at the first nesting level. The green curve is obtained for memory lifetime  $\tau_M=100~\mu s$  and the blue curve for  $\tau_M=1~ms$ .

#### 3.3. Comparison of optimized versus canonical protocol at any nesting level

In the canonical protocol all nesting levels operate on the same set of quantum memories and successive levels do not use new set of memories. An average remote entangled state at any level is obtained after the waiting time for that level. The next level receives the average state from the previous level as soon as it is obtained, i.e. no synchronization conditions are used. The average remote entangled state in the canonical protocol obtained at any nesting level i is given by an expression identical to equation (A.7) with  $\gamma^{C,(j)}$  replacing  $\gamma^{O,(j)}$ . Physically,  $\gamma^{C,(j)}$  can be understood as the degradation in the fidelity due to the waiting time at nesting level j. The EGR in the canonical protocol is given by (see appendix A)

$$R_{\rm DE}^{C,(i)} = \frac{p_{\rm S}}{(\prod_{j=1}^{i} \langle k \rangle_{j}) 2\tau_{C}} E \left[ \frac{1 + \prod_{j=1}^{i} \left[ \gamma^{C,(j)} \right]^{2^{i-j}}}{2} \right], \tag{10}$$

where  $\langle k \rangle_1 = (3-2p)/p(2-p)$  is the waiting time at the first nesting level due to the initial entanglement generation probability, and  $\langle k \rangle_j = (3-2p_S)/p_S(2-p_S) \forall j \geqslant 2$  is the waiting time due to the swapping success probability at the second nesting level and higher.

The manifold increase in the EGR in the optimized buffer time protocol compared to canonical protocol is seen at all nesting levels if the swapping success probability  $p_S$  is low, figure 5. On the other hand, if  $p_S$  is high then the canonical protocol can yield better rates for higher nesting levels. In OBP the probability factor on the right hand side of equation (9) scales as,  $p_{\text{out}}^{(i)} \sim (p_S)^i(p)^{2^i}$ , for  $p \ll 1$ , whereas the time to obtain a state goes as,  $n_{\rm in}^{(i)} n_{\rm out}^{(i)} 2\tau_C^{(i)} = 2^{i+1}\tau_C$ , for  $n_{\rm out}^{(i)} = 2$  taken as an example. This implies a low probability of obtaining a remote entangled state per unit time. In the CP, the probability factor  $p_S$  in the RHS of equation (10) is constant. However, the time to obtain a state,  $(\prod_{j=1}^{i} \langle k \rangle_j) 2\tau_C = \langle k \rangle_1 (\langle k \rangle_2)^{i-1} 2\tau_C$ , can diverge much faster than that in the OBP,  $2^{i+1}\tau_C$  in our example. This happens if the waiting time due to the swapping success probability,  $\langle k \rangle_i \geqslant 2 \quad \forall j \geqslant 2$ , which implies  $p_s \leqslant 0.64$ . In this case, the probability to obtain a remote entangled state per unit time in the CP can be even lower than that in the OBP. Moreover, the degradation in the fidelity,  $\prod_{i=1}^{i} [\gamma^{X,(j)}]^{2^{i-j}}$ , for X = O, C in the two cases has maximum contribution from initial nesting levels. As discussed earlier the OBP yields a 1/p-factor increase of EGR at the first nesting level itself. Therefore the advantage due to OBP persists for the first few nesting levels till a crossover nesting level even if  $p_S$  is high. At the crossover nesting level, the EGR of the optimized buffer time protocol becomes equal to or less than that of the canonical protocol. If  $p_S$  is low then the logarithm of the ratio of rates in the two protocols diverges with the nesting level. In practice, one can use the explicit expressions for the rates provided in equation (9) and (10) to numerically evaluate the performance of the protocols at each nesting level.

