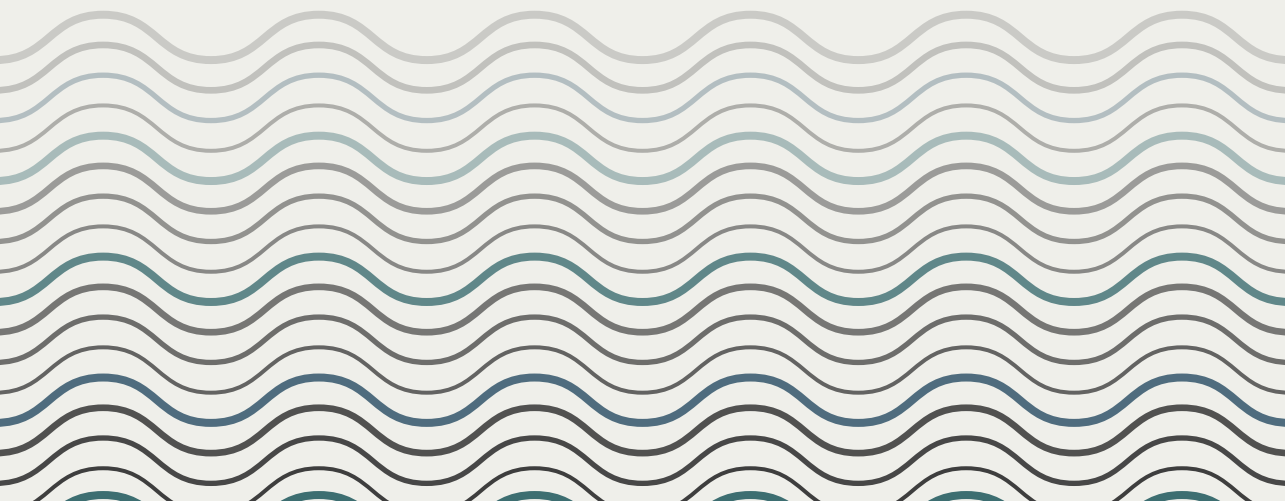
PERFORMANCE ANALYSIS OF NEAR-TERM QUANTUM NETWORKS

Bethany Davies



PERFORMANCE ANALYSIS OF NEAR-TERM QUANTUM NETWORKS

PERFORMANCE ANALYSIS OF NEAR-TERM QUANTUM NETWORKS

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof. dr. ir. H. Bijl
voorzitter van het College voor Promoties,
in het openbaar te verdedigen
op maandag 19 januari 2026 om 12.30 uur

door

Bethany Jane DAVIES

Master of Mathematics,
University of Cambridge, Verenigd Koninkrijk
geboren te Brighton, Verenigd Koninkrijk.

Dit proefschrift is goedgekeurd door de promotoren.

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof. dr. S.D.C. Wehner,	Technische Universiteit Delft, promotor
Prof. dr. ir. R. Hanson,	Technische Universiteit Delft, promotor
Onafhankelijke leden:	
Dr. R. Hai	Technische Universiteit Delft
Dr. W. Löffler	Universiteit Leiden
Prof. dr. M.M. de Weerdt	Technische Universiteit Delft
Dr. M. T. Wimmer	Technische Universiteit Delft
Prof. dr. ir. L.M.K. Vandersypen	Technische Universiteit Delft, reservelid

Printed by: Ridderprint | www.ridderprint.nl

Cover: Bethany Davies

ISBN 978-94-6537-131-3

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

CONTENTS

SUMMARY

Quantum networks hold the potential to enable new applications, such as secure key distribution, high-precision distributed sensing, and distributed quantum computing. A central functionality of a quantum network is the distribution of entanglement between remote parties. Since experimental implementations remain in an early stage, it is important to understand both the capabilities and limitations of near-term architectures. However, characterising quantum network performance is challenging, due to the complex, stochastic nature of even simple architectures. Analytical studies can therefore play a crucial role: they not only reduce computational cost but also reveal fundamental relationships between performance, the choice of entanglement distribution protocols, and properties of quantum network hardware. In this thesis, we develop analytical methods to study quantum network performance in several important scenarios.

We firstly analyse entanglement buffers, which are systems designed to generate and store high-quality entangled states to be consumed at any time. For this setting, we derive analytical expressions for two key performance metrics. The solutions are computationally efficient, make no restrictive assumptions about the entanglement purification protocol, and allow general insights: for example, that simple purification schemes can outperform more complex ones previously considered "optimal" in different contexts.

Then, we turn to the problem of entanglement packet generation, where multiple entangled states of sufficient quality must be established simultaneously between network users. The fast generation of entanglement packets is an essential capability for many quantum network protocols. We obtain analytical results for the entanglement packet generation rate under a constant entanglement generation scheme and later extend the analysis to adaptive schemes, where entanglement parameters are tuned dynamically. Using parameter regimes motivated by current experiments, we show that adaptivity can enhance the entanglement packet generation rate by up to a factor of twenty.

Finally, we examine a standard assumption in performance analyses: that the initial states in a quantum repeater chain can be approximated by a symmetrised, or "twirled", form. We investigate this assumption in the contexts of postselected and non-postselected entanglement swapping, where postselection is performed based on the Bell-state measurement outcomes at the repeaters. A central result is that, in many relevant cases, the twirled approximation is exact for non-postselected swapping. More generally, we provide a systematic framework to determine when the twirled approximation is valid for the initial states of a repeater chain.

SAMENVATTING

Quantumnetwerken bieden het vooruitzicht op nieuwe toepassingen, waaronder veilige sleutelverdeling, hoogprecieze gedistribueerde sensoren en gedistribueerde quantumcomputing. Een kernfunctionaliteit van een quantumnetwerk is de distributie van verstrengeling tussen ruimtelijk gescheiden partijen. Aangezien experimentele realisaties zich nog in een vroeg ontwikkelingsstadium bevinden, is het van belang zowel de mogelijkheden als de beperkingen van nabije-toekomstarchitecturen te doorgronden. Het karakteriseren van de prestaties van quantumnetwerken vormt echter een aanzienlijke uitdaging, vanwege de complexe en stochastische aard van zelfs relatief eenvoudige architecturen. Analytisch onderzoek speelt daarom een essentiële rol: het reduceert niet alleen de computationele complexiteit, maar legt tevens fundamentele verbanden bloot tussen netwerkprestaties, de gekozen protocollen voor verstrengelingsdistributie en de eigenschappen van quantumnetwerkhardware. In dit proefschrift worden analytische methoden ontwikkeld om de prestaties van quantumnetwerken in een aantal relevante scenario's systematisch te onderzoeken.

In het eerste deel analyseren wij verstrengelingsbuffers: systemen die zijn ontworpen voor het genereren en opslaan van verstrengelde toestanden van hoge kwaliteit, zodat deze op elk gewenst moment kunnen worden ingezet. Voor deze context leiden wij analytische uitdrukkingen af voor twee centrale prestatiemaatstaven. De verkregen oplossingen zijn computationeel efficiënt, maken geen beperkende aannames over het toegepaste verstrengelingszuiveringsprotocol en verschaffen algemene inzichten. Zo tonen zij onder meer aan dat relatief eenvoudige zuiveringsschema's in bepaalde gevallen betere prestaties leveren dan complexere schema's die in andere contexten als optimaal werden beschouwd.

Vervolgens behandelen wij het probleem van de generatie van verstrengelingspakketten, waarbij gelijktijdig meerdere verstrengelde toestanden van voldoende kwaliteit tussen netwerkgebruikers tot stand moeten worden gebracht. De snelle generatie van dergelijke verstrengelingspakketten is een essentiële vereiste voor vele quantumnetwerkprotocollen. Wij presenteren analytische resultaten voor de generatiesnelheid van verstrengelingspakketten binnen een constant verstrengelingsgeneratieschema en breiden deze analyse uit naar adaptieve schema's, waarin verstrengelingsparameters dynamisch worden geoptimaliseerd. Voor parameterregimes die zijn geïnspireerd door hedendaagse experimentele realisaties laten wij zien dat adaptiviteit de generatiesnelheid van verstrengelingspakketten met maximaal een factor twintig kan verhogen.

Ten slotte onderzoeken wij een gangbare aanname in prestatieanalyses, namelijk dat de begintoestanden in een quantumrepeaterketen kunnen worden benaderd door een gesymmetriseerde, ofwel "getwirlde", vorm. Deze aanname wordt geanalyseerd in de context van zowel postgeselecteerde als niet-postgeselecteerde verstrengelingsswapping, waarbij postselectie plaatsvindt op basis van de uitkomsten van Bell toestandsmetingen bij de repeaters. Een belangrijk resultaat is dat in veel relevante gevallen de

getwirlde benadering exact geldig is voor niet-postgeselecteerde swapping. Meer algemeen presenteren wij een systematisch raamwerk om vast te stellen onder welke voorwaarden de getwirlde benadering gerechtvaardigd is voor de begintoestanden van een repeaterketen.

1

INTRODUCTION

1.1. QUANTUM NETWORKS

A quantum network is an infrastructure that enables the transmission of quantum information between remote parties. A key example is the sharing of entangled quantum states between two or more parties [1, 2]. Once shared, entanglement can be used for applications that are otherwise not possible classically. Examples include a variety of cryptography applications such as the distribution of shared secret keys [3, 4] and blind quantum computation [5, 6], as well as tasks for high-precision distributed sensing, such as clock synchronisation [7, 8] and extending the baselines of telescopes [9].

On the physical level, entanglement generation schemes typically involve the transmission of quantum information encoded in a photonic state [10]. In order to generate entanglement between distantly-separated parties, the main challenge to overcome is *loss*, where photons may be absorbed or scattered on their way to the intended receiver and therefore do not arrive successfully. When travelling through optical fibre, which is one of the most promising forms of transmitting photons, the probability that a photon arrives successfully decays exponentially with distance. Therefore, in order to generate a single entangled state between two distantly-separated parties, on average one has to wait a time that increases exponentially with the distance between them, assuming that attempts are sequential. Since quantum network applications will likely require not just a single entangled state but the reliable delivery of many entangled states, generating entanglement via direct transmission is not a feasible strategy when distances are large ($\gtrsim 500$ km for performing quantum key distribution with fibre networks [10]).

To overcome loss, it is crucial to introduce *quantum repeaters*. The term *repeater* is borrowed from the classical world, where signal loss is also problematic in both wired and wireless networks. Essentially, a classical repeater reconstructs a classical signal (e.g. a string of 0s and 1s) based on the noisy signal received through error correction schemes. However, signal reconstruction is fundamentally different when instead quantum information is transmitted. This is because, unlike classical information, quantum information is disturbed through measurement and cannot be copied [11].

There exists a variety of proposals for quantum repeater architectures [12, 10]. The term *architecture* is very general, and in this context we use it to refer to the hardware and protocols used to generate long-distance entanglement. The architecture believed to be most implementable in the near-term future is sometimes referred to as a *first-generation* quantum repeater [13]. We will give more details about first-generation repeaters in the following section. Currently, several implementations of quantum networks and important subroutines exist in the form of proof-of-principle experiments and test beds [14, 15, 16, 17, 18]. There is currently much theoretical research that is devoted to understanding how best one can utilise noisy hardware in order to obtain the best possible performance [19]. Since even small (few-user) quantum networks are objects of significant complexity, evaluating high-level metrics such as the rate of generation of entangled states, or the quality of generated entanglement, is typically challenging and requires an approach involving a unique blend of mathematics, physics and computer science. In this thesis, we analytically study several scenarios that are of importance to near-term quantum networks. In the following section, we firstly motivate an abstract model encompassing near-term quantum networks that encompasses the scenarios studied in the thesis. We then go on to introduce the thesis contents.

1.1.1. MODELLING NEAR-TERM QUANTUM NETWORKS

Here, we explain the fundamental characteristics of near-term quantum networks, and motivate the mathematical models used in each chapter. We then go on to outline the main challenges for the performance analysis of quantum networks. From now on, we also refer to an entangled state shared between two parties as an *entangled link*, or just a *link*.

HERALDED ENTANGLEMENT GENERATION

When transmitting quantum information in the form of a photon over long distances, there is a significant probability of the photon being lost. When using the transmission of photons to mediate entanglement generation, it is therefore often necessary that a classical *heralding signal* is transmitted between the repeaters after each attempt to inform them about whether or not the entangled state has been generated successfully. Upon heralding success, the repeaters can continue with other network operations or applications that make use of the newly generated entangled state. We note that there are types of repeaters that suppress errors near-deterministically and therefore do not rely on heralding. However, these use highly sophisticated quantum operations, such as the reliable generation of highly entangled photonic graph states [20, 21, 22], and are therefore viewed as far-term. Such repeaters are therefore referred to as *third-generation* repeaters [13].

Moving back to the first-generation repeaters in which we are interested in this thesis, there are several different schemes for heralded entanglement generation [23, 24, 25, 26, 27]. These differ from each other in terms of the direction of quantum communication – e.g. from a central station to both end nodes, from both end nodes to a central station, or from one node directly to another. They also differ in the qubit encoding, which may be either single-rail or dual-rail (involve either one or two photons). Now, because heralded entanglement generation involves the transmission of quantum and

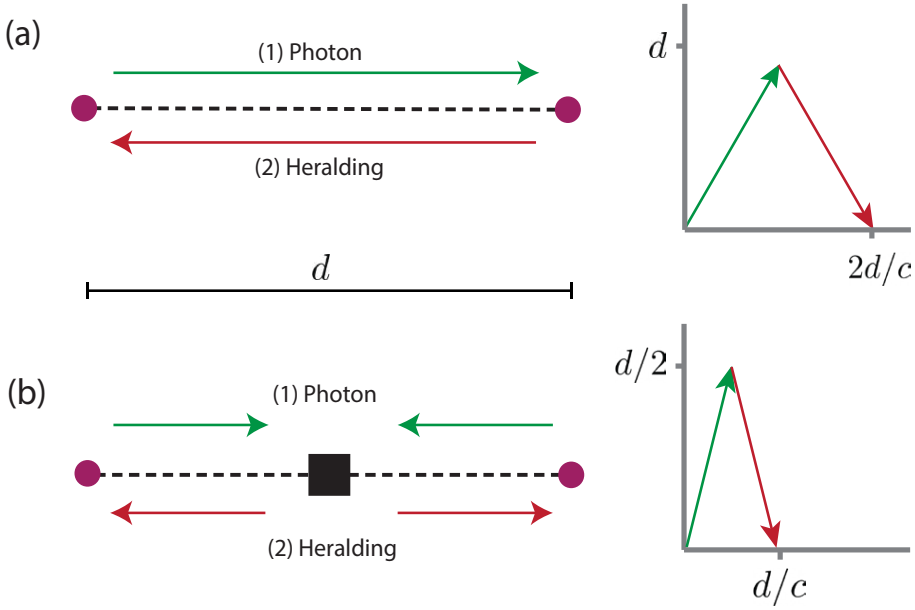


Figure 1.1: **Heralded entanglement generation takes up a fixed time interval.** Heralded entanglement generation schemes involve the transmission of quantum and classical information over some distance within a physical medium that could be e.g. optical fibre (dotted line). For example, this can be (a) the full distance between repeater nodes when the photon is sent from one node to another [31], or (b) the distance between the repeater nodes and a central station (black square) [23, 25]. In either case, the time until success or failure is proportional to the distance travelled by the signal divided by the speed of light, c.

classical information over long, fixed distances, a single attempt takes up a fixed time interval (see Figure 1.1). The length of the time interval is proportional to the time it takes light to travel between repeaters. Therefore, time is naturally divided into *discrete time steps*, where in each time step a single entanglement generation attempt is carried out. In each attempt, an entangled link is successfully generated with some probability of success, which is dependent on the entanglement generation protocol used and the hardware parameters of the physical system in question. For example, these parameters include the probability of successful emission of a photon and the probability of the photon successfully reaching the receiver. We also note that the abstraction of physical hardware is not only convenient for theoretical study, but also in practice. In particular, a network control architecture that coordinates high-level network operations such as entanglement swapping should be compatible with multiple possible candidates of the network hardware [28]. The time step may not only correspond to a single execution of a physical, link-level entanglement generation protocol, but for example batches of executions [29], or even end-to-end entanglement generation over a repeater chain [30]. In this way, the discretisation of time enables modularity of quantum network control protocols by abstracting not only the physical hardware, but also higher-level protocols for entanglement distribution.

ENTANGLEMENT SWAPPING

As explained in the previous section, if quantum repeaters are not employed, the entanglement generation rate decays exponentially with distance with a fibre-based implementation. First-generation repeaters overcome this limitation with a protocol known as *entanglement swapping* [32]. Consider a simple scenario with two end nodes and an intermediate (repeater) node placed between them. Suppose that the repeater node shares an entangled link with each end node. Then, an entanglement swap transforms the two shorter-distance links into a link shared between the end nodes, as depicted in Figure 1.2. This is achieved by applying local quantum measurements at the repeater, followed by classical communication of the measurement outcome to the end nodes.

To see how entanglement swapping can improve the entanglement generation rate, let us consider the following simple scenario, again illustrated in Figure 1.2. Suppose that the distance between the end nodes is d and that the attenuation length, which quantifies fibre loss, is d_{att}. The probability that a photon reaches the other end is therefore proportional to $e^{-\frac{d}{d_{\text{att}}}}$, and the average waiting time to generate entanglement with direct transmission scales as

$$T_{e2e} \propto \frac{1}{e^{-\frac{d}{d_{\text{att}}}}} = e^{\frac{d}{d_{\text{att}}}},$$

which increases exponentially with d. In particular, the entanglement generation rate without a repeater scales as

$$R_{e2e} \propto \frac{1}{T_{e2e}} = e^{-\frac{d}{d_{\text{att}}}}. \tag{1.1}$$

Now, let us consider the case with the repeater. Suppose that the left-hand link is generated first. Upon successful generation, the first link is stored in memory while the second link is being generated. When the second link is generated, an entanglement swap is performed immediately in order to obtain end-to-end entanglement. Assuming that the quantum memories are perfect (i.e. a link can be stored in memory for an unlimited amount of time), the average time until end-to-end entanglement is achieved is given by

$$T_{\text{rep}} = T_1 + T_2 \propto \frac{1}{e^{-\frac{d}{2d_{\text{att}}}}} + \frac{1}{e^{-\frac{d}{2d_{\text{att}}}}} = 2e^{\frac{d}{2d_{\text{att}}}}.$$

Here, T_1 is the time to generate the first link and T_2 is the time to generate the second link. With the repeater, the entanglement generation rate scales as

$$R_{\text{rep}} = \frac{1}{T_{\text{rep}}} \propto e^{-\frac{d}{2d_{\text{att}}}}. \tag{1.2}$$

In particular, when comparing (1.1) and (1.2) we see that placing a repeater in the middle can improve the scaling of the entanglement generation rate by a square-root factor. By a simple extension of the above argument, placing $n-1$ repeaters between the end nodes, also known as a *repeater chain*, would result in a scaling

$$R_{n,\text{rep}} \propto e^{-\frac{d}{nd_{\text{att}}}}.$$

This argument highlights that quantum repeaters based on entanglement swapping can lead to an improved scaling of the entanglement generation rate that goes beyond the

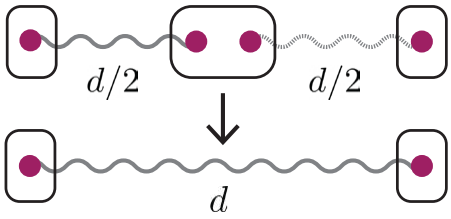


Figure 1.2: **Near-term quantum repeaters make use of entanglement swapping.** With local quantum operations and classical communication, entanglement swapping transforms two shorter-distance entangled states into a longer-distance entangled state. In a sequential repeater protocol, firstly the left-hand link is generated (solid wavy line), and then the right-hand link (dashed wavy line). As illustrated in the text, if memories are perfect then with the sequential repeater protocol and a single repeater, a quadratic improvement in the end-to-end entanglement generation rate is achieved.

limitations of direct transmission. We note that our argument was assuming that the repeater works perfectly: swapping may also be probabilistic [33, 34], or the entangled links may be subject to noise while stored, meaning that they are no longer entangled and the links have to be discarded [35].

NOISY MEMORIES

Above, we saw that quantum repeaters based on entanglement swapping can improve the end-to-end entanglement generation rate. However, this comes at the expense of constructing repeater stations in possession of quantum memories with a long memory lifetime. Developing memories with a long lifetime is experimentally challenging and much work is being carried out on improving the memory lifetime on various qubit platforms – see e.g. [36, 37, 38, 39]. A limited memory lifetime means that the quantum state stored in memory *decoheres* and therefore that the quality of stored entanglement decreases over time, meaning that the link will eventually be rendered useless. A closely related parameter to the memory lifetime is the *decoherence rate*, which is defined to be the inverse of the memory lifetime. To perform a realistic study of near-term quantum networks, decoherence must be taken into account. For example, consider again the simple entanglement swapping scenario from Figure 1.2. However, suppose now that the memory lifetime is limited. Then, it is possible that the first link expires before the second link is generated. If the first link expires, the entire protocol must start again by generating the first link [35]. Thus, a limited memory lifetime can have a negative effect on the end-to-end entanglement generation rate, as well as the quality. In particular, for a functional quantum repeater one generally expects that the decoherence rate should be lower than the entanglement generation rate [40]. A common way of combatting decoherence in a quantum network protocol is with *cut-offs*, where a state is discarded after its quality decays below a certain value [41, 42, 43, 44, 45]. This ensures that the delivered entanglement is of a quality high enough such that a target application can be executed successfully. Moreover, in general, quantum noise requires a large number of parameters to be specified (see e.g. Chapter 8 of [11]). It is therefore common in theoretical studies that noise is approximated to take a simplified form. Examples are Pauli

noise or depolarising noise, which are always in theory possible to obtain from a general noise channel by applying certain symmetrising operations [46, 47, 48]. Using such an approximation means that fewer parameters are required to fully specify the noise model, and that the system dynamics are often simpler to understand. With these simplifying assumptions, an analytical study is often more tractable and a simulation can be made be more efficient.

