Entanglement buffering with two quantum memories

Davies, Bethany; Iñesta, Álvaro G.; Wehner, Stephanie

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Entanglement buffering with two quantum memories

Bethany Davies[1,2,3], Álvaro G. Iñesta[1,2,3], and Stephanie Wehner[1,2,3]

[1]QuTech, Delft Univ. of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands

[2]EEMCS, Quantum Computer Science, Delft Univ. of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands

[3]Kavli Institute of Nanoscience, Delft Univ. of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands

Quantum networks crucially rely on the availability of high-quality entangled pairs of qubits, known as entangled links, distributed across distant nodes. Maintaining the quality of these links is a challenging task due to the presence of time-dependent noise, also known as decoherence. Entanglement purification protocols offer a solution by converting multiple low-quality entangled states into a smaller number of higher-quality ones. In this work, we introduce a framework to analyse the performance of entanglement buffering setups that combine entanglement consumption, decoherence, and entanglement purification. We propose two key metrics: the availability, which is the steady-state probability that an entangled link is present, and the average consumed fidelity, which quantifies the steady-state quality of consumed links. We then investigate a two-node system, where each node possesses two quantum memories: one for long-term entanglement storage, and another for entanglement generation. We model this setup as a continuous-time stochastic process and derive analytical expressions for the performance metrics. Our findings unveil a trade-off between the availability and the average consumed fidelity. We also bound these performance metrics for a buffering system that employs the well-known bilocal Clifford purification protocols. Importantly, our analysis demonstrates that, in the presence of noise, consistently purifying the buffered entanglement increases the average consumed fidelity, even when some buffered entanglement is discarded due to purification failures.

1 Introduction

The functionality of quantum network applications typically relies on the consumption of entangled pairs of qubits, also known as *entangled links*, that are shared among distant nodes [41]. The performance of quantum network applications does not only depend on the rate of production of entangled links, but also on their quality. In a quantum network, it is therefore a priority for high-quality entangled states to be readily available to network users. This is a challenging task, since entangled links are typically stored in memories that are subjected to time-dependent noise, meaning that the quality of stored entangled links decreases over time. This effect is known as decoherence.

A common way of overcoming the loss in quality of entangled links is to use *entanglement purification* protocols [3, 11, 12, 44]. An *m*-to-*n* entanglement purification protocol consumes *m* entangled quantum states of low quality and outputs *n* states with a higher
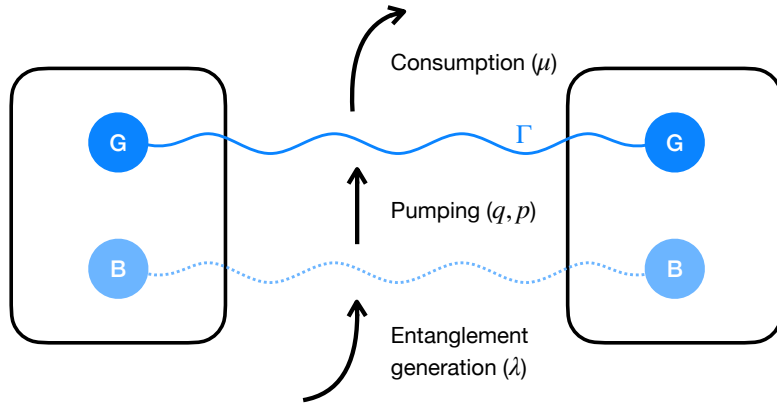
Figure 1: Illustration of the entanglement buffering system with two quantum memories (1G1B system). Each of the nodes has two memories (G and B). Memory G is used to store the buffered link. An entangled link is generated at a rate λ in memory B. If memory G is empty when the new link is generated in B, the link is immediately transferred to G. If memory G is occupied, the new link generated in B is immediately used to purify the buffered link with probability q (otherwise, the new link is discarded). The pumping protocol consumes the link in B to increase the quality of the buffered link in G, and it succeeds with probability p (otherwise, it destroys the link in G). The buffered link is consumed at a fixed rate μ. The quality of the entanglement stored in G decays exponentially with rate Γ. Formal definitions of the problem parameters can be found in Section 3.

quality, where typically $m > n$. The simplest form of purification schemes are 2-to-1, also known as *entanglement pumping* protocols. One downside of using purification is that there is typically a probability of failure, in which case the input entangled links must be discarded and nothing is produced.

In this work, we take a crucial step towards the design of high-quality entanglement buffering systems. The goal of the buffer is to make an entangled link available with a high quality, such that it can be consumed at any time for an application. We develop methods to analyse the performance of an entanglement buffering setup in a system with entanglement consumption, decoherence, and entanglement pumping. We introduce two metrics to study the performance: (i) the *availability*, which is the steady-state probability that a link is available, and (ii) the *average consumed fidelity*, which is the steady-state average quality of entangled links upon consumption. We measure the quality of quantum states with the fidelity, which is a well-known metric for this [32].

We use these metrics to study a two-node system where each of the nodes has two quantum memories, each of which can store a single qubit (see Figure 1). This system is of practical relevance since early quantum networks are expected to have a number of memories per node of this order (e.g. in [26] and [43], entanglement purification was demonstrated experimentally between two distant nodes, each with the capability of storing two qubits). We study a system where each node has one good (long-term) quantum memory, G, and one bad (short-term) memory, B, per node. We therefore refer to this entanglement buffering setup as the *1G1B system*. The good memories are used to store an entangled link between the nodes that can be consumed at any time. The bad memories are used to generate a new entangled link between the nodes. The new link may be used to pump the stored link with fresh entanglement.

Calculating the temporal evolution of the fidelity of an entangled link is generally a difficult task, since the fidelity depends on the history of operations that have been applied to the link in the past. By modelling the state of the 1G1B system as a continuous-time stochastic process, we are able to find analytical solutions for the availability and the

average consumed fidelity of the system. We illustrate the application of these results in a simplified scenario where purification has a linear action on the quality of the buffered link.

Our main contributions are the following:

- We propose two metrics to measure the performance of an entanglement buffering system: the availability and the average consumed fidelity.

- We provide a simple closed-form expression for the availability in the 1G1B system.

- We develop an analytical framework to calculate the average fidelity of the links consumed in a 1G1B system. We provide a closed-form expression for pumping schemes that increase the fidelity of the entangled link linearly with the initial fidelity.

Our main findings are the following:

- We confirm the intuition that, except in some edge cases, there is a trade-off between availability and average consumed fidelity: one must either consume low-quality entanglement at a higher rate, or high-quality entanglement at a lower rate.

- Consider a situation where bilocal Clifford protocols are employed (this is one of the most popular and well-studied classes of purification protocols [10]). Then, if the noise experienced by the quantum memories is above certain threshold, pumping the stored link with fresh entanglement always increases the average consumed fidelity, even if the stored link is often discarded due to a small probability of successful pumping. We provide an explicit expression for this noise threshold, which depends on the purification protocol employed and the fidelity of newly generated links.

The structure of the paper is the following. In Section 2, we provide a short overview of related work (an additional introduction to the quantum information concepts required to understand this work is provided in Appendix A). In Section 3, we explain the physical setup and provide a formal definition of the 1G1B system as a stochastic process. In Section 4, we define the performance metrics of interest and provide analytical expressions that enable their computation. In Section 5, we analyse the system in the case where the pumping protocol produces an output state whose fidelity is a linear function of the fidelity of one of the input states. In Section 5.1, we use these results to bound the performance of the 1G1B system, in the case where bilocal Clifford protocols are employed for entanglement pumping. Lastly, in Section 6, we discuss the implications of this work and future research directions.

2 Related work

The performance analysis of quantum networks is unique because of the trade-off between the rate of distribution of entangled links and the quality of distributed links, both of which are important for the functionality of networking applications. This leads to interesting stochastic problems, which are important to understand the parameter regimes of a possible architecture. For example, [37, 28, 23] deal with the problem of generating an end-to-end entangled link across a chain of quantum repeaters, where both the rate of production and the quality of the end-to-end links are quantities of interest. Another example is the problem of generating multiple entangled links between two users with a high quality, which is treated in [36, 9]. In these works, the time between successfully generated entangled links is modelled by a geometric distribution. However, the time taken

up by an entanglement generation attempt is generally small compared to other relevant time scales [34, 29]. Hence, a simplifying assumption that we make in this work is that the time between entanglement generation attempts is exponentially distributed. This is a common assumption in the quantum networking literature (see e.g. [39, 31, 7]), because it can enable the finding of closed-form relations between physical variables and protocol parameters. Here, we introduce and find expressions for the values of two key performance metrics in the steady state.

Previous work that incorporates entanglement purification schemes into the analysis of quantum network architectures typically involves numerical optimisation methods (see e.g. [40]), or only considers specific purification protocols [6]. By contrast, in this work we focus on presenting the purification protocol in a general way, and finding closed-form solutions for the performance metrics of interest (albeit for a simpler architecture). This is an important step towards an in-depth understanding of how one can expect purification to impact the performance of a near-term quantum network.

Other works have introduced the concept of entanglement buffering (preparing quantum links to be consumed at a later time) over a large-scale quantum network [24, 35]. To the best of our knowledge, the only work with a similar set-up to ours is [15], which was developed in parallel and independently of our work. There, the authors study the steady-state fidelities of a system involving two memories used for storage (good memories), and one memory used for generation (bad memory). This work differs from ours in multiple ways. For example, the analysis is done in discrete time and it is assumed that the fidelity takes a discrete set of values, whereas we do not make this assumption since we work in continuous time. Additionally, consumption of entanglement is not included in the system studied in [15], which may impact the steady-state behaviour.

Lastly, we note that previous work generally assumes a specific protocol for entanglement buffering between each pair of nodes, and does not address the following fundamental question: *what is the best way to buffer entanglement between two users in a quantum network?* To the best of our knowledge, we address this question for the first time.

3 The 1G1B (one good, one bad) system

We now define the 1G1B system. In Section 3.1, we describe and motivate the model of the system. In Section 3.2, we define the variables of interest precisely. This facilitates the definition of the performance metrics in Section 4. In Appendix A, we provide an introduction to the quantum background required.

3.1 System description

Below we provide a list of assumptions that model the 1G1B system, and provide motivation for each assumptions. An illustration of the system is given in Figure 1.

1. **Each of the nodes has two memories: one long-term memory (good, G) and one short-term memory (bad, B). The B memories are used to generate new entangled links. The G memories are used as long-term storage (entanglement buffer).**

   This is motivated by the fact that *storage* (G) and *communication* (B) qubits are often present in experimental scenarios, where the former is used to store entanglement and the latter is used to generate new links [2, 26, 27].

2. **New entangled links are generated in memory B according to a Poisson process with rate λ. New entangled links always have the form ρ_{new}.**

Physical entanglement generation attempts are typically probabilistic and heralded [5, 1]. In other words, the attempt can fail with some probability and, when this occurs, a failure flag is raised. Therefore, the generation of a single link may take multiple attempts. The time taken by an attempt is typically fixed (this is both the case in present-day quantum networks [34] and an assumption that is commonly made in the theoretical analysis of quantum networks [23, 24, 9]). Then, the time between attempts follows a geometric distribution. Since the probability of successful generation and the length of the time step is often small compared to other relevant time scales [34, 29], we use a continuous approximation, i.e. that the time between arrivals are exponentially distributed. This is a Poisson process (see e.g. Chapter 6.8 from [19]).

3. **When a link is newly generated in memory B, if memory G is empty (no link present), the new link is immediately placed there. If memory G is not empty, the nodes immediately either (i) attempt pumping with probability q, or (ii) discard the new link from memory B (probability $1 - q$).**

   This step is included because it may not always be a good idea to carry out pumping, due to there being a possibility of this failing.

4. **Links stored in memory G are Werner states.**

   Werner states take the simple form

   $$\rho_{\mathrm{W}} = F \left|\phi^+\right\rangle\!\left\langle\phi^+\right| + \frac{1-F}{3} \left|\psi^+\right\rangle\!\left\langle\psi^+\right| + \frac{1-F}{3} \left|\psi^-\right\rangle\!\left\langle\psi^-\right| + \frac{1-F}{3} \left|\phi^-\right\rangle\!\left\langle\phi^-\right|,$$

   where $\left\{\left|\phi^+\right\rangle, \left|\psi^+\right\rangle, \left|\phi^-\right\rangle, \left|\psi^-\right\rangle\right\}$ denote the Bell basis. A Werner state corresponds to maximally entangled state that has been subjected to isotropic noise. The state in the good memory is therefore fully described by one parameter: the fidelity F to the target state $\left|\phi^+\right\rangle$. Any state can be transformed into a Werner state with the same fidelity by applying extra noise, a process known as *twirling* [4, 21]. Hence, this assumption constitutes a worst-case model.

5. **While in memory G, states are subject to depolarising noise with memory lifetime $1/\Gamma$.**

   Depolarising noise can also be seen as a worst-case noise model [13]. After a time t in memory, this maps the state fidelity F to

   $$F \rightarrow e^{-\Gamma t} \left(F - \frac{1}{4}\right) + \frac{1}{4}.$$

6. **Consumption requests arrive according to a Poisson process with rate μ. When a consumption request arrives, if there is a stored link in memory G, it is immediately used for an application (and therefore removed from the memory). If there is no link available, the request is ignored.**

   This means that the time until the next consumption request arrives is independent of the arrival time of previous requests, and it is exponentially distributed. This assumption is commonly made in the performance analysis of queuing systems (see e.g. Chapter 14 from [38]).

7. **Assumptions about pumping:**

(a) **Pumping is carried out instantaneously.** This is because the execution time is generally much lower than the other timescales involved in the problem. For example, in state-of-the-art setups, an entangled link is generated approximately every 0.5 s [34], while entanglement pumping may take around $0.5 \cdot 10^{-3}$ s [26]. If the nodes are far apart, classical communication between them would only add a negligible contribution to the purification protocol (e.g. classical information takes less than 10^{-4} s to travel over 10 km of optical fiber).

(b) **Suppose that the link in memory G has fidelity F and the link in memory B is in state ρ_{new}. If pumping succeeds, the output link has fidelity $J(F, \rho_{\text{new}})$, and remains in the good memory. If pumping fails, all links are discarded from the system.** Here, the *jump function* $J(F, \rho_{\text{new}}) \in [0, 1]$ is dependent on the choice of purification protocol. Given the assumption that one of the links is a Werner state, the form of this function is

$$J(F, \rho_{\text{new}}) = \frac{\tilde{a}(\rho_{\text{new}})F + \tilde{b}(\rho_{\text{new}})}{p(F, \rho_{\text{new}})}, \tag{1}$$

with

$$p(F, \rho_{\text{new}}) = c(\rho_{\text{new}})F + d(\rho_{\text{new}}) \tag{2}$$

where $\tilde{a}, \tilde{b}, c, d$ are functions of ρ_{new}. Here, $p(F, \rho_{\text{new}})$ is the success probability of purification. See Appendix B for an explanation of why the jump function and success probability take this form.

(c) **Pumping succeeds with probability p, which is constant in the fidelity of memory G.** We see from the above that this is a special case, and that in general the probability of purification success varies linearly with the fidelity of the good memory. However, performing the analysis with a constant probability of success does allow us to find bounds on the operating regimes of the system by considering the best-case and worst-case values of p (see Section 5.1). Combining this with Assumption 7b, we see that this is effectively equivalent to setting $c(\rho_{\text{new}}) = 0$. The jump function is then linear in the fidelity of memory G, and can be written as
$$J(F, \rho_{\text{new}}) = a(\rho_{\text{new}})F + b(\rho_{\text{new}}),$$
where $a := \tilde{a}/p$ and $b := \tilde{b}/p$.

Implicit in the above is that the process of entanglement generation, pumping and consumption ((2),(3),(6) and (7b)) are independent. We provide a summary of the parameters involved in the 1G1B system in Table 1.

3.2 System definition

In this subsection, we define the state of the system mathematically, which will be the main object of study in the rest of this work. We view the state of the system as the number of rounds of pumping that the link in memory has undergone. From now on, when we refer to 1G1B, we refer to the stochastic process that evolves according to the following definition.

Table 1: Parameters of the 1G1B system. See main text for detailed explanations.

Hardware

λ	Rate of heralded entanglement generation (time between successful attempts is exponentially distributed with rate λ)
ρ_{new}	Entangled state produced after a successful entanglement generation
Γ	Rate of decoherence (fidelity of the entangled link decays exponentially over time with rate Γ)

Application

μ	Rate of consumption (specified by application)

Pumping protocol

q	Probability of attempting pumping immediately after a successful entanglement generation attempt (otherwise the new link is discarded)
p	Probability of successful pumping
$J(F, \rho_{\mathrm{new}})$	Jump function: fidelity of the output state following successful pumping (F is the fidelity of the Werner state stored in the good memory)

Definition 3.1 (1G1B system). Let $s(t)$ be the state of the 1G1B system at time t. This takes values

$$s(t) = \begin{cases} \emptyset & \text{if there is no link in memory,} \\ i \geq 0 & \text{if there is a link in memory which is the result of } i \text{ successful pumping rounds,} \end{cases} \tag{3}$$

where $i = 0$ corresponds to a link in memory that has not undergone any pumping. Assume that the system starts with no link, i.e. $s(0) = \emptyset$. The system transitions from state $\emptyset$ to state 0 when a new link is generated and placed in the good memory, which was previously empty. The rate of transition from $\emptyset$ to 0 is then given by the entanglement generation rate λ. Pumping success occurs when a new link is produced (rate λ), pumping is attempted (probability q), and pumping succeeds (probability p). Therefore, the transition from state i to $i + 1$ occurs with rate $\lambda q p$. The final allowed transition is from i to $\emptyset$ which occurs due to consumption or purification failure, which occurs with rate $\mu + \lambda q(1 - p)$.

We also refer to the state $i \geq 0$ as the ith *purification level*. Since the transitions between each state in 1G1B occur according to an exponential distribution with rate that is only dependent on the current state of the system, this is a continuous-time Markov chain (CTMC) on the state space $\{\emptyset, 0, 1, \ldots\}$. The resulting CTMC and the rate of transitions is depicted in Figure 2. This is the main object of study in our work.

Recall that we are also interested in the fidelity of the link in memory. This is dependent not only on the state $s(t) \in \{\emptyset, 0, 1, \ldots\}$, but also on the time spent in the states leading up to the current purification level. This motivates the following definition.

Definition 3.2. Suppose that $s(t) = i$. Then, we define random variable $\vec{T}(t)$ to be the length-$(i + 1)$ vector storing the times spent in the recent purification levels $0, 1, \ldots, i$ leading up to the current one, where time $T_j(t)$ was spent in the most recent visit to state
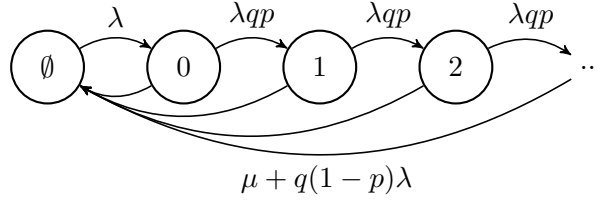
Figure 2: The transitions of the 1G1B system.

j ($j \le i-1$), and time $T_i(t)$ is the time spent so far in state i. See Figure 3 for a depiction of this.

We also need a framework with which to compute the fidelity at time t. Recalling assumption (5) of Section 3.1, we denote decoherence by the following.

Definition 3.3. Let $D_t : [0,1] \to [0,1]$ denote the action of depolarising noise on the state fidelity F. This has action

$$D_t[F] = e^{-\Gamma t}\left(F - \frac{1}{4}\right) + \frac{1}{4}.$$

We now formally define the jump function.

Definition 3.4. After successfully applying purification to a Werner state with fidelity F and a general two-qubit state ρ_{new}, the output state has fidelity $J(F, \rho_{\text{new}})$. We refer to J as the *jump function* of the protocol. The general form of this is given in (1).

We note that every purification protocol has a corresponding jump function. The exact form of J is dependent on the choice of pumping protocol, but in general is a continuous rational function of F, taking values in $[0,1]$.

We also need to compute the fidelity after many rounds of decoherence and pumping. This essentially means composing D_t and J.

Definition 3.5. Let $F^{(i)}(t_0, ..., t_i)$ denote the fidelity after spending time $t_0, ..., t_i$ in each purification level $0, 1, ..., i$. This may be defined recursively as

$$F^{(i)}(t_0, ..., t_i) = D_{t_i}\left[J(F^{(i-1)}(t_0, \ldots, t_{i-1}), \rho_{\text{new}})\right], \tag{4}$$

with $F^{(0)}(t_0) = D_{t_0}[F_{\text{new}}]$, where F_{new} is the fidelity of ρ_{new}.

Note that $F^{(i)}$ is a continuous and bounded function of its inputs, since the same is true for D_t and J. We are now equipped to define the fidelity of the system.