The advantage of the optimized buffer time protocol can be observed by plotting the logarithm of the ratio of EGRs,  $\log[\eta^{(i)}(p,\beta,n_{\mathrm{opt}})]$ , versus the distance between repeater nodes at the first nesting level as shown in figure 6. The repeater is placed midway, at a distance of  $L_0$  from either end node, figure 1. We assume that the entanglement generation probability varies with the distance as  $p=\mathrm{e}^{-L_0/L_a}$  with the attenuation length  $L_a=20$  kms. The memory quality parameter then varies as  $\beta=\mathrm{e}^{-(L_0/L_a)(L_a/c\tau_M)}$  where the speed of light in fiber is taken to be  $c=2\times10^8$  ms $^{-1}$ . We choose two different memory lifetimes with values of  $\tau_M=100~\mu\mathrm{s}$  (green curve) and  $\tau_M=1$  ms (blue curve). The latter shows that a hundred fold improvement in EGRs is obtained with

1 ms quantum memories when the repeater is at a distance of 100 kms whereas the increase is even higher in the former case.

#### 4. Discussion and conclusion

We presented a QR architecture and protocol that mitigates quantum memory decoherence. The protocol optimizes the buffer time of the quantum memories based on the operating point in parameter space. We showed the hierarchical optimization of the buffer time at all nesting levels. The resulting increase of EGRs by many orders of magnitude was demonstrated. Crucially, the improvement was achieved with state of art physical resources. For example, with current technology, entanglement generation probability of  $p \simeq 10^{-4}$ – $10^{-5}$ , memory lifetime of  $\tau_M \simeq 10^{-1}$  ms, swapping success probability  $p_S \simeq 0.5$  and transfer probability  $p_T \simeq 1$  are feasible [15–17]. If repeater stations are spaced at intervals of  $L_0 = 20$  km, corresponding to the attenuation length in optical fibers, the one-way classical communication time  $\tau_C = L_0/(2 \times 10^5 \, \mathrm{km \, s^{-1}}) = 10^{-1} \, \mathrm{ms}$  equals the memory lifetime. The memory quality parameter then is  $\beta = 0.135$ . In this region of parameter space the optimized buffer time protocol yields  $(10^4 - 10^5)$  increase in the EGR. The proposed optimized buffer time protocol performs particularly well in the technologically feasible parameter regions and could facilitate broad applications in the future development of quantum networks such as for interferometry [18] and secret sharing [19].

#### Acknowledgments

This work was supported in part by the Office of the Secretary of Defense, Quantum Science and Engineering Program. L J acknowledges support from the ARL-CDQI (W911NF-15-2-0067, W911NF-18-2-0237), NSF (EFMA-1640959), and the Packard Foundation (2013-39273).

### Appendix A. Derivation of the average state in the optimized buffer-time protocol and the Canonical protocol

The decoherence in a pair of identical quantum memories with a lifetime  $\tau_M$ , due to dephasing, occurs at the rate  $2/\tau_M$ . We will denote the phase decoherence superoperator by  $D_t$ . Thus for an initial stored state,  $\rho_0 = |\psi^-\rangle \langle \psi^-|$ , we have

$$\rho(t) = D_t(\rho_0) = P^-(t)\rho^+ + P^+(t)\rho^-, \tag{A.1}$$

with  $P^{\pm} := (1 \pm e^{-2t/\tau_M})/2$ .

We assume that our heralded scheme of entanglement swapping succeeds with a probability  $p_S$ . Under the swapping operation  $\hat{S}$  for a pair of 2-qubit states  $\rho(t_1)$  and  $\rho(t_2)$  stored in the two pairs of memories for times  $t_1$  and  $t_2$  we have the output state conditioned on heralding to be

$$\hat{S}[\rho(t_{1}), \rho(t_{2})] = \hat{S}[(P^{-}(t_{1})\rho^{+} + P^{+}(t_{1})\rho^{-})(P^{-}(t_{2})\rho^{+} + P^{+}(t_{2})\rho^{-})] 
= P^{-}(t_{1})P^{-}(t_{2})\hat{S}[\rho^{+}, \rho^{+}] + P^{-}(t_{1})P^{+}(t_{2})\hat{S}[\rho^{+}, \rho^{-}] 
+ P^{+}(t_{1})P^{-}(t_{2})\hat{S}[\rho^{-}, \rho^{+}] + P^{+}(t_{1})P^{+}(t_{2})\hat{S}[\rho^{-}, \rho^{-}] 
= (P^{-}(t_{1})P^{-}(t_{2}) + P^{+}(t_{1})P^{+}(t_{2}))\rho^{-} + (P^{+}(t_{1})P^{-}(t_{2}) + P^{-}(t_{1})P^{+}(t_{2}))\rho^{+} 
= P^{+}(t_{1} + t_{2})\rho^{-} + P^{-}(t_{1} + t_{2})\rho^{+},$$
(A.2)

where we have used the linearity of the swap operation in the second line and the equalities  $\hat{S}[\rho^+, \rho^+] = \hat{S}[\rho^-, \rho^-] = \rho^-, \hat{S}[\rho^+, \rho^-] = \hat{S}[\rho^-, \rho^+] = \rho^+$ .