COMBATTING NOISE WITH ENTANGLEMENT PURIFICATION

One potential avenue to improve the quality of noisy quantum states is with *entanglement purification* [49, 50]. Given n low-quality quantum states, an n-to-k entanglement purification protocol outputs k higher-quality entangled states with some success probability. An entanglement purification protocol is comprised of local operations, measurements and classical communication. The simplest protocols are two-to-one [49, 50, 51]. Similarly to heralded entanglement generation, entanglement purification requires classical communication between distant parties. This incurs extra time delays and therefore more decoherence. Indeed, combining entanglement purification with a heralded entanglement generation scheme can be viewed as a new heralded entanglement generation scheme [51]. The probability of success and the output fidelity depend both on the form of the initial states and on the specific entanglement purification protocol employed. The space of possible entanglement purification protocols is large and the best choice of protocol varies depending on the scenario in question, such as the number of initial states and noise present in the system [52, 53]. We note that encoding states in repeater nodes with quantum error correction schemes is a more advanced strategy to combat noise and would remove the need for extra classical communication between distant nodes [54]. However, quantum error correction requires powerful repeater nodes and is therefore viewed as a component of later-generation quantum repeaters [13, 10, 12].

DEMAND FOR ENTANGLEMENT

As a quantum network is constructed to generate entanglement between users in order to be consumed for an application, it is also important to incorporate demand from network users into the model and performance analysis. This can be used to understand the performance of the network as seen by users. Due to a lack of knowledge about the behaviour of network users, demands are typically assumed to arrive stochastically. A typical assumption in queueing systems is that demands for entanglement arrive according to a Poisson process [55]. In the Poisson case, the performance analysis is often much simpler in comparison to when demands arrive according to a general distribution. A performance analysis with the Poisson assumption is then viewed as a first step towards understanding the system for general inter-arrival times, and can be used to provide intuition about system behaviour in other more general cases. The format of demand varies with the system studied. For example, a demand may request the delivery of a single entangled pair [45, 44, 56, 57], an entanglement packet (multiple entangled pairs within a short space of time) [58, 28, 59], or that a continuous supply of entangled pairs is delivered at some rate over a given time period [60, 61].

1.2. PERFORMANCE ANALYSIS

It can be seen from the previous section that near-term quantum networks are inherently stochastic. There are several sources of stochasticity. Firstly, **heralded entanglement generation schemes** succeed probabilistically because there is a high probability that a photon is lost when sent across long distances. Hence, when using heralded schemes, successful entanglement generation can take multiple trials and the time until success is a random variable. Secondly, in entanglement purification, the **inherently probabilistic nature of quantum measurements** results in entanglement purification succeeding probabilistically. Lastly, we saw that **limited knowledge about network users** means that demands must be modelled as arriving stochastically. Due to the impact all of these factors, it can be challenging to understand relevant performance metrics of even the most simple architectures, because performance metrics have a potentially complex dependence on these underlying stochastic quantities. Common examples of performance metrics include the expectation value, variance, or full distribution of a random variable of interest, such as the time until an entangled state is delivered [62, 63, 58], or the fidelity of delivered states [56, 64, 65].

It is extremely beneficial to characterise performance metrics with *analytical methods*, by which we mean that mathematical tools are used to understand such quantities. If analytical methods are not used, a simulation must be performed to evaluate the performance metrics numerically, which can take a significant amount of time and computational resources. Due to the stochastic complexity of near-term networks, in many cases an analytical study is not feasible and therefore simulation is viewed as an integral component to evaluate performance. Several software packages have been developed specifically for the purpose of quantum network simulation – see e.g. [66, 67, 68, 69].

Much of this thesis is dedicated to an analytical study of performance metrics in several important scenarios. Carrying out an analytical study can save computational resources, as well as provide fundamental insights of the dependency of high-level performance metrics on network protocols and properties of the network hardware. If performance metrics are efficiently computable, they may then be subsequently optimised, to find fundamental limits on network performance.

The performance metrics considered in this thesis each fall into one of the following categories:

- **Rate metrics:** these impact the rate with which entanglement is delivered to network users. Examples include the expected waiting time for entanglement to be delivered to users after a request is submitted [62, 63, 58], the expected probability that a request will be served successfully [56], or the probability of successful entanglement purification (if one is evaluating purification performance) [53, 52].
- **Quality metrics:** these quantify the quality of delivered entanglement. Common examples include the (expected) fidelity of delivered states [56, 64, 65], or the output fidelity after successful entanglement purification [53, 52]. Instead of fidelity, one can also consider measures of entanglement such as the concurrence or negativity [70, 71].
- **Combined metrics:** these capture a mixture of the above metrics. Examples include measures of application performance, such as the secret-key rate [72, 4] or

the number of applications that can be executed per second [73]. The latter depends indirectly on the quality of states in memory, since state quality impacts the probability of successful application execution [73, 74]. One may also artificially construct a combined metric so that is in principle applicable to all applications, such as a measure of the entanglement quality multiplied by the entanglement generation rate [75].

In multi-user networks that contain many nodes in a complex topology, one may consider other metrics that take into account the number of users to whom entanglement is delivered, such as with percolation thresholds [76, 77] or the virtual neighbourhood size [78]. However, since this thesis is only concerned with two-user scenarios, we do not consider these metrics (although the scenarios considered may be subroutines in a larger network – see Figure 1.3). We note that combined metrics are important because there is typically a trade-off between entanglement generation rate and quality for all components of the networks, from hardware to software. Thus, finding the best balance between these factors should be achieved with a combined metric. However, in this thesis, we are mainly concerned with the first two forms of metric.

When carrying out a performance analysis, the choice of performance metric(s) depends on the system in question. For example, in the performance analysis of entanglement purification protocols, it is natural to compute the probability of successful purification and/or the fidelity of states post-purification [53, 52]. However, now suppose that one is analysing a protocol that generates end-to-end entanglement along a repeater chain that will be used for quantum key distribution, and the repeater protocol may involve entanglement purification as a subroutine. Then, one may instead wish to compute the secret-key rate achieved with the generated end-to-end entanglement [79, 35].

1.2.1. THESIS OUTLINE

In the previous sections, we introduced a generic model for near-term quantum networks. This thesis contains a selection of analytical studies of important scenarios occurring in a quantum network – see Figure 1.3 for a depiction. Here, we provide a brief introduction to the scenarios studied in this thesis, some of the results obtained, and their impact. Each chapter also includes its own introduction for readability.

ENTANGLEMENT BUFFERS

An entanglement buffer is a system that stores high-quality entangled links and ensures that they are readily available to consume for quantum protocols when needed. In Chapters 2 and 3, we study buffers that have a single memory with a long memory lifetime (a 'good memory'), and several memories that are used for entanglement generation that have a short memory lifetime (the 'bad memories'). Whenever new link(s) are generated in the bad memories, there is the chance to purify the link in the good memory. An important innovation we make in our study of this system is to take into account the impact of entanglement purification on state quality. Entanglement purification can be complex to incorporate because the probability of success and output fidelity of a protocol depend on the entangled link quality. The protocol performance therefore varies

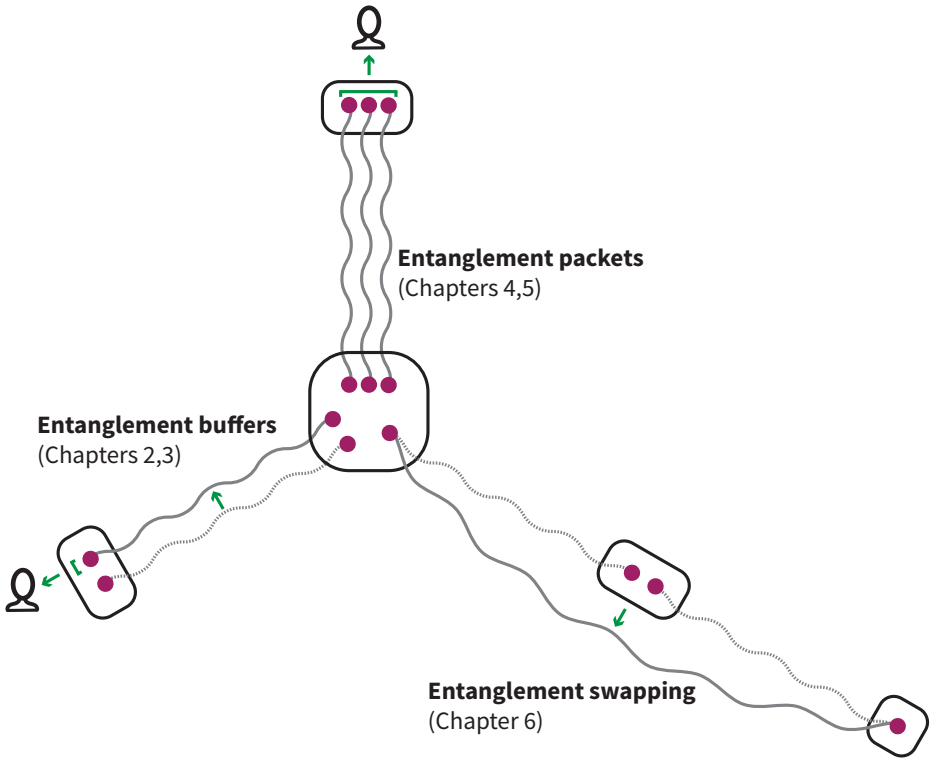


Figure 1.3: **Depiction of scenarios studied in the thesis.** The purple dots are quantum memories, the wavy lines are bipartite entangled states (links). Links may be consumed by network users for an application (depicted). See main text for more details of each chapter.

according to the state of the dynamically evolving system. In Chapters 2 and 3, we overcome this issue and derive solutions for two key performance metrics. Our solutions are closed-form and computationally efficient to evaluate, and keep the entanglement purification protocol completely general. In particular, in Chapter 3, we use our solutions to draw fundamental conclusions about purification policies, such as that more frequent purification always leads to an increased average consumed fidelity. Moreover, it enables us to easily compare different purification policies, and draw conclusions such as that simple purification policies tend to provide a better performance than complex purification policies. The difference between the two chapters is that Chapter 3 is a more general model than Chapter 2: there, we include multiple bad memories, a completely general entanglement purification policy, and carry out the analysis in discrete time. By contrast, in Chapter 2 we consider a system with a single bad memory, a certain class of entanglement purification policy, and carry out the analysis in continuous time. Thus, the reader only interested in the analytical solutions should jump straight to Chapter 3. However, we believe that the methodology presented in Chapter 2 is still highly relevant

for use in the field.

Key system ingredients:

- Two-node scenario.
- One good (long-lasting) memory, one or multiple bad (short-lasting) memories.
- Heralded, probabilistic entanglement generation.
- Decoherence.
- Entanglement purification.
- Entangled link consumption.

System goal:

- Make sure a high-quality link is available most of the time.

ENTANGLEMENT PACKET GENERATION

Many quantum network protocols require multiple high-fidelity entangled links to be available simultaneously. The entangled links may be subsequently processed for a quantum network application such as blind quantum computation [5], or to produce a higher-fidelity link with entanglement purification. Multiple simultaneously-existing links are also collectively referred to as an *entanglement packet* [28]. The generation of entanglement packets is a task of fundamental importance for a functional quantum network. In Chapters 4 and 5, we study the performance of entanglement packet generation. Firstly, in Chapter 4, we find analytical solutions for the average waiting time until an entanglement packet is generated, among other related quantities. Our solutions allow us to better understand the impact of several important parameters on performance. These include protocol parameters such as the number of entangled links required, and hardware parameters such as the probability of entangled link generation. Secondly, in Chapter 5 we consider a generalised version of the system in Chapter 4. In particular, instead of the entanglement generation parameters (probability of success and generated fidelity) being constant in each time step, we now allow the system to choose the entanglement generation parameters from a given set in each time step. This is motivated by existing entanglement generation schemes, where the probability of successful entanglement generation may be increased at the expense of the quality of the generated initial state, or vice-versa. By formulating the system as a Markov decision process, we go on to find policies that adaptively vary the system success probability in order to optimise the rate of entanglement packet generation. In the experimentally-motivated parameter regimes explored, our adaptive policies are found to outperform the constant-action policies (studied in Chapter 4) by a factor of up to 20. We conclude that, given an adjustable rate-fidelity trade-off present in the entanglement generation scheme, it can be highly advantageous to use adaptive protocols to boost the generation rate of entanglement packets.

Key system ingredients:

- Two-node scenario, multiple long-lasting memories.

- Heralded, probabilistic entanglement generation.
- Decoherence.

System goal:

- Quickly generate multiple links that exist simultaneously (entanglement packet) .

TWIRLED APPROXIMATIONS FOR ENTANGLEMENT SWAPPING

In our final study, presented in Chapter 6, we question an assumption that is commonly made in performance analyses: that the initial quantum states of a repeater chain are assumed to have a symmetrised form, known as a *twirled approximation*. It is called this way because any state may be transformed to such a form by implementing a twirling map, where random quantum gates are applied [47, 46]. Using such an approximation has many advantages: for example, twirled states require fewer parameters to be fully specified, which can simplify the analysis because there are fewer parameters to keep track of. Moreover, when a twirled approximation is used for the initial states of a repeater chain (before swapping), the symmetrised form is preserved, which simplifies further the calculation of the end-to-end state. However, using such approximations may also lead to an inaccuracy in the computation of the end-to-end state. Chapter 6 is devoted to quantifying this inaccuracy. We consider two scenarios: unconditional and conditional entanglement swapping. In conditional swapping, the end-to-end state after entanglement swapping is conditioned on the measurement outcome obtained when the swap is performed at each repeater. In unconditional swapping, the output state is a weighted average of all conditional outcomes. A key result is that in many important scenarios, the twirled approximation is exact for unconditional swapping. For conditional swapping, we find bounds on the difference in post-swap fidelity from what is obtained with the twirled approximation, for initial states with a general noisy form. Our study may be used to assess whether the twirled approximation is justified in a given situation.

Key system ingredients:

- Repeater chain.
- Noisy initial states.

System goal:

- Generate end-to-end state.

2

ENTANGLEMENT BUFFERING WITH TWO QUANTUM MEMORIES

Bethany Davies*, Álvaro G. Iñesta* and Stephanie Wehner

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.

*These authors contributed equally.

This chapter has been published separately at Davies, Bethany, Álvaro G. Iñesta, and Stephanie Wehner. "Entanglement buffering with two quantum memories." Quantum 8 (2024): 1458.

2.1. 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 [12]. 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 [49, 50, 80, 81]. An m-to-n entanglement purification protocol consumes m entangled quantum states of low quality and outputs n states with a higher 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 [11].

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 2.1). 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 [82] and [83], 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.

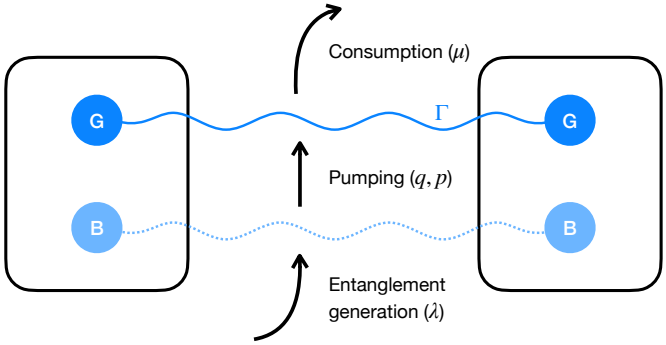


Figure 2.1: Illustration of the entanglement buffering system with two quantum memories (1G1B system). Each of the nodes has two memories (G and B). Memory G is used to store the buffered link. An entangled link is generated at a rate λ in memory B. If memory G is empty when the new link is generated 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 2.3.

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 [84]). 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 2.2, we provide a short overview of related work. In Section 2.3, we explain the physical setup and provide a formal definition of the 1G1B system as a stochastic process. In Section 2.5, we define the performance metrics of interest and provide analytical expressions that enable their computation. In Section 2.6, 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 2.6.1, 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 2.7, we discuss the implications of this work and future research directions.

2.2. 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, [44, 41, 45] 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 [43, 58]. 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 [14, 85]. 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. [86, 87, 88]), 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. [89]), or only considers specific purification protocols [90]. 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 [78, 91]. To the best of our knowledge, the only work with a similar set-up to ours is [65], 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 [65], 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?*

2.3. THE 1G1B (ONE GOOD, ONE BAD) SYSTEM

We now define the 1G1B system. In Section 2.3.1, we describe and motivate the model of the system. In Section 2.4, we define the variables of interest precisely. This facilitates the definition of the performance metrics in Section 2.5.

2.3.1. 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 2.1.

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 [92, 82, 93].

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 [94, 25]. 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 [14] and an assumption that is commonly made in the theoretical analysis of quantum networks [78, 45, 58]). 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 [14, 85], 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 [95]).

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 $\{|\phi^+\rangle, |\psi^+\rangle, |\phi^-\rangle, |\psi^-\rangle\}$ 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 $|\phi^+\rangle$. Any state can be transformed into a Werner state with the same fidelity by applying extra noise, a process known as *twirling* [47, 96]. 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 [46]. After a time t in memory, this maps the state fidelity F to

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

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 [55]).

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 [14], while entanglement pumping may take around $0.5 \cdot 10^{-3}$ s [82]. 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 $J(F, \rho_{\mathrm{new}})$, and remains in the good memory. If pumping fails, all links are discarded from the system.** Here, the *jump function* $J(F, \rho_{\mathrm{new}}) \in [0,1]$ 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

   $$J(F, \rho_{\mathrm{new}}) = \frac{\tilde{a}(\rho_{\mathrm{new}})F + \tilde{b}(\rho_{\mathrm{new}})}{p(F, \rho_{\mathrm{new}})}, \tag{2.1}$$

Table 2.1: 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
$J(F,\rho_{\text{new}})$	Jump function: fidelity of the output state following successful pumping (F is the fidelity of the Werner state stored in the good memory)

with

$$p(F,\rho_{\text{new}}) = c(\rho_{\text{new}})F + d(\rho_{\text{new}}) \tag{2.2}$$

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 2.8.1 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 2.6.1). 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 2.1.

2.4. 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.

Definition 2.1 (1G1B system)**.** Let $s(t)$ be the state of the 1G1B system at time t. This takes values

$$s(t) = \begin{cases} \emptyset \text{ if there is no link in memory,} \\ i \geq 0 \text{ if there is a link in memory which is the result of } i \text{ successful pumping rounds,} \end{cases} \tag{2.3}$$

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 λqp. 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 $\{\emptyset, 0, 1, ...\}$. The resulting CTMC and the rate of transitions is depicted in Figure 2.2. This is the main object of study in our work.

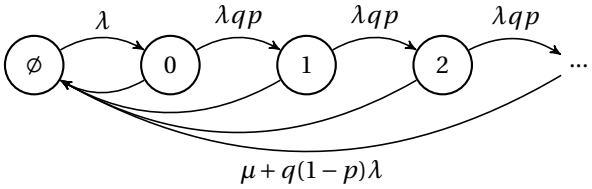


Figure 2.2: The transitions of the 1G1B system.

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

Definition 2.2. 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 j ($j \leq i-1$), and time $T_i(t)$ is the time spent so far in state i. See Figure 2.3 for a depiction of this.

We also need a framework with which to compute the fidelity at time t. Recalling assumption (5) of Section 2.3.1, we denote decoherence by the following.

Definition 2.3. 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 2.4. 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 (2.1).

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 2.5. 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

$$F^{(i)}(t_0, ..., t_i) = D_{t_i}\left[J(F^{(i-1)}(t_0, \ldots, t_{i-1}), \rho_{\text{new}})\right], \tag{2.4}$$

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

$$F(t) = \begin{cases} F^{(i)}(\vec{T}(t)) \text{ if } s(t) = i \geq 0, \\ 0, \text{ if } s(t) = \emptyset. \end{cases} \tag{2.5}$$

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:

$$F^{(i)}(t_0, \ldots, t_i) = D_{t_i}\left[J^{(i)}(F^{(i-1)}(t_0, \ldots, t_{i-1}), \rho_{\text{new}})\right], \tag{2.6}$$

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 (2.4). 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 2.6.1.

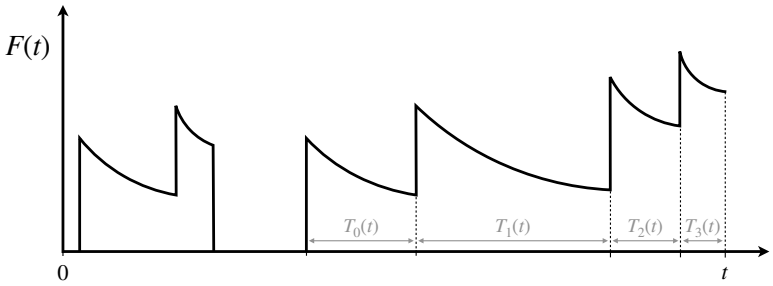


Figure 2.3: Example of the evolution of the fidelity of the buffered entanglement over time. The fidelity experiences a sudden boost every time a pumping protocol is successful. Then, it decays exponentially due to decoherence. Each state in the CTMC is identified by the number of times the current buffered 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.

2.5. 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.

2.5.1. 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

$$\pi_i := \lim_{t\to\infty} P(s(t) = i). \tag{2.7}$$

Then, we define our first performance metric as follows.

Definition 2.7 (Availability)**.** The availability A is defined as

$$A := 1 - \pi_\emptyset, \tag{2.8}$$

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 2.8.2. Moreover, it is possible to derive a closed-form expression for the availability, as stated in the proposition below.

Proposition 2.1. *Consider the 1G1B system (Definition 2.1). The availability is given by*

$$A = 1 - \pi_\emptyset = \frac{\lambda}{\lambda + \mu + \lambda q(1-p)}, \tag{2.9}$$

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

$$\pi_i = \frac{\lambda^{i+1} q^i p^i}{(\mu + \lambda q)^{i+1}} \pi_\emptyset. \tag{2.10}$$

See Appendix 2.8.2 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 2.8.2 for the general formula for the availability.