Definition 3.6. The fidelity of the 1G1B system at t is given by

$$F(t) = \begin{cases} F^{(i)}(\vec{T}(t)) \text{ if } s(t) = i \ge 0, \\ 0, \text{ if } s(t) = \emptyset. \end{cases} \tag{5}$$

Note that this formulation can also be adapted to incorporate a system where we apply a different pumping protocol in each state of the CTMC. In that case, we would employ a more general recurrence relation:

$$F^{(i)}(t_0, \ldots, t_i) = D_{t_i}\left[J^{(i)}(F^{(i-1)}(t_0, \ldots, t_{i-1}), \rho_{\text{new}})\right], \tag{6}$$
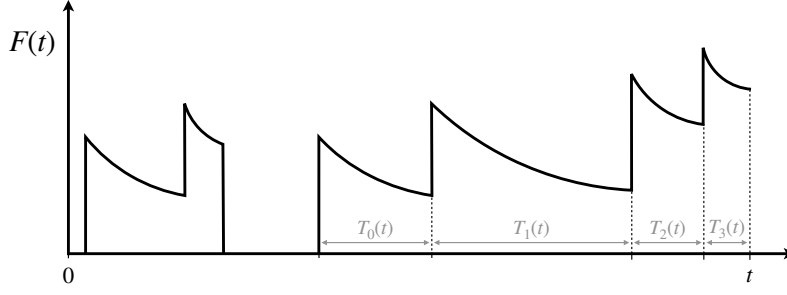
Figure 3: Example of the evolution of the fidelity of the buffered entanglement over time. The fidelity experiences a sudden boost every time a pumping protocol is successful. Then, it decays exponentially due to decoherence. Each state in the CTMC is identified by the number of times the current buffered link has been purified. If $s(t) = i$, the random variables $\{T_j(t) : j = 0, 1, ..., i-1\}$ are the times spent in each state of the CTMC immediately leading up to the current state, and $X_i(t)$ is the time so far spent in state i.

where the $J^{(i)}$ is the jump function corresponding to the pumping protocol applied in state i of the CTMC. For simplicity, however, we study recurrence relations of the form (4). This may be used to model the situation where the same pumping protocol is applied every time, or provide bounds for using multiple protocols, as we do in Section 5.1.

4 Performance metrics

In this Section, we define two metrics to evaluate the performance of an entanglement buffering system: the *availability* and the *average consumed fidelity*. We also provide analytical expressions for both metrics in the 1G1B system.

4.1 Availability

A natural measure for the quality of service provided to users is the probability that a consumption request may be served at any given time. If there is a link stored in the good memories, the consumption request is immediately served. However, if there is no entanglement available, the request is ignored. Letting $P(s(t) = i)$ be the probability that the system is in state i at time t, we define the steady-state distribution as

$$\pi_i := \lim_{t \to \infty} P(s(t) = i). \tag{7}$$

Then, we define our first performance metric as follows.

Definition 4.1 (Availability). The availability A is defined as

$$A := 1 - \pi_\emptyset, \tag{8}$$

which is the probability that there is a link in memory in the limit $t \to \infty$.

This definition can be applied to any entanglement buffering setup. In the 1G1B system, the availability is well-defined, as shown in Appendix C. Moreover, it is possible to derive a closed-form expression for the availability, as stated in the proposition below.

Proposition 1. *Consider the 1G1B system (Definition 3.1). The availability is given by*

$$A = 1 - \pi_\emptyset = \frac{\lambda}{\lambda + \mu + \lambda q(1-p)}, \tag{9}$$

and the rest of the steady-state distribution is given by

$$\pi_i = \frac{\lambda^{i+1}q^ip^i}{(\mu + \lambda q)^{i+1}}\pi_\emptyset. \tag{10}$$

See Appendix C for a proof of this proposition. We note that this can be derived in a straightforward manner using the balance equations for a CTMC. Instead, we use renewal theory, for two reasons. Firstly, this approach ties in neatly with the proof of the formula for the average fidelity (see the next subsection). Secondly, this approach provides a formula for the availability that is more general, as it also applies to the case where entanglement generation is described by a general random variable instead of being exponentially distributed. See Appendix C for the general formula for the availability.

4.2 Average consumed fidelity

The quality of service of an entanglement buffering system can also be measured in terms of the quality of the entanglement provided to the users. Therefore, the average fidelity of the entangled links upon consumption can be used as an additional metric to assess the performance of the system.

Definition 4.2 (Average consumed fidelity)**.** The *average consumed fidelity* is the average fidelity of the entangled link upon consumption, in the steady state. More specifically,

$$\overline{F} := \lim_{t \to \infty} \mathbb{E}\big[\, F(t) \,|\, s(t) \neq \emptyset \,\big]. \tag{11}$$

In the definition of $\overline{F}$, we condition on not being in $\emptyset$ since consumption events do not happen when there is no link present. As before, this performance metric can be applied to any entanglement buffering setup. In the case of the 1G1B system, it is possible to derive an analytical expression for $\overline{F}$ which explicitly depends on the steady-state distribution. The formula is given in the following theorem.

Theorem 1. *In the 1G1B system, the average consumed fidelity can be written as*

$$\overline{F} = \frac{1}{A}\sum_{i=0}^{\infty} c_i \pi_i, \tag{12}$$

where $\pi_i = \lim_{t \to \infty} P(s(t) = i)$, and

$$c_i = \mathbb{E}\left[F^{(i)}(Q_0, Q_1, ..., Q_i)\right] \tag{13}$$

where A is the availability, $Q_0, Q_1, \ldots, Q_i$ are i.i.d. random variables with $Q_0 \sim Exp(\mu + \lambda q)$, and $F^{(i)}$ is given in Definition 3.5.

Sketch proof of Theorem 1. A first step is to expand by conditioning on the value of $s(t)$,

$$\mathbb{E}[F(t)|s(t) \neq \emptyset] = \sum_{i=0}^{\infty} \mathbb{E}[F(t)|s(t) = i]P(s(t) = i|s(t) \neq \emptyset)$$

$$= \frac{1}{P(s(t) \neq \emptyset)}\sum_{i=0}^{\infty} \mathbb{E}[F(t)|s(t) = i]P(s(t) = i).$$
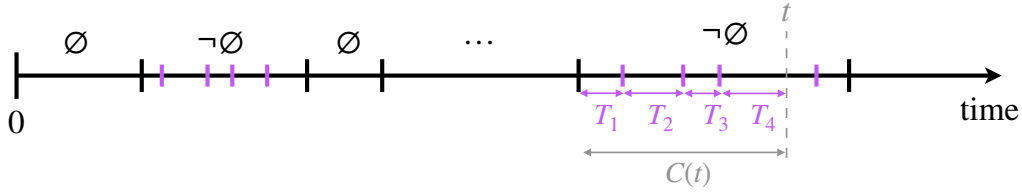
Figure 4: An example timeline of the 1G1B process. Black dashes are link generation and removal. Shorter purple dashes are pumping rounds. If there is a link present at time t, the random variable $C(t)$ is the total time spent so far in $\neg\emptyset$ (link present). Pumping rounds occur within the time $C(t)$ as a Poisson process with rate λqp. This may be used to characterise the distribution of $\vec{T}(t)$ in the limit $t \to \infty$, which is needed to prove Theorem 1.

In Proposition 5 (Appendix C.2), we show that, when $t \to \infty$, the limit can be brought inside of the sum, and so

$$\overline{F} = \lim_{t \to \infty} \mathbb{E}[F(t)|s(t) \neq \emptyset]$$
$$= \frac{1}{A} \sum_{i=0}^{\infty} \pi_i \cdot \lim_{t \to \infty} \mathbb{E}[F(t)|s(t) = i],$$

where we have used the definition of the steady-state distribution and the availability (see (7) and (8)). The values π_i may be computed using Proposition 1. The remaining work is then to show that

$$\lim_{t \to \infty} \mathbb{E}[F(t)|s(t) = i] = \mathbb{E}\left[F^{(i)}(Q_0, \ldots, Q_i)\right], \tag{14}$$

which essentially requires the characterisation of the limiting distribution of $\vec{T}(t)$, since from Definition 3.6 we recall that $\mathbb{E}[F(t)|s(t) = i] = \mathbb{E}\left[F^{(i)}(\vec{T}(t))|s(t) = i\right]$. This is achieved with the following result: conditional on $s(t) = i$, $\vec{T}(t) \to (Q_0, \ldots, Q_i)$ in distribution as $t \to \infty$, where the Q_j are i.i.d. random variables with $Q_0 \sim \text{Exp}(\mu + \lambda q)$. There are two main steps to show this (see Figure 4 for graphical intuition):

1. Let $C(t)$ be the total time spent so far in $\neg\emptyset$ (link in memory G) at the time t. The first step is to show that $C(t) \to C$ in distribution as $t \to \infty$, where $C \sim \text{Exp}(\mu + \lambda q(1-p))$. This is shown with renewal theory. For more details, see the results of Appendix C.1.

2. Characterise the limiting distribution of the time spent in each purification level *within* the time $C(t)$. These are the $T_j(t)$. We use the fact that pumping rounds occur as a Poisson process within the time $C(t)$. For more details, see the results of Appendix C.2.

Finally, since $F^{(i)}$ is a continuous function of its inputs, (14) follows. $\qquad\square$

For the full proof, see Appendix C. The particularly simple form of (13) can be attributed to the fact that in a CTMC, the time spent in a state is not influenced by the state to which the system transitions. As an example, in the CTMC from Figure 5, the time spent in state B before a transition does not depend on the transition itself, and this time is exponentially distributed with rate $r_{BA} + r_{BC}$. In the 1G1B system, the times spent in the states $j = 0, 1, \ldots, i-1$ leading up to state i are all exponentially distributed with
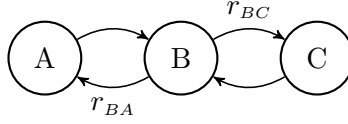
Figure 5: In a CTMC, the time spent in a state is independent of the transition that happens next. In this example, the time spent in state B before leaving is exponentially distributed with rate $r_{BA} + r_{BC}$.

rate $\lambda qp + \mu + \lambda q(1-p) = \mu + \lambda q$. As a consequence, the average fidelity after i successful purifications, c_i, does not depend on the probability of successful purification p.

Having systematic closed-form expressions for the functions $F^{(i)}$ enables the efficient computation of c_i and, therefore, $\overline{F}$. The calculation of $F^{(i)}$ in closed-form for a general J is quite involved, since the recurrence relation (4) becomes a rational difference equation with arbitrary coefficients. However, in the following sections we consider a jump function which is linear, for which it is possible to find a closed-form solution for $\overline{F}$.

5 Entanglement buffering with a linear jump function

In a purification protocol with a linear jump function, the output fidelity is a linear function of the fidelity of one of the input entangled links. When the probability of successful purification is constant with the fidelity of the good memory, as we assume in 1G1B, this implies that the jump function is linear. This is shown in Appendix B. In this Section, we compute a closed-form solution for the average consumed fidelity in a 1G1B system assuming a linear jump function. Then, we analyse the performance of the system using the performance metrics defined in Section 4 (availability and average consumed fidelity). In Section 5.1, we focus on bilocal Clifford protocols, an important type of purification scheme. For a given value of target availability, we provide upper and lower bounds on the average consumed fidelity that can be achieved by any bilocal Clifford protocol in the 1G1B system.

Purification protocols with linear jump functions are relevant for two main reasons:

(i) Purification protocols are generally more effective within some range of input fidelities (the increase in fidelity is larger when the input fidelities are within some interval). If the system operates within a small range of fidelities, one may approximate the true jump function with a linear jump function.

(ii) One can find linear jump functions that upper and lower bound a set of jump functions of interest. These may then be used to upper and lower bound a fidelity-based performance metric (such as the average consumed fidelity) of a system that has the freedom to employ any of these jump functions.

In Appendix E, we demonstrate (ii) in the case where bilocal Clifford protocols are employed in the 1G1B system. The output fidelity of a bilocal Clifford protocol can be upper and lower bounded by nontrivial linear functions when one of the input states is a Werner state (using some additional minor assumptions).

Consider a pumping scheme that takes as input a Werner state with fidelity F and an arbitrary state ρ_{new}. In the 1G1B system, these are the states in the good and the bad memories, respectively. A linear jump function can be written as

$$J(F, \rho_{\text{new}}) = a(\rho_{\text{new}})F + b(\rho_{\text{new}}),\tag{15}$$

with $0 \leq a(\rho_{\text{new}}) \leq 1$ and $(1 - a(\rho_{\text{new}}))/4 \leq b(\rho_{\text{new}}) \leq 1 - a(\rho_{\text{new}})$, as shown in Proposition 6. In what follows, we implicitly assume that a and b depend on ρ_{new}.

We now derive a closed-form solution for the average consumed fidelity of 1G1B when the jump function is linear, using Theorem 1. The formula requires knowledge of the steady state distribution $\{\pi_i : i = \emptyset, 0, 1, ...\}$, and the expected fidelities c_i, as defined in (13). Recall that we assume a constant p, and therefore the steady-state distribution is independent of the jump function. Hence, we can use the formula for π_i from Proposition 1. The work then lies in computing the c_i, which are dependent on the choice of jump function, recalling their definition in (13). From the same equation, we see that the first step to compute c_i is to find an explicit solution for the function $F^{(i)}$. The linear jump function (15) allows us to do this by solving the recurrence relation (4). The explicit form of $F^{(i)}$ is provided in the following proposition (see Appendix D.2 for a proof).

Proposition 2. *Consider a 1G1B system with $J(F, \rho_{\text{new}}) = aF + b$ and $F^{(0)}(t_0) = D_{t_0}(F_{\text{new}})$, where F_{new} is the fidelity of the state ρ_{new}. Then,*

$$F^{(i)}(t_0, ..., t_{i-1}, t_i) = \frac{1}{4} + \sum_{j=0}^{i} m_j^{(i)} e^{-\Gamma(t_j + t_{j+1} ... + t_i)} \tag{16}$$

where the constants $m_j^{(i)}$ are given by $m_0^{(0)} = F_{\text{new}} - \frac{1}{4}$, and

$$m_j^{(i)} = \begin{cases} a^{i-j}\left(\frac{a}{4} + b - \frac{1}{4}\right), & \text{if } j > 0, \\ a^i\left(F_{\text{new}} - \frac{1}{4}\right) & \text{if } j = 0. \end{cases} \tag{17}$$

for $i > 0$.

In the following Lemma, we use the formula for $F^{(i)}$ (found in Proposition 2) and combine this with Theorem 1 to derive a closed-form expression for c_i, and therefore for the average consumed fidelity.

Lemma 1. *Consider a 1G1B system with $J(F, \rho_{\text{new}}) = aF + b$ and $F^{(0)}(t_0) = D_{t_0}(F_{\text{new}})$, where F_{new} is the fidelity of the state ρ_{new}. Then, the average fidelity after $i \geq 0$ purification rounds is given by*

$$c_i = \frac{1}{4} + \left(F_{\text{new}} - \frac{1}{4}\right) \cdot a^i \gamma^{i+1} + \left(\frac{a}{4} + b - \frac{1}{4}\right) \gamma \frac{1 - a^i \gamma^i}{1 - a\gamma}, \tag{18}$$

where $\alpha = \mu + \lambda q$ and $\gamma = \alpha/(\alpha + \Gamma)$. Moreover, the average consumed fidelity is given by

$$\overline{F}_{\text{linear}} = \frac{\frac{1}{4}\Gamma + b\lambda q p + F_{\text{new}}\left(\mu + \lambda q(1 - p)\right)}{\Gamma + \mu + \lambda q(1 - pa)}. \tag{19}$$

The closed-form solution (19) is obtainable since $\overline{F} = \frac{1}{A}\sum_{i=0}^{\infty} \pi_i c_i$ is a geometric series with the linear jump function, as can be seen from the form of π_i and c_i as found in Proposition 1 and Equation 18. In the following proposition, we see how $\overline{F}$ varies with p and q.

Proposition 3. *The quantity $\overline{F}_{\text{linear}}$ has the following properties:*

(a) $\overline{F}_{\text{linear}}$ is a monotonic function of q;

(b) $\overline{F}_{\text{linear}}$ is a monotonic function of p;

We provide a proof of Lemma 1 and Proposition 3 in Appendix D.2. We now have closed-form expression for A and $\overline{F}_{\text{linear}}$, which allows for a thorough analysis of the performance of the 1G1B system with the linear jump function. In particular, the following conclusions may already be drawn.

- Result (a) from Proposition 3 implies that the average consumed fidelity is maximized for $q = 0$ or $q = 1$. Consider a 1G1B system with a fixed set of parameters and a pumping scheme with a linear jump function. If the pumping protocol is good enough (e.g. when $b \geq F_{\text{new}}(1-a)$, as explained in Appendix D.2), then pumping every time a link is generated ($q = 1$) maximises the average consumed fidelity. Sometimes, the pumping protocol chosen may impact the average consumed fidelity negatively and in that case one should never pump entanglement ($q = 0$) to increase the average consumed fidelity.

- Result (b) from Proposition 3 provides similar insights: a pumping protocol with a good jump function always benefits from a larger probability of success, i.e. $\overline{F}_{\text{linear}}$ is maximized for $p = 1$. When the protocol is detrimental, failure ($p = 0$) benefits the overall procedure, since it frees the good memory and allows for a fresh entangled link to be allocated there.

When the jump function is good (i.e. when $\overline{F}_{\text{linear}}$ is monotonically increasing in q), we observe a trade-off between $\overline{F}_{\text{linear}}$ and the availability A, which is a decreasing function of q, as can be seen from (9). This behaviour is shown in Figure 6. If we rarely purify (small q), a low-quality entangled state (small $\overline{F}_{\text{linear}}$) will be available most of the time (large A). In that case, the average consumed fidelity can be lower than the fidelity of newly generated links, since the entanglement is not being purified often enough to compensate the noise introduced by the memory over time (in Figure 6, the average consumed fidelity is below the dashed line for small q). When purification is performed more often (larger q), the quality of the stored entanglement will be higher (larger $\overline{F}_{\text{linear}}$), at the expense of a more limited availability (smaller A), since purification can fail and destroy the entanglement stored in the long-term memory. This trade-off disappears when the pumping scheme is deterministic ($p = 1$): the availability remains constant when varying q since purification will always succeed and the stored entanglement will not be destroyed. Note that, if the system is dominated by decoherence ($\Gamma \gg \lambda, \mu$), the average consumed fidelity will always be smaller than F_0.

As a validation check, we also implemented a Monte Carlo simulation of the 1G1B system, which provided the same availability and average consumed fidelity that we obtained analytically (our code is available at https://github.com/AlvaroGI/buffering-1G1B).

5.1 Operating regimes of bilocal Clifford protocols

In this Subsection, we study the operating regimes of the 1G1B system, under the assumption that the pumping protocol employed is a bilocal Clifford protocol [10, 25]. Firstly, we find upper and lower bounds for the availability. Then, for a desired value of the availability within these bounds, we find lower and upper bounds for the average consumed fidelity that can be provided by bilocal Clifford protocols. This analysis finds limits to the performance of the 1G1B buffering system.

Bilocal Clifford protocols are one of the most well-studied types of protocol [10, 25, 16]. One of their main advantages is that they are relatively simple to execute, since they
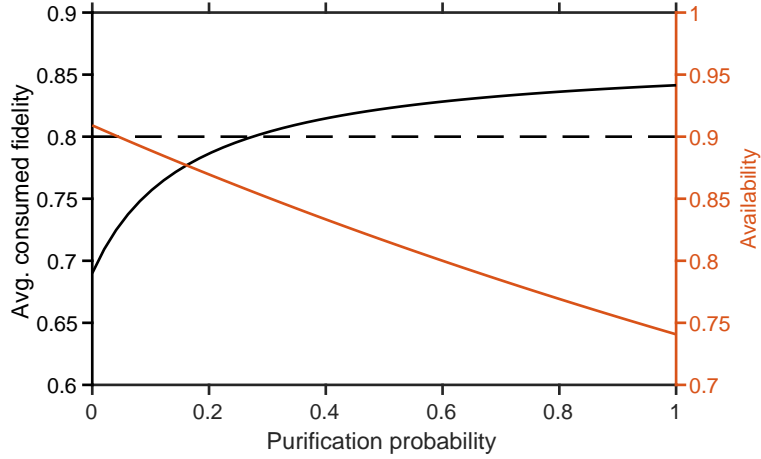
Figure 6: Trade-off between average consumed fidelity and availability. When the pumping is good enough (see discussion in main text), the average consumed fidelity $\overline{F}$ (black line) increases with increasing purification probability q, while the availability A (orange line) decreases. The dashed line corresponds to the fidelity of newly generated links ($F_{\text{new}} = 0.8$). Other parameters used in this example (times and rates in the same arbitrary units): $\lambda = 1$, $\mu = 0.1$, $p = 0.75$, $\Gamma = 1/40$, $J(F, \rho_{\text{new}}) = (1/3)F + (1 + F_{\text{new}})/3$, $\rho_{\text{new}} = F_{\text{new}} |\phi^+\rangle\langle\phi^+| + (1 - F_{\text{new}}) \left(|\psi^+\rangle\langle\psi^+| + |\psi^-\rangle\langle\psi^-| + |\phi^-\rangle\langle\phi^-| \right)/3$. This jump function corresponds to a linear approximation of a specific bilocal Clifford protocol (the DEJMPS protocol) in the high-fidelity regime [11].

involve a basic set of gates. To the best of our knowledge, bilocal Clifford circuits have been the only purification protocols implemented experimentally so far (see, e.g. [26, 43]). In Appendix E we provide further details on bilocal Clifford protocols.

Let us start our performance analysis by discussing the availability. The maximum value that can be achieved by any protocol (bilocal Clifford or not) is $\lambda/(\lambda + \mu)$, as can be seen from (9). This maximum value is obtained when there is no pumping or the pumping protocol succeeds deterministically, i.e. when $q = 0$ or $p = 1$. The availability is lower bounded by $\lambda/(2\lambda + \mu)$, and the minimum value is attained when a pumping protocol is always applied and it never succeeds, i.e., when $q = 1$ and $p = 0$.

To find bounds for the average consumed fidelity, we first need to bound the jump functions of all bilocal Clifford protocols, which we do in the following Lemma. We only consider nontrivial protocols, i.e. we do not consider bilocal Clifford protocols with $J(F, \rho_{\text{new}}) = F$ or $J(F, \rho_{\text{new}}) = F_{\text{new}}$, where F_{new} is the fidelity of ρ_{new}. The former trivial jump function corresponds to a protocol that leaves the buffered link untouched, while the second trivial jump function corresponds to a protocol that replaces the buffered link by the newly generated link.