The probabilistic process of charging a pair of memories with the state  $|\psi^-\rangle$  succeeds at some step number k. This can happen for possibly different step numbers  $k_1, k_2$  for the two pairs of memories shown in figure 1 of the main text. Assuming  $k_1 \leqslant k_2$ , the latter pair of memories still stores the state for a time  $t_2 = \tau_C$ , i.e. for one classical communication time. The earlier charged pair stores the state for a time  $t_1 = (k_2 - k_1)2\tau_C + \tau_C$ . Therefore, the correspondence of storage times to step numbers is

$$t_1 \to (k_2 - k_1) 2\tau_C + \tau_C$$
  

$$t_2 \to \tau_C.$$
(A.3)

Thus,  $(t_1 + t_2) \rightarrow (|k_2 - k_1| + 1)2\tau_C$  for all  $k_1, k_2$ .

In the OBP the average accessed state  $\rho^O$  is a probabilistically weighted sum of states obtained after swapping, the states that have been stored in the two pairs of quantum memories, subject to the condition on the charging step numbers  $|k_2 - k_1| \leq n$ . The probability distribution of successful entanglement generation has

the form  $\mathbb{P}(k_1, k_2) = (1 - p)^{k_1 + k_2 - 2} p^2$ . The form of the remote entangled state  $\rho^S(k_1, k_2)$  after the swap is given by equation (A.2) with  $(t_1 + t_2) \to (|k_2 - k_1| + 1)2\tau_C$  for all  $k_1, k_2$ . One also needs to account for the decoherence suffered in the two remote memories after a swapped state is obtained untill the memories are refreshed after n cycle times. This state obtained after the swap, equation (A.2), further decoheres in the remote quantum memories at the two ends for a time,  $t_n = \{2(n - \max(k_1, k_2)) + 1\}\tau_C$ , and leads to the state  $\rho^S_n(k_1, k_2) = D_{t_n}(\rho^S(k_1, k_2))$ . Thus the average state in the optimized protocol is given by

$$\rho^{O} = \frac{\sum_{k_{1}=1,k_{2}=1}^{n,n} \mathbb{P}(k_{1}, k_{2}) \rho_{n}^{S}(k_{1}, k_{2})}{(1 - (1 - p)^{n})^{2}} = \frac{1 + \gamma^{O}(p, \beta, n)}{2} \rho^{-} + \frac{1 - \gamma^{O}(p, \beta, n)}{2} \rho^{+}, \tag{A.4}$$

where  $\gamma^{O}(p, \beta, n)$  is given by

$$\gamma^{O}(p, \beta, n) = \beta^{3} \frac{p}{(1 - (1 - p)^{n})^{2}} \frac{f(p, \beta, n)}{(\beta^{2} - q)(\beta^{2} - q^{2})},$$
(A.5)

with  $f(p, \beta, n) = q^{2n}(\beta^2 + q(1-q-\beta^2)) + \beta^{2n}(2q^{n+2}-q^2+\beta^2-2q^n\beta^2+q(\beta^2-1)), q=(1-p),$  and  $\beta=e^{-2\tau_C/\tau_M}$ . Note that the function  $f(p,\beta,n)$  has  $q=\beta,\beta^2$  as roots and thus the factors in the denominator of equation (A.5) do not cause any singular behavior. For perfect quantum memories  $\beta=1$  so that  $\gamma^O(p,\beta,n)=0$ . For  $p=1,\gamma^O(p,\beta,n)=\beta^{2n+1}$  which is maximum for n=1 and  $\beta<1$ .