2.5.2. 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 2.8 (Average consumed fidelity)**.** The *average consumed fidelity* is the average fidelity of the entangled link upon consumption, in the steady state. More specifically,

$$\overline{F} := \lim_{t\to\infty} \mathbb{E}\big[\, F(t) \,|\, s(t) \neq \emptyset \,\big]. \tag{2.11}$$

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

$$\overline{F} = \frac{1}{A}\sum_{i=0}^{\infty} c_i \pi_i, \tag{2.12}$$

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

$$c_i = \mathbb{E}\left[F^{(i)}(Q_0, Q_1, ..., Q_i)\right] \tag{2.13}$$

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 2.5.*

Sketch proof of Theorem 2.1. A first step is to expand by conditioning on the value of $s(t)$,

$$\begin{aligned}\mathbb{E}[F(t)|s(t)\neq\emptyset] &= \sum_{i=0}^{\infty} \mathbb{E}[F(t)|s(t)=i]P(s(t)=i|s(t)\neq\emptyset)\\ &= \frac{1}{P(s(t)\neq\emptyset)}\sum_{i=0}^{\infty}\mathbb{E}[F(t)|s(t)=i]P(s(t)=i)\,.\end{aligned}$$

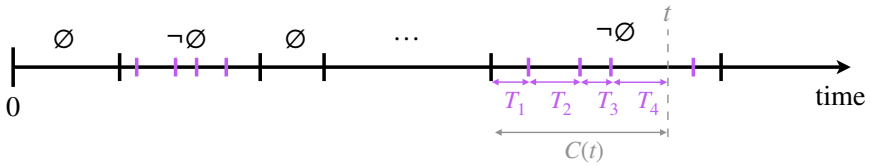


Figure 2.4: An example timeline of the 1G1B process. Black dashes are link generation and removal. Shorter purple dashes are pumping rounds. If there is a link present at time t, the random variable $C(t)$ is the total time spent so far in $\neg\emptyset$ (link present). Pumping rounds occur within the 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 2.1.

In Proposition 2.5 (Appendix 2.8.2), we show that, when $t \to \infty$, the limit can be brought inside of the sum, and so

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

where we have used the definition of the steady-state distribution and the availability (see (2.7) and (2.8)). The values π_i may be computed using Proposition 2.1. The remaining work is then to show that

$$\lim_{t\to\infty} \mathbb{E}\big[F(t)|s(t)=i\big] = \mathbb{E}\left[F^{(i)}(Q_0,\ldots,Q_i)\right], \tag{2.14}$$

which essentially requires the characterisation of the limiting distribution of $\vec{T}(t)$, since from Definition 2.6 we recall that $\mathbb{E}\big[F(t)|s(t)=i\big] = \mathbb{E}\big[F^{(i)}\left(\vec{T}(t)\right)\big|s(t)=i\big]$. 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 2.4 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 2.8.2.

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 2.8.2.

Finally, since $F^{(i)}$ is a continuous function of its inputs, (2.14) follows. □

For the full proof, see Appendix 2.8.2. The particularly simple form of (2.13) 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 2.5, 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 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.

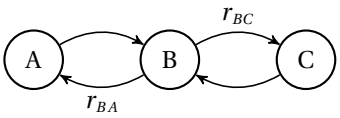


Figure 2.5: In a CTMC, the time spent in a state is independent of the transition that happens next. In this example, the time spent in state B before leaving is exponentially distributed with rate $r_{BA}+r_{BC}$.

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 (2.4) 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 $\overline{F}$.

2.6. 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 2.8.1. 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 2.5 (availability and average consumed fidelity). In Section 2.6.1, 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 2.8.4, 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

$$J(F,\rho_{\text{new}}) = a(\rho_{\text{new}})F + b(\rho_{\text{new}}), \tag{2.15}$$

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 2.6. 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 2.1. The formula requires knowledge of the steady state distribution $\{\pi_i : i = \emptyset, 0, 1, ...\}$, and the expected fidelities c_i, as defined in (2.13). 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 2.1. The work then lies in computing the c_i, which are dependent on the choice of jump function, recalling their definition in (2.13). 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 (2.15) allows us to do this by solving the recurrence relation (2.4). The explicit form of $F^{(i)}$ is provided in the following proposition (see Appendix 2.8.3 for a proof).

Proposition 2.2. *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,*

$$F^{(i)}(t_0,...,t_{i-1},t_i) = \frac{1}{4} + \sum_{j=0}^{i} m_j^{(i)} e^{-\Gamma(t_j+t_{j+1}...+t_i)} \tag{2.16}$$

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

$$m_j^{(i)} = \begin{cases} a^{i-j}\left(\frac{a}{4}+b-\frac{1}{4}\right), \text{ if } j>0, \\ a^i\left(F_{\text{new}}-\frac{1}{4}\right) \text{ if } j=0. \end{cases} \tag{2.17}$$

for $i > 0$.

In the following Lemma, we use the formula for $F^{(i)}$ (found in Proposition 2) and combine this with Theorem 2.1 to derive a closed-form expression for c_i, and therefore for the average consumed fidelity.

Lemma 2.1. *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*

$$c_i = \frac{1}{4} + \left(F_{\text{new}} - \frac{1}{4}\right)\cdot a^i\gamma^{i+1} + \left(\frac{a}{4}+b-\frac{1}{4}\right)\gamma\frac{1-a^i\gamma^i}{1-a\gamma}, \tag{2.18}$$

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

$$\overline{F}_{\text{linear}} = \frac{\frac{1}{4}\Gamma + b\lambda qp + F_{\text{new}}\left(\mu+\lambda q(1-p)\right)}{\Gamma+\mu+\lambda q(1-pa)}. \tag{2.19}$$

The closed-form solution (2.19) 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 2.1 and Equation 2.18. In the following proposition, we see how $\overline{F}$ varies with p and q.

Proposition 2.3. *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 2.1 and Proposition 2.3 in Appendix 2.8.3. 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 2.3 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 2.8.3), 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 2.3 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 (2.9). This behaviour is shown in Figure 2.6. 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 2.6, 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

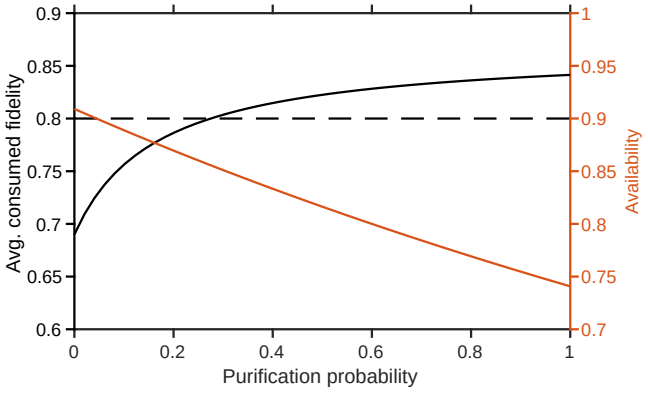


Figure 2.6: Trade-off between average consumed fidelity and availability. When the pumping is good enough (see discussion in main text), the average consumed fidelity $\overline{F}$ (black line) increases with increasing purification probability q, while the availability A (orange line) decreases. The 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 [50].

obtained analytically (our code is available at https://github.com/AlvaroGI/buffering-1G1B).

2.6.1. 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 [84, 97]. 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 [84, 97, 98]. One of their main advantages is that they are relatively simple to execute, since they 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. [82, 83]). In Appendix 2.8.4 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 (2.9). 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 2.2. *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:*

$$\rho_{\text{new}} = F_{\text{new}} \left|\Phi^+\right\rangle\!\left\langle\Phi^+\right| + \lambda_1 \left|\Psi^+\right\rangle\!\left\langle\Psi^+\right| + \lambda_2 |\Psi^-\rangle\langle\Psi^-| + \lambda_3 |\Phi^-\rangle\langle\Phi^-|, \tag{2.20}$$

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

$$F^* = \frac{2F_{\text{new}} - 1 + \sqrt{(2F_{\text{new}}-1)^2 - 2\lambda_{\min}(1 - 2F_{\text{new}} - 2\lambda_{\min})}}{2(2F_{\text{new}} - 1 + 2\lambda_{\min})}, \tag{2.21}$$

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:

$$a_{\text{l}} F + b_{\text{l}} \leq J(F,\rho_{\text{new}}) \tag{2.22}$$

where

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

$$b_{\text{l}} = \frac{F_{\text{new}} + \lambda_{\max}}{2} - \frac{a_{\text{l}}}{4}, \tag{2.23}$$

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

$$J(F,\rho_{\text{new}}) \leq a_{\text{u}} F + b_{\text{u}}, \tag{2.24}$$

with

$$a_{\text{u}} = \frac{4(1 - F_{\text{new}})}{3}, \text{ and } b_{\text{u}} = \frac{4F_{\text{new}} - 1}{3}. \tag{2.25}$$

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

$$p_{\text{l}} = \frac{1}{2}, \text{ and } p_{\text{u}} = F_{\text{new}} + \max(\lambda_1,\lambda_2,\lambda_3). \tag{2.26}$$

A proof of Lemma 2.2 can be found in Appendix 2.8.4. 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 [47, 96]. Note also that the lower bound for the jump function (2.22) 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 2.8.4.

If we regard F_{new} as a fixed parameter, the upper and lower bounds to the jump function (2.22) and (2.24) are linear in F, and the bounds to the success probability (2.26) are constant. It is now possible to find an upper and lower bound for the average consumed fidelity by combining Lemmas 2.1 and 2.2, as we do in the following corollary.

Corollary 2.1. *The average consumed fidelity of the 1G1B system when using any (non-trivial) bilocal Clifford protocol is lower bounded by*

$$\overline{F}_{\mathrm{l}} = \frac{\frac{1}{4}\Gamma + b_{\mathrm{l}}\lambda q p + F_{\mathrm{new}}\Big(\mu + \lambda q(1-p_{\mathrm{l}})\Big)}{\Gamma + \mu + \lambda q(1-p_{\mathrm{l}}a_{\mathrm{l}})}, \tag{2.27}$$

and upper bounded by

$$\overline{F}_{\mathrm{u}} = \frac{\frac{1}{4}\Gamma + b_{\mathrm{u}}\lambda q p + F_{\mathrm{new}}\Big(\mu + \lambda q(1-p_{\mathrm{u}})\Big)}{\Gamma + \mu + \lambda q(1-p_{\mathrm{u}}a_{\mathrm{u}})} \tag{2.28}$$

where a_{l}, b_{l}, p_{l}, a_{u}, b_{u}, and p_{u}, are given by (2.23), (2.25), and (2.26).

Now, we analyse the limits of the performance of the 1G1B system using the bounds on $\overline{F}$ from Corollary 2.1. 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 2.7(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 (2.22) and (2.26). 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 [84]). 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 2.7(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 2.7(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 2.7(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:

$$\Gamma > 4\mu p \frac{F_{\text{new}}(1-a)-b}{4F_{\text{new}}(1-p)+(4b+a)p-1}, \tag{2.29}$$

  where a, b, and p are given by the choice of purification protocol (see (2.15)). In Appendix 2.8.3 we compute this threshold analytically.

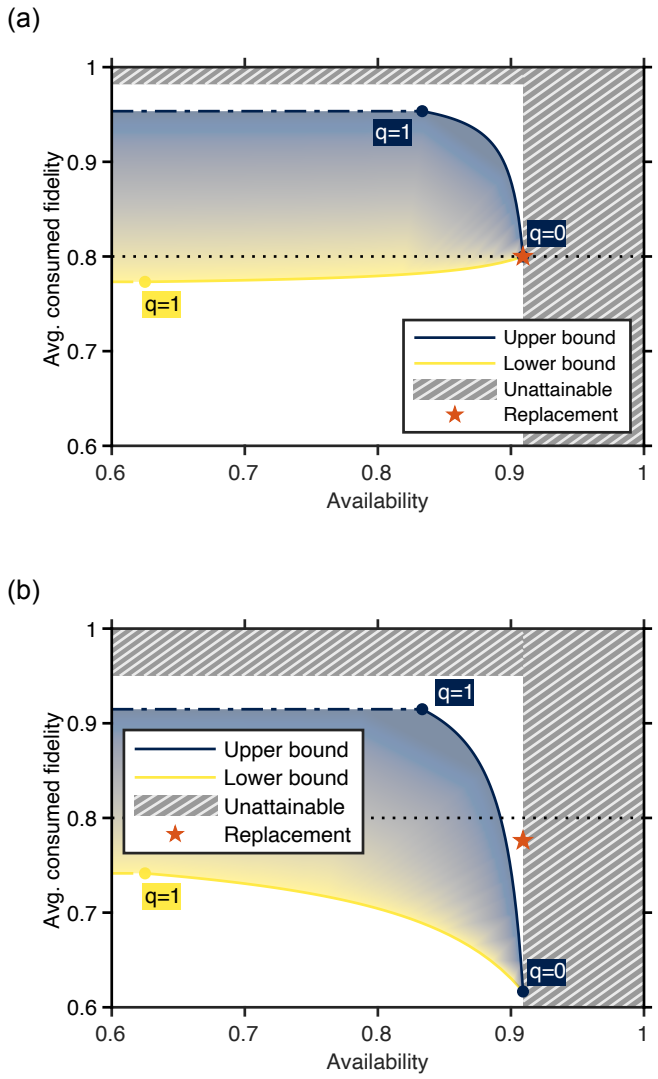


Figure 2.7: Noise in the memories decreases the average consumed fidelity but does not affect the availability. Bounds on the performance of a 1G1B system with a bilocal Clifford protocol, and with **(a)** noiseless memories ($\Gamma = 0$) or **(b)** noisy memories ($\Gamma = 5 \cdot 10^{-2}$ a.u.). For a given target availability, the average consumed fidelity is within the blue/yellow region (see Corollary 2.1). 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): $\lambda = 1$, $\mu = 0.1$, $F_{\text{new}} = 0.8$, $\rho_{\text{new}} = F_{\text{new}} |\phi^+\rangle\langle\phi^+| + (1 - F_{\text{new}}) \left(|\psi^+\rangle\langle\psi^+| + |\psi^-\rangle\langle\psi^-| \right)/2$.

2.7. 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.

2.8. APPENDIX

2.8.1. GENERAL FORM OF JUMP FUNCTION

In this Appendix, we explain the form (2.1) and (2.2) of the jump function and success probability for a general purification protocol, for two input states ρ_{W} and ρ_{new}, where

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

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 [11], the output state is then given by

$$\rho' = \frac{M_{\text{succ}}\Lambda\left(\rho_{\text{W}} \otimes \rho_{\text{new}}\right) M_{\text{succ}}^{\dagger}}{p(F, \rho_{\text{new}})}, \tag{2.30}$$

where

$$p(F, \rho_{\text{new}}) = \text{Tr}\left[M_{\text{succ}}\Lambda\left(\rho_{\text{W}} \otimes \rho_{\text{new}}\right) M_{\text{succ}}^{\dagger}\right]. \tag{2.31}$$

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

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

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 (2.1) and (2.2) for the jump function and success probability.

2.8.2. FORMULAE FOR PERFORMANCE METRICS

In this Appendix, we prove Proposition 2.1 and Theorem 2.1, which provide the formulae for our two performance metrics (availability and average consumed fidelity). Firstly, 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, we employ those results to prove Proposition 2.1 and Theorem 2.1.

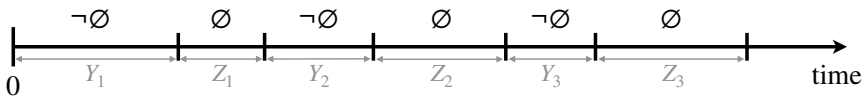


Figure 2.8: The simplified 1G1B process. The system alternates between the states $\neg\emptyset$ (link in memory G) and $\emptyset$ (no link in memory G). The system starts in $\neg\emptyset$. The times spent in $\neg\emptyset$ and $\emptyset$ are denoted by Y_i and Z_i, respectively.

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 2.8.

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

Definition 2.9 (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 2.10. A *renewal process* $\{N = N(t) : t \geq 0\}$ is a process such that

$$N(t) = \max\{n : A_n \leq t\} \tag{2.32}$$

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 $\mathrm{d}g(x) \equiv g'(x)\mathrm{d}x$ for differentiable functions g.

Definition 2.11. 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 [95].

Theorem 2.2 (Renewal Theorem/ Theorem 10.1.11 from [95])**.** *Consider a renewal process as given in Definition 2.10. 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*

$$f(t) = H(t) + \int_0^t f(t-x)\mathrm{d}F_X(x). \tag{2.33}$$

Then, a solution is

$$f(t) = H(t) + \int_0^t H(t-x)\mathrm{d}m(x). \tag{2.34}$$

If H is bounded on finite intervals then f is bounded on finite intervals, and (2.34) is the unique solution of (2.33) 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 2.2, and is useful when taking the infinite limit.

Theorem 2.3 (Key renewal theorem/Theorem 11.2.7 from [95])**.** *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 2.4. *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,*

$$\lim_{t\to\infty} p(t) = \frac{\mathbb{E}(Y_1)}{\mathbb{E}(Y_1) + \mathbb{E}(Z_1)}. \tag{2.35}$$

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

$$p(t) = P(r(t) = \neg\emptyset \cap X_1 > t) + P(r(t) = \neg\emptyset \cap X_1 < t). \tag{2.36}$$

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

$$P(r(t) = \neg\emptyset | X_1 = x) = p(t-x), \tag{2.37}$$

since the process starts afresh at time x. Then, (2.36) becomes

$$p(t) = 1 - F_Y(t) + \int_0^t p(t-x)\mathrm{d}F_X(x), \tag{2.38}$$

where $\mathrm{d}F_X(x) \equiv F'_X(x)\mathrm{d}x$. We now see that this is of the form (2.33) with $H(t) = 1 - F_Y(t)$, and so by Theorem 2.2,

$$p(t) = 1 - F_Y(t) + \int_0^t (1 - F_Y(t-x))\,\mathrm{d}m(x). \tag{2.39}$$

Taking the infinite limit,

$$\lim_{t\to\infty} p(t) = 1 - 1 + \lim_{t\to\infty} \int_0^t (1 - F_Y(t-x))\mathrm{d}m(x). \tag{2.40}$$

It can be seen that $H(t) = 1 - F_Y(t)$ satisfies the conditions (a)-(c) required by Theorem 2.3, so we may apply this Theorem to take the limit:

$$\lim_{t\to\infty} p(t) = \frac{1}{\mathbb{E}[X_1]} \int_0^\infty (1 - F_Y(x))\mathrm{d}x \tag{2.41}$$

$$= \frac{1}{\mathbb{E}[X_1]} \int_0^\infty P(Y_1 > x)\mathrm{d}x = \frac{\mathbb{E}[Y_1]}{\mathbb{E}[X_1]}. \tag{2.42}$$

Finally, using $\mathbb{E}[X_1] = \mathbb{E}[Y_1 + Z_1] = \mathbb{E}[Y_1] + \mathbb{E}[Z_1]$ suffices to show (2.35). □

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

$$C(t) = \begin{cases} t - A_{N(t)}, \text{ if } r(t) = \neg\emptyset, \\ t - A_{N(t)} - Y_{N(t)+1}, \text{ if } r(t) = \emptyset. \end{cases} \tag{2.43}$$

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 2.9 for an illustration of this concept. In the following Lemma, we characterise the distribution of $C(t)$, conditional on being in the state $\neg\emptyset$.

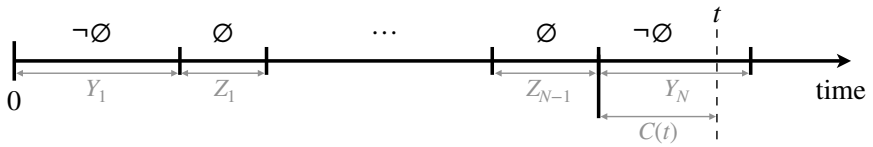


Figure 2.9: Current lifetime of the simplified 1G1B process. The random variable $C(t)$ denotes the time spent so far in the current state at time t. This is most interesting when $r(t) = \neg\emptyset$, because it tells us the age of a link in memory.

Lemma 2.3. *Consider the simplified 1G1B system. The limiting distribution of $C(t)$ conditional on there being a link is given by*

$$\lim_{t\to\infty} P(C(t) > x | r(t) = \neg\emptyset) = \frac{1}{\mathbb{E}[Y_1]} \int_x^\infty (1 - F_Y(s))\mathrm{d}s, \tag{2.44}$$

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

Proof. Writing

$$P(C(t) > x | r(t) = \neg\emptyset) = \frac{P(C(t) > x \cap r(t) = \neg\emptyset)}{P(r(t) = \neg\emptyset)}, \tag{2.45}$$

we see that the bottom of the fraction has already been dealt with in Proposition 2.4. We therefore focus on

$$G(t,x) := P(C(t) > x \cap r(t) = \neg\emptyset). \tag{2.46}$$

Conditioning on X_1, we see that

$$G(t,x) = P(C(t) > x \cap r(t) = \neg\emptyset \cap X_1 > t) + P(C(t) > x \cap r(t) = \neg\emptyset \cap X_1 \le t). \tag{2.47}$$

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

$$P(C(t) > x \cap r(t) = \neg\emptyset | X_1 = y) = G(t-y). \tag{2.48}$$

Then, noting that $G(t,x) = 0$ for $t < x$, (2.47) becomes

$$G(t,x) = \mathbb{1}_{\{t\ge x\}}(1 - F_Y(t)) + \int_0^t G(t-y)\mathrm{d}F_X(y), \tag{2.49}$$

which is in the form of (2.33) with $H(t) = \mathbb{1}_{\{t\ge x\}}(1 - F_Y(t))$. Then, by Theorem 2.2, $G(t,x)$ is given by

$$G(t,x) = \mathbb{1}_{\{t\ge x\}}(1 - F_Y(t)) + \int_0^t \mathbb{1}_{\{t-y\ge x\}}(1 - F_Y(t-y))\mathrm{d}m(y) \tag{2.50}$$

which has limit

$$\lim_{t\to\infty} G(t,x) = 0 + \lim_{t\to\infty} \int_0^{t-x} (1 - F_Y(t-y))\mathrm{d}m(y) \tag{2.51}$$

$$= \lim_{s\to\infty} \int_0^s (1 - F_Y(s+x-y))\mathrm{d}m(y), \tag{2.52}$$

letting $s = t - x$. Then, noting that $g(s) = 1 - F_Y(s+x)$ satisfies conditions (a)-(c) of Theorem 2.3, we may apply this to find

$$\lim_{t\to\infty} G(t,x) = \frac{1}{\mathbb{E}[X_1]}\int_0^\infty g(s)\mathrm{d}s = \frac{1}{\mathbb{E}[X_1]}\int_0^\infty (1-F_Y(s+x))\mathrm{d}s \tag{2.53}$$

$$= \frac{1}{\mathbb{E}[X_1]}\int_x^\infty (1-F_Y(s))\mathrm{d}s. \tag{2.54}$$

From Proposition 2.4, 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 (2.54) as follows:

$$\lim_{t\to\infty} P(C(t) > x | r(t) = \neg\emptyset) = \frac{1}{\mathbb{E}[Y_1]}\int_x^\infty (1-F_Y(s))\mathrm{d}s, \tag{2.55}$$

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,

$$\lim_{t\to\infty} P(C(t) > x | r(t) = \neg\emptyset) = \beta\int_x^\infty e^{-\beta s}\mathrm{d}s = e^{-\beta x}, \tag{2.56}$$

and so conditional on there being a link, the current lifetime approaches an exponential distribution. □

We have now characterised the availability (Proposition 2.4) and current lifetime (Lemma 2.3) for the simplified 1G1B system. However, note that both Proposition 2.4 and Lemma 2.3 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 2.3. The result below states that Theorem 2.3 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 [95] or [99].