Lemma 2. *Let $J(F, \rho_{\text{new}})$ be the jump function of a nontrivial bilocal Clifford protocol ($J(F, \rho_{\text{new}}) \neq F$ and $J(F, \rho_{\text{new}}) \neq F_{\text{new}}$, where F_{new} is the fidelity of ρ_{new}). Assume ρ_{new} is a Bell-diagonal state:*

$$\rho_{\text{new}} = F_{\text{new}} \left|\Phi^+\right\rangle\!\left\langle\Phi^+\right| + \lambda_1 \left|\Psi^+\right\rangle\!\left\langle\Psi^+\right| + \lambda_2 \left|\Psi^-\right\rangle\!\left\langle\Psi^-\right| + \lambda_3 \left|\Phi^-\right\rangle\!\left\langle\Phi^-\right|, \qquad (20)$$

with $F_{\text{new}} + \lambda_1 + \lambda_2 + \lambda_3 = 1$. Let us define F^* as

$$F^* = \frac{2F_{\text{new}} - 1 + \sqrt{(2F_{\text{new}} - 1)^2 - 2\lambda_{\min}(1 - 2F_{\text{new}} - 2\lambda_{\min})}}{2(2F_{\text{new}} - 1 + 2\lambda_{\min})}, \qquad (21)$$

where $\lambda_{\min} = \min(\lambda_1, \lambda_2, \lambda_3)$. *Then, for all* $F \in [\frac{1}{4}, F^*]$, *the jump function is lower bounded as follows:*

$$a_{\mathrm{l}} F + b_{\mathrm{l}} \leq J(F, \rho_{\mathrm{new}}) \tag{22}$$

where

$$a_{\mathrm{l}} = \frac{2(4F^* - 1)\left[2F_{\mathrm{new}} - (F_{\mathrm{new}} + \lambda_{\min})(F_{\mathrm{new}} + \lambda_{\max})\right] + 4(\lambda_{\max} - \lambda_{\min})(1 - F^*)}{(4F^* - 1)\left[(4F_{\mathrm{new}} + 4\lambda_{\max} - 2)F^* + 2 - F_{\mathrm{new}} - \lambda_{\max}\right]}, \ and$$

$$b_{\mathrm{l}} = \frac{F_{\mathrm{new}} + \lambda_{\max}}{2} - \frac{a_{\mathrm{l}}}{4}, \tag{23}$$

with $\lambda_{\max} = \max\{\lambda_1, \lambda_2, \lambda_3\}$. *For* $F \in [1/4, 1]$, *the jump function is upper bounded as*

$$J(F, \rho_{\mathrm{new}}) \leq a_{\mathrm{u}} F + b_{\mathrm{u}}, \tag{24}$$

with

$$a_{\mathrm{u}} = \frac{4(1 - F_{\mathrm{new}})}{3}, \ and \quad b_{\mathrm{u}} = \frac{4F_{\mathrm{new}} - 1}{3}. \tag{25}$$

Moreover, the success probability of the protocol is bounded by $p_{\mathrm{l}} \leq p \leq p_{\mathrm{u}}$, *where*

$$p_{\mathrm{l}} = \frac{1}{2}, \ and \quad p_{\mathrm{u}} = F_{\mathrm{new}} + \max(\lambda_1, \lambda_2, \lambda_3). \tag{26}$$

A proof of Lemma 2 can be found in Appendix E. We show this by considering properties of the jump functions of bilocal Clifford protocols, which may be found explicitly. Note that, despite the fact that we assume that newly generated entangled links are Bell-diagonal, other forms of the density matrix are also valid in practice, since they can be brought to Bell-diagonal form by adding extra noise [4, 21]. Note also that the lower bound for the jump function (22) only applies when the fidelity of the buffered link is below F^*, but this is always the case in the 1G1B system, as shown in Appendix E.

If we regard F_{new} as a fixed parameter, the upper and lower bounds to the jump function (22) and (24) are linear in F, and the bounds to the success probability (26) are constant. It is now possible to find an upper and lower bound for the average consumed fidelity by combining Lemmas 1 and 2, as we do in the following corollary.

Corollary 1. *The average consumed fidelity of the 1G1B system when using any (non-trivial) bilocal Clifford protocol is lower bounded by*

$$\overline{F}_{\mathrm{l}} = \frac{\frac{1}{4}\Gamma + b_{\mathrm{l}}\lambda q p + F_{\mathrm{new}}\left(\mu + \lambda q(1 - p_{\mathrm{l}})\right)}{\Gamma + \mu + \lambda q(1 - p_{\mathrm{l}} a_{\mathrm{l}})}, \tag{27}$$

and upper bounded by

$$\overline{F}_{\mathrm{u}} = \frac{\frac{1}{4}\Gamma + b_{\mathrm{u}}\lambda q p + F_{\mathrm{new}}\left(\mu + \lambda q(1 - p_{\mathrm{u}})\right)}{\Gamma + \mu + \lambda q(1 - p_{\mathrm{u}} a_{\mathrm{u}})} \tag{28}$$

where a_{l}, b_{l}, p_{l}, a_{u}, b_{u}, *and* p_{u}, *are given by (23), (25), and (26).*

Now, we analyse the limits of the performance of the 1G1B system using the bounds on $\overline{F}$ from Corollary 1. Let us start with a 1G1B system with perfect memories, i.e. with $\Gamma = 0$. This corresponds to an ideal situation that we can use as a benchmark: once we introduce noise, the average consumed fidelity will be lower than in this ideal case. Figure 7(a) shows the achievable combinations of average consumed fidelity and availability for $F_{\mathrm{new}} = 0.8$, generation rate $\lambda = 1$, and consumption rate $\mu = 0.1$. Below, we list some important observations that may be drawn from this Figure:

- The regions shaded in grey correspond to **unattainable values** of average fidelity and availability, and they apply to any pumping scheme (bilocal Clifford or not). The average consumed fidelity cannot be larger than the one provided by a hypothetical protocol with jump function $J(F, \rho_{\mathrm{new}}) = 1$ and probability of success $p = 1$, which is applied with probability $q = 1$ (however, such a protocol does not exist).

- The performance of a 1G1B system that uses any **bilocal Clifford protocol** is contained within the **region shaded in blue and yellow**. The yellow/blue line corresponds to a hypothetical protocol with jump function and success probability saturating the lower/upper bounds from (22) and (26). For a fixed target availability, the blue line provides an upper bound on the maximum average consumed fidelity that can be achieved by using bilocal Clifford protocols. Here, we observe again the tradeoff between both performance metrics: if our target availability is very close to the maximum value, we cannot increase the average consumed fidelity beyond F_{new} (dotted line); but as we decrease the desired availability, we can achieve a higher consumed fidelity until we reach a maximum value.

- As a reference, we show the performance of the **replacement protocol** (red star): in such a protocol, every time a new link is generated in the bad memory, the link in the good memory is replaced by the new one, without any form of purification. The replacement protocol is not bilocal Clifford because success is always declared (in bilocal Clifford circuits, success depends on some measurement outcomes [10]). This simple protocol achieves maximum availability, given by $A = \lambda/(\lambda + \mu)$. However, since no purification is performed, this protocol cannot increase the fidelity above the initial value F_{new}. In the absence of decoherence, the replacement protocol is equivalent to applying no purification at all ($q = 0$).

In Figure 7(b), we perform a similar analysis for a 1G1B system in which the good memory has a finite lifetime, i.e. $\Gamma > 0$. This is a more realistic scenario. The following observations may be drawn from this Figure:

- **Imperfect memories decrease the average consumed fidelity** but do not affect the availability. The availability is unaffected by the decoherence experienced by the entangled links, and therefore can take the same range of values as in Figure 7(a).

- The replacement protocol no longer provides an average fidelity F_{new}. Instead, the average fidelity is lower than F_{new} since the quality of the state stored in the good memory decreases over time and is never increased beyond F_{new} due to the absence of purification. However, the **replacement protocol performs better than no pumping at all** ($q = 0$). This is because the system can improve its fidelity every time a new link is produced, instead of waiting for a consumption event.

- In the presence of noise, the lower and upper bounds for bilocal Clifford protocols also shift towards lower values of average fidelity. Both the upper and lower bounds take their minimum value at $q = 0$. This means that, in the presence of noise, any pumping protocol will increase the average consumed fidelity, i.e. **any pumping ($q > 0$) is better than no pumping ($q = 0$)**, even if it succeeds with the lowest-possible probability. This is in contrast to when there is no noise (Figure 7(a)), where the lower bound takes its minimum at $q = 1$ and no such conclusion can be drawn. In fact, this conclusion (any pumping is better than no pumping) always applies when the amount of noise, Γ, is above the following threshold:

$$\Gamma > 4\mu p \frac{F_{\mathrm{new}}(1 - a) - b}{4F_{\mathrm{new}}(1 - p) + (4b + a)p - 1}, \tag{29}$$

where a, b, and p are given by the choice of purification protocol (see (15)). In Appendix D.3 we compute this threshold analytically.
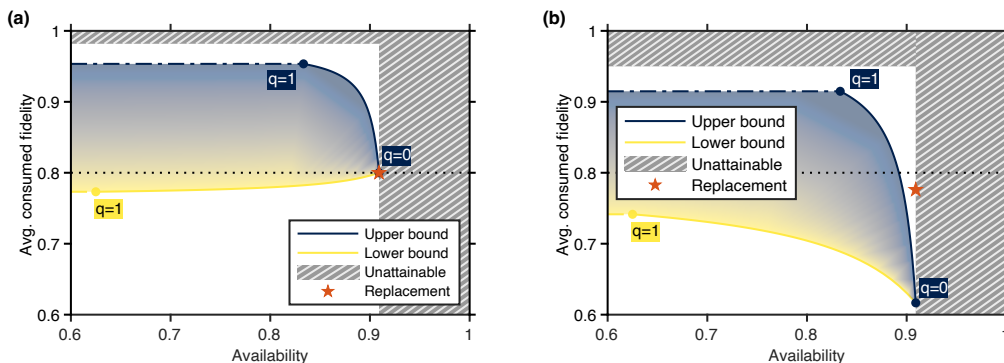

Figure 7: Noise in the memories decreases the average consumed fidelity but does not affect the availability. Bounds on the performance of a 1G1B system with a bilocal Clifford protocol, and with **(a)** noiseless memories ($\Gamma = 0$) or **(b)** noisy memories ($\Gamma = 5 \cdot 10^{-2}$ a.u.). For a given target availability, the average consumed fidelity is within the blue/yellow region (see Corollary 1). Availability is maximized for $q = 0$ (q is the probability of purification after successful entanglement generation), and it decreases for increasing q. White regions cannot be achieved by bilocal Clifford protocols. Striped regions cannot be achieved by any pumping protocol. Red star: performance of the replacement protocol (buffered link is replaced by new links). Dotted line: fidelity of newly generated entangled links. Parameters used in this example (times and rates in the same arbitrary units): $\lambda = 1$, $\mu = 0.1$, $F_{\mathrm{new}} = 0.8$, $\rho_{\mathrm{new}} = F_{\mathrm{new}} |\phi^+\rangle\langle\phi^+| + (1 - F_{\mathrm{new}}) (|\psi^+\rangle\langle\psi^+| + |\psi^-\rangle\langle\psi^-|) / 2$.

6 Conclusions and outlook

Our work sheds light on how to buffer high-quality entanglement shared among remote nodes in a quantum network. We have proposed two metrics to measure the performance of an entanglement buffering system: the availability and the average consumed fidelity. The availability corresponds to the fraction of time in which entanglement is available for consumption. The average consumed fidelity measures the quality of the entanglement upon consumption. We have used these performance metrics to analyse the 1G1B system, an entanglement buffering setup that uses two quantum memories per node. One of these memories has a finite lifetime and is used to buffer the entanglement, while the other memory is only used for entanglement generation. Entanglement generated in the bad memory can be used to pump the entanglement stored in the good memory. We have modelled the system as a continuous-time stochastic process and derived analytical expressions for both performance metrics. Our results confirm the intuition that, except in some edge cases, there is a trade-off between consuming entanglement at a high rate (high availability) and consuming high-quality entanglement (high average consumed fidelity). Remarkably, we found that, in a practical scenario (i.e. when the pumping protocol is bilocal Clifford and there is noise in the good memory), pumping the buffered entanglement is better than no pumping in terms of average consumed fidelity, even if the pumping has some probability of failure.

An assumption that allows us to find analytical solutions for our performance metrics is that the success probability of purification is constant over time. The model would be more realistic if the probability of successful purification was dependent on the state fidelity at that time, since this is the case for most protocols (in particular, the probability

of successful purification is typically lower for input states with lower fidelity). This may mean that, realistically, the computation of the average fidelity when *conditioning* on successful purification may bias the system towards higher fidelity. However, we believe the comparison of our model (constant success probability) with a more realistic one incorporating this effect (success probability dependent on $F(t)$) to be beyond the scope of this work, since we expect this to greatly complicate the analysis of the problem.

Our proposed metrics can be used to evaluate the performance of other entanglement buffering systems. An interesting extension of this work would be to compare the performance of the 1G1B system to a bipartite entanglement buffering setup with n quantum memories per node. In such a system, one could employ more advanced purification protocols that consume more than two entangled states. We also expect that the mathematical framework developed in this work can be used to initiate the performance analysis of more complex systems. We leave this as future work.

Acknowledgements

References

[1] Sean D. Barrett and Pieter Kok. Efficient high-fidelity quantum computation using matter qubits and linear optics. *Physical Review A*, 71:060310, Jun 2005.

[2] Simon C Benjamin, Daniel E Browne, Joe Fitzsimons, and John JL Morton. Brokered graph-state quantum computation. *New Journal of Physics*, 8(8):141, 2006.

[3] Charles H Bennett, Gilles Brassard, Sandu Popescu, Benjamin Schumacher, John A Smolin, and William K Wootters. Purification of noisy entanglement and faithful teleportation via noisy channels. *Physical Review Letters*, 76(5):722, 1996.

[4] Charles H Bennett, David P DiVincenzo, John A Smolin, and William K Wootters. Mixed-state entanglement and quantum error correction. *Physical Review A*, 54(5):3824, 1996.

[5] Hannes Bernien, Bas Hensen, Wolfgang Pfaff, Gerwin Koolstra, Machiel S Blok, Lucio Robledo, Tim H Taminiau, Matthew Markham, Daniel J Twitchen, Lilian Childress, et al. Heralded entanglement between solid-state qubits separated by three metres. *Nature*, 497(7447):86–90, 2013.

[6] Sylvia Bratzik, Silvestre Abruzzo, Hermann Kampermann, and Dagmar Bruß. Quantum repeaters and quantum key distribution: The impact of entanglement distillation on the secret key rate. *Physical Review A*, 87(6):062335, 2013.

[7] Aparimit Chandra, Wenhan Dai, and Don Towsley. Scheduling quantum teleportation with noisy memories. In *2022 IEEE International Conference on Quantum Computing and Engineering (QCE)*, pages 437–446. IEEE, 2022.

[8] Richard Cleve, Daniel Gottesman, and Hoi-Kwong Lo. How to share a quantum secret. *Physical Review Letters*, 83(3):648, 1999.

[9] Bethany Davies, Thomas Beauchamp, Gayane Vardoyan, and Stephanie Wehner. Tools for the analysis of quantum protocols requiring state generation within a time window. *IEEE Transactions on Quantum Engineering*, 2024.

[10] Jeroen Dehaene, Maarten Van den Nest, Bart De Moor, and Frank Verstraete. Local permutations of products of bell states and entanglement distillation. *Physical Review A*, 67(2):022310, 2003.

[11] David Deutsch, Artur Ekert, Richard Jozsa, Chiara Macchiavello, Sandu Popescu, and Anna Sanpera. Quantum privacy amplification and the security of quantum cryptography over noisy channels. *Physical Review Letters*, 77(13):2818, 1996.

[12] Wolfgang Dür, H-J Briegel, Juan Ignacio Cirac, and Peter Zoller. Quantum repeaters based on entanglement purification. *Physical Review A*, 59(1):169, 1999.

[13] Wolfgang Dür, Marc Hein, J Ignacio Cirac, and H-J Briegel. Standard forms of noisy quantum operations via depolarization. *Physical Review A*, 72(5):052326, 2005.

[14] Artur K Ekert. Quantum cryptography based on Bell's theorem. *Physical Review Letters*, 67(6):661, 1991.

[15] Karim Elsayed, Wasiur R KhudaBukhsh, and Amr Rizk. On the fidelity distribution of link-level entanglements under purification. *arXiv preprint arXiv:2310.18198*, 2023.

[16] Kenneth Goodenough, Sébastian de Bone, Vaishnavi L Addala, Stefan Krastanov, Sarah Jansen, Dion Gijswijt, and David Elkouss. Near-term n to k distillation protocols using graph codes. *IEEE J. on Selected Areas in Comms.*, 42(7):1830-1849, 2024.

[17] Daniel Gottesman. The heisenberg representation of quantum computers. *arXiv preprint quant-ph/9807006*, 1998.

[18] Daniel Gottesman. Theory of fault-tolerant quantum computation. *Physical Review A*, 57(1):127, 1998.

[19] Geoffrey Grimmett and David Stirzaker. *Probability and random processes*. Oxford university press, 2020.

[20] Mark Hillery, Vladimír Bužek, and André Berthiaume. Quantum secret sharing. *Physical Review A*, 59(3):1829, 1999.

[21] Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. General teleportation channel, singlet fraction, and quasidistillation. *Physical Review A*, 60(3):1888, 1999.

[22] Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. Separability of mixed states: necessary and sufficient conditions. *Physics Letters A*, 223(1):1–8, 1996.

[23] Álvaro G Iñesta, Gayane Vardoyan, Lara Scavuzzo, and Stephanie Wehner. Optimal entanglement distribution policies in homogeneous repeater chains with cutoffs. *npj Quantum Information*, 9(1):46, 2023.

[24] Álvaro G Iñesta and Stephanie Wehner. Performance metrics for the continuous distribution of entanglement in multiuser quantum networks. *Physical Review A*, 108(5):052615, 2023.

[25] Sarah Jansen, Kenneth Goodenough, Sébastian de Bone, Dion Gijswijt, and David Elkouss. Enumerating all bilocal clifford distillation protocols through symmetry reduction. *Quantum*, 6:715, 2022.

[26] Norbert Kalb, Andreas A Reiserer, Peter C Humphreys, Jacob JW Bakermans, Sten J Kamerling, Naomi H Nickerson, Simon C Benjamin, Daniel J Twitchen, Matthew Markham, and Ronald Hanson. Entanglement distillation between solid-state quantum network nodes. *Science*, 356(6341):928–932, 2017.

[27] Yuan Lee, Eric Bersin, Axel Dahlberg, Stephanie Wehner, and Dirk Englund. A quantum router architecture for high-fidelity entanglement flows in quantum networks. *npj Quantum Information*, 8(1):75, 2022.

[28] Boxi Li, Tim Coopmans, and David Elkouss. Efficient optimization of cut-offs in quantum repeater chains. In *2020 IEEE International Conference on Quantum Computing and Engineering (QCE)*, pages 158–168. IEEE, 2020.

[29] Jian-Long Liu, Xi-Yu Luo, Yong Yu, Chao-Yang Wang, Bin Wang, Yi Hu, Jun Li, Ming-Yang Zheng, Bo Yao, Zi Yan, et al. A multinode quantum network over a metropolitan area. *arXiv preprint arXiv:2309.00221*, 2023.

[30] Kosto V Mitov, Edward Omey, Kosto V Mitov, and Edward Omey. *Renewal processes.* Springer, 2014.

[31] Philippe Nain, Gayane Vardoyan, Saikat Guha, and Don Towsley. On the analysis of a multipartite entanglement distribution switch. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 4(2):1–39, 2020.

[32] Michael A Nielsen and Isaac Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition.* Cambridge University Press, 2010.

[33] Asher Peres. Separability criterion for density matrices. *Physical Review Letters*, 77(8):1413, 1996.

[34] Matteo Pompili, Sophie LN Hermans, Simon Baier, Hans KC Beukers, Peter C Humphreys, Raymond N Schouten, Raymond FL Vermeulen, Marijn J Tiggelman, Laura dos Santos Martins, Bas Dirkse, et al. Realization of a multinode quantum network of remote solid-state qubits. *Science*, 372(6539):259–264, 2021.

[35] Shahrooz Pouryousef, Nitish K Panigrahy, and Don Towsley. A quantum overlay network for efficient entanglement distribution. In *IEEE INFOCOM 2023-IEEE Conference on Computer Communications*, pages 1–10. IEEE, 2023.

[36] Ludmiła Praxmeyer. Reposition time in probabilistic imperfect memories. *arXiv preprint arXiv:1309.3407*, 2013.

[37] Evgeny Shchukin, Ferdinand Schmidt, and Peter van Loock. Waiting time in quantum repeaters with probabilistic entanglement swapping. *Physical Review A*, 100(3):032322, 2019.

[38] Piet Van Mieghem. *Performance analysis of complex networks and systems.* Cambridge University Press, 2014.

[39] Gayane Vardoyan, Philippe Nain, Saikat Guha, and Don Towsley. On the capacity region of bipartite and tripartite entanglement switching. *ACM Transactions on Modeling and Performance Evaluation of Computing Systems*, 8(1-2):1–18, 2023.