The state in equation (A.4) is the output of the first nesting level and is denoted as  $\rho^{O, (1)}$ . Second nesting level receives the average output state from the first nesting level and outputs a probabilistically weighted average state with the probability defined by the input probability for the second level. The probability distribution at any nesting level, i, is given by

$$\mathbb{P}^{O,(i)}(k_1, k_2) = (p_{\text{in}}^{(i)})^2 (q_{\text{in}}^{(i)})^{k_1 + k_2 - 2} \tag{A.6}$$

with  $q_{\rm in}^{(i)}=(1-p_{\rm in}^{(i)})$ . The average state obtained in the optimistic protocol at any nesting level depends on the sequence of values of the parameter sets  $\bar{\tau}_C=\{\tau_C^{(1)},\,\tau_C^{(2)},\,...,\,\tau_C^{(i)}\}$ ,  $\bar{\beta}=\{\beta^{(1)},\,\beta^{(2)},\,...,\,\beta^{(i)}\}$ ,  $\bar{n}_{\rm in}=\{n_{\rm in}^{(1)},\,n_{\rm in}^{(2)},\,...,\,n_{\rm in}^{(i)}\}$ , the initial charging success probability p, the swapping success probability  $p_S$ , and the transfer success probability  $p_T$ . This state is obtained by iterating the process of averaging over the states received from the previous nesting level and normalizing by the appropriate probability normalization factor resulting in

$$\rho_{\text{out}}^{O,(i)} = \frac{1 + \prod_{j=1}^{i} \left[ \gamma^{O,(j)} \right]^{2^{i-j}}}{2} \rho^{-} + \frac{1 - \prod_{j=1}^{i} \left[ \gamma^{O,(j)} \right]^{2^{i-j}}}{2} \rho^{+}. \tag{A.7}$$

In the above expression the explicit dependence of  $\gamma^{O,\,(\,j)}$  on the relevant parameter values are suppressed for brevity. The actual expression for which is

$$\gamma^{O,(j)} = \mathcal{N}^{-1} \sum_{k_1, k_2 = 1}^{n_{\text{out}}^{(j)}} \mathbb{P}^{O,(j)}(k_1, k_2) [\beta^{(j)}]^{n_{\text{in}}^{(j)} \{2(n_{\text{out}}^{(j)} - k_1) + 2\} + 1}, \tag{A.8}$$

where  $\mathcal{N} = \sum_{k_1, k_2=1}^{n_{\text{out}}^{(j)}} \mathbb{P}^{(j)}(k_1, k_2)$ .

In the CP, the average obtained state is a probabilistically weighted sum of states over all possible storage times in the two pairs of memories. The state obtained after the swap operation is given by equation (A.2) with  $(t_1 + t_2) \rightarrow (|k_2 - k_1| + 1)2\tau_C$  for all  $k_1, k_2$ . The obtained swapped state decoheres during time  $\tau_C$  in the two remote memories resulting in the state,  $\rho_r^S(k_1, k_2) = D_{\tau_C}(\rho^S(k_1, k_2))$ , since the results of the swap operation have to be communicated to the end nodes. Therefore, the CP average state is

$$\rho^{C} = \sum_{k=1}^{\infty,\infty} \mathbb{P}(k_{1}, k_{2}) \rho_{r}^{S}(k_{1}, k_{2}) = \frac{1 + \gamma^{C}(p, \beta)}{2} \rho^{-} + \frac{1 - \gamma^{C}(p, \beta)}{2} \rho^{+}, \tag{A.9}$$

where the function  $\gamma^{C}(p, \beta)$  is given by

$$\gamma^{C}(p,\beta) = \beta^{3} p^{2} \frac{(1-\beta^{2}q) + 2\beta^{2}q}{(1-q^{2})(1-\beta^{2}q)}$$
(A.10)

with  $\beta = e^{-2\tau_C/\tau_M}$  and q = 1 - p. Again, for  $\beta = 1$ ,  $\gamma^C(p, \beta) = 1$  and for p = 1,  $\gamma^C(p, \beta) = \beta^3$ . The expected number of steps to obtain an entangled state in both involved segments is given by

$$\langle k \rangle = \sum_{k=1}^{\infty} k p^2 q^{2(k-1)} + 2 \sum_{k_1=1}^{\infty} p^2 q^{k_1-2} (\sum_{k_2=k_1+1}^{\infty} k_2 q^{k_2}) = \frac{(3-2p)}{p(2-p)}.$$
 (A.11)