Definition 2.13. 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 2.14. 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 2.4 (Key renewal theorem for delayed renewal processes/Theorem 1.20 of [99])**.** *Consider a delayed renewal process $N^{\mathrm{d}}(t)$. If $g : [0,\infty) \to [0,\infty)$ satisfies the same conditions (a)-(c) of Theorem 2.3, then*

$$\lim_{t\to\infty}\int_0^t g(t-x)\mathrm{d}m^{\mathrm{d}}(x) = \frac{1}{\mathbb{E}[X_2]}\int_0^\infty g(x)\mathrm{d}x. \tag{2.57}$$

A consequence of Theorem 2.4 is that even for delayed renewal processes, the limiting distribution is the same as for the non-delayed case. Therefore, the results of Proposition 2.4 and Lemma 2.3 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 2.2. *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,*

$$\lim_{t\to\infty} p(t) = \frac{\mathbb{E}[Y_1]}{\mathbb{E}[Y_1]+\mathbb{E}[Z_1]} = \frac{\lambda}{\lambda+\beta}, \tag{2.58}$$

and the distribution of the current lifetime of a link satisfies

$$\lim_{t\to\infty} P\left(C(t) > x | r(t) = \neg\emptyset\right) = \frac{1}{\mathbb{E}[Y_1]}\int_x^\infty (1 - F_Y(s))\mathrm{d}s = e^{-\beta x}. \tag{2.59}$$

Recalling that $\beta = \mu + \lambda q(1-p)$, we see that the formula for the availability in Proposition 2.1 is already shown by (2.58).

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 2.1), as well as the average consumed fidelity (Theorem 2.1).

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 2.1), 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 2.12). 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 2.2, we see that $C(t)$ converges in distribution to $C \sim \mathrm{Exp}(\beta)$. 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 2.10 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 2.4. *Consider a renewal process $\tilde{N}(t)$ with arrival times $S_0 = 0$, $S_n = \sum_{i=1}^n R_i$, with $R_1 \sim Exp(\delta)$. Let $C \sim Exp(\beta)$ 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,*

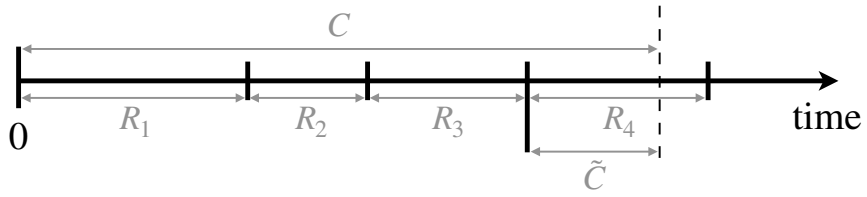


Figure 2.10: Number of pumping rounds. We are interested in the number of pumping rounds that have been carried out while a link is in memory. Here, C is the (limiting) distribution of the current lifetime in memory (see Figure 2.9), and R_i is the time between the $(i-1)$th and ith pumping round.

1. *The distribution of M is given by*

$$P(M \geq m) = \left(\frac{\delta}{\beta+\delta}\right)^m, \tag{2.60}$$

or equivalently

$$P(M = m) = \left(\frac{\delta}{\beta+\delta}\right)^m \left(\frac{\beta}{\beta+\delta}\right). \tag{2.61}$$

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

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

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

where we have used $P(C > R_1|R_1 = x) = P(C > x) = e^{-\beta x}$ and $R_1 \sim \text{Exp}(\delta)$. Then, assuming (2.60),

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

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

$$\begin{aligned} P(C > R_{m+1} + S_m) &= \int_0^\infty \int_0^\infty \mathrm{d}F_R(r)\mathrm{d}F_{S_m}(s) P(C > r+s) \\ &= \int_0^\infty \int_0^\infty \mathrm{d}F_R(r)\mathrm{d}F_{S_m}(s) P(C > r) P(C > s) \\ &= P(C > R_{m+1}) P(C > S_m). \end{aligned} \tag{2.62}$$

Finally, (2.61) follows from

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

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

$$P(E_1 \cap E_2 \cap E_3) = P(E_1 \cap E_2) - P(E_1 \cap E_2 \cap \neg E_3), \tag{2.63}$$

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

$$E_1 = \left\{R_i > x_i \ \forall i = 1, \ldots, m+1\right\}, \ E_2 = \left\{C \geq x_{m+1} + \sum_{i=1}^{m} R_i\right\}, \ E_3 = \left\{C < \sum_{i=1}^{m+1} R_i\right\}.$$

Now,

$$\begin{aligned} E_1 \cap E_2 \cap E_3 &= \left\{R_1 > x_1, \ldots, R_m > x_m, R_{m+1} > x_{m+1} \cap \sum_{i=1}^{m+1} R_i > C \geq x_{m+1} + \sum_{i=1}^{m} R_i\right\} \\ &\overset{a}{=} \left\{R_1 > x_1, \ldots, R_m > x_m, \tilde{C} > x_{m+1} \cap \sum_{i=1}^{m+1} R_i > C \geq \sum_{i=1}^{m} R_i\right\} \\ &\overset{b}{=} \left\{R_1 > x_1, \ldots, R_m > x_m, \tilde{C} > x_{m+1} \cap M = m\right\}, \end{aligned}$$

where in (a) we have used the definition of $\tilde{C}$, and in (b) we have used the definition of M. Then, by (2.63), we see that

$$\begin{aligned} &P\left(R_1 > x_1, \ldots, R_m > x_m, \tilde{C} > x_{m+1} \cap M = m\right) \\ &\qquad = P\left(R_1 > x_1, \ldots, R_{m+1} > x_{m+1} \cap C \geq x_{m+1} + \sum_{i=1}^{m} R_i\right) \\ &\qquad\qquad - P\left(R_1 > x_1, \ldots, R_{m+1} > x_{m+1} \cap C \geq \sum_{i=1}^{m+1} R_i\right). \end{aligned} \tag{2.64}$$

By the independence of the R_i, this is equivalent to

$$(2.64) = \left[P\left(C \geq x_{m+1} + \sum_{i=1}^{m} R_i \middle| R_1 > x_1, ..., R_{m+1} > x_{m+1} \right) \right.$$
$$\left. - P\left(C \geq \sum_{i=1}^{m+1} R_i \middle| R_1 > x_1, ..., R_{m+1} > x_{m+1} \right) \right] \prod_{i=1}^{m+1} P\left(R_i > x_i\right), \quad (2.65)$$

and we now use the memoryless property of the exponential distribution (see the argument leading to (2.62)) to rewrite as

$$(2.64) = \left[P\left(C \geq x_{m+1} \middle| R_{m+1} > x_{m+1}\right) \prod_{i=1}^{m} P\left(C \geq R_i \middle| R_i > x_i\right) \right.$$
$$\left. - \prod_{i=1}^{m+1} P\left(C \geq R_i \middle| R_i > x_i\right) \right] \prod_{i=1}^{m+1} P\left(R_i > x_i\right), \quad (2.66)$$

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

$$(2.64) = \left[P\left(C \geq x_{m+1} \cap R_{m+1} > x_{m+1}\right) - P\left(C \geq R_{m+1} > x_{m+1}\right)\right) \right] \prod_{i=1}^{m} P\left(C \geq R_i > x_i\right),$$
$$= P\left(R_{m+1} > C \geq x_{m+1}\right) \prod_{i=1}^{m} P\left(C \geq R_i > x_i\right),$$

where we have again made use of (2.63) 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

$$(2.64) = \frac{\beta}{\beta + \delta} e^{-(\beta+\delta)x_{m+1}} \cdot \prod_{i=1}^{m} \left[\frac{\delta}{\beta + \delta} e^{-(\beta+\delta)x_i} \right]$$
$$= \frac{\beta}{\beta + \delta} \cdot \left(\frac{\delta}{\beta + \delta} \right)^m \prod_{i=1}^{m+1} e^{-(\beta+\delta)x_i} = P(M = m) \prod_{i=1}^{m+1} e^{-(\beta+\delta)x_i}.$$

It therefore follows that

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

which suffices to show the second result.

□

Recalling that $C(t)$ converges in distribution to C, we now adapt Lemma 2.4 to apply to $C(t)$. In order to do this, we use the following result (for a proof, see Chapter 7 of [95]).

Theorem 2.5 (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 2.3. *Suppose that $C(t)$ and X are independent random variables, and $C(t)$ converges in distribution to C as $t \to \infty$. Then,*

$$\lim_{t\to\infty} P(C(t) > X) = P(C > X). \tag{2.67}$$

Proof. Consider a sequence of times $\{t_n\}_{n\geq 1}$ such that $0 < t_1 < t_2 < ...$ 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 2.5, $C_n - X \to C - X$ 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. □

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 2.4. *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 2.4 still hold in the limit $u \to \infty$. In particular,*

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

$$\lim_{u\to\infty} P(M(u) \geq m) = P(M \geq m) = \left(\frac{\delta}{\beta+\delta}\right)^m. \tag{2.68}$$

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

$$\lim_{u\to\infty} P(X_1 > x_1, ..., X_m > x_m, \tilde{C}(u) > x_{m+1} | M(u) = m) = \prod_{i=1}^{m+1} e^{-(\beta+\delta)x_i}, \tag{2.69}$$

Proof. 1. Making use of Corollary 2.3, 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 (2.65), only replacing C with $C(u)$ and M with $M(u)$, to show that

$$P\left(R_1 > x_1, \ldots, R_m > x_m, \tilde{C}(u) > x_{m+1} \cap M(u) = m\right)$$
$$= \left[P\left(C(u) \geq x_{m+1} + \sum_{i=1}^{m} R_i \middle| R_1 > x_1, \ldots, R_{m+1} > x_{m+1} \right) \right.$$
$$\left. - P\left(C(u) \geq \sum_{i=1}^{m+1} R_i \middle| R_1 > x_1, \ldots, R_{m+1} > x_{m+1} \right) \right] \prod_{i=1}^{m+1} P\left(R_1 > x_i\right).$$

By Corollary 2.3, in the limit $u \to \infty$ this satisfies

$$\lim_{u\to\infty} P(R_1 > x_1, \ldots, R_m > x_m, \tilde{C}(u) > x_{m+1} \cap M(u) = m)$$
$$= \left[P\left(C \geq x_{m+1} + \sum_{i=1}^{m} R_i \middle| R_1 > x_1, \ldots, R_{m+1} > x_{m+1} \right) \right.$$
$$\left. - P\left(C \geq \sum_{i=1}^{m+1} R_i \middle| R_1 > x_1, \ldots, R_{m+1} > x_{m+1} \right) \right] \prod_{i=1}^{m+1} P\left(R_1 > x_i\right) = (2.65).$$

It then follows that

$$\lim_{u\to\infty} P(R_1 > x_1, \ldots, R_m > x_m, \tilde{C}(u) > x_{m+1} \middle| M(u) = m)$$
$$= \lim_{u\to\infty} \frac{P(R_1 > x_1, \ldots, R_m > x_m, \tilde{C}(u) > x_{m+1} \cap M(u) = m)}{P(M(u) = m)}$$
$$= \frac{P(R_1 > x_1, \ldots, R_m > x_m, \tilde{C} > x_{m+1} \cap M = m)}{P(M = m)} = \prod_{i=1}^{m+1} e^{-(\beta+\delta)x_i},$$

by Lemma 2.4.

□

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 2.2). This leads to the following results.

Corollary 2.5. *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 $Q_0 \sim \text{Exp}(\beta + \delta)$.*

We now continue with the formulae for the performance metrics. The availability in the 1G1B system was given in Proposition 2.1 and the average consumed fidelity was given in Theorem 2.1. Next, we prove both of them.

Proof of Proposition 2.1. From Corollary 2.2, 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 2.4 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 2.4, we obtain

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

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 (2.58)) and ties in more neatly with the derivation of the formula for the average consumed fidelity, as we will see below. □

The following proposition will be helpful in the proof of Theorem 2.1 (formula for average consumed fidelity).

Proposition 2.5. *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*

$$\lim_{t\to\infty} \sum_{i=0}^{\infty} e_i(t) p_i(t) = \sum_{i=0}^{\infty} c_i \pi_i. \tag{2.70}$$

Proof of proposition 2.5. To show (2.70), it suffices to show that for any $\epsilon > 0$, there exists a T such that for all $t > T$,

$$\left| \sum_{i=0}^{\infty} e_i(t) p_i(t) - \sum_{i=0}^{\infty} c_i \pi_i \right| < \epsilon. \tag{2.71}$$

We firstly bound the sum using the triangle inequality,

$$\begin{aligned} \left| \sum_{i=0}^{\infty} e_i(t) p_i(t) - \sum_{i=0}^{\infty} c_i \pi_i \right| &= \left| \sum_{i=0}^{\infty} e_i(t)(p_i(t) - \pi_i) + (e_i(t) - c_i)\pi_i \right| \\ &\leq \underbrace{\sum_{i=0}^{\infty} e_i(t) |p_i(t) - \pi_i|}_{(A)} + \underbrace{\sum_{i=0}^{\infty} |e_i(t) - c_i| \pi_i}_{(B)}. \end{aligned} \tag{2.72}$$

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

$$\lim_{t\to\infty} \sum_{i=0}^{\infty} |p_i(t) - \pi_i| = 0. \tag{2.73}$$

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 $\forall\, t > T_1$,

$$\begin{aligned} \left|1 - \sum_{i=0}^{N} p_i(t)\right| &= \left|1 - \sum_{i=0}^{N} \left(\pi_i - (\pi_i - p_i(t))\right)\right| \\ &< \left|1 - \sum_{i=0}^{N} \pi_i\right| + \sum_{i=0}^{N} |\pi_i - p_i(t)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \end{aligned} \tag{2.74}$$

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, $\forall\, t > T$,

$$\begin{aligned} \sum_{i=0}^{\infty} |p_i(t) - \pi_i| &= \sum_{i=0}^{N} |p_i(t) - \pi_i| + \sum_{i>N} |p_i(t) - \pi_i| \\ &< N \cdot \frac{\epsilon}{N} + \sum_{i>N} p_i(t) + \sum_{i>N} \pi_i \\ &< \epsilon + \epsilon + \frac{\epsilon}{2}, \end{aligned}$$

from (2.74). This suffices to show (2.73). 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.

$$\lim_{t \to \infty} \sum_{i=0}^{\infty} |e_i(t) - c_i| \pi_i = 0. \tag{2.75}$$

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,

$$\begin{aligned} \sum_{i=0}^{\infty} |e_i(t) - c_i| \pi_i &= \sum_{i=0}^{N} |e_i(t) - c_i| \pi_i + \sum_{i>N} |e_i(t) - c_i| \pi_i \\ &< \left(\sum_{i=0}^{N} |e_i(t) - c_i|\right) \cdot \left(\sum_{i=0}^{N} \pi_i\right) + \sum_{i>N} \pi_i \\ &< \epsilon(1 - \epsilon) + \epsilon \ \forall\ t > T, \end{aligned}$$

which shows (2.75). □

Combining these results ($(A) \to 0$ and $(B) \to 0$) in (2.72)) suffices to show Proposition 2.5. We are now ready to prove Theorem 2.1 (formula for average consumed fidelity).

Proof of Theorem 2.1. We firstly expand out the average consumed fidelity (Definition 2.8) as a sum by conditioning on the value of $s(t)$,

$$\begin{aligned}\overline{F} := \mathbb{E}[F(t)|s(t) \neq \emptyset] &= \sum_{i=0}^{\infty} \mathbb{E}[F(t)|s(t) = i] P(s(t) = i|s(t) \neq \emptyset) \\ &= \frac{1}{P(s(t) \neq \emptyset)} \sum_{i=0}^{\infty} \mathbb{E}[F(t)|s(t) = i] P(s(t) = i)\,. \end{aligned} \tag{2.76}$$

Recall that we are interested in the limit $t \to \infty$ of the above. Note that from Proposition 2.1, we know the limiting values of $P(s(t) \neq \emptyset)$ and $P(s(t) = i)$. We now claim that

$$\lim_{t\to\infty} \mathbb{E}[F(t)|s(t) = i] = \mathbb{E}\left[F^{(i)}(Q_0, Q_1, ..., Q_i)\right] \tag{2.77}$$

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 2.5. We use the following result:

Theorem 2.6 (Theorem 7.2.19 of [95])**.** *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 2.6). As mentioned in Section 2.3, $F^{(i)}$ is a continuous and bounded function. Therefore, (2.77) follows by combining Theorem 2.6 and Corollary 2.5.

From Proposition 2.5, we therefore see that

$$\begin{aligned}\lim_{t\to\infty} \mathbb{E}[F(t)|s(t) \neq \emptyset] &= \lim_{t\to\infty} \frac{1}{P(s(t) \neq \emptyset)} \sum_{i=0}^{\infty} \mathbb{E}[F(t)|s(t) = i] P(s(t) = i) \\ &= \frac{1}{A} \sum_{i=0}^{\infty} \lim_{t\to\infty} \mathbb{E}[F(t)|s(t) = i] \lim_{t\to\infty} P(s(t) = i) \\ &= \frac{1}{A} \sum_{i=0}^{\infty} c_i \pi_i,\end{aligned}$$

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

2.8.3. AVERAGE CONSUMED FIDELITY WITH A LINEAR JUMP FUNCTION

In this Appendix we focus on linear jump functions. Firstly, we provide bounds for the coefficients of a linear jump function. Then, we prove Proposition 2.2, and 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 2.1). We also show that $\overline{F}_{\text{linear}}$ is monotonic in the probability of pumping q and the probability of successful pumping p (Proposition 2.3). Lastly, we discuss in which situations $\overline{F}_{\text{linear}}$ is monotonically increasing in q, and we compute the noise threshold (2.29) discussed in Section 2.6.1 (above this threshold, any purification is better than no purification).

BOUNDS ON THE PARAMETERS OF A LINEAR JUMP FUNCTION

Proposition 2.6. *Consider a jump function that is linear with the fidelity of one of the input states, i.e.,*

$$J(F,\rho) = a(\rho)F + b(\rho), \tag{2.78}$$

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

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

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

DERIVATION OF $\overline{F}_{\text{linear}}$

Proof of Proposition 2.2. 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

$$F^{(i)}(t_0,...,t_{i-1},t_i) = \frac{1}{4} + \sum_{j=0}^{i} m_j^{(i)} e^{-\Gamma(t_j+t_{j+1}...+t_i)} \tag{2.79}$$

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

$$F^{(0)}(t_0) = D_{t_0}(F_{\text{new}}) = e^{-\Gamma t_0}\left(F_{\text{new}} - \frac{1}{4}\right) + \frac{1}{4}, \tag{2.80}$$

from which we see that $m_0^{(0)} = F_{\text{new}} - \frac{1}{4}$. If we assume that (2.79) is true for some i, using the recursive relation from (2.4) we can show that (2.79) is also true for $i+1$:

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

from which it follows that

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

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 $j > 0$. □

Proof of Lemma 2.1. 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 (2.13)). Using the expression for $F^{(i)}$ from Proposition 2.2 (also given in (2.79)), we can evaluate c_i as follows

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

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:

$$c_i = \frac{1}{4} + \left(F_{\text{new}} - \frac{1}{4}\right) \cdot a^i \gamma^{i+1} + \gamma\left(\frac{a}{4} + b - \frac{1}{4}\right) \frac{1 - a^i \gamma^i}{1 - a\gamma}. \tag{2.81}$$

The final formula for the average fidelity may then be computed with the results of Proposition 2.1 and Theorem 2.1 as

$$\begin{aligned}
\overline{F}_{\text{linear}} &= \lim_{t \to \infty} \mathbb{E}(F(t) | s(t) \ne \emptyset) = \frac{1}{1 - \pi_\emptyset} \sum_{i=0}^{\infty} c_i \pi_i \\
&= \frac{1}{4} + \frac{\gamma}{1 - a\gamma} \cdot \left(\frac{a}{4} + b - \frac{1}{4}\right) + \frac{\gamma}{(1 - \pi_\emptyset)} \cdot \left(F_{\text{new}} - \frac{1}{4} - \frac{\frac{a}{4} + b - \frac{1}{4}}{1 - a\gamma}\right) \sum_{i=0}^{\infty} \pi_i (a\gamma)^i,
\end{aligned} \tag{2.82}$$

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 2.1, we may evaluate the sum as a geometric series,

$$\begin{aligned}\sum_{i=0}^{\infty}\pi_i(a\gamma)^i &= \frac{\lambda}{\mu+\lambda q}\sum_{i=0}^{\infty}\left(\frac{\lambda q p a\gamma}{\mu+\lambda q}\right)^i\pi_\emptyset \\ &= \frac{\lambda}{\mu+\lambda q}\cdot\frac{1}{1-\frac{\lambda q p a\gamma}{\mu+\lambda q}}\pi_\emptyset \\ &= \frac{\lambda}{\mu+\lambda q-\lambda q p a\gamma}\pi_\emptyset.\end{aligned}$$

We may now substitute this into (2.82) to obtain a closed-form solution for the average fidelity,

$$\begin{aligned}\overline{F}_{\text{linear}} &= \frac{1}{4}+\frac{\gamma}{1-a\gamma}\cdot\left(\frac{a}{4}+b-\frac{1}{4}\right)+\gamma\cdot\left(F_{\text{new}}-\frac{1}{4}-\frac{\frac{a}{4}+b-\frac{1}{4}}{1-a\gamma}\right)\frac{\lambda}{\mu+\lambda q-\lambda q p a\gamma}\frac{\pi_\emptyset}{1-\pi_\emptyset} \\ &= \frac{1}{4}+\frac{\gamma}{1-a\gamma}\cdot\left(\frac{a}{4}+b-\frac{1}{4}\right)+\gamma\cdot\left(F_{\text{new}}-\frac{1}{4}-\frac{\frac{a}{4}+b-\frac{1}{4}}{1-a\gamma}\right)\frac{\mu+\lambda q(1-p)}{\mu+\lambda q-\lambda q p a\gamma} \\ &= \frac{\frac{1}{4}\Gamma+b\lambda q p+F_{\text{new}}\big(\mu+\lambda q(1-p)\big)}{\Gamma+\mu+\lambda q(1-pa)},\end{aligned} \tag{2.83}$$

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). □

Proof of Proposition 2.3. To show (a), we compute the partial derivative of the average consumed fidelity with respect to q:

$$\frac{\partial\overline{F}_{\text{linear}}}{\partial q} = \lambda\frac{\Gamma\big(4F_{\text{new}}(1-p)+(4b+a)p-1\big)+4\mu p\big(b-F_{\text{new}}(1-a)\big)}{4\big(\Gamma+\mu+\lambda q(1-ap)\big)^2}. \tag{2.84}$$

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:

$$\frac{\partial\overline{F}_{\text{linear}}}{\partial p} = \lambda q\frac{4(b-F_{\text{new}})(\Gamma+\mu+\lambda q)+a\big(\Gamma+4F_{\text{new}}(\mu+\lambda q)\big)}{4\big(\Gamma+\mu+\lambda q(1-ap)\big)^2}. \tag{2.85}$$

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

□

NOISE THRESHOLD

In the previous Section, we showed that $\overline{F}_{\text{linear}}$ is monotonic in q and p (Proposition 2.3). 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 (2.84) can be written as follows:

$$\frac{\partial \overline{F}_{\text{linear}}}{\partial q} = \frac{\lambda}{x^2}\left(\Gamma y + 4\mu p z\right), \tag{2.86}$$

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 ($\Gamma > -4\mu p z/y$), the derivative is positive and the pumping is always beneficial, even if it succeeds with a very small probability.