[40] Michelle Victora, Spyros Tserkis, Stefan Krastanov, Alexander Sanchez de la Cerda, Steven Willis, and Prineha Narang. Entanglement purification on quantum networks. *Physical Review Research*, 5(3):033171, 2023.

[41] Stephanie Wehner, David Elkouss, and Ronald Hanson. Quantum internet: A vision for the road ahead. *Science*, 362(6412):eaam9288, 2018.

[42] Reinhard F Werner. Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model. *Physical Review A*, 40(8):4277, 1989.

[43] Haoxiong Yan, Youpeng Zhong, Hung-Shen Chang, Audrey Bienfait, Ming-Han Chou, Christopher R Conner, Étienne Dumur, Joel Grebel, Rhys G Povey, and Andrew N Cleland. Entanglement purification and protection in a superconducting quantum network. *Physical Review Letters*, 128(8):080504, 2022.

[44] Pei-Shun Yan, Lan Zhou, Wei Zhong, and Yu-Bo Sheng. Advances in quantum entanglement purification. *Science China Physics, Mechanics & Astronomy*, 66(5):250301, 2023.

A Quantum preliminaries

Here, we introduce the quantum information concepts needed in this work. For a general reference on quantum information theory, see, e.g., ref. [32].

In classical computer science, information is generally stored in the form of *bits*, discrete binary variables that can take value zero or one. In quantum information, bits are generalized to *qubits*. Qubits describe systems that can be in a linear combination of two different states (say, state zero and state one). As a consequence, a qubit must not be described as a binary variable but as a vector $|\psi\rangle \in \mathbb{C}^2$ with unit norm. We represent these vectors using the Dirac notation, and we refer to $|\cdot\rangle$ as a ket and to its conjugate transpose $\langle\cdot| \equiv |\cdot\rangle^\dagger$ as a bra. The state of a *pure n*-qubit system is then described by a d-dimensional vector $|\psi\rangle \in \mathbb{C}^d$, with $d = 2^n$. The elements of the basis of $\mathbb{C}^d$ are usually labeled $|\mathbf{x}\rangle$, with $\mathbf{x} \in \{0,1\}^n$. For example, the basis of the two-qubit space $\mathbb{C}^4$ can be written as $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$.

Let us now consider a two-qubit system $|\psi\rangle \in \mathbb{C}^4$. When one of the qubits is in state $|\psi\rangle_1 \in \mathbb{C}^2$ and the other qubit is in state $|\psi\rangle_2 \in \mathbb{C}^2$, we can write the joint state of the system as the tensor product of the individual qubits: $|\psi\rangle = |\psi\rangle_1 \otimes |\psi\rangle_2$. This is called a *product state*. When the two qubits are *entangled*, it is not possible to describe their joint state as a tensor product. One of the intuitive effects of the entanglement is that, if both qubits are measured, the measurement outcomes will be correlated. This is the basis for many quantum networking applications, such as quantum key distribution [14] and quantum secret sharing [20, 8].

Noise and quantum operations, such as *gates* and *measurements*, modify the state of a quantum system. When dealing with noisy systems, it is generally more convenient to represent quantum states using the *density matrix* formalism instead of kets and bras. The density matrix of a *pure* state $|\psi\rangle$ can be written as the outer product $\rho = |\psi\rangle\langle\psi|$. When the system is not pure, it is called *mixed*, and it is written as a mixture of pure states: $\rho = \sum_i \alpha_i |\psi_i\rangle\langle\psi_i|$, with $\alpha_i \in [0,1]$ and $\sum_i \alpha_i = 1$. Intuitively, this corresponds to a system with some degree of uncertainty from a classical point of view. For example, consider a device that prepares the state $|\psi_1\rangle$ with probability α and the state $|\psi_2\rangle$ with probability $1 - \alpha$. We can write the output state of the device as a mixed state $\alpha |\psi_1\rangle\langle\psi_1| + (1 - \alpha) |\psi_2\rangle\langle\psi_2|$.

We refer to a two-qubit state ρ as an *entangled link* (in our setup, each qubit is located at a distant node, and the entanglement can be regarded as "link" connecting them). Entangled links can be entangled to different degrees. The *Bell states* are examples of two-qubit maximally entangled (pure) states:

$$\left|\phi^+\right\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}, \ \left|\psi^+\right\rangle = \frac{|01\rangle + |10\rangle}{\sqrt{2}}, \ \left|\psi^-\right\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}, \ \left|\phi^-\right\rangle = \frac{|00\rangle - |11\rangle}{\sqrt{2}}. \tag{30}$$

Measurements on the individual qubits of any of these states yield maximally correlated outcomes. All two-qubit maximally entangled states are equivalent, since they can be mapped via single-qubit operations to each other. Therefore, we can measure the *quality of the entanglement* of a two-qubit state by measuring how close the state is to one of the Bell states, say $|\phi^+\rangle$. Formally, we do this using the *fidelity* of the state:

$$F(\rho) = \left\langle \Phi^+ \middle| \rho \middle| \Phi^+ \right\rangle, \tag{31}$$

where ρ is the density matrix of an arbitrary two-qubit state.

Lastly, an important type of mixed entangled state is the *Bell-diagonal state*:

$$\rho_{\mathrm{BD}} = F_{\mathrm{BD}} \left|\phi^+\right\rangle\!\left\langle\phi^+\right| + \lambda_1 \left|\psi^+\right\rangle\!\left\langle\psi^+\right| + \lambda_2 \left|\psi^-\right\rangle\!\left\langle\psi^-\right| + \lambda_3 \left|\phi^-\right\rangle\!\left\langle\phi^-\right|,$$

with $F_{\mathrm{BD}}, \lambda_1, \lambda_2, \lambda_3 \in [0,1]$ subjected to the normalization constraint $F_{\mathrm{BD}}+\lambda_1+\lambda_2+\lambda_3 = 1$. The fidelity of this state is F_{BD}. Bell-diagonal states are relevant because any two-qubit state can be transformed into Bell-diagonal form while preserving the fidelity by applying extra noise, a process known as *twirling* [4, 21]. A specific instance of Bell-diagonal state is the Werner state [42]:

$$\rho_{\mathrm{w}} = F \left|\phi^+\right\rangle\!\left\langle\phi^+\right| + \frac{1-F}{3} \left|\psi^+\right\rangle\!\left\langle\psi^+\right| + \frac{1-F}{3} \left|\psi^-\right\rangle\!\left\langle\psi^-\right| + \frac{1-F}{3} \left|\phi^-\right\rangle\!\left\langle\phi^-\right|, \qquad (32)$$

with fidelity $F \in [0,1]$. The Werner state corresponds to a maximally entangled state that has been subjected to isotropic noise.

B   General form of jump function

In this Appendix, we explain the form (1) and (2) of the jump function and success probability for a general purification protocol, for two input states ρ_{W} and ρ_{new}, where

$$\rho_{\text{W}} = F \left|\phi^+\right\rangle\!\left\langle\phi^+\right| + \frac{1-F}{3} \left|\psi^+\right\rangle\!\left\langle\psi^+\right| + \frac{1-F}{3} \left|\psi^-\right\rangle\!\left\langle\psi^-\right| + \frac{1-F}{3} \left|\phi^-\right\rangle\!\left\langle\phi^-\right|$$

is a Werner state and ρ_{new} is a general two-qubit state. Suppose that the purification protocol is described by a sequence of (possibly noisy) quantum operations that are described by a CPTP map Λ, and the final measurement outcome that signals success has measurement operator M_{succ}. From e.g. Chapter 2.4 of [32], the output state is then given by

$$\rho' = \frac{M_{\text{succ}}\Lambda\left(\rho_{\text{W}} \otimes \rho_{\text{new}}\right) M_{\text{succ}}^\dagger}{p(F, \rho_{\text{new}})}, \tag{33}$$

where

$$p(F, \rho_{\text{new}}) = \text{Tr}\left[M_{\text{succ}}\Lambda\left(\rho_{\text{W}} \otimes \rho_{\text{new}}\right) M_{\text{succ}}^\dagger\right]. \tag{34}$$

We next rewrite the Werner state as

$$\begin{aligned}
\rho_{\text{W}} &= F \left|\phi^+\right\rangle\!\left\langle\phi^+\right| + (1-F)\rho^\perp \\
&= \rho^\perp + F \left(\left|\phi^+\right\rangle\!\left\langle\phi^+\right| - \rho^\perp\right)
\end{aligned}$$

where

$$\rho^\perp = \frac{1}{3}\left(\left|\psi^+\right\rangle\!\left\langle\psi^+\right| + \left|\psi^-\right\rangle\!\left\langle\psi^-\right| + \left|\phi^-\right\rangle\!\left\langle\phi^-\right|\right),$$

and p is the probability of success. We therefore have

$$\begin{aligned}
M_{\text{succ}}\Lambda\left(\rho_{\text{W}} \otimes \rho_{\text{new}}\right) M_{\text{succ}}^\dagger &= M_{\text{succ}}\Lambda\left(\rho^\perp \otimes \rho_{\text{new}}\right) M_{\text{succ}}^\dagger \\
&\quad + F \cdot M_{\text{succ}}\Lambda\left(\left(\left|\phi^+\right\rangle\!\left\langle\phi^+\right| - \rho^\perp\right) \otimes \rho_{\text{new}}\right) M_{\text{succ}}^\dagger,
\end{aligned}$$

and taking the trace of the above yields

$$p(F, \rho_{\text{new}}) = d(\rho_{\text{new}}) + F \cdot c(\rho_{\text{new}}),$$

where c and d are obtained from the choice of purification protocol, i.e. from Λ and M_{succ}. Similarly, the output fidelity of upon success is given by

$$\left\langle\phi^+\right| \rho' \left|\phi^+\right\rangle = \frac{F \cdot \tilde{a}(\rho_{\text{new}}) + \tilde{b}(\rho_{\text{new}})}{p(F, \rho_{\text{new}})},$$

where

$$\begin{aligned}
\tilde{a}(\rho_{\text{new}}) &= \left\langle\phi^+\right| M_{\text{succ}}\Lambda\left(\left(\left|\phi^+\right\rangle\!\left\langle\phi^+\right| - \rho^\perp\right) \otimes \rho_{\text{new}}\right) M_{\text{succ}}^\dagger \left|\phi^+\right\rangle, \\
\tilde{b}(\rho_{\text{new}}) &= \left\langle\phi^+\right| M_{\text{succ}}\Lambda\left(\rho^\perp \otimes \rho_{\text{new}}\right) M_{\text{succ}}^\dagger \left|\phi^+\right\rangle.
\end{aligned}$$

This confirms the form (1) and (2) for the jump function and success probability.

C Formulae for performance metrics

In this Appendix, we prove Proposition 1 and Theorem 1, which provide the formulae for our two performance metrics (availability and average consumed fidelity). First, in C.1, we describe the stochastic process in the 1G1B setup in a simplified form and we provide some intermediate results that are necessary for the main proofs. Then, in C.2, we employ the results from C.1 to prove Proposition 1 and Theorem 1.

C.1 Simplified 1G1B

We now only view the 1G1B system as taking one of two states: $\emptyset$ (no entangled link in memory G), or $\neg\emptyset$ (link in memory G). The system then alternates between these two states. For an illustration, see Figure 8.
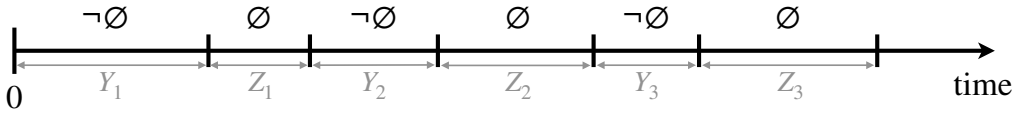

Figure 8: The simplified 1G1B process. The system alternates between the states $\neg\emptyset$ (link in memory G) and $\emptyset$ (no link in memory G). The system starts in $\neg\emptyset$. The times spent in $\neg\emptyset$ and $\emptyset$ are denoted by Y_i and Z_i, respectively.

More formally, the simplified 1G1B process is the following.

Definition C.1 (Simplified 1G1B). Let $r(t) \in \{\emptyset, \neg\emptyset\}$ denote the state of simplified 1G1B at time t. Suppose that $r(0) = \neg\emptyset$, i.e. the system starts when there is a link in memory F. Let Y_1 be the time until this first link is removed, and let Z_1 be the time for which the system is empty until a fresh link is produced again. Let $\{Y_i\}_{i\geq 1}$ be the times spent in $\neg\emptyset$ until the link was removed from memory G (due to consumption or failed purification), and $\{Z_i\}_{i\geq 1}$ be the times which the system spent in $\emptyset$ until a link was produced. Then, according to our model of 1G1B, the Y_i are i.i.d. and exponentially distributed with rate $\beta := \mu + \lambda q(1 - p)$, and the Z_i are i.i.d and exponentially distributed with rate λ.

Recall that λ is the rate of generation of new entangled links, μ is the rate of consumption of links in memory G, q is the probability of immediately using new links for pumping, and p is the probability of successful pumping.

We will write the distribution functions as $F_Y(t) = P(Y_1 \leq t) = 1 - e^{-\beta t}$, and $F_Z(t) = P(Z_1 \leq t) = 1 - e^{-\lambda t}$. The process $X_i := Y_i + Z_i$ defines a *renewal process*, which we introduce with the following definition.

Definition C.2. A *renewal process* $\{N = N(t) : t \geq 0\}$ is a process such that

$$N(t) = \max\{n : A_n \leq t\} \tag{35}$$

where $A_0 = 0$, $A_n = X_1 + ... + X_n$ for $n \geq 1$, and X_i is a sequence of i.i.d. and strictly positive random variables.

The value A_n is referred to as the nth *arrival time* of the process, and the values X_i are known as the *interarrival times*. From now on, we also use $A_0 = 0$, $A_n = X_1 + ... + X_n$ to denote the nth time at which a fresh link is produced, causing the system to move from $\emptyset$ into $\neg\emptyset$.

The *renewal function* is central to renewal theory, which we define below. Throughout, we use the convention $dg(x) \equiv g'(x)dx$ for differentiable functions g.

Definition C.3. Let $N(t)$ be a renewal process. Then, the *renewal function* is $m(t) :=$ $\mathbb{E}[N(t)]$.

We will derive formulae for the availability and average consumed fidelity using this mathematical framework. An important result that we will use in order to do this is the renewal theorem, which we state below. This result assumes that the X_i are not *arithmetic*. If X_1 is arithmetic, this essentially means that X_1 only takes values in a set $\{mk : m = 0, \pm 1, ...\}$, with $k > 0$. For more details of arithmetic random variables, see Chapter 10 of [19].

Theorem 2 (Renewal Theorem/ Theorem 10.1.11 from [19])**.** *Consider a renewal process as given in Definition C.2. Let F_X be the distribution function of the random variable X_1, where X_1 is not arithmetic. Let $H(t)$ be a bounded function. Consider solutions f to the renewal-type equation*

$$f(t) = H(t) + \int_0^t f(t-x)\mathrm{d}F_X(x). \tag{36}$$

Then, a solution is

$$f(t) = H(t) + \int_0^t H(t-x)\mathrm{d}m(x). \tag{37}$$

If H is bounded on finite intervals then f is bounded on finite intervals, and (37) is the unique solution of (36) with this property.

The renewal-type equation often arises when studying renewal processes, as we will see further on. The following result may be derived using Theorem 2, and is useful when taking the infinite limit.

Theorem 3 (Key renewal theorem/Theorem 11.2.7 from [19])**.** *If $g : [0, \infty) \to [0, \infty)$ is such that*

(a) *$g(t) \geq 0$ for all t,*

(b) *$\int_0^\infty g(t)\mathrm{d}t < \infty$,*

(c) *g is a non-increasing function,*

then

$$\lim_{t \to \infty} \int_0^t g(t-x)\mathrm{d}m(x) = \frac{1}{\mathbb{E}[X_1]} \int_0^\infty g(x)\mathrm{d}x,$$

whenever X_1 is not arithmetic.

We are now partially equipped to show the formulae for the availability and average fidelity. Next, we show a set of intermediate results that we will need for the main proofs.

Proposition 4. *Let $p(t) = P(r(t) = \neg\emptyset)$ be the probability that a link is available at time t in the simplified 1G1B process. Then,*

$$\lim_{t \to \infty} p(t) = \frac{\mathbb{E}(Y_1)}{\mathbb{E}(Y_1) + \mathbb{E}(Z_1)}. \tag{38}$$

Proof. We proceed by conditioning on the value of X_1. Now,

$$p(t) = P(r(t) = \neg\emptyset \cap X_1 > t) + P(r(t) = \neg\emptyset \cap X_1 < t). \tag{39}$$

Notice that the event $\{r(t) = \neg\emptyset \cap X_1 > t\}$ occurs if and only if $Y_1 > t$. Further, if $x < t$, then

$$P(r(t) = \neg\emptyset | X_1 = x) = p(t - x), \tag{40}$$

since the process starts afresh at time x. Then, (39) becomes

$$p(t) = 1 - F_Y(t) + \int_0^t p(t - x) \mathrm{d}F_X(x), \tag{41}$$

where $\mathrm{d}F_X(x) \equiv F_X'(x)\mathrm{d}x$. We now see that this is of the form (36) with $H(t) = 1 - F_Y(t)$, and so by Theorem 2,

$$p(t) = 1 - F_Y(t) + \int_0^t \left(1 - F_Y(t - x)\right) \mathrm{d}m(x). \tag{42}$$

Taking the infinite limit,

$$\lim_{t\to\infty} p(t) = 1 - 1 + \lim_{t\to\infty} \int_0^t (1 - F_Y(t - x))\mathrm{d}m(x). \tag{43}$$

It can be seen that $H(t) = 1 - F_Y(t)$ satisfies the conditions (a)-(c) required by Theorem 3, so we may apply this Theorem to take the limit:

$$\lim_{t\to\infty} p(t) = \frac{1}{\mathbb{E}[X_1]} \int_0^\infty (1 - F_Y(x))\mathrm{d}x \tag{44}$$

$$= \frac{1}{\mathbb{E}[X_1]} \int_0^\infty P(Y_1 > x)\mathrm{d}x = \frac{\mathbb{E}[Y_1]}{\mathbb{E}[X_1]}. \tag{45}$$

Finally, using $\mathbb{E}[X_1] = \mathbb{E}[Y_1 + Z_1] = \mathbb{E}[Y_1] + \mathbb{E}[Z_1]$ suffices to show (38). $\qquad\square$

Recall that the average fidelity of the system at a given time t is dependent on the time spent in each purification level leading up to this point. Therefore, in order to understand the average fidelity we first of all look at the *current lifetime* in this simplified setting.

Definition C.4 (Current lifetime). Consider the simplified $1G1B$ system. Let $C(t)$ be the time spent so far in a state at time t. More formally,

$$C(t) = \begin{cases} t - A_{N(t)}, & \text{if } r(t) = \neg\emptyset, \\ t - A_{N(t)} - Y_{N(t)+1}, & \text{if } r(t) = \emptyset. \end{cases} \tag{46}$$

The first case $(r(t) = \neg\emptyset)$ is of most interest here, because it corresponds to when a link is in memory and is subject to decoherence. See Figure 9 for an illustration of this concept. In the following Lemma, we characterise the distribution of $C(t)$, conditional on being in the state $\neg\emptyset$.

Lemma 3. *Consider the simplified $1G1B$ system. The limiting distribution of $C(t)$ conditional on there being a link is given by*

$$\lim_{t\to\infty} P\left(C(t) > x | r(t) = \neg\emptyset\right) = \frac{1}{\mathbb{E}[Y_1]} \int_x^\infty (1 - F_Y(s))\mathrm{d}s, \tag{47}$$

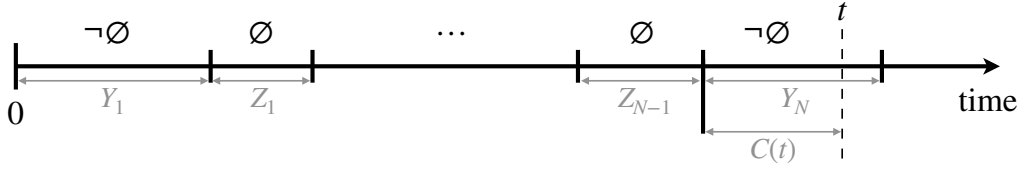*which is an exponential distribution with parameter β when $Y_1 \sim Exp(\beta)$.*

Figure 9: Current lifetime of the simplified 1G1B process. The random variable $C(t)$ denotes the time spent so far in the current state at time t. This is most interesting when $r(t) = \neg\emptyset$, because it tells us the age of a link in memory.