The average state after any nesting level in the CP is obtained by iterating the averaging procedure at every nesting level. The probability distribution at the first nesting level is determined by the entanglement generation probability. For the second and higher nesting levels the probability distribution is determined by the swapping

success probability. Thus

$$\mathbb{P}^{C,(1)}(k_1, k_2) = p^2 q^{k_1 + k_2 - 2}, \, \mathbb{P}^{C,(i)}(k_1, k_2) = p_S^2 q_S^{k_1 + k_2 - 2}.$$
(A.12)

By iterating the averaging procedure at each nesting level using the above probability distributions we get the form of the average state for any nesting level

$$\rho_{\text{out}}^{C,(i)} = \frac{1 + \prod_{j=1}^{i} \left[ \gamma^{C,(j)} \right]^{2^{i-j}}}{2} \rho^{-} + \frac{1 - \prod_{j=1}^{i} \left[ \gamma^{C,(j)} \right]^{2^{i-j}}}{2} \rho^{+}, \tag{A.13}$$

where

$$\gamma^{C,(j)} = \sum_{k_1, k_2=1}^{\infty, \infty} \mathbb{P}^{C,(j)}(k_1, k_2) [\beta^{(j)}]^3 [\beta^{(1)}]^2 \prod_{l=0}^{(j-1)} \langle k \rangle_l (k_2 - k_l), \tag{A.14}$$

with 
$$\langle k \rangle_0 = 1$$
,  $\langle k \rangle_1 = (3 - 2p)/p(2 - p)$ ,  $\langle k \rangle_l = (3 - 2p_s)/p_s(2 - p_s)$  for  $l \ge 2$ .

with  $\langle k \rangle_0 = 1$ ,  $\langle k \rangle_1 = (3-2p)/p(2-p)$ ,  $\langle k \rangle_l = (3-2p_S)/p_S(2-p_S)$  for  $l \geqslant 2$ . In both the OBP and CP, if the initial entangled state across the elementary segments has a fidelity of f, i.e.  $\rho_0 = f \rho^- + (1-f) \rho^+$ , then the output states equation (A.7) and equation (A.13) include the fidelity factor f in their coefficients. Thus for the OBP  $\gamma^{O,\,(1)} \to f \gamma^{O,\,(1)}$ , while for the CP  $\gamma^{C,\,(1)} \to f \gamma^{C,\,(1)}$ . Our results reman unchanged for any value of the initial fidelity.

For a repeater with  $N_m$  nesting levels, OBP requires  $2(2^{N_m+2}-2)$  memories whereas CP requires  $2^{N_m+2}$ quantum memories. Thus, the OBP requires at most twice as many quantum memories as the CP. The EGR per memory used is still higher by orders of magnitude in the OBP in the relevant regions of parameter space.

#### Appendix B. Ratio of rates and optimal wait-window size

The OBP provides higher EGRs than the CP for most regions of the  $(p, \beta)$ -parameter space, as shown in figure 2 of the main text. The size of the optimal wait-window in terms of cycle time depends on the operating point in the  $(p, \beta)$  parameter space, figure (B1).

We identify several regions:

- Long-lived quantum memories, low entanglement generation probabilities  $(p \to 0, \beta \to 1)$ . In this region, figure (B2) suggests the scaling of  $n_{\rm opt} \sim 1/p$ .
- Short-lived quantum memories, high entanglement generation probabilities ( $p \to 1, \beta \to 0$ ). In this region, the best schedule of course is the rapid reset strategy with  $n_{\text{opt}} = 1$ .
- Short-lived memories, low entanglement generation probabilities ( $p \to 0, \beta \to 0$ ). In this region, figure (B3) suggests that rapid resetting with  $n_{\rm opt}=1$  constitutes the best schedule.
- Long-lived quantum memories, high entanglement generation probabilities  $(p \to 1, \beta \to 1)$ . In this region, the best schedule also turns out to be the rapid reset strategy  $n_{\rm opt}=1$ . For  $\beta\simeq 1$  the CP provides better EGRs than the OBP.