2.8.4. 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 2.2, 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 [84, 97], which are a well-known type of purification scheme. We provide more details about this type of protocol in the next section.
- 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

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

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

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

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

Note that any bipartite state can be brought to Bell-diagonal form while preserving the fidelity by means of twirling (adding extra noise) [47, 96].

- We only consider newly generated states with fidelity to some Bell state larger than $1/2$, i.e., we assume $F_{\text{BD}} > 1/2$ (note that $F_{\text{BD}} > 1/2$ is equivalent to $\lambda_i > 1/2$ for some i, since the states are equivalent upon some Pauli corrections). 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 2.3).

In this Appendix, we firstly provide a formal definition of bilocal Clifford protocols. Then, we prove Lemma 2.2, where bounds are found for the jump function and success probability of bilocal Clifford protocols in a system with the above restrictions.

BILOCAL CLIFFORD PROTOCOLS

Bilocal Clifford protocols [84, 97] 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 [100, 101]. 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 [98]. This type of protocol also includes well-known purification protocols, such as DEJMPS [50].

LINEAR BOUNDS TO THE PERFORMANCE OF BILOCAL CLIFFORD PROTOCOLS

Here, we prove Lemma 2.2, where bounds on the jump function and success probability of every bilocal Clifford protocol are found. Consider pumping two states of the form

$$\rho_{\text{W}} = F\left|\Phi^+\right\rangle\left\langle\Phi^+\right| + \frac{1-F}{3}\left(\left|\Psi^+\right\rangle\left\langle\Psi^+\right| + \left|\Psi^-\right\rangle\left\langle\Psi^-\right| + \left|\Phi^-\right\rangle\left\langle\Psi^-\right|\right) \tag{2.87}$$

$$\rho_{\text{BD}} = F_{\text{BD}}\left|\Phi^+\right\rangle\left\langle\Phi^+\right| + \lambda_1\left|\Psi^+\right\rangle\left\langle\Psi^+\right| + \lambda_2\left|\Psi^-\right\rangle\left\langle\Psi^-\right| + \lambda_3\left|\Phi^-\right\rangle\left\langle\Psi^-\right|. \tag{2.88}$$

Using the methods from [97], 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 [97, 98]. There are only seven protocols that provide a unique combination of J and p, as shown in Table 2.2. We refer to the i-th jump function and success probability as $J_i(F, \rho_{\text{BD}})$ and $p_i(F, \rho_{\text{BD}})$, for $i = 1, ..., 7$.

Table 2.2: The jump function and success probability for all 2-1 bilocal Clifford protocols, with input states given in (2.87) and (2.88) .

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 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 2.2.

Proof of Lemma 2.2. We firstly show the linear lower bound (i.e. the formulae given in (2.23)). 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 $\lambda_{\mathrm{min}} = \lambda_3$ and $\lambda_{\mathrm{max}} = \lambda_1$. 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 2.7, Corollory 2.6, Proposition 2.8** – the formula for F^* is derived (Equation 2.21). 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 2.6) satisfies $F(t) \leq F^*$.

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

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

(**Proposition 2.7 and Corollary 2.6**).

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

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

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

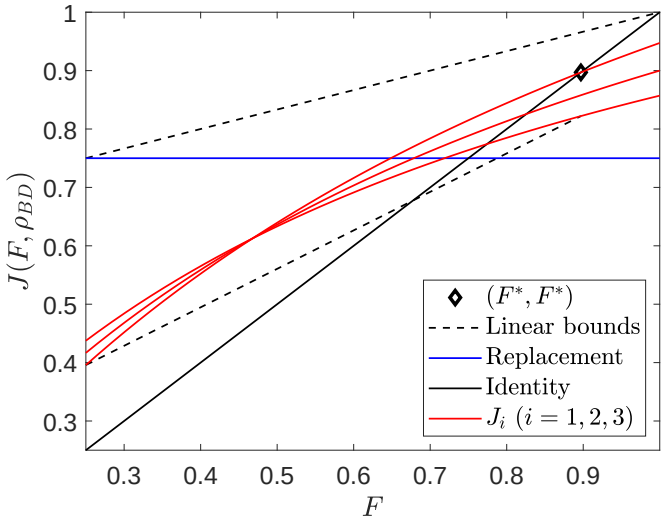


Figure 2.11: Linear bounds for the jump function of bilocal Clifford protocols (black dashed lines). The jump functions shown are J_1-J_3 (red lines), J_4-J_6 (identity operation, black line), and J_7 (probabilistic replacement, blue line). F^* is the highest fidelity achievable by pumping a low-fidelity 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, $F_{\text{BD}} = 0.75$ and $\rho_{\text{BD}} = (0.75, 0.125, 0.833, 0.0417)$.

4. For $i = 1,2,3$, $J_i(F,\rho_{\text{BD}})$ is a concave function of F (**Proposition 2.9**).

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

$$\left(F^*, J_1(F^*,\rho_{\text{BD}})\right), \quad \left(\frac{1}{4}, J_3\left(\frac{1}{4},\rho_{\text{BD}}\right)\right), \tag{2.89}$$

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 $\lambda_{\text{max}} = \lambda_1$ and $\lambda_{\text{min}} = \lambda_3$, 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 2.2 (in (2.23)). When choosing the points in (2.89), 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 2.11 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 2.2 (in (2.25)). 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 2.7**).
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 2.6**).
3. For $i = 1,2,3$, J_i is an increasing and concave function of F (**Proposition 2.9**).
4. Consider the tangent to J_3 at $F = 1$. This lies below J_{UB} in the range $F \in [1/4, 1]$ (**Proposition 2.10**).

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

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

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

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

□

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

Proposition 2.7. *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

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

which simplifies to

$$(\lambda_1-\lambda_2)\left((2-16F_{\text{BD}})F^2+(8F_{\text{BD}}-4)F+2-F_{\text{BD}}\right)=0. \tag{2.90}$$

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

$$F = \frac{4F_{\text{BD}}-2\pm 3\sqrt{2F_{\text{BD}}}}{2(8F_{\text{BD}}-1)} \tag{2.91}$$

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

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

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

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

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

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

□

We now continue with the following corollary.

Corollary 2.6. *Suppose that $\lambda_1 \geq \lambda_2 \geq \lambda_3$. Then, for $F \geq F_{int}$,*

$$J_3(F, \rho_{\text{BD}}) \geq J_2(F, \rho_{\text{BD}}) \geq J_1(F, \rho_{\text{BD}}), \tag{2.92}$$

Proof. From Proposition 2.7, J_1, J_2 and J_3 will not intersect again for $F > F_{\text{int}}$. Therefore, their ordering remains the same for all $F > F_{\text{int}}$. The jump function outputting the largest fidelity in this range will therefore also have the largest limit as $F \to \infty$. We see that

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

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 (2.92). □

From Proposition 2.7 and Corollary 2.6, 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 2.8. *The largest fidelity obtainable by pumping a low-fidelity Werner state with ρ_{BD} and bilocal Clifford protocols is*

$$F^* = \frac{2F_{\text{BD}} - 1 + \sqrt{(2F_{\text{BD}}-1)^2 + 2\lambda_{\min}(2F_{\text{BD}} - 1 + 2\lambda_{\min})}}{2(2F_{\text{BD}} - 1 + 2\lambda_{\min})}, \tag{2.93}$$

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

$$\begin{aligned} & F^* = J_i(F^*,\rho_{\text{BD}}) \\ \Leftrightarrow\ & F^* = \frac{(\lambda_i - 3F_{\text{BD}})F^* - \lambda_i}{(2 - 4F_{\text{BD}} - 4\lambda_i)F^* - 2 + F_{\text{BD}} + \lambda_i} \\ \Leftrightarrow\ & 0 = (2 - 4F_{\text{BD}} - 4\lambda_i)F^2 + (4F_{\text{BD}} - 2)F + \lambda_i, \end{aligned}$$

which has solutions

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

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 (2.93). □

Proposition 2.9. *For any ρ_{BD} with $F_{\text{BD}} > 1/2$, , for $i = 1,2,3$ $J_i(F,\rho_{\text{BD}})$ 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,

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{ad-bc}{(cx+d)^2}, \quad \frac{\mathrm{d}^2y}{\mathrm{d}x^2} = -2c\frac{ad-bc}{(cx+d)^3}. \tag{2.94}$$

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,

$$\begin{aligned} a &= 3F_{\mathrm{BD}}-\lambda_i > \frac{3}{2}-\frac{1}{2} = 1 \\ b &= \lambda_i < \frac{1}{2} \\ c &= 4(F_{\mathrm{BD}}+\lambda_i)-2 \leq 4\cdot 1-2 = 2 \\ d &= 2-(F_{\mathrm{BD}}+\lambda_i) \geq 2-1 = 1 \end{aligned}$$

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 (2.94) that

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

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

Proposition 2.10. *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 $J_{\tan}(F,\rho_{\mathrm{BD}})$. Then, this lies below J_{UB} for all $F\in[1/4,1]$, i.e.*

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

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 2.2.

Proof. We firstly compute the formula for the tangent to J_i at $F=1$. Recalling the formula (2.94), 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 $J_i(1,\rho_{\text{BD}}) = F_{\text{BD}}/(F_{\text{BD}}+\lambda_i)$. We note that at $F=1$,

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

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

$$J_{\text{UB}}\left(\frac{1}{4},\rho_{\text{BD}}\right) \geq J_{\text{tan}}\left(\frac{1}{4},\rho_{\text{BD}}\right), \tag{2.95}$$

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

$$\begin{aligned} J_{\text{UB}}\left(\frac{1}{4},\rho_{\text{BD}}\right) - J_{\text{tan}}\left(\frac{1}{4},\rho_{\text{BD}}\right) &= F_{\text{BD}} - \left(\frac{2F_{\text{BD}}}{3(F_{\text{BD}}+\lambda_3)^2} - \frac{1}{3}\right)\left(-\frac{3}{4}\right) - \frac{F_{\text{BD}}}{F_{\text{BD}}+\lambda_3} \\ &= F_{\text{BD}} - \frac{1}{4} + \frac{F_{\text{BD}}}{2(F_{\text{BD}}+\lambda_3)^2} - \frac{F_{\text{BD}}}{F_{\text{BD}}+\lambda_3}. \end{aligned}$$

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

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

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

$$\lambda_3 \in \left[0, \frac{1-F_{\text{BD}}}{3}\right], \quad \text{and} \quad x \in \left[F_{\text{BD}}, \frac{1+2F_{\text{BD}}}{3}\right]. \tag{2.96}$$

To prove the proposition, it therefore suffices to show positivity of h for x in the range (2.96). We start by establishing the monotonicity of h:

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

since $x = F_{\text{BD}}+\lambda_3 \leq 1$. We therefore see that h is decreasing. To show that h is positive in the range (2.96), it therefore suffices to show that

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

We have

$$\begin{aligned} h\left(\frac{1+2F_{\text{BD}}}{3}\right) &= F_{\text{BD}} - \frac{1}{4} + \frac{9F_{\text{BD}}}{2(1+2F_{\text{BD}})^2} - \frac{3F_{\text{BD}}}{1+2F_{\text{BD}}} \\ &= F_{\text{BD}} - \frac{1}{4} + 6F_{\text{BD}}\left(\frac{\frac{3}{4} - \frac{1}{2}(1+2F_{\text{BD}})}{(1+2F_{\text{BD}})^2}\right) \\ &= F_{\text{BD}} - \frac{1}{4} + 6F_{\text{BD}}\frac{\frac{1}{4} - F_{\text{BD}}}{(1+2F_{\text{BD}})^2} \\ &= \left(F_{\text{BD}} - \frac{1}{4}\right)\left(1 - \frac{6F_{\text{BD}}}{(1+2F_{\text{BD}})^2}\right). \end{aligned}$$

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

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

which holds. We therefore see that

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

for all x in the range (2.96), and therefore

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

This suffices to show the proposition. □

ADDITIONAL PROOFS

Lemma 2.5. *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 2.5. 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 [102, 103]. 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, $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$, 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

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

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., $F(\rho, |\phi^+\rangle) \equiv \langle\phi^+| \rho |\phi^+\rangle = \lambda_0$), we conclude that the state is entangled iff the fidelity to one of the Bell states is larger than $1/2$. □

2.9. 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 2.8.1). 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 2.8.1). 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 2.8.1).

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 2.6). 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):

$$\overline{F} \approx \overline{F}' \equiv \frac{\sum_{i=1}^{N_{\text{samples}}} F_i(t_{\text{sim}})}{N'_{\text{samples}}}, \tag{2.97}$$

where

$$N'_{\text{samples}} = \sum_{i=1}^{N_{\text{samples}}} \mathbb{1}_{F_i(t_{\text{sim}})>0} \tag{2.98}$$

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:

$$\varepsilon_F = \sqrt{\frac{\sum_{i=1}^{N'_{\text{samples}}} \left(F_i(t_{\text{sim}}) - \overline{F}'\right)^2}{N'_{\text{samples}}\left(N'_{\text{samples}} - 1\right)}}, \tag{2.99}$$

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}:

$$A \approx A' \equiv \frac{1}{N_{\text{samples}}} \sum_{i=0}^{N_{\text{samples}}} \mathbb{1}_{F_i(t_{\text{sim}})>0}, \tag{2.100}$$

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

$$\varepsilon_A = \sqrt{\frac{A'(1-A')}{N_{\text{samples}}}}. \tag{2.101}$$

Next, we study again the example from Figure 2.7, and we compare the bounds discussed in the main text with the results from our simulation. In Figure 2.12, we show the same lower and upper bounds (yellow and dark blue lines, respectively) from Figure 2.7. We simulated three buffering systems, each of them using the unique bilocal Clifford protocols 1, 2, and 3 from Table 2.2 (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 2.12 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, $\rho_{\text{new}} = F_{\text{new}} |\phi^+\rangle\langle\phi^+| + (1-F_{\text{new}}) \left(|\psi^+\rangle\langle\psi^+| + |\psi^-\rangle\langle\psi^-|\right)/2$. 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.

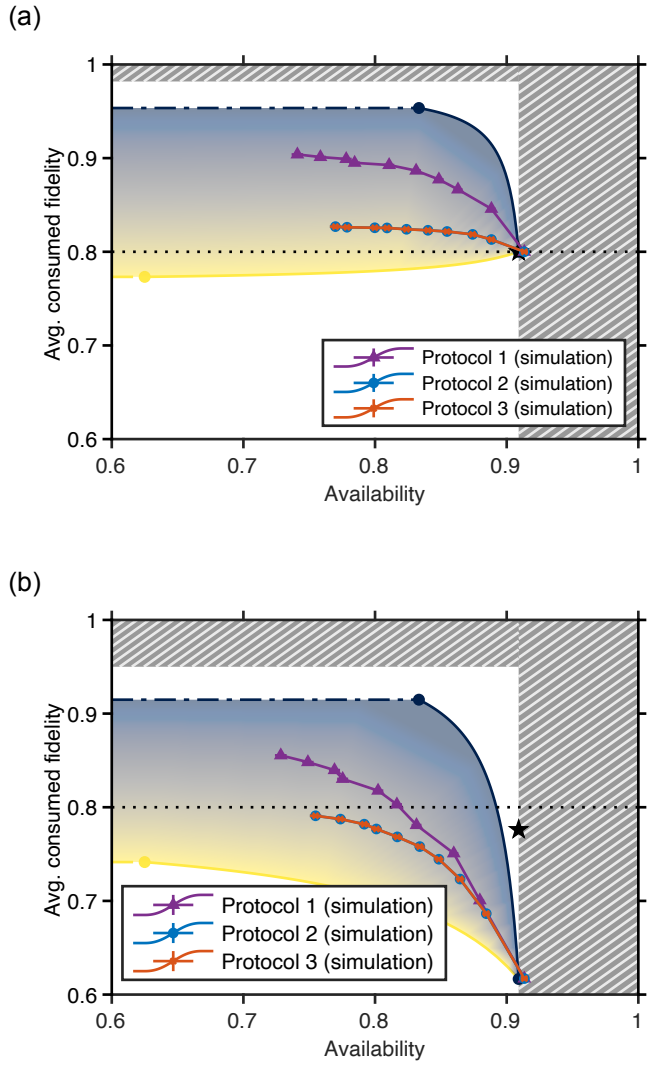


Figure 2.12: Bounds derived assuming a constant probability of success still apply when the assumption is lifted. **(a)** Noiseless memories ($\Gamma = 0$) or **(b)** noisy memories ($\Gamma = 5 \cdot 10^{-2}$ a.u.). For a given target availability, the average consumed fidelity is within the blue/yellow region (see Corollary 2.1). 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 2.2, ($q = 0$ 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): $\lambda = 1$, $\mu = 0.1$, $F_{\text{new}} = 0.8$, $\rho_{\text{new}} = F_{\text{new}} |\phi^+\rangle\langle\phi^+| + (1 - F_{\text{new}}) \left(|\psi^+\rangle\langle\psi^+| + |\psi^-\rangle\langle\psi^-|\right)/2$. Numerical parameters used in the simulation: $t_{\text{sim}} = 50$, $N_{\text{samples}} = 10^4$.

3

ENTANGLEMENT BUFFERING WITH MULTIPLE QUANTUM MEMORIES

Álvaro G. Iñesta*, Bethany Davies*, Sounak Kar and Stephanie Wehner

Entanglement buffers are systems that maintain high-quality entanglement, ensuring it is readily available for consumption when needed. In this work, we study the performance of a two-node buffer, where each node has one long-lived quantum memory for storing entanglement and multiple short-lived memories for generating fresh entanglement. Newly generated entanglement may be used to purify the stored entanglement, which degrades over time. Stored entanglement may be removed due to failed purification or consumption. We derive analytical expressions for the system performance, which is measured using the entanglement availability and the average fidelity upon consumption. Our solutions are computationally efficient to evaluate, and they provide fundamental bounds to the performance of purification-based entanglement buffers. We show that purification must be performed as frequently as possible to maximise the average fidelity of entanglement upon consumption, even if this often leads to the loss of high-quality entanglement due to purification failures. Moreover, we obtain heuristics for the design of good purification policies in practical systems. A key finding is that simple purification protocols, such as DEJMPS, often provide superior buffering performance compared to protocols that maximise output fidelity.

*These authors contributed equally.

This chapter has been released separately at https://arxiv.org/abs/2502.20240.

3.1. INTRODUCTION

Entanglement is a fundamental resource for many quantum network applications, including some quantum key distribution protocols [3, 104], distributed quantum sensing [105, 106, 107, 108], and coordination tasks where communication is either prohibited or insufficiently fast [109, 110]. Pre-distributing entanglement between remote parties would eliminate the need to generate and distribute entangled states on demand, saving time and resources [111, 112, 91, 78]. However, entanglement degrades over time due to decoherence, preventing long-term storage.

Entanglement buffers are systems that store entanglement until it is needed for an application. Passive buffers, which store entanglement in quantum memories, are constrained by the coherence time of these memories [113]. To overcome this limitation, purification-based entanglement buffers have been proposed [56, 65]. These systems store entangled states and employ purification protocols to ensure the states remain high quality, mitigating the effects of decoherence. Purification protocols take m low-quality entangled states as input and produce n higher-quality states as output, typically with $m > n$ [49, 50, 80, 81]. These protocols often involve some probability of failure, in which case all the input states are lost and no entanglement is produced. Here, we focus on purification-based buffers.

As proposed in ref. [56], the performance of an entanglement buffer can be measured with two quantities: the availability (probability that entanglement is available for consumption when requested, see Definition 3.2) and the average consumed fidelity (average quality of entanglement at the time of consumption, see Definition 3.3). As well as having practical utility, entanglement buffers are a useful theoretical tool in order to understand the impact of several important interacting processes that occur in a quantum network: ongoing generation, purification, and consumption of entanglement. Of major interest is the impact of the entanglement purification protocol on the performance of the system. Since the success probability of entanglement purification typically depends on the fidelity of the input states, any rate and fidelity metrics are inherently coupled in systems making use of purification. This coupling adds complexity to analytical calculations. Consequently, most analytical studies on the performance of quantum networking systems exclude purification, and its impact on performance is typically explored with numerical methods [114, 89]. Nevertheless, as is a main result in this work, for entanglement buffering systems closed-form solutions are obtainable for a fully general purification protocol. One may then efficiently compute the performance of a particular purification policy, as well as make formal statements about how often purification should be applied to the buffered entanglement.