Proof. Writing

$$P(C(t) > x | r(t) = \neg\emptyset) = \frac{P(C(t) > x \cap r(t) = \neg\emptyset)}{P(r(t) = \neg\emptyset)}, \tag{48}$$

we see that the bottom of the fraction has already been dealt with in Proposition 4. We therefore focus on

$$G(t, x) := P(C(t) > x \cap r(t) = \neg\emptyset). \tag{49}$$

Conditioning on X_1, we see that

$$G(t, x) = P(C(t) > x \cap r(t) = \neg\emptyset \cap X_1 > t) + P(C(t) > x \cap r(t) = \neg\emptyset \cap X_1 \leq t). \tag{50}$$

Now, the event $\{C(t) > x \cap r(t) = \neg\emptyset \cap X_1 > t\}$ occurs if and only if $Y_1 > t > x$. Moreover, if $y < t$ then the process starts afresh from time y, and

$$P(C(t) > x \cap r(t) = \neg\emptyset | X_1 = y) = G(t - y). \tag{51}$$

Then, noting that $G(t, x) = 0$ for $t < x$, (50) becomes

$$G(t, x) = \mathbb{1}_{\{t \geq x\}}(1 - F_Y(t)) + \int_0^t G(t - y) \mathrm{d}F_X(y), \tag{52}$$

which is in the form of (36) with $H(t) = \mathbb{1}_{\{t \geq x\}}(1 - F_Y(t))$. Then, by Theorem 2, $G(t, x)$ is given by

$$G(t, x) = \mathbb{1}_{\{t \geq x\}}(1 - F_Y(t)) + \int_0^t \mathbb{1}_{\{t - y \geq x\}}(1 - F_Y(t - y)) \mathrm{d}m(y) \tag{53}$$

which has limit

$$\lim_{t \to \infty} G(t, x) = 0 + \lim_{t \to \infty} \int_0^{t - x} (1 - F_Y(t - y)) \mathrm{d}m(y) \tag{54}$$

$$= \lim_{s \to \infty} \int_0^s (1 - F_Y(s + x - y)) \mathrm{d}m(y), \tag{55}$$

letting $s = t - x$. Then, noting that $g(s) = 1 - F_Y(s + x)$ satisfies conditions (a)-(c) of Theorem 3, we may apply this to find

$$\lim_{t \to \infty} G(t, x) = \frac{1}{\mathbb{E}[X_1]} \int_0^\infty g(s) \mathrm{d}s = \frac{1}{\mathbb{E}[X_1]} \int_0^\infty (1 - F_Y(s + x)) \mathrm{d}s \tag{56}$$

$$= \frac{1}{\mathbb{E}[X_1]} \int_x^\infty (1 - F_Y(s)) \mathrm{d}s. \tag{57}$$

From Proposition 4, we observe that

$$\mathbb{E}[X_1] = \frac{\mathbb{E}[Y_1]}{\lim_{t \to \infty} P(r(t) = \neg\emptyset)}.$$

We can use this to rewrite (57) as follows:

$$\lim_{t\to\infty} P\left(C(t) > x | r(t) = \neg\emptyset\right) = \frac{1}{\mathbb{E}[Y_1]} \int_x^\infty (1 - F_Y(s))\mathrm{d}s, \tag{58}$$

which we notice is only dependent on the distribution of Y_1. In the case $Y_1 \sim \mathrm{Exp}(\beta)$, as considered in the 1G1B system,

$$\lim_{t\to\infty} P\left(C(t) > x | r(t) = \neg\emptyset\right) = \beta \int_x^\infty e^{-\beta s}\mathrm{d}s = e^{-\beta x}, \tag{59}$$

and so conditional on there being a link, the current lifetime approaches an exponential distribution. $\qquad\square$

We have now characterised the availability (Proposition 4) and current lifetime (Lemma 3) for the simplified 1G1B system. However, note that both Proposition 4 and Lemma 3 assumed that the system starts in the state $r(0) = \neg\emptyset$. This was necessary in order to satisfy all of the conditions (a)-(c) of Theorem 3. The result below states that Theorem 3 still holds, even if the renewal process is *delayed*, which means that the first arrival has a different distribution to the others. For more details of delayed renewal processes, see [19] or [30].

Definition C.5. Let $\{X_i\}_{i\geq 1}$ be independent positive random variables such that $\{X_i\}_{i\geq 2}$ have the same distribution. Let $A_0 = 0$, $A_n = \sum_{i=1}^n X_i$, and $N^\mathrm{d} = \max\{n : A_n \leq t\}$. Then, $N^\mathrm{d}(t)$ is a delayed renewal process.

Definition C.6. Let N^d be a delayed renewal process. Then, $m^\mathrm{d}(t) := \mathbb{E}[N^\mathrm{d}(t)]$ is the delayed renewal function.

Theorem 4 (Key renewal theorem for delayed renewal processes/Theorem 1.20 of [30])**.** *Consider a delayed renewal process $N^\mathrm{d}(t)$. If $g : [0, \infty) \to [0, \infty)$ satisfies the same conditions (a)-(c) of Theorem 3, then*

$$\lim_{t\to\infty} \int_0^t g(t-x)\mathrm{d}m^\mathrm{d}(x) = \frac{1}{\mathbb{E}[X_2]} \int_0^\infty g(x)\mathrm{d}x. \tag{60}$$

A consequence of Theorem 4 is that even for delayed renewal processes, the limiting distribution is the same as for the non-delayed case. Therefore, the results of Proposition 4 and Lemma 3 hold even when the distribution of X_1 is not the same as $\{X_i\}_{i\geq 2}$. In particular, they still hold when the process starts in $\emptyset$. This is summarised with the following corollary.

Corollary 2. *Consider the simplified 1G1B process, now altered to start in $r(0) = \emptyset$. Let Z_0 be the time for which the system is empty until the first fresh link is produced. Let Y_1 be the time in which this link is present in memory until it is removed again, and so on. Let the probability of finding a link at time t be $p(t) = P(r(t) = \neg\emptyset)$. Then,*

$$\lim_{t\to\infty} p(t) = \frac{\mathbb{E}[Y_1]}{\mathbb{E}[Y_1] + \mathbb{E}[Z_1]} = \frac{\lambda}{\lambda + \beta}, \tag{61}$$

and the distribution of the current lifetime of a link satisfies

$$\lim_{t\to\infty} P\left(C(t) > x | r(t) = \neg\emptyset\right) = \frac{1}{\mathbb{E}[Y_1]} \int_x^\infty (1 - F_Y(s))\mathrm{d}s = e^{-\beta x}. \tag{62}$$

Recalling that $\beta = \mu + \lambda q(1 - p)$, we see that the formula for the availability in Proposition 1 is already shown by (61).
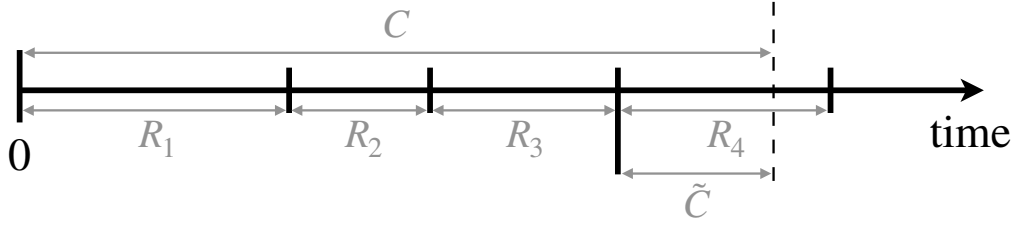
Figure 10: Number of pumping rounds. We are interested in the number of pumping rounds that have been carried out while a link is in memory. Here, C is the (limiting) distribution of the current lifetime in memory (see Figure 9), and R_i is the time between the $(i-1)$th and ith pumping round.

C.2 Availability and average consumed fidelity in 1G1B

Here, we compute the availability and the rest of the steady-state distribution of the 1G1B system (Proposition 1), as well as the average consumed fidelity (Theorem 1).

In order to calculate the average fidelity, we not only need the time spent in $\neg\emptyset$, but also the times spent in each pumping level leading up to the current one.

From 1G1B (Definition 3.1), one may define a simplified 1G1B system as

$$r(t) = \begin{cases} \neg\emptyset \text{ if } s(t) \geq 0 \\ \emptyset \text{ if } s(t) = \emptyset. \end{cases}$$

For the characterisation of the fidelity of the link in memory at time t, $F(t)$, we are interested in the successful pumping attempts that occur in the the time interval $[A_{N(t)}, A_{N(t)} + C(t))$, where $C(t)$ is the current lifetime (Definition C.4). In 1G1B, the successful pumping attempts are a Poisson process with rate $\delta := \lambda pq$. Since the rate is constant for all t, the number of successful pumping attempts within the interval $[A_{N(t)}, A_{N(t)} + C(t))$ has the identical distribution as the number of successful pumping attempts in the time interval $[0, C(t))$. From Corollary 2, we see that $C(t)$ converges in distribution to $C \sim \text{Exp}(\beta)$. In the following Lemma, we characterise the number of successful pumping attempts that occur within the time C, and the time spent between each pair of consecutive pumping rounds. See Figure 10 for an illustration. An observation that we use below is that within the time interval $[0, C)$, the times at which pumping occurs form a separate renewal process, which is convenient for notation.

Lemma 4. *Consider a renewal process $\tilde{N}(t)$ with arrival times $S_0 = 0$, $S_n = \sum_{i=1}^n R_i$, with $R_1 \sim \text{Exp}(\delta)$. Let $C \sim \text{Exp}(\beta)$ be independent of the R_i. Let $M = N(C)$ be the number of arrivals that have occurred by time C. Let $\tilde{C} := C - S_M$ be the current lifetime at time C. Then,*

1. *The distribution of M is given by*

$$P(M \geq m) = \left(\frac{\delta}{\beta + \delta}\right)^m, \tag{63}$$

   or equivalently

$$P(M = m) = \left(\frac{\delta}{\beta + \delta}\right)^m \left(\frac{\beta}{\beta + \delta}\right). \tag{64}$$

2. *Conditional on $M = m$, the random variables $(R_1, ..., R_m, \tilde{C})$ are mutually independent and identically distributed as $\text{Exp}(\beta + \delta)$.*

Proof. 1. We proceed by induction. Letting $F_R := P(R \leq x)$ We have

$$P(M \geq 1) = P(C > R_1) = \int_0^\infty P(C > R_1 | R_1 = x) \mathrm{d}F_R(x)$$
$$= \int_0^\infty e^{-\beta x} \cdot \delta e^{-\delta x} \mathrm{d}x = \frac{\delta}{\delta + \beta},$$

where we have used $P(C > R_1 | R_1 = x) = P(C > x) = e^{-\beta x}$ and $R_1 \sim \mathrm{Exp}(\delta)$. Then, assuming (63),

$$P(M \geq m + 1) = P(C > S_{m+1})$$
$$= P(C > R_{m+1} + S_m)$$
$$\stackrel{a}{=} P(C > R_{m+1})P(C > S_m),$$
$$= P(C > R_1)P(M \geq m)$$
$$\stackrel{b}{=} \left(\frac{\delta}{\beta + \delta}\right) \left(\frac{\delta}{\beta + \delta}\right)^m = \left(\frac{\delta}{\beta + \delta}\right)^{m+1}.$$

In step (b), we have used the inductive assumption. In step (a) we have made use of the memoryless property of the exponential distribution: Since R_{m+1} and S_m are positive and independent random variables, this has as a consequence

$$P(C > R_{m+1} + S_m) = \int_0^\infty \int_0^\infty \mathrm{d}F_R(r) \mathrm{d}F_{S_m}(s) P(C > r + s)$$
$$= \int_0^\infty \int_0^\infty \mathrm{d}F_R(r) \mathrm{d}F_{S_m}(s) P(C > r) P(C > s)$$
$$= P(C > R_{m+1}) P(C > S_m). \tag{65}$$

Finally, (64) follows from

$$P(M = m) = P(M \geq m) - P(M \geq m + 1)$$
$$= \left(\frac{\delta}{\beta + \delta}\right)^m - \left(\frac{\delta}{\beta + \delta}\right)^{m+1}$$
$$= \left(\frac{\delta}{\beta + \delta}\right)^m \left(\frac{\beta}{\beta + \delta}\right).$$

2. We firstly note that for any events E_1, E_2, E_3, it holds that

$$P(E_1 \cap E_2 \cap E_3) = P(E_1 \cap E_2) - P(E_1 \cap E_2 \cap \neg E_3), \tag{66}$$

where $\neg E$ denotes the complement of the event E. Now, consider the events

$$E_1 = \Big\{ R_i > x_i \; \forall i = 1, \dots, m+1 \Big\}, \quad E_2 = \Big\{ C \geq x_{m+1} + \sum_{i=1}^m R_i \Big\}, \quad E_3 = \Big\{ C < \sum_{i=1}^{m+1} R_i \Big\}.$$

Now,

$$E_1 \cap E_2 \cap E_3 = \Big\{ R_1 > x_1, \dots, R_m > x_m, R_{m+1} > x_{m+1} \cap \sum_{i=1}^{m+1} R_i > C \geq x_{m+1} + \sum_{i=1}^m R_i \Big\}$$
$$\stackrel{a}{=} \Big\{ R_1 > x_1, \dots, R_m > x_m, \tilde{C} > x_{m+1} \cap \sum_{i=1}^{m+1} R_i > C \geq \sum_{i=1}^m R_i \Big\}$$
$$\stackrel{b}{=} \Big\{ R_1 > x_1, \dots, R_m > x_m, \tilde{C} > x_{m+1} \cap M = m \Big\},$$

where in (a) we have used the definition of $\tilde{C}$, and in (b) we have used the definition of M. Then, by (66), we see that

$$P\left(R_1 > x_1, \ldots, R_m > x_m, \tilde{C} > x_{m+1} \cap M = m\right)$$
$$= P\left(R_1 > x_1, \ldots, R_{m+1} > x_{m+1} \cap C \geq x_{m+1} + \sum_{i=1}^{m} R_i\right)$$
$$- P\left(R_1 > x_1, \ldots, R_{m+1} > x_{m+1} \cap C \geq \sum_{i=1}^{m+1} R_i\right). \quad (67)$$

By the independence of the R_i, this is equivalent to

$$(67) = \left[P\left(C \geq x_{m+1} + \sum_{i=1}^{m} R_i \middle| R_1 > x_1, ..., R_{m+1} > x_{m+1}\right)\right.$$
$$\left. - P\left(C \geq \sum_{i=1}^{m+1} R_i \middle| R_1 > x_1, ..., R_{m+1} > x_{m+1}\right)\right] \prod_{i=1}^{m+1} P\left(R_i > x_i\right), \quad (68)$$

and we now use the memoryless property of the exponential distribution (see the argument leading to (65)) to rewrite as

$$(67) = \left[P\left(C \geq x_{m+1} | R_{m+1} > x_{m+1}\right) \prod_{i=1}^{m} P\left(C \geq R_i | R_i > x_i\right)\right.$$
$$\left. - \prod_{i=1}^{m+1} P\left(C \geq R_i | R_i > x_i\right)\right] \prod_{i=1}^{m+1} P\left(R_i > x_i\right), \quad (69)$$

which, using that $P(C_i \geq R_i | R_i > x_i)P(R_i > x_i) = P(C_i \geq R_i > x_i)$, becomes

$$(67) = \left[P\left(C \geq x_{m+1} \cap R_{m+1} > x_{m+1}\right) - P\left(C \geq R_{m+1} > x_{m+1})\right)\right] \prod_{i=1}^{m} P\left(C \geq R_i > x_i\right),$$
$$= P\left(R_{m+1} > C \geq x_{m+1}\right) \prod_{i=1}^{m} P\left(C \geq R_i > x_i\right),$$

where we have again made use of (66) to rewrite the factor on the left. Now,

$$P(C \geq R_1 > x_1) = \int_{x_1}^{\infty} P(C \geq y) \mathrm{d}F_R(y)$$
$$= \int_{x_1}^{\infty} e^{-\beta y} \cdot \delta e^{-\delta y} = \frac{\delta}{\beta + \delta} e^{-(\beta+\delta)x_1},$$

and by symmetry

$$P(R_1 \geq C > x_{m+1}) = \frac{\beta}{\beta + \delta} e^{-(\beta+\delta)x_{m+1}}.$$

We therefore see that

$$(67) = \frac{\beta}{\beta + \delta} e^{-(\beta+\delta)x_{m+1}} \cdot \prod_{i=1}^{m} \left[\frac{\delta}{\beta + \delta} e^{-(\beta+\delta)x_i}\right]$$
$$= \frac{\beta}{\beta + \delta} \cdot \left(\frac{\delta}{\beta + \delta}\right)^m \prod_{i=1}^{m+1} e^{-(\beta+\delta)x_i} = P(M = m) \prod_{i=1}^{m+1} e^{-(\beta+\delta)x_i}.$$

It therefore follows that

$$P\left(R_1 > x_1, \ldots, R_m > x_m, \tilde{C} > x_{m+1} | M = m\right) = \prod_{i=1}^{m+1} e^{-(\beta+\delta)x_i},$$

which suffices to show the second result.

$\square$

Recalling that $C(t)$ converges in distribution to C, we now adapt Lemma 4 to apply to $C(t)$. In order to do this, we use the following result (for a proof, see Chapter 7 of [19]).

Theorem 5 (Continuous mapping theorem). *Let $\{X_n\}$ be a sequence of random variables taking values in $\mathbb{R}^k$. If $X_n \to X$ in distribution as $n \to \infty$ and $g : \mathbb{R}^k \to \mathbb{R}^l$ is continuous, then $g(X_n) \to g(X)$ in distribution as $n \to \infty$.*

Corollary 3. *Suppose that $C(t)$ and X are independent random variables, and $C(t)$ converges in distribution to C as $t \to \infty$. Then,*

$$\lim_{t\to\infty} P\left(C(t) > X\right) = P(C > X). \tag{70}$$

Proof. Consider a sequence of times $\{t_n\}_{n \geq 1}$ such that $0 < t_1 < t_2 < \ldots$ and $\lim_{n\to\infty} t_n = \infty$. Let $C_n := C(t_n)$. Then, $C_n \to C$ in distribution. Moreover, since C_n and X are independent for all n, the pair $(C_n, -X) \to (C, -X)$ in distribution. Now, the function $g : \mathbb{R}^2 \to \mathbb{R}$, with $g(x, y) = x + y$ is continuous. Then, by Theorem 5, $C_n - X \to C - X$ in distribution, and so

$$\lim_{n\to\infty} P(C_n - X > 0) = \lim_{n\to\infty} P(C(t_n) - X > 0) = P(C - X > 0).$$

Since this is true for all such sequences $\{t_n\}$, the result follows. $\square$

In the following corollary, we let the current lifetime be dependent on the parameter u to avoid confusion with the time of the renewal process (which is denoted as t).

Corollary 4. *Consider a renewal process $N(t)$ with arrival times $S_0 = 0$, $S_n = \sum_{i=1}^{n} R_i$, with $R_1 \sim Exp(\delta)$. Suppose that $C(u)$ converges in distribution to $C \sim Exp(\beta)$ as $u \to \infty$. Let $M(u) = N(C(u))$ be the number of arrivals that have occurred by time $C(u)$. Let $\tilde{C}(u) := C(u) - S_{M(u)}$ be the current lifetime at time $C(u)$. Then, the results of Lemma 4 still hold in the limit $u \to \infty$. In particular,*

1. *The limiting distribution of $M(u)$ is that of M,*

$$\lim_{u\to\infty} P(M(u) \geq m) = P(M \geq m) = \left(\frac{\delta}{\beta+\delta}\right)^m. \tag{71}$$

2. *Conditional on $M(u) = m$, the random variables $(R_1, \ldots, R_m, \tilde{C})$ converge in distribution to mutually independent and identically distributed $Exp(\beta + \delta)$ as $u \to \infty$, i.e.*

$$\lim_{u\to\infty} P(X_1 > x_1, \ldots, X_m > x_m, \tilde{C}(u) > x_{m+1} | M(u) = m) = \prod_{i=1}^{m+1} e^{-(\beta+\delta)x_i}, \tag{72}$$

Proof. 1. Making use of Corollary 3, we have

$$\lim_{u\to\infty} P(M(u) \geq m) = \lim_{u\to\infty} P(C(u) > S_m) = P(C > S_m) = \left(\frac{\delta}{\beta+\delta}\right)^m.$$

2. One may use exactly the same arguments as were used to obtain (68), only replacing C with $C(u)$ and M with $M(u)$, to show that

$$P\left(R_1 > x_1, \ldots, R_m > x_m, \tilde{C}(u) > x_{m+1} \cap M(u) = m\right)$$

$$= \left[P\left(C(u) \geq x_{m+1} + \sum_{i=1}^m R_i \Big| R_1 > x_1, ..., R_{m+1} > x_{m+1}\right)\right.$$

$$\left. - P\left(C(u) \geq \sum_{i=1}^{m+1} R_i \Big| R_1 > x_1, ..., R_{m+1} > x_{m+1}\right)\right] \prod_{i=1}^{m+1} P\left(R_1 > x_i\right).$$

By Corollary 3, in the limit $u \to \infty$ this satisfies

$$\lim_{u\to\infty} P(R_1 > x_1, ..., R_m > x_m, \tilde{C}(u) > x_{m+1} \cap M(u) = m)$$

$$= \left[P\left(C \geq x_{m+1} + \sum_{i=1}^m R_i \Big| R_1 > x_1, ..., R_{m+1} > x_{m+1}\right)\right.$$

$$\left. - P\left(C \geq \sum_{i=1}^{m+1} R_i \Big| R_1 > x_1, ..., R_{m+1} > x_{m+1}\right)\right] \prod_{i=1}^{m+1} P\left(R_1 > x_i\right) = (68).$$

It then follows that

$$\lim_{u\to\infty} P(R_1 > x_1, ..., R_m > x_m, \tilde{C}(u) > x_{m+1} \Big| M(u) = m)$$

$$= \lim_{u\to\infty} \frac{P(R_1 > x_1, ..., R_m > x_m, \tilde{C}(u) > x_{m+1} \cap M(u) = m)}{P(M(u) = m)}$$

$$= \frac{P(R_1 > x_1, ..., R_m > x_m, \tilde{C} > x_{m+1} \cap M = m)}{P(M = m)} = \prod_{i=1}^{m+1} e^{-(\beta+\delta)x_i},$$

by Lemma 4.

$\square$

For the case when $C(u)$ is the current lifetime of simplified 1G1B, the random variable $(R_1, \ldots, R_m, \tilde{C}(u))$ by definition has the same distribution as $\vec{T}(u)$. Recall that $\vec{T}(u)$ contains the times spent in each purification level leading up to the current one at time u in 1G1B (Definition 3.2). This leads to the following results.