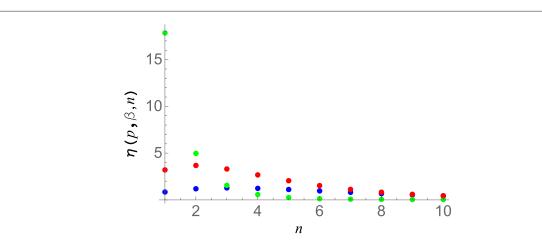
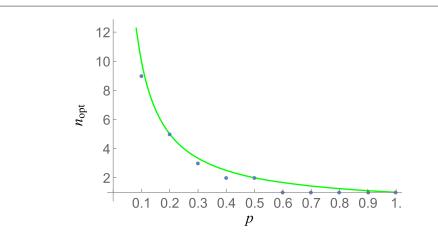
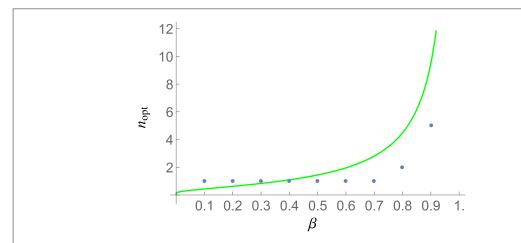


Figure B1. The optimal wait-window size in the optimized buffer time protocol depends on the location in parameter space. Shown here is the ratio of entanglement generation rates for  $(p, \beta) = (0.1, 0.9)$  in blue, (0.1, 0.4) in green and (0.05, 0.8) in red. The sequences of blue, red and green dots have its maximum at  $n_{\text{opt}} = 3, 2, 1$  respectively.



**Figure B2.** Scaling of the optimal wait-window size  $n_{\text{opt}}$  (blue dots) with the charging success probability p for a fixed value of  $\beta = 0.99$ . A scaling of  $n_{\text{opt}} \sim 1/p$  is observed as  $p \to 0$  while  $n_{\text{opt}} = 1$  for p > .5.



**Figure B3.** Scaling of the optimal wait-window size  $n_{\rm opt}$  (blue dots) with the memory parameter  $\beta$  for a fixed value of p=0.01. A scaling of  $n_{\rm opt}\sim -1/\log(\beta)$  is observed as  $\beta\to 1$  while  $n_{\rm opt}=1$  for  $\beta\leqslant .7$ .

#### Appendix C. Ratio of EGRs for nesting level N=1

The OBP and the CP can be compared with respect to the EGR which is the maximum rate of distillable entanglement. Here we compare the distillable EGRs in the two protocols at the first nesting level, i=1, in the  $p\to 0$ ,  $\beta\to 0$  region. We find approximations for the rates

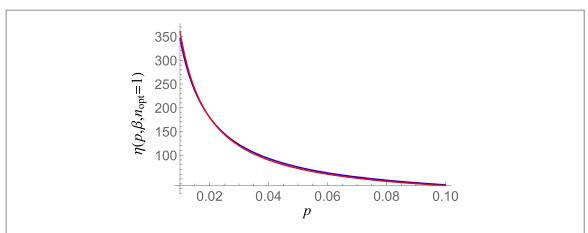
$$R_{DE}^{C}(p, \beta) = \frac{p_{S}}{\tau_{C}} \frac{p(2-p)}{3-2p} H[x_{1}(p, \beta)],$$

$$R_{DE}^{O}(p, \beta, n) = \frac{p_{S}}{\tau_{C}} \frac{(1-(1-p)^{n})^{2}}{n} H[x_{2}(p, \beta, n)],$$
(C.1)