Here, we study the *1GnB system*: a purification-based entanglement buffer with one good (long-lived) memory and n bad (short-lived) memories. The good memory can store entanglement, which can be consumed at any time by an application. In contrast, bad memories can generate entanglement concurrently but cannot store it; they act as communication qubits. For instance, carbon-13 nuclear spins in diamond can serve as good memories with coherence times up to 1 min [115], while electron spins in nitrogen-vacancy centers may function as communication qubits, with coherence times generally below 1 s [116].

Each time entanglement is generated in some of the bad memories, the system may

choose to immediately use it to purify the entanglement stored in the good memory. If purification is not attempted, the newly generated entanglement is discarded. We illustrate the 1GnB system in Figure 3.1. Note that the physical platform must enable easy access to stored entanglement for consumption and purification. However, network activities, such as repeated entanglement generation attempts and purification, may introduce additional noise, reducing memory lifetimes. For example, in ref. [14], even when the carbon-13 nuclear spin used as a storage qubit is protected from network noise by applying stronger magnetic fields, it exhibits a shortened lifetime of approximately 11.6 ms.

The 1GnB buffering system is a generalisation of the 1G1B system that was originally proposed in [56]. 1G1B is a system with only one good quantum memory and one bad memory. Here, we generalise the work in [56] in three main ways. Firstly, we now consider several (n) bad memories. Including several bad memories in our model now means that there is the possibility of generating multiple entangled links in the same entanglement generation attempt, for example via frequency [117, 113, 118] or time multiplexing [119, 120], which are commonly proposed ways of improving the rate of entanglement generation [121, 122, 123]. Moreover, the simultaneous generation of multiple links opens up the use of stronger purification protocols, thereby providing an improvement to system fidelity metrics as well as the rate. Note again that the physical implementation of the buffer must allow for such multiplexing and for purification of the generated entanglement. The second generalisation from previous work is that we now model the system in discrete time rather than continuous time, which is more accurate to real-world systems, as entanglement generation typically happens in discrete attempts (see e.g. refs. [25, 124, 94, 125]). Finally, we now derive our solutions for a fully arbitrary purification protocol. In particular, the solutions for performance metrics presented in ref. [56] only apply for purification protocols with a constant probability of success (i.e. the success probability must be independent of the fidelity of the buffered quantum state). However, in this work, we remove this assumption and derive closed-form solutions for the availability and the average consumed fidelity of buffers that use arbitrary purification protocols. This is in contrast to [65], where although performance metrics are derived analytically and the probability of success is not necessarily constant, their computation requires solving a linear system of equations, which has dimension that scales with system parameters such as the memory lifetime.

In this chapter, we firstly provide analytical expressions for the availability, A, and the average consumed fidelity, $\overline{F}$, of the 1GnB system (see model description in Section 3.2). Then, we use these expressions to find fundamental limits to the performance of entanglement buffers. Lastly, we investigate how the 1GnB system should be operated: because there is a large amount of freedom in the choice of purification protocols, it is not clear what purification strategies should be employed to maximise A and $\overline{F}$. For example, would it be beneficial to use a purification subroutine that provides a larger fidelity boost (which could increase $\overline{F}$) if this comes at the cost of a higher probability of failure (which means losing high-quality entanglement more frequently, decreasing A and maybe also $\overline{F}$)? Our main findings are the following:

- MONOTONIC PERFORMANCE – We show that, to maximise the average consumed fidelity, purification must be performed as much as possible, i.e. every time en-

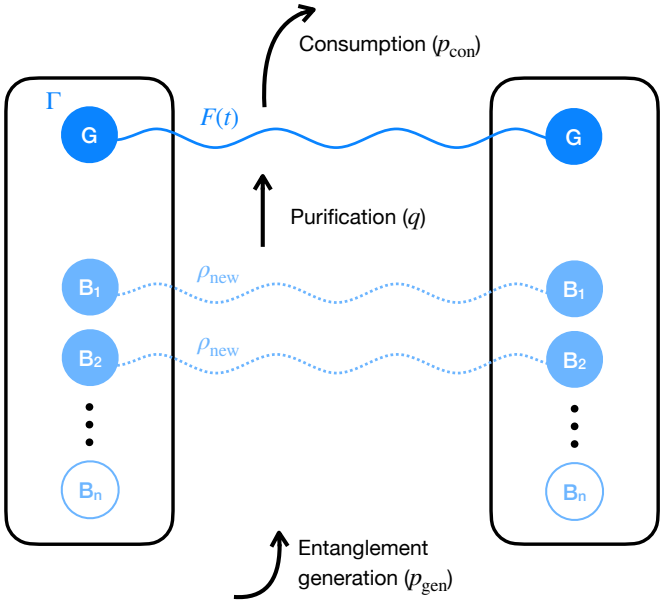


Figure 3.1: **Illustration of the 1GnB buffering system.** Entanglement generation is attempted in every bad memory (B_1, …, B_n) simultaneously in each time slot. Each memory succeeds with probability p_{gen}. The good memory, G, stores entanglement, which decoheres at rate Γ. When G is full and new entanglement is generated in any of the B memories, a purification subroutine is applied with probability q. Entanglement is consumed from G with probability p_{con} in each time slot.

tanglement is generated in any of the bad memories. This holds even if the purification protocol has a large probability of failure. Nevertheless, there is a tradeoff between both performance metrics, since the availability decreases when purification is performed more frequently.

- FUNDAMENTAL BOUNDS — We provide upper and lower bounds for the availability and the average consumed fidelity of a 1GnB system, which constitute fundamental limits to the impact that a purification policy can have on the performance.

- SIMPLE CAN BE BETTER THAN OPTIMAL — Simple purification protocols can greatly outperform advanced purification protocols that maximise the fidelity of the output entangled state. For example, we find that a buffering system using the 2-to-1 purification protocol from ref. [50] (known as DEJMPS) can outperform a system using the n-to-1 optimal bilocal Clifford protocol from ref. [97], in terms of both availability and average consumed fidelity.

3.2. THE 1GnB SYSTEM

In this section, we provide a short description of the entanglement buffering setup (see Figure 3.1). The goal of the system is to buffer bipartite entanglement shared between two nodes. These nodes could be, for example, two end users in a quantum network or

two processors in a quantum computing cluster. We refer to bipartite entanglement as an *entangled link* between the two nodes. In the 1GnB system:

- Each node has *one long-lived memory* (good, G) and n *short-lived memories* (bad, B).
- The G memories are used to store the entangled link. We assume *the link stored in memory is a Werner state* (any bipartite state can be transformed into a Werner state with the same fidelity by applying extra noise, a process known as *twirling* [47, 96]). Such a state can be parametrised with its fidelity to the target maximally entangled state, F.
- The entangled link stored in G is subject to *depolarising noise* with memory lifetime $1/\Gamma$, which causes an exponential decay in fidelity with rate Γ. That is, if the link in memory has an initial fidelity F, after time t this reduces to

$$F \mapsto \left(F - \frac{1}{4}\right) e^{-\Gamma t} + \frac{1}{4}. \tag{3.1}$$

- Before each entanglement generation attempt, the system checks if a new *consumption request* has arrived. The arrival of a new consumption request in each time step occurs with probability p_{con}. If there is a link stored in memory G when a consumption request arrives, the link is immediately consumed and therefore removed from the memory. This takes up the entire time step. If there is no link available, the request is discarded and the system proceeds with the entanglement generation attempt.
- The B memories are used to generate new entangled links. In the literature, these are usually called communication or broker qubits [92]. This communication qubit can be, for example, the electron spin in a nitrogen-vacancy center [94, 126, 93]. Every time step that is not taken up by consumption, *entanglement generation* is attempted in all n bad memories simultaneously, e.g. using frequency or spatial multiplexing, and each of them independently generates an entangled link with probability p_{gen}. This means that, after each multiplexed attempt, the number of successfully generated links follows a binomial distribution with parameters (n, p_{gen}). Each of these new links is of the form ρ_{new}, which is an arbitrary state that depends on the entanglement generation protocol employed (see e.g. refs. [25, 127, 124, 128]).
- When $k \geq 1$ entangled links are generated in the B memories and the G memory is empty, one of the links is *transferred* to the G memory. If the G memory is occupied, the new links may be used to *purify* the link in memory. The system decides to attempt purification with probability q. If the system does not decide to purify, the new links are discarded. If the system decides to attempt purification and this succeeds, then the resultant link in the G memory is twirled, converting it into the form of a Werner state with the same fidelity.

Table 3.1 summarises all variables of the system. Next, we discuss how to model the purification strategy.

3.2.1. PURIFICATION POLICY

The main degree of freedom in the 1GnB system is the choice of purification protocol. This is given by the purification policy.

Definition 3.1. The *purification policy* π is a function that indicates the purification protocol that must be used when k links are generated in the B memories,

$$\pi : k \in \{1, \ldots, n\} \mapsto \pi(k) \in \mathscr{P}_{k+1}, \tag{3.2}$$

where $\mathscr{P}_m$ is the set of all m-to-1 purification protocols.

Protocol $\pi(k)$ of purification policy π is the $(k+1)$-to-1 purification protocol that is used when k new links are generated in the B memories (examples of basic protocols can be found in refs. [50, 49, 84]; see ref. [129] for a survey). The purification protocol updates the fidelity of the buffered link from F to $J_k(F)$, where

$$J_k(F) = \frac{1}{4} + \frac{a_k(\rho_{\text{new}})\left(F - \frac{1}{4}\right) + b_k(\rho_{\text{new}})}{c_k(\rho_{\text{new}})\left(F - \frac{1}{4}\right) + d_k(\rho_{\text{new}})}. \tag{3.3}$$

We call J_k the *jump function of protocol* $\pi(k)$. The protocol succeeds with probability

$$p_k(F) = c_k(\rho_{\text{new}})\left(F - \frac{1}{4}\right) + d_k(\rho_{\text{new}}), \tag{3.4}$$

otherwise all of the links (including the buffered one) are discarded and the G memory becomes empty. In Appendix B of ref. [56], the forms (3.3) and (3.4) for the output fidelity and success probability are justified, given that the buffered link is a Werner state with fidelity F and any other input state is given by the same arbitrary density matrix ρ_{new}. We therefore see that the action of any purification protocol on the fidelity of the buffered link is determined by the four parameters $a_k(\rho_{\text{new}})$, $b_k(\rho_{\text{new}})$, $c_k(\rho_{\text{new}})$, $d_k(\rho_{\text{new}})$. In Appendix 3.6.3, we discuss the values that these coefficients can take. As an example, we also provide the explicit form of these coefficients for the well-known 2-to-1 DEJMPS protocol[50].

Lastly, note that purification policy π employs protocol $\pi(k)$ when k new links are generated. However, this does not mean that all the new links are used in the protocol. For example, a policy may simply replace the link in memory with a newly generated link and ignore the rest of the new links.

3.2.2. FIDELITY OF THE BUFFERED ENTANGLEMENT

Given the system description, we now view 1GnB as a discrete-time stochastic process. In particular, at time t the state of the system is the fidelity $F(t)$ of the buffered link, as this is the only quantity that can change over time. If there is no link in the buffered memory at time t, we let $F(t) = 0$. This is for notational convenience, as recalling the decoherence (3.1), one can never reach zero fidelity if there is a link present.

We now outline the characteristic behaviours of $F(t)$ when moving from time t to time $t+1$.

Let us consider first $F(t) = 0$. If entanglement generation is unsuccessful, in the next time step the fidelity will remain at that value: $F(t+1) = 0$. If entanglement generation

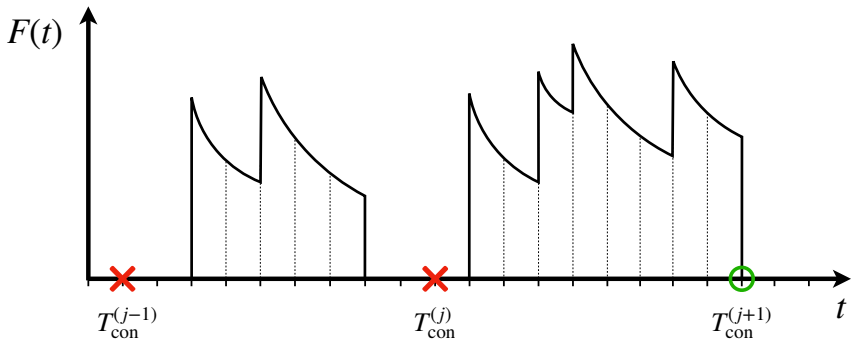


Figure 3.2: **Example dynamics of the 1GnB system.** Here, the fidelity $F(t)$ of the link in the G memory is plotted against time. The vertical lines represent discretisation of time. The jumps in fidelity occur when the link is purified successfully. In between purifications, the link is subject to decoherence and the fidelity decreases. The link in the G memory is removed due to either failed purification or consumption. When there is no link in memory, $F(t) = 0$. The j-th consumption request arrives at time $T_{\text{con}}^{(j)}$. The green tick (red crosses) represent when a consumption request is (is not) served.

is successful, in the next time step the fidelity will be F_{new}, where $F_{\text{new}} = \langle\Phi_{00}|\,\rho_{\text{new}}\,|\Phi_{00}\rangle$ is the fidelity of freshly generated links. We will assume that $F_{\text{new}} > 1/4$.

If $F(t) > 0$, then in the next time step this could evolve in one of the following ways: (i) if no purification is attempted then the fidelity simply decoheres by one unit of time according to (3.1); (ii) if k new links are generated and purification is successfully performed, the fidelity decoheres by one time step and is then mapped according to the corresponding jump function (3.3); (iii) if a consumption request has arrived or if purification fails, the link is removed and the system becomes empty.

In Figure 3.2, we illustrate an example of how the fidelity may evolve.

In the following subsection, we define the two performance metrics: the availability and the average consumed fidelity. We then present simple closed-form solutions for these two performance metrics in the 1GnB system.

3.2.3. BUFFERING PERFORMANCE

The first step towards the design of useful entanglement buffers is to determine a suitable way to measure performance. Here, we define two performance metrics for entanglement buffers – these quantities were proposed in ref. [56], where they were used to study the 1G1B system. Then, we provide exact, closed-form expressions for these two performance metrics in the 1GnB system.

Our first metric is the *availability*. A user is able to consume entanglement only when there is a link available in memory G at the time of requesting the entanglement. Therefore, an important performance measure is the probability that entanglement is available when a consumption request arrives.

Definition 3.2 (Availability)**.** The availability A is the probability that there is an entan-

Table 3.1: Parameters of the 1G1B system. See main text for further details.

Hardware	
n	Number of short-lived memories
p_{gen}	Probability of successful entanglement generation attempt
ρ_{new}	Bipartite entangled state produced after a successful entanglement generation
Γ	Rate of decoherence
Application	
p_{con}	Probability of consumption request
Purification policy	
q	Probability of attempting purification immediately after a successful entanglement generation attempt (otherwise the new links are discarded)
$J_k(F)$	Jump function. Given a buffered link with fidelity F, $J_k(F)$ is the fidelity immediately following a successful purification using k newly generated links. Rational function with coefficients a_k, b_k, c_k, d_k – see (3.3).
$p_k(F)$	Probability of successful purification using k newly generated links. Linear function with coefficients c_k, d_k – see (3.4).

gled link present in memory G when a consumption request arrives. This is defined as

$$A = \lim_{m\to\infty} \frac{1}{m} \sum_{j=1}^{m} \mathbb{1}_{\text{link exists}}(T_{\text{con}}^{(j)}), \tag{3.5}$$

where $T_{\text{con}}^{(j)}$ is the arrival time of the j-th consumption request, and $\mathbb{1}_{\text{link exists}}(t)$ is and indicator function that takes the values one if there is a link stored in memory G at time t, and zero otherwise.

The availability may be seen as a rate metric: it determines the rate at which entanglement can be consumed. The second performance metric is the *average consumed fidelity*, which captures the average quality of consumed entanglement.

Definition 3.3 (Average consumed fidelity)**.** The average consumed fidelity is the average fidelity of the entangled link upon consumption, conditional on a link being present. More specifically,

$$\overline{F} = \lim_{m\to\infty} \frac{\sum_{j=1}^{m} \mathbb{1}_{\text{link exists}}\left(T_{\text{con}}^{(j)}\right) \cdot F^{-}(T_{\text{con}}^{(j)})}{\sum_{j=1}^{m} \mathbb{1}_{\text{link exists}}(T_{\text{con}}^{(j)})}, \tag{3.6}$$

where

$$F^{-}(t) = \begin{cases} e^{-\Gamma}\left(F(t-1) - \frac{1}{4}\right) + \frac{1}{4}, \text{ if } F(t-1) > 0, \\ 0, \text{ if } F(t-1) = 0. \end{cases} \tag{3.7}$$

is the fidelity of the link stored in memory G at the end of the previous timestep at time $t-1$ (and therefore consumed at time t), and $T_{\text{con}}^{(j)}$ is the arrival time of the j-th consumption request.

The indicator function in the numerator of (3.6) is included for clarity, but is not necessary: if there is no link in memory at time t, then $F(t) = 0$ by definition.

We note that the Definitions 3.2 and 3.3 are presented differently to how they were in ref. [56]. This is because the new definitions have a clearer operational meaning, as they are from the viewpoint of the consumer. However, in Appendix 3.6.1 we show that these metrics are equivalent for the 1GnB system.

As our first main result, we derive analytical solutions for the availability and the average consumed fidelity in the 1GnB system

Theorem 3.1 (Formula for the availability)**.** *The availability of the 1GnB system is given by*

$$A = \frac{\mathbb{E}[T_{\text{occ}}]}{\mathbb{E}[T_{\text{gen}}] + \mathbb{E}[T_{\text{occ}}]} \ a.s. \tag{3.8}$$

where T_{gen} is the time to generate new entangled links and T_{occ} is the time from when the G memory becomes occupied until it is emptied due to consumption or to failed purification. The expected values are given by

$$\mathbb{E}[T_{\text{gen}}] = \frac{1}{1-\left(1-p_{\text{gen}}\right)^n} \tag{3.9}$$

and

$$\mathbb{E}[T_{\text{occ}}] = \frac{1-\tilde{A}+\tilde{C}(F_{\text{new}}-\frac{1}{4})}{\left[(1-\tilde{A})(1-\tilde{D})-\tilde{B}\tilde{C}\right]\tilde{P}}, \tag{3.10}$$

with

$$\begin{aligned}
\tilde{P} &:= p_{\text{con}} + q\left(1-\left(1-p_{\text{gen}}\right)^n\right)(1-p_{\text{con}}),\\
\tilde{A} &:= \frac{q(1-p_{\text{con}})\tilde{a}}{e^{\Gamma}-\left(1-q+q\left(1-p_{\text{gen}}\right)^n\right)(1-p_{\text{con}})},\\
\tilde{B} &:= \frac{q(1-p_{\text{con}})\tilde{b}}{p_{\text{con}}+q\left(1-\left(1-p_{\text{gen}}\right)^n\right)(1-p_{\text{con}})},\\
\tilde{C} &:= \frac{q(1-p_{\text{con}})\tilde{c}}{e^{\Gamma}-\left(1-q+q\left(1-p_{\text{gen}}\right)^n\right)(1-p_{\text{con}})},\\
\tilde{D} &:= \frac{q(1-p_{\text{con}})\tilde{d}}{p_{\text{con}}+q\left(1-\left(1-p_{\text{gen}}\right)^n\right)(1-p_{\text{con}})},
\end{aligned}$$

and

$$\tilde{a} := \sum_{k=1}^{n} a_k \cdot \binom{n}{k} (1-p_{\text{gen}})^{n-k} p_{\text{gen}}^k,$$
$$\tilde{b} := \sum_{k=1}^{n} b_k \cdot \binom{n}{k} (1-p_{\text{gen}})^{n-k} p_{\text{gen}}^k,$$
$$\tilde{c} := \sum_{k=1}^{n} c_k \cdot \binom{n}{k} (1-p_{\text{gen}})^{n-k} p_{\text{gen}}^k,$$
$$\tilde{d} := \sum_{k=1}^{n} d_k \cdot \binom{n}{k} (1-p_{\text{gen}})^{n-k} p_{\text{gen}}^k.$$

Proof. See Appendix 3.6.2. □

From Theorem 3.1, we see that the availability depends on all the parameters of the system (listed in Table 3.1), including the noise level Γ. The latter may come as a surprise, since one would expect noise to have an impact on the average consumed fidelity but maybe not on the availability, which is only affected by processes that fill or deplete the G memory. These processes are entanglement generation, failed purification, and consumption. In our model, the probability of failed purification depends via (3.4) on the fidelity of the buffered link, which is in turn affected by the level of noise. As a consequence, noise has an indirect effect on the availability.

Theorem 3.2 (Formula for the average consumed fidelity)**.** *The average consumed fidelity of the 1GnB system is given by*

$$\overline{F} = \frac{\tilde{w} F_{\text{new}} + \tilde{x}}{\tilde{y} F_{\text{new}} + \tilde{z}} \quad a.s. \tag{3.11}$$

with

$$\tilde{w} := p_{\text{con}} + q(1-p_{\text{con}}) \left(p_{\text{gen}}^* + \frac{1}{4}\tilde{c} - \tilde{d} \right),$$
$$\tilde{x} := \frac{1}{4} \left[e^{\Gamma} - 1 + q(1-p_{\text{con}}) \left(-\tilde{a} + 4\tilde{b} - \frac{1}{4}\tilde{c} + \tilde{d} \right) \right],$$
$$\tilde{y} := q(1-p_{\text{con}})\tilde{c},$$
$$\tilde{z} := e^{\Gamma} - 1 + p_{\text{con}} + q(1-p_{\text{con}}) \left(p_{\text{gen}}^* - \tilde{a} - \frac{1}{4}\tilde{c} \right),$$

where $p_{\text{gen}}^* = 1-(1-p_{\text{gen}})^n$*, and* $\tilde{a}$*,* $\tilde{b}$*,* $\tilde{c}$*, and* $\tilde{d}$ *are given in Theorem 3.1.*

Proof. See Appendix 3.6.2. □

We note that both A and $\overline{F}$ have been defined as random variables in Definitions 3.2 and 3.3. However, as shown in Theorems 3.1 and 3.2, these quantities are almost surely deterministic functions of the system parameters. For clarity and convenience, we will adopt a slight abuse of notation and treat A and $\overline{F}$ as deterministic functions. This convention will be maintained throughout the remainder of the text.