Corollary 5. *Conditional on $s(t) = i$, $\vec{T}(t)$ converges in distribution to $(Q_0, \ldots, Q_i)$ as $t \to \infty$, where the Q_j are i.i.d. random variables with $Q_0 \sim \mathrm{Exp}(\beta + \delta)$.*

We now continue with the formulae for the performance metrics. The availability in the 1G1B system was given in Proposition 1 and the average consumed fidelity was given in Theorem 1. Next, we prove both of them.

Proof of Proposition 1. From Corollary 2, we see that

$$A = \lim_{t\to\infty} P(s(t) = \neg\emptyset) = \frac{\lambda}{\lambda + \mu + \lambda q(1-p)}.$$

Further, for $i \geq 0$

$$P(s(t) = i) = P(s(t) = i | s(t) \neq \emptyset) \cdot P(s(t) \neq \emptyset).$$

Letting $C(t)$ denote the current lifetime of simplified 1G1B at time t, and $M(t)$ denote the number of purifications that have occurred within this time, by Corollary 4 it follows that

$$P(s(t) = i) = P(M(t) = i) \cdot P(s(t) \neq \emptyset) \to P(M = i) \cdot A$$

as $t \to \infty$. Recalling the distribution of M as found in Lemma 4, we obtain

$$\lim_{t \to \infty} P(s(t) = i) = \left(\frac{\lambda q p}{\mu + \lambda q} \right)^i \cdot \frac{\mu + \lambda q (1 - p)}{\mu + \lambda q} \cdot A$$

$$= \frac{\lambda^{i+1} q^i p^i}{(\mu + \lambda q)^{i+1}} \cdot \frac{\mu + \lambda q (1 - p)}{\lambda + \mu + \lambda q (1 - p)}.$$

We note that this result can also be derived with the global balance equations of a CTMC. Here, we chose to use the derivation with renewal theory since it offers a more general formula for the availability (see (61)) and ties in more neatly with the derivation of the formula for the average consumed fidelity, as we will see below. $\qquad \square$

The following proposition will be helpful in the proof of Theorem 1 (formula for average consumed fidelity).

Proposition 5. *Let $\{p_i(t)\}_{i \geq 0}$ and $\{e_i(t)\}_{i \geq 0}$ be such that for all i, $\lim_{t \to \infty} p_i(t) = \pi_i$ and $\lim_{t \to \infty} e_i(t) = c_i$. Suppose also that for all t, $0 \leq e_i(t) \leq 1$, $0 \leq p_i(t) \leq 1$ and $\sum_{i=0} p_i(t) = 1$. Then*

$$\lim_{t \to \infty} \sum_{i=0}^{\infty} e_i(t) p_i(t) = \sum_{i=0}^{\infty} c_i \pi_i. \tag{73}$$

Proof of proposition 5. To show (73), it suffices to show that for any $\epsilon > 0$, there exists a T such that for all $t > T$,

$$\left| \sum_{i=0}^{\infty} e_i(t) p_i(t) - \sum_{i=0}^{\infty} c_i \pi_i \right| < \epsilon. \tag{74}$$

We firstly bound the sum using the triangle inequality,

$$\left| \sum_{i=0}^{\infty} e_i(t) p_i(t) - \sum_{i=0}^{\infty} c_i \pi_i \right| = \left| \sum_{i=0}^{\infty} e_i(t) (p_i(t) - \pi_i) + (e_i(t) - c_i) \pi_i \right|$$

$$\leq \underbrace{\sum_{i=0}^{\infty} e_i(t) |p_i(t) - \pi_i|}_{(A)} + \underbrace{\sum_{i=0}^{\infty} |e_i(t) - c_i| \pi_i}_{(B)}. \tag{75}$$

We then show that $(A) \to 0$ and $(B) \to 0$ as $t \to \infty$. We firstly show that

$$\lim_{t \to \infty} \sum_{i=0}^{\infty} |p_i(t) - \pi_i| = 0. \tag{76}$$

Note that, since $\sum_{i=0}^{\infty} p_i(t) = 1$, it follows that $\sum_{i=0}^{\infty} \pi_i = 1$. Then, choose N such that

$$\sum_{i=0}^{N} \pi_i > 1 - \frac{\epsilon}{2}$$

and choose T_1 such that

$$\sum_{i=0}^{N} |p_i(t) - \pi_i| < \frac{\epsilon}{2}, \forall t > T_1$$

which is possible since the sum is finite. Then $\forall t > T_1$,

$$\left| 1 - \sum_{i=0}^{N} p_i(t) \right| = \left| 1 - \sum_{i=0}^{N} (\pi_i - (\pi_i - p_i(t))) \right|$$

$$< \left| 1 - \sum_{i=0}^{N} \pi_i \right| + \sum_{i=0}^{N} |\pi_i - p_i(t)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \tag{77}$$

Now, choose T_2 such that $\forall t > T_2$,

$$|p_i(t) - \pi_i| < \frac{\epsilon}{N}, \forall i = 0, ..., N$$

and let $T = \max\{T_1, T_2\}$. Then, $\forall t > T$,

$$\sum_{i=0}^{\infty} |p_i(t) - \pi_i| = \sum_{i=0}^{N} |p_i(t) - \pi_i| + \sum_{i>N} |p_i(t) - \pi_i|$$

$$< N \cdot \frac{\epsilon}{N} + \sum_{i>N} p_i(t) + \sum_{i>N} \pi_i$$

$$< \epsilon + \epsilon + \frac{\epsilon}{2},$$

from (77). This suffices to show (76). Combined with the fact that the e_i are bounded, it follows that $(A) \to 0$. We now show that $(B) \to 0$, i.e.

$$\lim_{t \to \infty} \sum_{i=0}^{\infty} |e_i(t) - c_i|\pi_i = 0. \tag{78}$$

To show this, let $\epsilon > 0$. Choose N such that $\sum_{i=0}^{N} \pi_i > 1 - \epsilon$. Choose T such that

$$\sum_{i=0}^{N} |e_i(t) - c_i| < \epsilon, \forall t > T.$$

This is possible since the LHS is a finite sum. Then,

$$\sum_{i=0}^{\infty} |e_i(t) - c_i|\pi_i = \sum_{i=0}^{N} |e_i(t) - c_i|\pi_i + \sum_{i>N} |e_i(t) - c_i|\pi_i$$

$$< \left(\sum_{i=0}^{N} |e_i(t) - c_i| \right) \cdot \left(\sum_{i=0}^{N} \pi_i \right) + \sum_{i>N} \pi_i$$

$$< \epsilon(1 - \epsilon) + \epsilon \ \forall t > T,$$

which shows (78). $\qquad\square$

Combining these results $((A) \to 0$ and $(B) \to 0)$ in (75)) suffices to show Proposition 5. We are now ready to prove Theorem 1 (formula for average consumed fidelity).

Proof of Theorem 1. We firstly expand out the average consumed fidelity (Definition 4.2) as a sum by conditioning on the value of $s(t)$,

$$\overline{F} := \mathbb{E}[F(t)|s(t) \neq \emptyset] = \sum_{i=0}^{\infty} \mathbb{E}[F(t)|s(t) = i]P(s(t) = i|s(t) \neq \emptyset)$$

$$= \frac{1}{P(s(t) \neq \emptyset)} \sum_{i=0}^{\infty} \mathbb{E}[F(t)|s(t) = i]P(s(t) = i). \tag{79}$$

Recall that we are interested in the limit $t \to \infty$ of the above. Note that from Proposition 1, we know the limiting values of $P(s(t) \neq \emptyset)$ and $P(s(t) = i)$. We now claim that

$$\lim_{t \to \infty} \mathbb{E}[F(t)|s(t) = i] = \mathbb{E}\left[F^{(i)}(Q_0, Q_1, ..., Q_i)\right] \tag{80}$$

where $Q_0, Q_1, \ldots, Q_i$ are i.i.d. random variables with $Q_0 \sim \text{Exp}(\mu + \lambda q)$, and $F^{(i)}$ is given in Definition 3.5. We use the following result:

Theorem 6 (Theorem 7.2.19 of [19])**.** *Let X_n be a sequence of random variables. Then, $X_n \to X$ in distribution if and only if $\mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)]$ for all bounded continuous functions g.*

Recall that conditional on $s(t) = i$, we have $F(t) = F^{(i)}\left(\vec{T}(t)\right)$ (from Definition 3.6). As mentioned in Section 3, $F^{(i)}$ is a continuous and bounded function. Therefore, (80) follows by combining Theorem 6 and Corollary 5.

From Proposition 5, we therefore see that

$$\lim_{t \to \infty} \mathbb{E}[F(t)|s(t) \neq \emptyset] = \lim_{t \to \infty} \frac{1}{P(s(t) \neq \emptyset)} \sum_{i=0}^{\infty} \mathbb{E}[F(t)|s(t) = i]P(s(t) = i)$$

$$= \frac{1}{A} \sum_{i=0}^{\infty} \lim_{t \to \infty} \mathbb{E}[F(t)|s(t) = i] \lim_{t \to \infty} P(s(t) = i)$$

$$= \frac{1}{A} \sum_{i=0}^{\infty} c_i \pi_i,$$

where $c_i = \mathbb{E}\left[F^{(i)}(Q_0, Q_1, ..., Q_i)\right]$ and $\pi_i = \lim_{t \to \infty} P(s(t) = i)$. $\qquad \square$

D Average consumed fidelity with a linear jump function

In this Appendix we focus on linear jump functions. In D.1, we provide bounds for the coefficients of a linear jump function. In D.2, we first prove Proposition 2. Then, we use that Proposition to derive the average consumed fidelity in a 1G1B system that uses a pumping protocol with a linear jump function, which we denote by $\overline{F}_{\text{linear}}$ (i.e., we show Lemma 1). We also show that $\overline{F}_{\text{linear}}$ is monotonic in the probability of pumping q and the probability of successful pumping p (Proposition 3). Lastly, in D.3, we discuss in which situations $\overline{F}_{\text{linear}}$ is monotonically increasing in q, and we compute the noise threshold (29) discussed in Section 5.1 (above this threshold, any purification is better than no purification).

D.1 Bounds on the parameters of a linear jump function

Proposition 6. *Consider a jump function that is linear with the fidelity of one of the input states, i.e.,*

$$J(F, \rho) = a(\rho)F + b(\rho), \tag{81}$$

where F is the fidelity of one of the input states and ρ is the second input state. Then, the coefficients $a(\rho)$ and $b(\rho)$ must satisfy

$$0 \le a(\rho) \le 1 \quad \text{and} \quad \frac{1}{4}(1 - a(\rho)) \le b(\rho) \le 1 - a(\rho).$$

Proof. First, we require $J(F, \rho) \le 1$, which is equivalent to $b \le 1 - a$. We also require $J(F, \rho) \ge 1/4$, which leads to $b \ge (1 - a)/4$. By imposing that the upper bound on b has to be larger than the lower bound, we find that $a \le 1$. Finally, since we want jump functions that increase with increasing F, we want $a \ge 0$. $\qquad\square$

D.2 Derivation of $\overline{F}_{\text{linear}}$

Proof of Proposition 2. Here, we consider a 1G1B system with $J(F, \rho_{\text{new}}) = aF + b$ and $F^{(0)}(t_0) = D_{t_0}(F_{\text{new}})$, where F_{new} is the fidelity of the state ρ_{new}. Our goal is to find an analytical solution for the fidelity of the entangled link after i consecutive successful purifications, $F^{(i)}(t_0, ..., t_{i-1}, t_i)$. The time passed between purification j and $j+1$ is given by t_j. After the i-th purification the system spent time t_i without any transitions (i.e., no purification or consumption events). We show in this proof that $F^{(i)}$ is given by

$$F^{(i)}(t_0, ..., t_{i-1}, t_i) = \frac{1}{4} + \sum_{j=0}^{i} m_j^{(i)} e^{-\Gamma(t_j + t_{j+1}... + t_i)} \tag{82}$$

where the constants $m_j^{(i)}$ are given by $m_0^{(0)} = F_{\text{new}} - \frac{1}{4}$, and

$$m_j^{(i)} = \begin{cases} a^{i-j}\left(\frac{a}{4} + b - \frac{1}{4}\right), & \text{if } j > 0, \\ a^i\left(F_{\text{new}} - \frac{1}{4}\right) & \text{if } j = 0. \end{cases}$$

for $i > 0$.

We proceed by induction. For $i = 0$, we have

$$F^{(0)}(t_0) = D_{t_0}(F_{\text{new}}) = e^{-\Gamma t_0}\left(F_{\text{new}} - \frac{1}{4}\right) + \frac{1}{4}, \tag{83}$$

from which we see that $m_0^{(0)} = F_{\text{new}} - \frac{1}{4}$. If we assume that (82) is true for some i, using the recursive relation from (4) we can show that (82) is also true for $i + 1$:

$$
\begin{aligned}
F^{(i+1)}(t_0, ..., t_i) &= D_{t_{i+1}}\left(J(F^{(i)}, \rho)\right) \\
&= D_{t_{i+1}}\left(aF^{(i)} + b\right) \\
&= e^{-\Gamma t_{i+1}}\left(aF^{(i)} + b - \frac{1}{4}\right) + \frac{1}{4} \\
&= \frac{1}{4} + \left(\frac{a}{4} + b - \frac{1}{4}\right)e^{-\Gamma t_{i+1}} + \sum_{j=0}^{i} am_j^{(i)} e^{-\Gamma(t_j + ... + t_i + t_{i+1})},
\end{aligned}
$$

from which it follows that

$$
\begin{aligned}
m_j^{(i+1)} &= am_j^{(i)} \ \ (0 \le j \le i) \\
m_{i+1}^{(i+1)} &= \frac{a}{4} + b - \frac{1}{4}
\end{aligned}
$$

Then, by the inductive assumption, $m_0^{(i+1)} = a^{i+1}\left(F_{\text{new}} - \frac{1}{4}\right)$, and $m_j^{(i+1)} = a^{i+1-j}\left(\frac{a}{4} + b - \frac{1}{4}\right)$ for $j > 0$. $\qquad\square$

Proof of Lemma 1. Here we consider a 1G1B system with $J(F, \rho_{\text{new}}) = aF + b$ and $F^{(0)}(t_0) = D_{t_0}(F_{\text{new}})$, where F_{new} is the fidelity of the state ρ_{new}. Our goal is to find a closed-form solution for the average fidelity after $i \ge 0$ purification rounds, c_i, and for the average consumed fidelity, $\overline{F}_{\text{linear}}$.

We defined c_i as the average value of $F^{(i)}$ (see (13)). Using the expression for $F^{(i)}$ from Proposition 2 (also given in (82)), we can evaluate c_i as follows

$$
\begin{aligned}
c_i &:= \int_0^\infty dt_i f_\alpha(t_i) ... \int_0^\infty dt_0 f_\alpha(t_0) F^{(i)}(t_0, ..., t_{i-1}, t_i) \\
&= \int_0^\infty dt_i f_\alpha(t_i) ... \int_0^\infty dt_0 f_\alpha(t_0) \left[\frac{1}{4} + \sum_{j=0}^{i} m_j^{(i)} e^{-\Gamma(t_j + ... + t_{i-1} + t_i)}\right] \\
&= \frac{1}{4} + \sum_{j=0}^{i} m_j^{(i)}\left(\frac{\alpha}{\alpha + \Gamma}\right)^{i-j+1} \\
&= \frac{1}{4} + \left(F_{\text{new}} - \frac{1}{4}\right) \cdot a^i \gamma^{i+1} + \gamma\left(\frac{a}{4} + b - \frac{1}{4}\right)\sum_{j=1}^{i} a^{i-j}\gamma^{i-j},
\end{aligned}
$$

where $\alpha = \mu + \lambda q$, $f_\alpha(t_i) = \alpha e^{-\alpha t_i}$ (since the times t_i are exponentially distributed with rate α), $\gamma = \alpha/(\alpha + \Gamma)$. Using the fact that this is a geometric series, we may now obtain a closed-form solution for c_i:

$$
c_i = \frac{1}{4} + \left(F_{\text{new}} - \frac{1}{4}\right) \cdot a^i \gamma^{i+1} + \gamma\left(\frac{a}{4} + b - \frac{1}{4}\right)\frac{1 - a^i \gamma^i}{1 - a\gamma}. \tag{84}
$$

The final formula for the average fidelity may then be computed with the results of Proposition 1 and Theorem 1 as

$$
\begin{aligned}
\overline{F}_{\text{linear}} &= \lim_{t \to \infty} \mathbb{E}\left(F(t)|s(t) \ne \emptyset\right) = \frac{1}{1 - \pi_\emptyset}\sum_{i=0}^{\infty} c_i \pi_i \\
&= \frac{1}{4} + \frac{\gamma}{1 - a\gamma}\cdot\left(\frac{a}{4} + b - \frac{1}{4}\right) + \frac{\gamma}{(1 - \pi_\emptyset)}\cdot\left(F_{\text{new}} - \frac{1}{4} - \frac{\frac{a}{4} + b - \frac{1}{4}}{1 - a\gamma}\right)\sum_{i=0}^{\infty} \pi_i(a\gamma)^i,
\end{aligned}
$$
$$\tag{85}$$

where the constant terms are no longer in the sum since

$$\frac{1}{1 - \pi_\emptyset} \sum_{i=0}^\infty \pi_i = 1,$$

by the normalisation of the steady state distribution. Recalling the distribution of π from Proposition 1, we may evaluate the sum as a geometric series,

$$\sum_{i=0}^\infty \pi_i (a\gamma)^i = \frac{\lambda}{\mu + \lambda q} \sum_{i=0}^\infty \left(\frac{\lambda q p a \gamma}{\mu + \lambda q} \right)^i \pi_\emptyset$$

$$= \frac{\lambda}{\mu + \lambda q} \cdot \frac{1}{1 - \frac{\lambda q p a \gamma}{\mu + \lambda q}} \pi_\emptyset$$

$$= \frac{\lambda}{\mu + \lambda q - \lambda q p a \gamma} \pi_\emptyset.$$

We may now substitute this into (85) to obtain a closed-form solution for the average fidelity,

$$\overline{F}_{\text{linear}} = \frac{1}{4} + \frac{\gamma}{1 - a\gamma} \cdot \left(\frac{a}{4} + b - \frac{1}{4} \right) + \gamma \cdot \left(F_{\text{new}} - \frac{1}{4} - \frac{\frac{a}{4} + b - \frac{1}{4}}{1 - a\gamma} \right) \frac{\lambda}{\mu + \lambda q - \lambda q p a \gamma} \frac{\pi_\emptyset}{1 - \pi_\emptyset}$$

$$= \frac{1}{4} + \frac{\gamma}{1 - a\gamma} \cdot \left(\frac{a}{4} + b - \frac{1}{4} \right) + \gamma \cdot \left(F_{\text{new}} - \frac{1}{4} - \frac{\frac{a}{4} + b - \frac{1}{4}}{1 - a\gamma} \right) \frac{\mu + \lambda q(1 - p)}{\mu + \lambda q - \lambda q p a \gamma}$$

$$= \frac{\frac{1}{4}\Gamma + b\lambda q p + F_{\text{new}}\left(\mu + \lambda q(1 - p) \right)}{\Gamma + \mu + \lambda q(1 - p a)}, \tag{86}$$

which completes the closed-form solutions for our two performance metrics in this set-up (in the last step we used Mathematica to simplify the expression). $\qquad\square$

Proof of Proposition 3. To show (a), we compute the partial derivative of the average consumed fidelity with respect to q:

$$\frac{\partial \overline{F}_{\text{linear}}}{\partial q} = \lambda \frac{\Gamma\left(4F_{\text{new}}(1 - p) + (4b + a)p - 1 \right) + 4\mu p(b - F_{\text{new}}(1 - a))}{4\left(\Gamma + \mu + \lambda q(1 - ap) \right)^2}. \tag{87}$$

Since the sign of the derivative does not depend on q, we conclude that $\overline{F}_{\text{linear}}$ is monotonic in q.

To show (b), we proceed similarly:

$$\frac{\partial \overline{F}_{\text{linear}}}{\partial p} = \lambda q \frac{4(b - F_{\text{new}})(\Gamma + \mu + \lambda q) + a\left(\Gamma + 4F_{\text{new}}(\mu + \lambda q) \right)}{4\left(\Gamma + \mu + \lambda q(1 - ap) \right)^2}. \tag{88}$$

Since the sign of this derivative does not depend on p, we conclude that $\overline{F}_{\text{linear}}$ is monotonic in p.

$$\square$$

D.3 Noise threshold

In the previous Section, we showed that $\overline{F}_{\text{linear}}$ is monotonic in q and p (Proposition 3). Nevertheless, note that $\overline{F}_{\text{linear}}$ can be monotonically increasing or decreasing in q and in p depending on the values of the other parameters. For a pumping protocol with a good enough jump function, $\overline{F}_{\text{linear}}$ becomes increasing in q. A sufficient condition is for the jump function to satisfy $b \geq F_{\text{new}}(1-a)$, as we show next. The partial derivative with respect to q from (87) can be written as follows:

$$\frac{\partial \overline{F}_{\text{linear}}}{\partial q} = \frac{\lambda}{x^2}(\Gamma y + 4\mu pz), \tag{89}$$

where $x = 2(\Gamma + \mu + \lambda q(1 - ap))$, $y = 4F_{\text{new}}(1-p) + (4b+a)p - 1$, and $z = b - F_{\text{new}}(1-a)$. Using the fact that $b \geq (1-a)/4$, we find that $y \geq 0$. A sufficient condition for the partial derivative to be positive is that $z \geq 0$, i.e., if $b \geq F_{\text{new}}(1-a)$, then the average consumed fidelity is monotonically increasing in q. Moreover, we can conclude that, if the noise is above certain threshold ($\Gamma > -4\mu pz/y$), the derivative is positive and the pumping is always beneficial, even if it succeeds with a very small probability.