where  $x_2=(1+\sqrt{1-(\gamma^O(p,\beta,n))^2})/2$ , and  $x_1=(1+\sqrt{1-(\gamma^C(p,\beta))^2})/2$ . In the  $p\to 0, \beta\to 0$  region both  $\gamma^C, \gamma^O\ll 1$  so that  $x_1(p,\beta)\approx 1-(\gamma^C(p,\beta))^2/4$ ,  $x_2(p,\beta,n)\approx 1-(\gamma^O(p,\beta,n))^2/4$ ,  $\gamma^C(p,\beta)\approx \beta^3 p/2$ ,  $\gamma^O(p,\beta,n)\approx \beta^3/n^2$ . We now use the property of binary entropy that, H(1-x)=H(x),  $0\leqslant x\leqslant 1$  and a small x approximation,  $H(x)\approx x\log_2(e/x)$ ,  $x\to 0$ . Further, we approximate  $p=(2-p)/3-2p\approx 2p/3$  and  $(1-(1-p)^n)^2/n\approx np^2$ . We also know from our numerical investigations (and analytical results) that in this region  $n_{\rm opt}=1$ . Putting all this together we get

$$\eta(p, \beta, n_{\text{opt}} = 1) = \frac{R_{\text{DE}}^{O}(p, \beta, n_{\text{opt}} = 1)}{R_{\text{DE}}^{C}(p, \beta)} \\
\approx \frac{6}{p} \frac{\log_2(4e) - 6\log_2(\beta)}{\log_2(16e) - 6\log_2(\beta) - 2\log_2(p)}.$$
(C.2)

For p,  $\beta$  values of technological relevance the above ratio is well approximated as 1/p, shown in figure (C1).



**Figure C1.** Scaling of the entanglement generation rate ratio using the optimized buffer time protocol and the canonical protocol: blue curve—the exact rate expressions from (C.1) for  $\beta = .1$  and  $p \in [1\%, 10\%]$ ; red curve—approximation as 3.6/p.

#### **ORCID** iDs

Siddhartha Santra https://orcid.org/0000-0001-6695-3929

#### References

- [1] Meter R V 2014 Quantum Networking (New York: Wiley)
- [2] Briegel H J, Dür W, Cirac J I and Zoller P 1998 Phys. Rev. Lett. 81 5932-5
- [3] Duan L M, Lukin M D, Cirac J I and Zoller P 2001 Nature 414 413
- [4] Sangouard N, Simon C, de Riedmatten H and Gisin N 2011 Rev. Mod. Phys. 83 33-80
- [5] Razavi M, Piani M and Lütkenhaus N 2009 Phys. Rev. A 80 032301
- [6] Hartmann L, Kraus B, Briegel H J and Dür W 2007 Phys. Rev. A 75 032310
- [7] Collins O A, Jenkins S D, Kuzmich A and Kennedy T A B 2007 Phys. Rev. Lett. 98 060502
- [8] Jiang L, Taylor J M, Nemoto K, Munro W J, Van Meter R and Lukin M D 2009 Phys. Rev. A 79 032325
- [9] Munro W J, Stephens A M, Devitt S J, Harrison K A and Nemoto K 2012 Nat. Photon. 777
- [10] Muralidharan S, Kim J, Lütkenhaus N, Lukin M D and Jiang L 2014 Phys. Rev. Lett. 112 250501
- [11] Jiang L, Taylor J M, Khaneja N and Lukin M D 2007 Proc. Natl Acad. Sci. 104 17291–6
- [12] Bennett CH, Brassard G, Popescu S, Schumacher B, Smolin J A and Wootters W K 1996 Phys. Rev. Lett. 76 722–5
- [13] Simon C et al 2010 Eur. Phys. J. D 58 1–22
- [14] Bernardes N K, Praxmeyer L and van Loock P 2011 Phys. Rev. A 83 012323
- [15] Cernoch A, Soubusta J, Bartuskova L, Dusek M and Fiurasek J 2008 Phys. Rev. Lett. 100 180501
- [16] Heshami K, England DG, Humphreys PC, Bustard PJ, Acosta VM, Nunn J and Sussman BJ 2016 J. Mod. Opt. 63 2005–28
- [17] Rozpędek F, Goodenough K, Ribeiro J, Kalb N, Vivoli V C, Reiserer A, Hanson R, Wehner S and Elkouss D 2018 Quantum Sci. Technol. 3 034002
- [18] Santra S, Kirby B T, Malinovsky V S and Brodsky M 2018 J. Mod. Opt. (https://doi.org/10.1080/09500340.2018.1511864)
- [19] Hillery M 1999 Phys. Rev. A 59 1829