3.3. BUFFERING SYSTEM DESIGN

In this section, we discuss our main findings after analysing the performance of the 1GnB system. In Subsection 3.3.1, we study the impact of a general purification protocol on the system performance. In particular, it is shown that the availability and the average consumed fidelity are monotonic in the parameter q that determines how frequently the system attempts purification. In the remaining subsections, we investigate how the choice of purification policy impacts the performance of the buffering system, and we provide heuristic rules for the design of a good purification policy.

3.3.1. MONOTONIC PERFORMANCE

Each time a B memory successfully generates entanglement, there is the opportunity to purify the buffered link. This is controlled by the parameter q, which is the probability that, after some fresh links are successfully generated, they are used to attempt purification (otherwise they are discarded). If purification is never attempted ($q = 0$), the fidelity of the buffered link will never be increased, although the buffered link will never be lost to failed purification. If purification is always attempted ($q = 1$), the availability and average consumed fidelity might be affected as follows:

- Purifying more often means risking the loss of buffered entanglement more frequently, since purification can fail. This suggests availability may be decreasing in q. However, many purification protocols have a probability of success that is increasing in the fidelity of the buffered link, F. This means that, when purification is applied more frequently to maintain a high-fidelity link, subsequent purification attempts are more likely to succeed. Consequently, it is not clear that the availability is decreasing in q.

- The fidelity of the buffered link increases after applying several purification rounds. However, if purification is applied too greedily, we may lose a high-quality link and we would have to restart the system with a lower-quality link. If a consumption request then arrives, it would only be able to consume low-quality entanglement. Hence, it is not clear that the average consumed fidelity is increasing in q.

In the following, we address the previous discussion and show that, if purification is always attempted ($q = 1$), the availability is actually minimised, while the average consumed fidelity is maximised. More generally, we show that A and $\overline{F}$ are both monotonic in q, given some reasonable conditions on the jump functions J_k. The following results (Propositions 3.1 and 3.2) may be used to answer an important question about the 1GnB system: *how frequently should we purify the buffered state in order to maximise A (or $\overline{F}$)?* That is, *what value of q optimises our performance metrics?*

Proposition 3.1. *The availability is a non-increasing function of q, i.e.*

$$\frac{\partial A}{\partial q} \leq 0. \tag{3.12}$$

Proof. See Appendix 3.6.4. □

As previously explained, the monotonicity of the availability in q is not a trivial result, and it has fundamental implications. It allows us to derive upper and lower bounds that apply to 1GnB systems using *any* purification policy.

Corollary 3.1. *The availability is bounded as*

$$\frac{p^*_{\text{gen}} \cdot (\gamma + p_{\text{con}})}{\xi + \xi' \cdot p^*_{\text{gen}} + \xi'' \cdot (p^*_{\text{gen}})^2} \leq A \leq \frac{p^*_{\text{gen}}}{p^*_{\text{gen}} + p_{\text{con}}}, \tag{3.13}$$

with $p^*_{\text{gen}} := 1 - (1 - p_{\text{gen}})^n$, $\gamma := e^{\Gamma} - 1$, $\xi := \gamma p_{\text{con}} + p^2_{\text{con}}$, $\xi' := 1 + 2\gamma + (2 - \gamma) p_{\text{con}} - 2p^2_{\text{con}}$, *and* $\xi'' := 2(1 - p_{\text{con}})^2$. *Moreover, the upper bound is tight, and for any purification policy is achieved when* $q = 0$.

Proof. See Appendix 3.6.4. □

We refer to p^*_{gen} as the *effective generation probability*, since it is the probability that at least one new link is generated in a single (multiplexed) attempt.

The upper bound from (3.13) only depends on the effective generation probability and the probability of consumption. This bound is achievable with any purification policy: to maximise the availability, it suffices to never purify ($q = 0$). A special case are deterministic policies (those with $p_k(F) = 1$, $\forall k$), which achieve this bound for any q. This upper bound coincides with the tight upper bound found in previous work for a 1G1B system [56]. Note that the 1G1B analysis from ref. [56] was done in continuous time, where rates were used instead of probabilities. In this framework, the maximum availability was $\lambda/(\lambda + \mu)$, where λ was the (non-multiplexed) entanglement generation rate and μ was the consumption rate.

Unlike the upper bound, we note that the lower bound from (3.13) is not yet shown to be tight. We believe that the availability at $q = 1$ of a policy that always fails purification ($c_k = d_k = 0$, $\forall k$) constitutes a tight lower bound for any other purification policy. We leave this proof as future work.

Figure 3.3 shows the upper and lower bounds for the availability from (3.13) versus p^*_{gen} for two different noise levels. As discussed, only the lower bound is affected by noise. In particular, we have observed that the gap between the bounds is reduced when the noise level increases. Another remarkable feature is that, when p^*_{gen} approaches zero, both upper and lower bounds are equal to $p^*_{\text{gen}}/p_{\text{con}}$ to first order in p_{gen}. Hence, in the limit of small effective generation probabilities, the availability also satisfies

$$A \approx \frac{p^*_{\text{gen}}}{p_{\text{con}}}. \tag{3.14}$$

Proposition 3.2. *The average consumed fidelity is a non-decreasing function of* q, *i.e.,*

$$\frac{\partial \overline{F}}{\partial q} \geq 0, \tag{3.15}$$

if $J_k(F_{\text{new}}) \geq F_{\text{new}}$, $\forall k \in \mathbb{N}$.

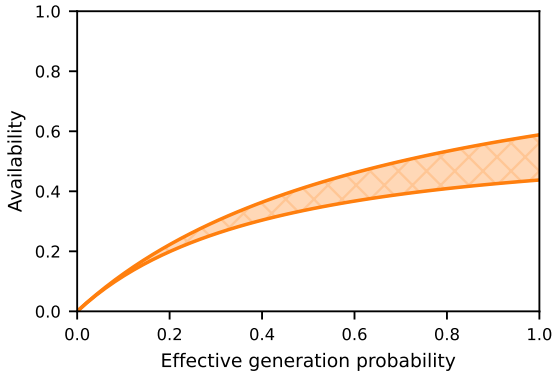


Figure 3.3: **The upper bound on the availability is tight and it converges to the lower bound in the limit of small generation probabilities.** Upper and lower bounds on the availability from (3.13), versus the effective generation probability $p^*_{\text{gen}} = 1-(1-p_{\text{gen}})^n$. The availability can only take values within the shaded region. In this example we use $\Gamma = 1$ and $p_{\text{con}} = 0.7$.

Proof. See Appendix 3.6.5. □

As previously explained, the monotonicity of $\overline{F}$ in q is not a trivial result. In fact, this behaviour is only certain for purification policies composed of protocols that can increase the fidelity of a newly generated link. That is, when k new links are generated, the protocol applied satisfies $J_k(F_{\text{new}}) \geq F_{\text{new}}$. This is a reasonable condition: otherwise, we would be applying purification protocols that decrease the fidelity of new links.

Proposition 3.2 also allows us to derive useful upper and lower bounds for $\overline{F}$ that apply to 1GnB systems using any purification policy.

Corollary 3.2. *The average consumed fidelity is bounded as*

$$\frac{\gamma + 4F_{\text{new}}p_{\text{con}}}{4\gamma + 4p_{\text{con}}} \leq \overline{F} \leq \frac{\gamma + 4F_{\text{new}}p_{\text{con}} + 3(1-p_{\text{con}})p^*_{\text{gen}}}{4\gamma + 4p_{\text{con}}}, \tag{3.16}$$

with $\gamma := e^{\Gamma} - 1$. *Moreover, the lower bound is tight, and for any purification policy is achieved when* $q = 0$.

Proof. See Appendix 3.6.5. □

We see that the tight lower bound from (3.16) does not depend on the number of memories n, the probability of successful entanglement generation p_{gen}, or the purification policy. This is because this bound corresponds to $q = 0$. In such a case, no purification is applied, and the consumed fidelity only depends on the initial fidelity (F_{new}) and the amount of decoherence experienced until consumption (given by Γ and p_{con}).

The bounds on $\overline{F}$ can be used to determine if the parameters of the system need an improvement to meet specific quality-of-service requirements. For example, let us consider Figure 3.4, which shows the bounds for $p_{\text{con}} = 0.7$ and two different values of Γ. If noise is strong ($\Gamma = 1$ in this example), we observe that values of p^*_{gen} below 0.5 yield

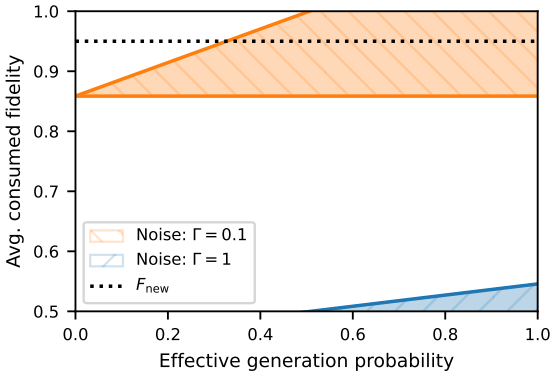


Figure 3.4: **The upper bound on the average consumed fidelity marks unachievable values for any purification policy.** Upper and lower bounds on the average consumed fidelity $\overline{F}$ from (3.16), versus the effective generation probability $p^*_{\text{gen}} = 1-(1-p_{\text{gen}})^n$. $\overline{F}$ can only take values within the shaded region. In this example we use $p_{\text{con}} = 0.7$.

$\overline{F} < 1/2$, which means that, on average, the consumed link will not be entangled [56]. Hence, if the consumption request rate is $p_{\text{con}} = 0.7$, we need to increase p^*_{gen} beyond 0.5 (by increasing the number of B memories, n, or the probability of successful entanglement generation, p_{gen}) or to decrease the noise experienced in memory G in order to provide a useful average state. When the noise level is $\Gamma = 0.1$, Figure 3.4 shows that $\overline{F} > 0.85$. Moreover, for $p^*_{\text{gen}} > 0.3$, the upper bound is above F_{new}, which means that a smart choice of purification policy may allow us to buffer entanglement with $\overline{F} > F_{\text{new}}$. Ultimately, this means that, in this regime, an entanglement buffer with faulty memories may be able to keep entanglement at higher fidelities than a perfect memory.

3.4. CHOOSING A PURIFICATION POLICY

In previous studies of entanglement buffering, the choice of purification policy was restricted by the properties of the system. For example, in ref. [56] the 1G1B system was studied, where only 2-to-1 purification protocols can be implemented, and the jump function was assumed to be linear in the fidelity of the buffered link. Other works include simplifying assumptions (e.g. in ref. [65], a buffer is studied that employs the purification protocol proposed in ref. [130]). The 1GnB buffering system offers more freedom in the choice of purification protocols. In a 1GnB buffer, each entanglement generation attempt is multiplexed and can generate up to n new links at a time. When $k \leq n$ new links are produced, any $(k+1)$-to-1 purification protocol can in principle be implemented. This provides an extra knob that can be used to tune the performance of the system to the desired values. In this section, we investigate the impact that specific purification policies have on the system and we provide guidelines on how to choose a suitable purification policy. Note that an exhaustive optimisation problem would be extremely computationally expensive to solve due to the large space of purification policies – optimising over a_k, b_k, c_k, d_k is not easy, since it is not certain that every combination

of those parameters corresponds to an implementable purification circuit.

3.4.1. SIMPLE POLICIES: IDENTITY, REPLACEMENT, AND CONCATENATION

There are two trivial deterministic policies ($p_k = 1, \forall k$) that we will use as a baseline:

- In the *identity policy*, the system does not perform any operation on the buffered link, which yields an output fidelity $J_k(F) = F$, $\forall k > 0$. This is equivalent to setting $q = 0$. As discussed in Section 3.3.1, the identity policy therefore maximises the availability and minimises the average consumed fidelity.
- In the *replacement policy*, the system replaces the buffered entangled link by a new link, yielding an output fidelity $J_k(F) = F_{\text{new}}$, $\forall k > 0$. This corresponds to $a_k = 0$, $b_k = F_{\text{new}} - 1/4$, $c_k = 0$, and $d_k = 1$. Since this policy is deterministic, from the discussion in Section 3.3.1 we find that the replacement policy also provides maximum availability for any value of q. Since $\overline{F}$ is maximised for $q = 1$ (Proposition 3.2), we will only consider a replacement policy that always chooses to replace the link in memory when a new link is generated. That is, the replacement policy implicitly assumes $q = 1$.

Another simple strategy is the *DEJMPS policy*. This policy consists in applying the well-known 2-to-1 DEJMPS purification protocol [50] using the buffered link and a newly generated link as inputs. If more than one link is successfully generated, we use only one of them and discard the rest. We provide the purification coefficients a_k, b_k, c_k, and d_k for this policy in Appendix 3.6.3. One of the main drawbacks of the DEJMPS policy is that it does not take full advantage of the multiplexed entanglement generation, as it only uses one of the newly generated links and discards the rest. A technique that could improve the performance of the policy is *concatenation*, which consists in applying DEJMPS to all links (the buffered one and the newly generated ones) consecutively until only one link remains, which will be stored in memory G. Note that the concatenation of DEJMPS subroutines can be applied using different orders of the links (see Figure 3.5). The order determines the output fidelity and probability of success [131], which affects the performance of the buffering system. In what follows, we consider the *concatenated DEJMPS* policy, where DEJMPS is applied sequentially to all the newly generated links and the buffered link is used in the last application of DEJMPS, as in Figure 3.5a. In our analysis, we found that different orderings provided qualitatively similar behaviour of our two performance metrics (see Appendix 3.6.7 for further details).

Figure 3.6 shows the performance of several policies: identity, replacement, DEJMPS, and concatenated DEJMPS $\times N$. The latter is a policy that applies DEJMPS sequentially up to N times and discards any extra links: if $k \leq N$ links are generated then k concatenations are performed, and if $k > N$ links are generated, N concatenations are performed. We note that concatenated DEJMPS $\times 1$ is just the same as the DEJMPS policy. DEJMPS and concatenated DEJMPS are plotted for $q \in [0, 1]$. The maximum average consumed fidelity is indicated with a dot, and it is achieved when $q = 1$. The first observation from this figure is that a higher level of concatenation decreases the availability. This is because it requires multiple DEJMPS subroutines to succeed, which decreases the overall probability of successful purification. However, a higher level of concatenation can significantly increase the average consumed fidelity $\overline{F}$. For example, the maximum $\overline{F}$ that

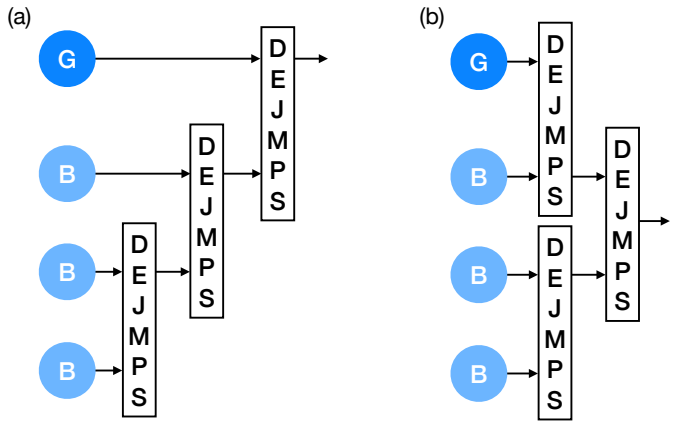


Figure 3.5: **The ordering in a concatenated policy matters.** Example of two different orderings when the buffered link (G) and three newly generated links (B) are used. We call ordering (a) "concatenated DEJMPS". Ordering (b) is often called "nested" [30].

DEJMPS can achieve is 0.915, while concatenated DEJMPS ×2 leads to $\overline{F} = 0.937$ (for $q = 1$). Nevertheless, for the parameter values explored, we also find that increasing the number of concatenations beyond two often reduces both A and $\overline{F}$. This behaviour is shown more explicitly in Figure 3.7, where we plot the maximum $\overline{F}$ versus the maximum number of concatenations N. In this example, the number of B memories is $n = 10$, and therefore it is only possible to perform up to 10 concatenated applications of DEJMPS. We observe that $\overline{F}$ is maximised for two concatenations. The same was observed for different parameter values – in some edge cases, $\overline{F}$ increases with more concatenations, although the increase is marginal (see Appendix 3.6.7 for further details). In conclusion, this result shows that even if many new links are successfully generated in parallel, it can sometimes be beneficial to use only one or two of them for purification while discarding the rest.

3.4.2. SIMPLE POLICIES CAN OUTPERFORM COMPLEX POLICIES

In the previous section, we found that implementing a simple 2-to-1 protocol, even when multiple links are generated in the B memories, can provide a better performance than using all of the newly generated links for purification with concatenated 2-to-1 protocols. A follow-up question arises: *what if we employ more sophisticated* $(k+1)$*-to-1 protocols instead of simply concatenating 2-to-1 protocols? Can we then improve the performance of the buffer?* This is the question that we explore now.

Much recent work has focused on the search for optimal purification protocols [53, 52, 97], where optimal protocols are typically defined as those which maximise the output fidelity, or in some cases the success probability. Here, we evaluate the performance of a 1GnB system with some of these protocols, and we find a surprising result: simple protocols like DEJMPS can vastly outperform these more complex protocols in terms of buffering performance. In particular, we consider the bilocal Clifford protocols that maximise the output fidelity, given in ref. [97]. We refer to this policy as the *optimal bilo-*

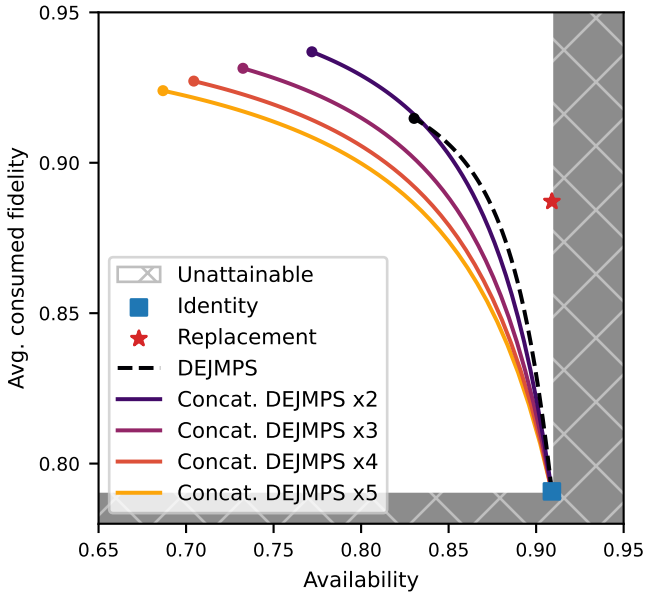


Figure 3.6: **Concatenating simple purification policies decreases A but may increase $\overline{F}$.** Performance of 1GnB systems with different purification policies, in terms of availability A and average consumed fidelity $\overline{F}$. The shaded area corresponds to unattainable values of A and $\overline{F}$ (see (3.13) and (3.16)). Lines and markers show the combinations of A and $\overline{F}$ achievable by different purification policies: identity (square marker), replacement (star marker), DEJMPS (dashed line), and concatenated DEJMPS (solid lines). Concatenation can boost $\overline{F}$ (e.g. the maximum $\overline{F}$ of twice-concatenated DEJMPS is larger than DEJMPS), but excessive concatenation may eventually lead to a drop in $\overline{F}$. Parameter values used in this example: $n = 10$, $p_{\text{gen}} = 0.5$, ρ_{new} is a Werner state with $F_{\text{new}} = 0.9$, $p_{\text{con}} = 0.1$, and $\Gamma = 0.02$.

cal Clifford (optimal-bC) policy. In Appendix 3.6.3, we discuss the details of this policy and provide its purification coefficients a_k, b_k, c_k, and d_k.

Figure 3.8 shows the performance of the optimal-bC policy in comparison to DEJMPS and twice-concatenated DEJMPS. The optimal-bC policy provides a significantly lower availability, A, without providing any advantage in average consumed fidelity, $\overline{F}$. In other words, for any desired A, using DEJMPS or twice-concatenated DEJMPS always provides a larger $\overline{F}$ than the optimal-bC policy. If we want to increase A as much as possible, the replacement policy is better than any other, as discussed earlier. We say that the performance of DEJMPS, twice-concatenated DEJMPS and replacement forms the *Pareto frontier* [132], which informally is the set of best achievable values for A and $\overline{F}$ for this collection of protocols. We tested different parameter combinations and found that the Pareto frontier was often made of DEJMPS, concatenated DEJMPS and replacement. The reason for these simple policies to outperform the optimal-bC policy is that the optimal bilocal Clifford protocols maximise the output fidelity at the expense of a reduced probability of success. At some point, the sacrifice in the probability of success can outweigh the benefit of a larger output fidelity, thereby reducing the overall performance of the buffer in terms of both A and $\overline{F}$.

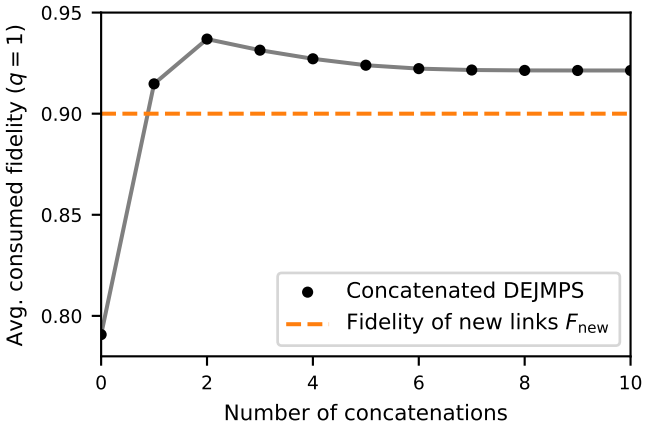


Figure 3.7: **Excessive concatenation worsens the performance.** Maximum average consumed fidelity $\overline{F}$ achieved by a purification policy that concatenates DEJMPS a limited number of times. Zero concatenations corresponds to an identity policy (no purification is performed). One concatenation corresponds to the DEJMPS policy. Excessive concatenation may decrease $\overline{F}$. Parameter values used in this example: $n = 10$, $p_{\text{gen}} = 0.5$, ρ_{new} is a Werner state with $F_{\text{new}} = 0.9$, $p_{\text{con}} = 0.1$, and $\Gamma = 0.02$.