E Bounds for the performance of bilocal Clifford protocols

In this Appendix, we find bounds to the output fidelity and the probability of success of 2-to-1 purification protocols. In particular, we show Lemma 2, where upper and lower bounds on the jump function and the success probability of any bilocal Clifford protocol, taking as input a Werner state ρ_{W} and a Bell-diagonal state ρ_{BD}. We define the fidelity of a state ρ as $F(\rho, |\phi^+\rangle) = \langle\phi^+|\,\rho\,|\phi^+\rangle$, where $|\phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ is one of the Bell states. We find the bounds for a system with the following restrictions.

- We consider 2-to-1 purification protocols, i.e., protocols that take two bipartite entangled states as input and output a single bipartite state. This allows us to use these bounds directly for the analysis of the 1G1B system.

- We restrict the pumping protocols to bilocal Clifford protocols [10, 25], which are a well-known type of purification scheme. We provide more details about this type of protocol in E.1.

- We assume that one of the input states is a Werner state (in the 1G1B system, this is the state in the good memory, which suffers from depolarising noise) and the other input state is Bell-diagonal (in the 1G1B system, this is the state generated via heralded entanglement generation and placed in the bad memory). Mathematically, the input states can be written, respectively, as

$$\rho_{\mathrm{W}} = F\left|\phi^+\right\rangle\!\!\left\langle\phi^+\right| + \frac{1-F}{3}\left|\psi^+\right\rangle\!\!\left\langle\psi^+\right| + \frac{1-F}{3}\left|\psi^-\right\rangle\!\!\left\langle\psi^-\right| + \frac{1-F}{3}\left|\phi^-\right\rangle\!\!\left\langle\phi^-\right|,$$

$$\rho_{\mathrm{BD}} = F_{\mathrm{BD}}\left|\phi^+\right\rangle\!\!\left\langle\phi^+\right| + \lambda_1\left|\psi^+\right\rangle\!\!\left\langle\psi^+\right| + \lambda_2\left|\psi^-\right\rangle\!\!\left\langle\psi^-\right| + \lambda_3\left|\phi^-\right\rangle\!\!\left\langle\phi^-\right|,$$

with $F, F_{\mathrm{BD}}, \lambda_1, \lambda_2, \lambda_3 \in [0,1]$ subjected to the normalization constraint $F_{\mathrm{BD}} + \lambda_1 + \lambda_2 + \lambda_3 = 1$, and with the Bell states defined as

$$\left|\phi^+\right\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}, \; \left|\psi^+\right\rangle = \frac{|01\rangle + |10\rangle}{\sqrt{2}}, \; \left|\psi^-\right\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}, \; \left|\phi^-\right\rangle = \frac{|00\rangle - |11\rangle}{\sqrt{2}}.$$

  Note that any bipartite state can be brought to Bell-diagonal form while preserving the fidelity by means of twirling (adding extra noise) [4, 21].

- We only consider newly generated states with fidelity to some Bell state larger than 1/2, i.e., we assume $F_{\mathrm{BD}} > 1/2$ (note that $F_{\mathrm{BD}} > 1/2$ is equivalent to $\lambda_i > 1/2$ for some i, since the states are equivalent upon some Pauli corrections). As shown in E.3, this is a necessary and sufficient condition for the existence of entanglement (otherwise, the state is not useful for purification).

- We assume the Werner state has fidelity $F > 1/4$, since the good memory is initially occupied with a state with fidelity larger than 1/2, and this fidelity can decay at most to 1/4 due to depolarising noise (see Definition 3.3).

In E.1, we provide a formal definition of bilocal Clifford protocols. Then, in E.2, we prove Lemma 2, where bounds are found for the jump function and success probability of bilocal Clifford protocols in a system with the above restrictions.

E.1 Bilocal Clifford protocols

Bilocal Clifford protocols [10, 25] take n bipartite states as input and outputs a single bipartite state. They consist of the following steps:

1. $C^T \otimes C^\dagger$ is applied to the state, where C is some Clifford circuit. A Clifford circuit consists of Hadamard gates, phase gates S, and CNOTs [18, 17]. If the state is held by two separate parties, one of them applies C^T and the other one applies $C^\dagger$.

2. All of the qubit pairs except one are measured (in a 2-to-1 protocol, one qubit pair is measured and the other one is kept).

3. Depending on the parity of the measurement outcomes, success or failure is declared. Local unitaries may be performed after a success.

One of the main advantages of bilocal Clifford protocols is that they are relatively simple to execute in practice, since they involve a basic set of gates. Additionally, any stabilizer code C can be mapped to a bilocal Clifford circuit that applies $C^T \otimes C^\dagger$, allowing the analysis of bilocal Clifford circuits from a quantum error-correction perspective [16]. This type of protocol also includes well-known purification protocols, such as DEJMPS [11].

E.2 Linear bounds to the performance of bilocal Clifford protocols

In this Appendix, we prove Lemma 2, where bounds on the jump function and success probability of every bilocal Clifford protocol are found. Consider pumping two states of the form

$$\rho_{\mathrm{W}} = F \left| \Phi^+ \right\rangle \left\langle \Phi^+ \right| + \frac{1-F}{3} \left(\left| \Psi^+ \right\rangle \left\langle \Psi^+ \right| + \left| \Psi^- \right\rangle \left\langle \Psi^- \right| + \left| \Phi^- \right\rangle \left\langle \Phi^- \right| \right) \tag{90}$$

$$\rho_{\mathrm{BD}} = F_{\mathrm{BD}} \left| \Phi^+ \right\rangle \left\langle \Phi^+ \right| + \lambda_1 \left| \Psi^+ \right\rangle \left\langle \Psi^+ \right| + \lambda_2 \left| \Psi^- \right\rangle \left\langle \Psi^- \right| + \lambda_3 \left| \Phi^- \right\rangle \left\langle \Phi^- \right|. \tag{91}$$

Using the methods from [25], we can find the analytical expressions for the output fidelity and success probability for every bilocal Clifford protocol. The restriction to bilocal Clifford protocols and Bell-diagonal states allows us to do this enumeration of analytical functions efficiently [25, 16]. There are only seven protocols that provide a unique combination of J and p, as shown in Table 2. We refer to the i-th jump function and success probability as $J_i(F, \rho_{\mathrm{BD}})$ and $p_i(F, \rho_{\mathrm{BD}})$, for $i = 1, ..., 7$.

Table 2: The jump function and success probability for all 2-1 bilocal Clifford protocols, with input states given in (90) and (91) .

Protocol	Jump function	Success probability
1	$\frac{(4\lambda_1+3\lambda_2+3\lambda_3-3)F-\lambda_1}{(4\lambda_2+4\lambda_3-2)F-\lambda_2-\lambda_3-1}$	$\frac{2}{3}(1-2\lambda_2-2\lambda_3)F+\frac{1}{3}(1+\lambda_2+\lambda_3)$
2	$\frac{(3\lambda_1+4\lambda_2+3\lambda_3-3)F-\lambda_2}{(4\lambda_1+4\lambda_3-2)F-\lambda_1-\lambda_3-1}$	$\frac{2}{3}(1-2\lambda_3-2\lambda_1)F+\frac{1}{3}(1+\lambda_3+\lambda_1)$
3	$\frac{(3\lambda_1+3\lambda_2+4\lambda_3-3)F-\lambda_3}{(4\lambda_1+4\lambda_2-2)F-\lambda_1-\lambda_2-1}$	$\frac{2}{3}(1-2\lambda_1-2\lambda_2)F+\frac{1}{3}(1+\lambda_1+\lambda_2)$
4	F	$F_{\mathrm{BD}} + \lambda_1$
5	F	$F_{\mathrm{BD}} + \lambda_2$
6	F	$F_{\mathrm{BD}} + \lambda_3$
7	F_{BD}	$\frac{2}{3}F+\frac{1}{3}$

We see that for these particular input states, J_4, J_5 and J_6 produce no change in the fidelity of ρ_{W}. They also have a non-unity success probability. It would therefore be
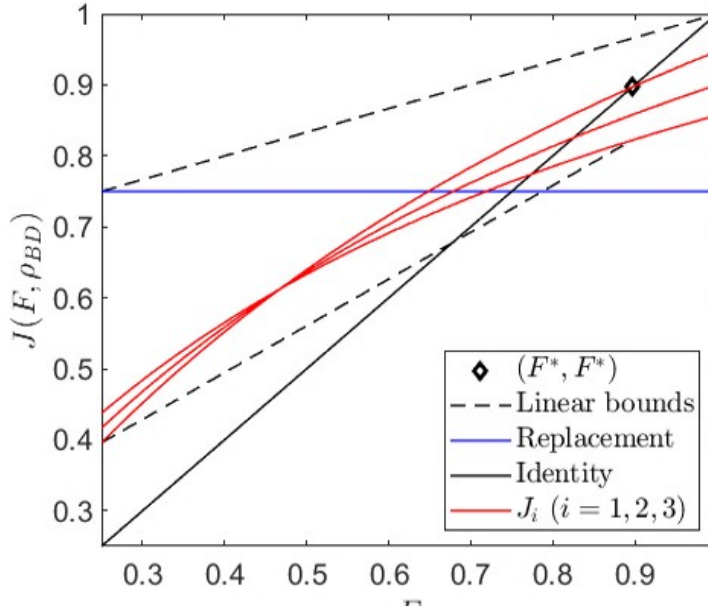
Figure 11: Linear bounds for the jump function of bilocal Clifford protocols (black dashed lines). The jump functions shown are J_1-J_3 (red lines), J_4-J_6 (identity operation, black line), and J_7 (probabilistic replacement, blue line). F^* is the highest fidelity achievable by pumping a low-fidelity Werner state with the fixed Bell-diagonal state ρ_{BD}. The lower bound holds in the range $[1/4, F^*]$. The upper bound holds in the range $[1/4, 1]$. Here, $F_{\mathrm{BD}} = 0.75$ and $\rho_{\mathrm{BD}} = (0.75, 0.125, 0.833, 0.0417)$.

advantageous to simply perform no action instead of attempting Protocols 4-6. Similarly, J_7 assumes the fidelity of the Bell-diagonal state, which is the same change as performing replacement. Since replacement can be achieved with probability one, it does not make sense to perform Protocol 7. Therefore, the only remaining 'non-trivial' protocols are Protocols 1-3. In the following, we therefore find bounds for the jump function of Protocols 1-3. Notice that there is symmetry in the λ_i: J_2 and p_2 can be obtained by permuting $(\lambda_1, \lambda_2, \lambda_3)$ in J_1 and p_1, and similarly for J_3 and p_3.

In the following, we show Lemma 2.

Proof of Lemma 2. We firstly show the linear lower bound (i.e. the formulae given in (23)). We assume that $\lambda_1 \geq \lambda_2 \geq \lambda_3$. Then, by symmetry in the λ_i, one may retrieve the bound by setting $\lambda_{\min} = \lambda_3$ and $\lambda_{\max} = \lambda_1$. In order to show this bound, we make use of the following collection of results. It is important to note that when showing all of the following results, ρ_{BD} is fixed.

1. **Proposition 7, Corollory 6, Proposition 8** – the formula for F^* is derived (Equation 21). This is the maximum achievable fidelity achievable in the 1G1B system, with fixed Bell-diagonal input state ρ_{BD}. Therefore, at any given time t, the fidelity $F(t)$ of the stored link in the 1G1B system (see Definition 3.6) satisfies $F(t) \leq F^*$.

2. At $F = F^*$, Protocol 3 provides the best output fidelity,

$$J_1 (F^*, \rho_{\mathrm{BD}}) \leq J_2 (F^*, \rho_{\mathrm{BD}}) \leq J_3 (F^*, \rho_{\mathrm{BD}})$$

(**Proposition 7 and Corollary 6**).

3. At $F = 1/4$, Protocol 1 provides the best output fidelity, i.e.

$$J_3\left(F^*, \rho_{\text{BD}}\right) \leq J_2\left(F^*, \rho_{\text{BD}}\right) \leq J_1\left(F^*, \rho_{\text{BD}}\right),$$

since $J_i(1/4, \rho_{\text{BD}}) = (F_{\text{BD}} + \lambda_i)/2$.

4. For $i = 1, 2, 3$, $J_i(F, \rho_{\text{BD}})$ is a concave function of F (**Proposition 9**).

In particular, the third result means that any straight line taken between two points on J_i must lie below the curve itself. The linear lower bound is the linear function connecting the points

$$\left(F^*, J_1(F^*, \rho_{\text{BD}})\right), \quad \left(\frac{1}{4}, J_3\left(\frac{1}{4}, \rho_{\text{BD}}\right)\right), \tag{92}$$

which is given by

$$J_{\text{LB}}(F, \rho_{\text{BD}}) = \left(\frac{J_1(F^*, \rho_{\text{BD}}) - \frac{F_{\text{BD}} + \lambda_3}{2}}{F^* - \frac{1}{4}}\right)\left(F - \frac{1}{4}\right) + \frac{F_{\text{BD}} + \lambda_3}{2},$$

where we have used the fact that $J_i(1/4, \rho_{\text{BD}}) = (F_{\text{BD}} + \lambda_i)/2$. Letting $\lambda_{\max} = \lambda_1$ and $\lambda_{\min} = \lambda_3$, this may be rearranged into the form

$$J_{\text{LB}}(F, \rho_{\text{new}}) = a_{\text{l}}F + b_{\text{l}},$$

with a_{l} and b_{l} given in Lemma 2 (in (23)). When choosing the points in (92), we are joining the line corresponding to the lowest of the J_i for both $F = 1/4$ and $F = F^*$. By the concavity property, this is therefore a lower bound for all of the J_i in the region $[1/4, F^*]$. See Figure 11 for an illustration of this lower bound.

We now show the upper bound. We choose this to be the linear function connecting the points $(1/4, F_{\text{BD}})$ and $(1, 1)$, which is given by

$$J_{\text{UB}}(F, \rho_{\text{new}}) = \left(\frac{1 - F_{\text{BD}}}{1 - \frac{1}{4}}\right)\left(F - \frac{1}{4}\right) + F_{\text{BD}},$$

and may be rearranged into the form

$$J_{\text{UB}}(F, \rho_{\text{new}}) = a_{\text{u}}F + b_{\text{u}},$$

with a_{u} and b_{u} given in Lemma 2 (in (25)). We show that this is an upper bound with the following steps. Again, for ease of notation, we exploit the symmetry in λ_i and assume that $\lambda_1 \geq \lambda_2 \geq \lambda_3$.

1. In the domain $F > 0$, the jump functions J_1, J_2 and J_3 intersect at the same point F_{int}. Moreover, for $i = 1, 2, 3$, $J_i(F_{\text{int}}, \rho_{\text{BD}}) = \sqrt{\frac{F_{\text{BD}}}{2}} < F_{\text{BD}}$. (**Proposition 7**).

2. In the domain $F \in [F_{\text{int}}, 1]$, the jump function outputting the highest-fidelity outcome out of protocols 1-3 is J_3 (**Corollary 6**).

3. For $i = 1, 2, 3$, J_i is an increasing and concave function of F (**Proposition 9**).

4. Consider the tangent to J_3 at $F = 1$. This lies below J_{UB} in the range $F \in [1/4, 1]$ (**Proposition 10**).

By result (3) from the above list (concavity), we see that the tangent to J_3 at $F = 1$ upper bounds J_3 for all F. By result (2) from the above list, this also upper bounds J_1 and J_2 in the range $F \in [F_{\text{int}}, 1]$. Therefore, by result (4), J_{UB} upper bounds J_1, J_2 and J_3 in the range $F \in [F_{\text{int}}, 1]$. Moreover, for $F < F_{\text{int}}$, by results (1) and (3), $J_i(F, \rho_{\text{BD}}) \leq \sqrt{\frac{F_{\text{BD}}}{2}} < F_{\text{BD}} \leq J_{\text{UB}}(F, \rho_{\text{BD}})$, by the definition of J_{UB} (J_{UB} runs through the point $(1/4, F_{\text{BD}})$ and is increasing). This suffices to show that the upper bound holds.

Finally, we show the bounds for p_i. Recalling that $F_{\text{BD}} + \lambda_1 + \lambda_2 + \lambda_3 = 1$, we have

$$\frac{\partial}{\partial F} p_1(F, \rho_{\text{BD}}) = \frac{2}{3}(1 - \lambda_2 - \lambda_3) = \frac{2}{3}(2F_{\text{BD}} + 2\lambda_1 - 1)$$
$$\geq \frac{2}{3}(2F_{\text{BD}} - 1) > 0.$$

Therefore, $p_1(F, \rho_{\text{BD}})$ is an increasing function of F. By symmetry, p_2 and p_3 are also increasing functions of F. Since the fidelity $F(t)$ of the 1G1B system always lies in the region $F(t) \in [1/4, F^*]$, it follows that at any point in time, the success probability p of purification may be bounded with

$$p_i\left(\frac{1}{4}, \rho_{\text{BD}}\right) \leq p \leq p_i\left(F^*, \rho_{\text{BD}}\right).$$

$\square$

Below are the collection of results that were used to show the bounds on the jump functions.

Proposition 7. *In the domain $F > 0$, jump functions 1-3 intersect exactly once at the same point F_{int}, such that $J_i(F_{int}, \rho_{\text{new}}) = \sqrt{\frac{F_{\text{BD}}}{2}} < F_{\text{BD}}$.*

Proof. We firstly compute the intersection point of jump functions 1 and 2. This occurs at the F value which satisfies

$$\frac{(4\lambda_1 + 3\lambda_2 + 3\lambda_3 - 3)F - \lambda_1}{(4\lambda_2 + 4\lambda_3 - 2)F - \lambda_2 - \lambda_3 - 1} = \frac{(3\lambda_1 + 4\lambda_2 + 3\lambda_3 - 3)F - \lambda_2}{(4\lambda_1 + 4\lambda_3 - 2)F - \lambda_1 - \lambda_3 - 1},$$

or alternatively, recalling that $F_{\text{BD}} + \lambda_1 + \lambda_2 + \lambda_3 = 1$,

$$\frac{(\lambda_1 - 3F_{\text{BD}})F - \lambda_1}{(2 - 4F_{\text{BD}} - 4\lambda_1)F - 2 + F_{\text{BD}} + \lambda_1} = (1 \leftrightarrow 2),$$

where to obtain the RHS we exchange labels 1 and 2 of the LHS. This is equivalent to

$$((\lambda_1 - 3F_{\text{BD}})F - \lambda_1)((2 - 4F_{\text{BD}} - 4\lambda_2)F - 2 + F_{\text{BD}} + \lambda_2) - (1 \leftrightarrow 2) = 0,$$

which simplifies to

$$(\lambda_1 - \lambda_2)\left((2 - 16F_{\text{BD}})F^2 + (8F_{\text{BD}} - 4)F + 2 - F_{\text{BD}}\right) = 0. \tag{93}$$

Then, if $\lambda_1 \neq \lambda_2$, the points of intersection depend only on F_{BD} and therefore are symmetric in λ_1, λ_2 and λ_3. The points of intersection are given by

$$F = \frac{4F_{\text{BD}} - 2 \pm 3\sqrt{2F_{\text{BD}}}}{2(8F_{\text{BD}} - 1)} \tag{94}$$

and recalling that $F_{\mathrm{BD}} \in (1/2, 1]$, the solution lying in the domain of interest $(F > 0)$ is

$$F_{\mathrm{int}} = \frac{4F_{\mathrm{BD}} - 2 + 3\sqrt{2F_{\mathrm{BD}}}}{2(8F_{\mathrm{BD}} - 1)}.$$

Then, since F_{int} is symmetric in λ_1, λ_2 and λ_3, all jump functions intersect at the point F_{int}. One may also show that

$$J_i(F_{\mathrm{int}}, \rho_{\mathrm{new}}) = \sqrt{\frac{F_{\mathrm{BD}}}{2}},$$

e.g. using software such as Mathematica. Since $F_{\mathrm{BD}} < 1/2$, we have

$$\sqrt{\frac{1}{2}} < \sqrt{F_{\mathrm{BD}}} \Leftrightarrow \sqrt{\frac{F_{\mathrm{BD}}}{2}} < F_{\mathrm{BD}}.$$

$\square$

We now continue with the following corollary.

Corollary 6. *Suppose that $\lambda_1 \geq \lambda_2 \geq \lambda_3$. Then, for $F \geq F_{int}$,*

$$J_3(F, \rho_{\mathrm{BD}}) \geq J_2(F, \rho_{\mathrm{BD}}) \geq J_1(F, \rho_{\mathrm{BD}}), \tag{95}$$

Proof. From Proposition 7, J_1, J_2 and J_3 will not intersect again for $F > F_{\mathrm{int}}$. Therefore, their ordering remains the same for all $F > F_{\mathrm{int}}$. The jump function outputting the largest fidelity in this range will therefore also have the largest limit as $F \to \infty$. We see that

$$\lim_{F \to \infty} J_i(F, \rho_{\mathrm{BD}}) = \frac{3F_{\mathrm{BD}} - \lambda_i}{4F_{\mathrm{BD}} + \lambda_i - 2},$$

which is a decreasing function of λ_i. Therefore, $\lambda_3 = \min\{\lambda_1, \lambda_2, \lambda_3\}$ gives the largest limit, and J_1 satisfies (95). $\square$

From Proposition 7 and Corollary 6, we know which of J_1, J_2 and J_3 provide the best fidelity for $F \in [1/2, 1]$. With the following proposition, we see that for some lower fidelities, it is better to replace with the bad link rather than choose to pump.

Proposition 8. *The largest fidelity obtainable by pumping a low-fidelity Werner state with ρ_{BD} and bilocal Clifford protocols is*

$$F^* = \frac{2F_{\mathrm{BD}} - 1 + \sqrt{(2F_{\mathrm{BD}} - 1)^2 + 2\lambda_{\min}(2F_{\mathrm{BD}} - 1 + 2\lambda_{\min})}}{2(2F_{\mathrm{BD}} - 1 + 2\lambda_{\min})}, \tag{96}$$

where $\lambda_{\min} = \min\{\lambda_1, \lambda_2, \lambda_3\}$.

Proof. Consider applying pumping protocol $i \in \{1, 2, 3\}$. This stops improving the Werner state fidelity at the value of F such that

$$F^* = J_i(F^*, \rho_{\mathrm{BD}})$$

$$\Leftrightarrow F^* = \frac{(\lambda_i - 3F_{\mathrm{BD}})F^* - \lambda_i}{(2 - 4F_{\mathrm{BD}} - 4\lambda_i)F^* - 2 + F_{\mathrm{BD}} + \lambda_i}$$

$$\Leftrightarrow 0 = (2 - 4F_{\mathrm{BD}} - 4\lambda_i)F^2 + (4F_{\mathrm{BD}} - 2)F + \lambda_i,$$

which has solutions

$$F = \frac{2F_{\mathrm{BD}} - 1 \pm \sqrt{(2F_{\mathrm{BD}} - 1)^2 + 2\lambda_i(2F_{\mathrm{BD}} - 1 + 2\lambda_i)}}{2(2F_{\mathrm{BD}} - 1 + 2\lambda_i)},$$

one of which is positive and one negative. Recalling that for $F > \frac{1}{2}$, the jump function taking the largest value is J_i with $\lambda_i = \lambda_{\min}$, means that the maximum fidelity achievable is (96). □

Proposition 9. *For any ρ_{BD} with $F_{\mathrm{BD}} > 1/2$, , for $i = 1, 2, 3$ $J_i(F, \rho_{\mathrm{BD}})$ is a strictly concave and increasing function of F.*

Proof. We differentiate J_i. Firstly, consider derivatives of functions of the form

$$y = \frac{ax + b}{cx + d}.$$

This may be rewritten as

$$y = \frac{a}{c} + \frac{b - \frac{ad}{c}}{cx + d}.$$

Then,

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{ad - bc}{(cx + d)^2}, \qquad \frac{\mathrm{d}^2 y}{\mathrm{d}x^2} = -2c\frac{ad - bc}{(cx + d)^3}. \tag{97}$$

To check the sign of these functions, we must therefore check the sign of $ad - bc$. Recalling that J_i may be rewritten as

$$J_i(F, \rho_{\mathrm{BD}}) = \frac{(3F_{\mathrm{BD}} - \lambda_i)F + \lambda_i}{(4F_{\mathrm{BD}} + 4\lambda_i - 2)F + 2 - F_{\mathrm{BD}} - \lambda_i},$$

in this case,

$$a = 3F_{\mathrm{BD}} - \lambda_i > \frac{3}{2} - \frac{1}{2} = 1$$

$$b = \lambda_i < \frac{1}{2}$$

$$c = 4(F_{\mathrm{BD}} + \lambda_i) - 2 \leq 4 \cdot 1 - 2 = 2$$

$$d = 2 - (F_{\mathrm{BD}} + \lambda_i) \geq 2 - 1 = 1$$

and it follows that $ad - bc > 1 \cdot 1 - 2 \cdot 1/2 = 0$. Then, since

$$c = 4F_{\mathrm{BD}} + 4\lambda_i - 2 > 4 \cdot \frac{1}{2} + 4\lambda_i - 2 = 4\lambda_i \geq 0,$$

it follows from (97) that

$$\frac{\partial}{\partial F}J_i(F, \rho_{\mathrm{BD}}) > 0, \qquad \frac{\partial^2}{\partial F^2}J_i(F, \rho_{\mathrm{BD}}) < 0.$$

Therefore, J_i is a strictly concave and increasing function of F. □

Proposition 10. *Suppose that $\lambda_1 \geq \lambda_2 \geq \lambda_3$. Consider the tangent to $J_3(F, \rho_{\mathrm{BD}})$ at $F = 1$. Denote this by $J_{\mathrm{tan}}(F, \rho_{\mathrm{BD}})$. Then, this lies below J_{UB} for all $F \in [1/4, 1]$, i.e.*

$$J_{\mathrm{tan}}(F, \rho_{\mathrm{BD}}) \leq J_{\mathrm{UB}}(F, \rho_{\mathrm{BD}}),$$

where

$$J_{\mathrm{UB}}(F, \rho_{\mathrm{BD}}) = \frac{4(1 - F_{\mathrm{BD}})}{3}F + \frac{4F_{\mathrm{BD}} - 1}{3}$$

is the linear upper bound from Lemma 2.

Proof. We firstly compute the formula for the tangent to J_i at $F = 1$. Recalling the formula (97), this has gradient

$$\frac{\partial}{\partial F} J_3(F, \rho_{\mathrm{BD}})\big|_{F=1} = \frac{ad - bc}{(c + d)^2} = \frac{6F_{\mathrm{BD}} - 3(F_{\mathrm{BD}} + \lambda_3)^2}{(3(F_{\mathrm{BD}} + \lambda_3))^2} = \frac{2F_{\mathrm{BD}}}{3(F_{\mathrm{BD}} + \lambda_3)^2} - \frac{1}{3}.$$

Since the tangent runs through the point $(1, J_3(1, \rho_{\mathrm{BD}}))$, it has formula

$$J_{\tan}(F, \rho_{\mathrm{BD}}) = \left(\frac{2F_{\mathrm{BD}}}{3(F_{\mathrm{BD}} + \lambda_3)^2} - \frac{1}{3} \right)(F - 1) + \frac{F_{\mathrm{BD}}}{F_{\mathrm{BD}} + \lambda_3},$$

where we have used $J_i(1, \rho_{\mathrm{BD}}) = F_{\mathrm{BD}}/(F_{\mathrm{BD}} + \lambda_i)$. We note that at $F = 1$,

$$J_{\mathrm{UB}}(1, \rho_{\mathrm{BD}}) = 1 \geq \frac{F_{\mathrm{BD}}}{F_{\mathrm{BD}} + \lambda_3} = J_{\tan}(1, \rho_{\mathrm{BD}}).$$

Therefore, to show the proposition, it suffices to show that

$$J_{\mathrm{UB}}\left(\frac{1}{4}, \rho_{\mathrm{BD}} \right) \geq J_{\tan}\left(\frac{1}{4}, \rho_{\mathrm{BD}} \right), \tag{98}$$

since both J_{UB} and $J_{\tan}$ are linear in F and therefore intersect at most once. Now,

$$J_{\mathrm{UB}}\left(\frac{1}{4}, \rho_{\mathrm{BD}} \right) - J_{\tan}\left(\frac{1}{4}, \rho_{\mathrm{BD}} \right) = F_{\mathrm{BD}} - \left(\frac{2F_{\mathrm{BD}}}{3(F_{\mathrm{BD}} + \lambda_3)^2} - \frac{1}{3} \right)\left(-\frac{3}{4} \right) - \frac{F_{\mathrm{BD}}}{F_{\mathrm{BD}} + \lambda_3}$$

$$= F_{\mathrm{BD}} - \frac{1}{4} + \frac{F_{\mathrm{BD}}}{2(F_{\mathrm{BD}} + \lambda_3)^2} - \frac{F_{\mathrm{BD}}}{F_{\mathrm{BD}} + \lambda_3}.$$

Now, let $x := F_{\mathrm{BD}} + \lambda_3$, and

$$h(x) := F_{\mathrm{BD}} - \frac{1}{4} + \frac{F_{\mathrm{BD}}}{2x^2} - \frac{F_{\mathrm{BD}}}{x}.$$

By the assumption that $\lambda_3 = \min\{\lambda_1, \lambda_2, \lambda_3\}$ and the condition $F_{\mathrm{BD}} + \lambda_1 + \lambda_2 + \lambda_3 = 1$, it follows that

$$\lambda_3 \in \left[0, \frac{1 - F_{\mathrm{BD}}}{3} \right], \quad \text{and} \quad x \in \left[F_{\mathrm{BD}}, \frac{1 + 2F_{\mathrm{BD}}}{3} \right]. \tag{99}$$

To prove the proposition, it therefore suffices to show positivity of h for x in the range (99). We start by establishing the monotonicity of h:

$$\frac{\partial}{\partial x} h(x) = -\frac{F_{\mathrm{BD}}}{x^3} + \frac{F_{\mathrm{BD}}}{x^2} = -\frac{F_{\mathrm{BD}}}{x^3}(1 - x) \leq 0,$$

since $x = F_{\mathrm{BD}} + \lambda_3 \leq 1$. We therefore see that h is decreasing. To show that h is positive in the range (99), it therefore suffices to show that

$$h\left(\frac{1 + 2F_{\mathrm{BD}}}{3} \right) \geq 0.$$

We have

$$h\left(\frac{1 + 2F_{\mathrm{BD}}}{3} \right) = F_{\mathrm{BD}} - \frac{1}{4} + \frac{9F_{\mathrm{BD}}}{2(1 + 2F_{\mathrm{BD}})^2} - \frac{3F_{\mathrm{BD}}}{1 + 2F_{\mathrm{BD}}}$$

$$= F_{\mathrm{BD}} - \frac{1}{4} + 6F_{\mathrm{BD}}\left(\frac{\frac{3}{4} - \frac{1}{2}(1 + 2F_{\mathrm{BD}})}{(1 + 2F_{\mathrm{BD}})^2} \right)$$

$$= F_{\mathrm{BD}} - \frac{1}{4} + 6F_{\mathrm{BD}}\frac{\frac{1}{4} - F_{\mathrm{BD}}}{(1 + 2F_{\mathrm{BD}})^2}$$

$$= \left(F_{\mathrm{BD}} - \frac{1}{4} \right)\left(1 - \frac{6F_{\mathrm{BD}}}{(1 + 2F_{\mathrm{BD}})^2} \right).$$

Then, since $F_{\mathrm{BD}} > \frac{1}{2}$, we have

$$h\left(\frac{1 + 2F_{\mathrm{BD}}}{3}\right) > 0 \Leftrightarrow 1 - \frac{6F_{\mathrm{BD}}}{(1 + 2F_{\mathrm{BD}})^2} > 0$$
$$\Leftrightarrow (1 + 2F_{\mathrm{BD}})^2 > 6F_{\mathrm{BD}}$$
$$\Leftrightarrow 4F_{\mathrm{BD}}^2 - 2F_{\mathrm{BD}} + 1 > 0$$
$$\Leftrightarrow (1 - 2F_{\mathrm{BD}})^2 + 2F_{\mathrm{BD}} > 0,$$

which holds. We therefore see that

$$h(x) \geq h\left(\frac{1 + 2F_{\mathrm{BD}}}{3}\right) > 0$$

for all x in the range (99), and therefore

$$J_{\mathrm{UB}}\left(\frac{1}{4}, \rho_{\mathrm{BD}}\right) - J_{\mathrm{tan}}\left(\frac{1}{4}, \rho_{\mathrm{BD}}\right) > 0.$$

This suffices to show the proposition. $\qquad\square$

E.3 Additional proofs

Lemma 5. *A Bell-diagonal state*

$$\rho = \lambda_0 \left|\phi^+\right\rangle\!\left\langle\phi^+\right| + \lambda_1 \left|\psi^+\right\rangle\!\left\langle\psi^+\right| + \lambda_2 \left|\psi^-\right\rangle\!\left\langle\psi^-\right| + \lambda_3 \left|\phi^-\right\rangle\!\left\langle\phi^-\right|,$$

with $\lambda_0 + \lambda_1 + \lambda_2 + \lambda_3 = 1$, is entangled if and only if $\lambda_i > 1/2$ for some i.

Proof of Lemma 5. We will analyse the entanglement of a Bell-diagonal state using the Peres-Horodecki criterion, which states that a bipartite, 2×2 dimensional quantum state ρ is entangled if and only if the partial transpose of ρ has at least one negative eigenvalue [33, 22]. A Bell-diagonal state can be written in the Bell basis as

$$\rho = \lambda_0 \left|\phi^+\right\rangle\!\left\langle\phi^+\right| + \lambda_1 \left|\psi^+\right\rangle\!\left\langle\psi^+\right| + \lambda_2 \left|\psi^-\right\rangle\!\left\langle\psi^-\right| + \lambda_3 \left|\phi^-\right\rangle\!\left\langle\phi^-\right|.$$

In the computational basis, $\{\left|00\right\rangle, \left|01\right\rangle, \left|10\right\rangle, \left|11\right\rangle\}$, the Bell-diagonal state can be written as

$$\rho = \begin{pmatrix} \lambda_0 + \lambda_3 & 0 & 0 & \lambda_0 - \lambda_3 \\ 0 & \lambda_1 + \lambda_2 & \lambda_1 - \lambda_2 & 0 \\ 0 & \lambda_1 - \lambda_2 & \lambda_1 + \lambda_2 & 0 \\ \lambda_0 - \lambda_3 & 0 & 0 & \lambda_0 + \lambda_3 \end{pmatrix}.$$

The partial transpose of this density matrix is given by

$$\rho^{\mathrm{PT}} = \begin{pmatrix} \lambda_0 + \lambda_3 & 0 & 0 & \lambda_1 - \lambda_2 \\ 0 & \lambda_1 + \lambda_2 & \lambda_0 - \lambda_3 & 0 \\ 0 & \lambda_0 - \lambda_3 & \lambda_1 + \lambda_2 & 0 \\ \lambda_1 - \lambda_2 & 0 & 0 & \lambda_0 + \lambda_3 \end{pmatrix}.$$

The eigenvalues of the partial transpose are $\xi_i = 1 - 2\lambda_i$, $i = 1, 2, 3, 4$. One of the eigenvalues is negative iff $\lambda_i > 1/2$ for some i. Therefore, according to the Peres-Horodecki criterion, the state is entangled iff $\lambda_i > 1/2$ for some i. Since these λ_i correspond to the fidelity of ρ to one of the Bell states (e.g., $F(\rho, \left|\phi^+\right\rangle) \equiv \left\langle\phi^+\right| \rho \left|\phi^+\right\rangle = \lambda_0$), we conclude that the state is entangled iff the fidelity to one of the Bell states is larger than $1/2$. $\quad\square$

F Numerical simulations

In our analytical calculations, we assumed a purification protocol with constant success probability (which implies a linear jump function, as shown in Appendix B). This allowed us to derive bounds for the performance of any 1G1B entanglement buffering system that uses bilocal Clifford protocols. However, the success probability of these purification protocols is in general linear in the fidelity of the buffered state (see Appendix B). In this Appendix, we compare the analytical bounds, which assume a constant success probability, to the actual values obtained via a simulation that considers the true (linear, non-constant) success probability.

Our discrete-event simulation keeps track of the buffered link, which decoheres until an event is triggered. These events could correspond to a consumption request (which consumes the buffered memory) or a successful entanglement generation (which is followed by pumping, with probability q). When purification is performed, it succeeds with a probability that depends linearly on the fidelity of the buffered link (see Appendix B).

To compute the average consumed fidelity and the availability, we run the simulation N_{samples} times. In each realization i of the process, we let the system evolve over t_{sim} units of time until convergence to a steady state, and we record the fidelity of the buffered link $F_i(t_{\text{sim}})$ (if the memory is empty, the fidelity is set to zero, as was specified in Definition 3.6). Then, we estimate the average consumed fidelity as the average fidelity of the buffered link at t_{sim} (conditional on the buffered link being present):

$$\overline{F} \approx \overline{F}' \equiv \frac{\sum_{i=1}^{N_{\text{samples}}} F_i(t_{\text{sim}})}{N'_{\text{samples}}}, \tag{100}$$

where

$$N'_{\text{samples}} = \sum_{i=1}^{N_{\text{samples}}} \mathbb{1}_{F_i(t_{\text{sim}})>0} \tag{101}$$

is the number of samples in which $F_i(t_{\text{sim}}) > 0$ ($\mathbb{1}$ is the indicator function). We measure the error in the estimate using the standard error:

$$\varepsilon_F = \sqrt{\frac{\sum_{i=1}^{N'_{\text{samples}}} \left(F_i(t_{\text{sim}}) - \overline{F}'\right)^2}{N'_{\text{samples}}\left(N'_{\text{samples}} - 1\right)}}, \tag{102}$$

which corresponds to the square root of the unbiased sample variance divided by the number of samples. The availability is estimated as the proportion of samples in which there is a buffered link at time t_{sim}:

$$A \approx A' \equiv \frac{1}{N_{\text{samples}}} \sum_{i=0}^{N_{\text{samples}}} \mathbb{1}_{F_i(t_{\text{sim}})>0}, \tag{103}$$

Note that A' is the average of a binary random variable. We can therefore model this random variable as Bernoulli-distributed with probability of success A'. This yields a variance $A'(1 - A')$, which allows us to compute the standard error as

$$\varepsilon_A = \sqrt{\frac{A'(1 - A')}{N_{\text{samples}}}}. \tag{104}$$

Next, we study again the example from Figure 7, and we compare the bounds discussed in the main text with the results from our simulation. In Figure 12, we show the same lower and upper bounds (yellow and dark blue lines, respectively) from Figure 7. We simulated three buffering systems, each of them using the unique bilocal Clifford protocols 1, 2, and 3 from Table 2 (we neglect protocols 4-7 since they are trivial). We emphasise that these simulations consider the true probabilities of success (which are linear but non-constant in the fidelity of the buffered link) and the true jump functions (rational in the fidelity of the buffered link) of the purification protocols. Figure 12 shows the availability and average consumed fidelity attained by each of these systems, for different values of q. We first note that protocols 2 (blue circles) and 3 (red crosses) are equivalent. This is due to the symmetry of the newly generated state considered in this example, $\rho_{\text{new}} = F_{\text{new}} |\phi^+\rangle\langle\phi^+| + (1 - F_{\text{new}}) (|\psi^+\rangle\langle\psi^+| + |\psi^-\rangle\langle\psi^-|) /2$. More importantly, the performance of the simulated systems lies within the analytical bounds, which were derived assuming a constant probability of success. This serves as empirical evidence that our simplified model is still useful when lifting the assumption about a constant probability of success, and can guide the design of more complex and realistic buffering systems.
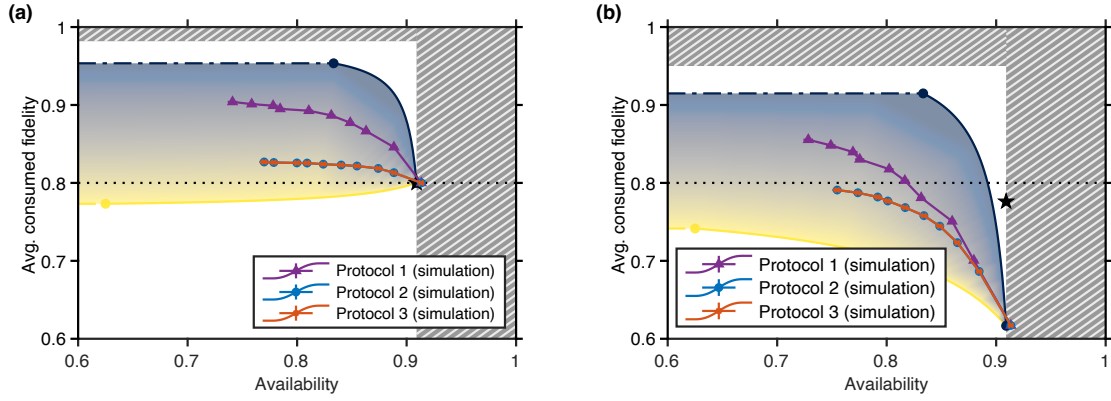

Figure 12: Bounds derived assuming a constant probability of success still apply when the assumption is lifted. **(a)** Noiseless memories ($\Gamma = 0$) or **(b)** noisy memories ($\Gamma = 5 \cdot 10^{-2}$ a.u.). For a given target availability, the average consumed fidelity is within the blue/yellow region (see Corollary 1). Availability is maximized for $q = 0$ (q is the probability of purification after successful entanglement generation), and it decreases for increasing q. White regions cannot be achieved by bilocal Clifford protocols. Striped regions cannot be achieved by any pumping protocol. Black star: performance of the replacement protocol (buffered link is replaced by new links). Dotted line: fidelity of newly generated entangled links. Solid lines with markers: performance of the 1G1B system obtained via simulation, using the true jump functions and true probabilities of success of purification protocols 1, 2, and 3 from Table 2, ($q = 0$ for the rightmost data point, decreasing in intervals of 0.111 until reaching $q = 1$ in the leftmost data point). The simulation considers a linear probability of success, unlike the analytical calculations, in which this probability is assumed to be constant. Parameters used in this example (times and rates in the same arbitrary units): $\lambda = 1$, $\mu = 0.1$, $F_{\text{new}} = 0.8$, $\rho_{\text{new}} = F_{\text{new}} |\phi^+\rangle\langle\phi^+| + (1 - F_{\text{new}}) (|\psi^+\rangle\langle\psi^+| + |\psi^-\rangle\langle\psi^-|) /2$. Numerical parameters used in the simulation: $t_{\text{sim}} = 50$, $N_{\text{samples}} = 10^4$.