Our comparison between simple and optimal purification protocols is by no means an exhaustive study. However, it shows that purification protocols that maximise only the output fidelity (or probability of success) must not be blindly used in more complex systems involving many impacting factors such as decoherence and consumption, such as entanglement buffers. In fact, we find that discarding some of the newly generated links and applying a 2-to-1 protocol can provide larger A and $\overline{F}$ than using all of the links in a more sophisticated purification subroutine. Note that this does not mean that multiplexed entanglement generation is not useful: even if we only employ 2-to-1 protocols, multiplexing boosts the effective entanglement generation rate, which allows for a more frequent purification of the buffered link.

Additionally, we also tested other complex policies that use (suboptimal) k-to-1 protocols, such as the *513 EC policy*, which uses a 5-to-1 protocol based on a $[[5,1,3]]$ quantum error correcting code. In Appendix 3.6.6, we explain this policy in detail and show that it can outperform DEJMPS and twice-concatenated DEJMPS in some parameter regions.

3.4.3. FLAGS CAN IMPROVE PERFORMANCE

As discussed in the previous sections, concatenating protocols multiple times does not necessarily improve the performance of the buffer (neither in terms of A nor $\overline{F}$). The reason is that, when concatenating, a single failure in one of the purification subroutines (in our examples, DEJMPS) leads to failure of the whole concatenated protocol. This can be easily solved: instead of considering the concatenated protocol as a black box that only succeeds when all subroutines succeed, *what if we condition the execution of each subroutine on the success/failure of previous subroutines?* Consider for example the con-

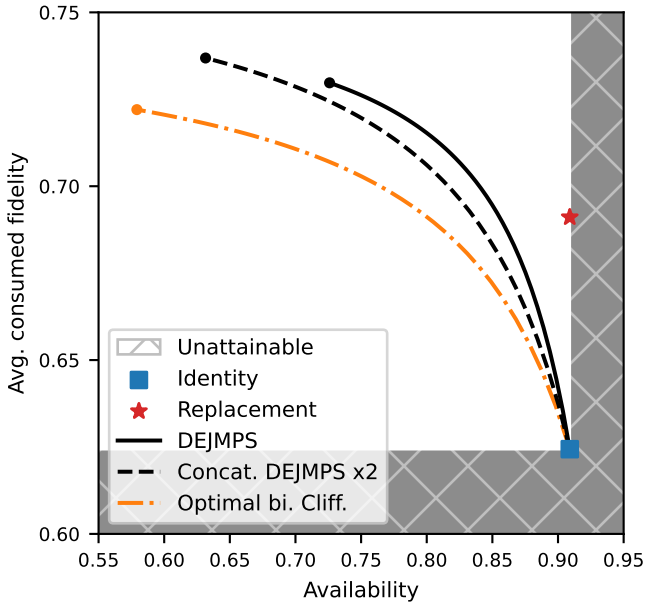


Figure 3.8: **Simple policies perform better despite discarding freshly generated entanglement.** Performance of 1GnB systems with different purification policies, in terms of availability A and average consumed fidelity $\overline{F}$. The shaded area corresponds to unattainable values of A and $\overline{F}$ (see (3.13) and (3.16)). Lines and markers show the combinations of A and $\overline{F}$ achievable by different purification policies: identity (square marker), replacement (star marker), DEJMPS (solid line), twice-concatenated DEJMPS (dashed line), and optimal-bC (dotted line). Parameter values used in this example: $n = 5$, $p_{\text{gen}} = 0.8$, ρ_{new} is a Werner state with $F_{\text{new}} = 0.7$, $p_{\text{con}} = 0.1$, and $\Gamma = 0.02$.

catenated protocol from Figure 3.5a. If any of the DEJMPS subroutines fails, the whole protocol fails and the buffered link has to be discarded. However, we can fix this by raising a failure flag whenever any of the first two subroutines fails. If this flag is raised, the third subroutine is not executed and we leave the buffered link untouched. The flagged version of a concatenated protocol has a larger probability of success, but can also have a lower output fidelity. This means that it is not clear a priori what is the impact of flags on the buffer performance. We now analyse a simple case in which we conclude that flags can be either beneficial or detrimental depending on the values of system parameters such as the level of noise Γ, and not only on the purification policy itself.

Let us consider a policy that operates as follows. For simplicity, we assume that newly generated states ρ_{new} are Werner states with fidelity F_{new}. When k new links are generated and there is already a link stored in memory G:

1. If $k = 1$, we apply the replacement protocol, which has coefficients $a_1 = 0$, $b_1 = F_{\text{new}} - 1/4$, $c_1 = 0$, and $d_1 = 1$.

2. If $k \geq 2$, we apply the DEJMPS protocol to two of the fresh links and discard the rest. Then, we replace the link in memory with the output from the DEJMPS subroutine,

without checking whether it was successful or not. This means that the output fidelity of the protocol is the same as the output fidelity from the DEJMPS subroutine. Since replacement is deterministic, the success probability of this protocol is also the same as the success probability of the DEJMPS subroutine. The purification coefficients for $k \geq 2$ are therefore given by $a_k = 0$, $b_k = a(\rho_{\text{new}}) \cdot (F_{\text{new}} - 1/4) + b(\rho_{\text{new}})$, $c_k = 0$, and $d_k = c(\rho_{\text{new}}) \cdot (F_{\text{new}} - 1/4) + d(\rho_{\text{new}})$, where a, b, c, and d are the coefficients of the DEJMPS protocol (given in Appendix 3.6.3).

Now, let us consider a flagged variant of the previous policy, with coefficients a'_k, b'_k, c'_k, and d'_k. It works as follows:

1. When $k = 1$, we apply the replacement protocol.

2. When $k \geq 2$ links are generated, the DEJMPS protocol is applied to two of the fresh links, and the rest are discarded. Then, the link in memory is replaced with the output from the DEJMPS subroutine, but only if the subroutine succeeds (otherwise, the buffered link is left untouched). This protocol is now fully deterministic, since the buffered link is never removed from memory. Consequently, $c'_k = 0$, and $d'_k = 1$. The output fidelity of this protocol can be computed as the weighted average of the original fidelity of the link in memory and the output fidelity of the DEJMPS subroutine – the first term must be weighted by the probability of failure of the subroutine, and the second term by the probability of success. Then, the remaining purification coefficients can be computed as $a'_k = 1 - c(\rho_{\text{new}}) \cdot (F_{\text{new}} - 1/4) - d(\rho_{\text{new}})$ and $b'_k = a(\rho_{\text{new}}) \cdot (F_{\text{new}} - 1/4) + b(\rho_{\text{new}})$, where a, b, c, and d are the coefficients of the DEJMPS protocol (given in Appendix 3.6.3).

By introducing the flags, we have created a protocol with probability of success $p'_k = 1 \geq p_k$, where p_k is the probability of success of the original protocol. However, it can be shown that the output fidelity of the flagged protocol is $J'_k(F) \leq J_k(F)$, where J_k is the jump function of the original protocol. This holds when DEJMPS can improve the fidelity of the newly generated links, i.e. when $J(F_{\text{new}}) \geq F_{\text{new}}$, where J is the jump function of DEJMPS. The opposite regime is not interesting, since DEJMPS is decreasing the fidelity of the links and we would be better off not purifying.

As shown in the previous example, internal flags increase the probability of success of purification protocols, which should boost the availability of the buffer. However, flags may have the side effect of reducing the output fidelity, and therefore it is not clear what is their impact on the average consumed fidelity. In Figure 3.9, we show the performance of a 1GnB system using the policy described above, versus the level of noise in memory G. We show A (orange lines) and $\overline{F}$ (black lines) for the original policy (solid lines) and the flagged policy (dashed lines). As expected, the availability is larger for the flagged policy. The behaviour of $\overline{F}$ is more interesting. When the level of noise is low, the flagged policy provides better performance, since it prevents high-quality entanglement from being lost to a failed purification. However, when noise is strong, flagging becomes detrimental in terms of $\overline{F}$: the buffer is likely to store low-quality entanglement due to the strong noise, and flags prevent the buffered link from being discarded earlier due to failed purification and being replaced by a fresh link. Note that other strategies,

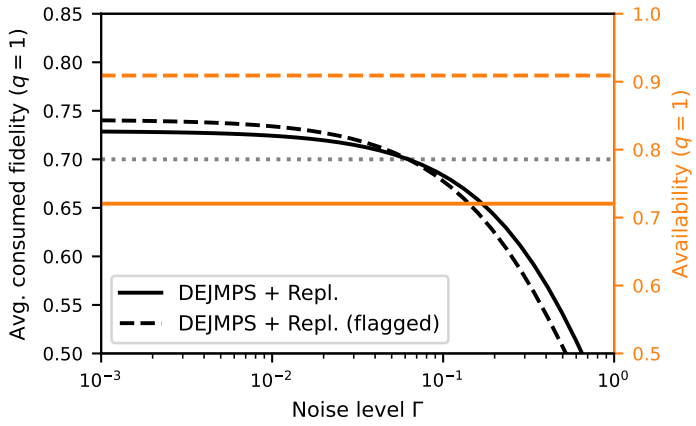


Figure 3.9: **Flagged protocols boost the availability but may decrease the average consumed fidelity.** Availability A and average consumed fidelity $\overline{F}$ versus the noise level Γ, for a 'DEJMPS + Replacement' policy and its flagged version. In the first policy, the buffered link is lost when a DEJMPS subroutine fails. The second policy incorporates a flag that prevents this from happening – it succeeds deterministically at the expense of a lower output fidelity. The flagged policy yields larger A, but may decrease $\overline{F}$ in some parameter regimes (e.g. when Γ is large). Parameter values used in this example: $n = 2$, $p_{\text{gen}} = 1$, ρ_{new} is a Werner state [135] with $F_{\text{new}} = 0.7$, and $p_{\text{con}} = 0.1$.

such as using the output state regardless of the success or failure flag [133] and using hyperentangled states [134], can also be employed for designing deterministic purification protocols.

In conclusion, internal flags are a solid tool to improve the availability of entanglement buffers based on concatenated purification protocols. However, they can decrease the average consumed fidelity in some parameter regimes. Hence, flagged purification policies should not be assumed to be better than their non-flagged counterparts, and their performance should be carefully evaluated before being adopted.

3.5. OUTLOOK

In this chapter, we have studied the behaviour of entanglement buffers with one long-lived memory and n short-lived memories (1GnB system). In particular, we have provided analytical expressions for the two main performance measures: the availability and the average consumed fidelity. These expressions provide valuable insights, such as the fundamental limits to the performance of 1GnB systems discussed earlier.

Since our analytical solutions are not computationally expensive to evaluate, we expect our buffering setup to be easy to incorporate in more complex network architectures, such as quantum repeater chains or even large-scale quantum networks. Additionally, larger buffering systems with multiple long-lived memories, e.g. an mGnB setup, can be implemented with multiple 1GnB systems in parallel.

Due to the vast freedom in the choice of purification policy, there are multiple ways in which our analysis of purification strategies for entanglement buffers can be extended.

Notably, determining the optimal ordering in which simple protocols should be applied to newly generated links (e.g. concatenated, nested [30], or banded [131]) is left as future work. Additionally, finding policies that optimise availability or average consumed fidelity remains an important open question.

3.6. APPENDIX

3.6.1. A NOTE ON THE VIEWPOINT

In this appendix, we provide three further ways to compute the performance metrics A and $\overline{F}$. The initial (and most natural) definitions of the performance metrics (see Definitions 3.2 and 3.3) consist in averages from the viewpoint of the network user, who consumed the links. In Lemma 3.1, we show that the averaging may only be done over a single cycle of the renewal process. In Lemma 3.3, we show that the performance metrics can also be computed as limiting values when time goes to infinity. Lastly, in Lemma 3.2, it is shown that one may compute the metrics by averaging over time, regardless of consumption arrival times.

We denote the arrival time of the j-th consumption request as $T_{\text{con}}^{(j)}$. From now on, we write

$$\mathbb{1}_{\text{l.e.}}(F) \equiv \mathbb{1}_{\text{link exists}}(F) = \begin{cases} 1 \text{ if } F > 0, \\ 0 \text{ if } F = 0. \end{cases} \tag{3.17}$$

In the following, we let $\mathbb{N}_0$ denote the natural numbers containing zero, and $\mathbb{N} = \mathbb{N}_0 \setminus \{0\}$. Recall that $F = \{F(t), t \in \mathbb{N}_0\}$ is a discrete-time stochastic process. The value $F(t)$ is defined to be the fidelity at the *beginning* of the time step $[t, t+1)$. Then, since consumption removes the link from the G memory, at each consumption time we have $F(T_{\text{con}}^{(j)}) = 0$. However, the consumed fidelity at this time depends on the value of the fidelity at time $T_{\text{con}}^{(j)} - 1$. We therefore introduce some new notation to more easily treat this issue.

In order to do this, we firstly note that associated with F there is an equivalent continuous-time stochastic process $\{F_{\text{cont}}(s) : s \geq 0\}$ that is obtained from F with the following procedure: given $t \in \mathbb{N}$,

(i) if $F(t) > 0$, then for $s \in [t, t+1)$, $F_{\text{cont}}(s)$ may be deduced by applying decoherence (3.1) to $F(t)$;

(ii) if $F(t) = 0$, then $F_{\text{cont}}(s) = 0$ for $s \in [t, t+1)$.

Conversely, F may be obtained from F_{cont} by taking its values at integer times.

From F_{cont}, for $t \in \mathbb{N}$, we define another discrete time process F^-,

$$F^- := \{F_{\text{cont}}(t^-) : t \in \mathbb{N}\}, \tag{3.18}$$

where t^- denotes taking the left-hand limit. In particular, the consumed fidelity $F^-(t)$ takes the value of the fidelity at the *end* of the time step $[t-1, t)$. The values of F^- may also be deduced directly from F as

$$F^-(t) = \begin{cases} e^{-\Gamma}\left(F(t-1) - \frac{1}{4}\right) + \frac{1}{4}, \text{ if } F(t-1) > 0, \\ 0, \text{ if } F(t-1) = 0. \end{cases} \tag{3.19}$$

We note that the evolution of $\{F^-(t), t \in \mathbb{N}\}$ may be deduced directly from $\{F(t), t \in \mathbb{N}\}$ via (3.19), and vice-versa. The value $F^-(t)$ may be interpreted as the state of the system 'just before' time t, and $F(t)$ the state 'just after'. Each completely captures the behaviour of the 1GnB system.

We then restate the original definitions 3.2 and 3.3 of availability, A, and average consumed fidelity, $\overline{F}$, below.

Definition 3.4 (Performance metrics, viewpoint of network user)**.** We have

$$A = \lim_{m\to\infty} \frac{1}{m} \sum_{j=1}^{m} \mathbb{1}_{\text{l.e.}}\left(F^-(T_{\text{con}}^{(j)})\right), \tag{3.20}$$

and

$$\overline{F} = \lim_{m\to\infty} \frac{\sum_{j=1}^{m} F^-(T_{\text{con}}^{(j)}) \cdot \mathbb{1}_{\text{l.e.}}\left(F^-(T_{\text{con}}^{(j)})\right)}{\sum_{j=1}^{m} \mathbb{1}_{\text{l.e.}}\left(F^-(T_{\text{con}}^{(j)})\right)}. \tag{3.21}$$

We now present a second way to compute the performance metrics, which is the form that is used to derive the solutions for A and $\overline{F}$ in Theorems 3.1 and 3.2 (see Appendix 3.6.2). To show this result, we use the fact that $F(t)$ is a regenerative process. Informally, every time the link in the G memory is removed from the system, the process 'starts again', in the sense that the stochastic properties from that point onwards are the same as when starting from any other time when the G memory is empty. This stems from the fact that entangled link generation and consumption request arrivals are assumed to be Markovian.

Definition 3.5 (Regenerative process, informal)**.** A regenerative process $\{X(t), t \geq 0\}$ is a stochastic process with the following properties: there exists a random variable $V_1 > 0$ such that

(i) $\{X(t+V_1), t \geq 0\}$ is independent of $\{X(t), t \leq V_1\}$ and V_1;

(ii) $\{X(t+V_1), t \geq 0\}$ is stochastically equivalent to $\{X(t), t \geq 0\}$ (i.e. these two processes have the same joint distributions).

For a formal definition of a regenerative process, see e.g. [136]. If the process is regenerative, it may also be shown that there is a sequence of regeneration cycles $V_0 = 0, \{V_k\}$ such that the sequence regenerates at each cycle, i.e. $\{X(t), t \geq 0\}$ and $\{X(t+V_k), t \geq 0\}$ are stochastically equivalent.

We now show that our process F is regenerative. Let us assume the system starts when a new link is freshly generated and moved to the G memory, such that $F(0) = F_{\text{new}}$. The system then evolves as follows: the link in the G memory may undergo some purification rounds, between which it is subject to decoherence, and then is eventually removed from the G memory after time $T_{\text{occ}}^{(1)}$ due to either purification failure or consumption. The time $T_{\text{occ}}^{(1)}$ is the time during which the G memory is occupied. In particular,

$$T_{\text{occ}}^{(1)} := \min\{t : F(t) = 0\}. \tag{3.22}$$

After the link is removed, the system will then attempt entanglement generation until a successful generation. Let the time from which the G memory is emptied until a new link is produced be $T_{\text{gen}}^{(1)}$. By the assumption that entanglement generation attempts are independent and Bernoulli, $T_{\text{gen}}^{(1)} \sim \text{Geo}(1-(1-p_{\text{gen}})^n)$. When a fresh link is generated at time $t = T_{\text{occ}}^{(1)} + T_{\text{gen}}^{(1)}$, we have $F(T_{\text{occ}}^{(1)} + T_{\text{gen}}^{(1)}) = F_{\text{new}}$ and, from this time on, the process behaves equivalently to how it did from time $t = 0$. Letting $V_1 = T_{\text{occ}}^{(1)} + T_{\text{gen}}^{(1)}$, we see that $F(t)$ is regenerative. All regeneration cycles $\{V_k\}$ may each be split into two phases: we have $V_k = T_{\text{occ}}^{(k)} + T_{\text{gen}}^{(k)}$, where $T_{\text{occ}}^{(k)}$ is the time during which the memory is occupied, and $T_{\text{gen}}^{(k)}$ is the time during which the memory is empty and entanglement generation is being attempted. We note that since F^- is in one-to-one correspondence with F via (3.19), then F^- is also regenerative with the same cycle lengths.

For the following results, we note two important properties of the process $\{V_k\}$. Firstly, the mean cycle length $\mathbb{E}[V_1] = \mathbb{E}[T_{\text{occ}}^{(1)}] + \mathbb{E}[T_{\text{gen}}^{(1)}]$ is finite: this may be seen by the fact that $T_{\text{gen}}^{(1)}$ is geometrically distributed (and therefore $\mathbb{E}[T_{\text{gen}}^{(1)}] < \infty$) and that $T_{\text{occ}}^{(1)}$ is bounded above by the time until the next consumption request, which is geometrically distributed, and so $\mathbb{E}[T_{\text{occ}}^{(1)}] \leq \mathbb{E}[T_{\text{con}}^{(1)}] < \infty$. The second important property is that the $\{V_k\}$ are *aperiodic*, which means that V_1 takes values in a set of integers that have greatest common denominator equal to one. Again, this may be seen by the fact that consumption and entanglement generation are assumed to be geometric. If $p_{\text{gen}} < 1$, the value of V_1 has a non-zero probability of taking any value in $\mathbb{N} \setminus \{1\}$ and therefore satisfies this property. The same holds if $p_{\text{gen}} = 1$, and there is a non-zero probability of either no purification or successful purification. The cases where the $\{V_k\}$ are periodic may be accounted for separately:

(A) If $p_{\text{gen}} = 1$ and $p_{\text{con}} = 1$, a link will deterministically be generated when in the empty state, and deterministically consumed in the following time step. The fidelity $F(t)$ then deterministically alternates between 0 and F_{new}, and the cycle length is always two. We therefore have

$$A = \frac{1}{2}, \quad \overline{F} = e^{-\Gamma}\left(F_{\text{new}} - \frac{1}{4}\right) + \frac{1}{4}. \tag{3.23}$$

(B) If $p_{\text{gen}} = 1$, $q = 1$ and $c_k = d_k = 0$, then we have deterministic link generation, and the system always decides to purify. However, purification always fails. The fidelity then again deterministically alternates between 0 and F_{new}, and the cycle length is two. We note that even if purification is always attempted and always fails, then if a consumption request arrives, this will take priority over purification and the link will be consumed with fidelity $e^{-\Gamma}\left(F_{\text{new}} - \frac{1}{4}\right) + \frac{1}{4}$. Then, $\overline{F}$ will also take this value. Moreover, by applying the PASTA property in discrete time [137], we have $A = 1/2$. Our metrics then take the values (3.23), as in case (A).

We note that our formulae, as given in Theorems 3.1 and 3.2, still hold for the above cases. The solutions for edge case (A) are obtained by inputting $p_{\text{gen}} = 1$ and $p_{\text{con}} = 1$. Edge case (B) can be dealt with in the same way: take $p_{\text{gen}} = 1$, $q = 1$ and the limit $c_k, d_k \to 0$. Note that the jump function (3.3) must still be well-defined, and so necessarily we must also take $a_k, b_k \to 0$. We then obtain (3.23). Although the proof in the general case may not be immediately applied in these cases, our formula still holds.

Lemma 3.1 (Performance metrics, single cycle)**.** *Suppose that the 1GnB system parameters are not in edge cases (A) or (B). The performance metrics in Definition 3.4 may be written in terms of the properties of a single cycle:*

$$A = \frac{\mathbb{E}[T_{\text{occ}}^{(1)}]}{\mathbb{E}[T_{\text{occ}}^{(1)}] + \mathbb{E}[T_{\text{gen}}^{(1)}]} \text{ a.s.} \tag{3.24}$$

and

$$\overline{F} = \mathbb{E}[F^{-}(T_{\text{occ}}^{(1)})|C_1] \text{ a.s.} \tag{3.25}$$

where C_1 is the event where the first link is removed due to consumption (and not failed purification), or equivalently $C_1 \equiv \{T_{\text{occ}}^{(1)} = T_{\text{con}}^{(1)}\}$.

Proof. Let F_∞^- be a random variable with distribution given by

$$\mathrm{P}\left(F_\infty^- \in B\right) = \lim_{t\to\infty} \frac{1}{t} \sum_{s=1}^{t} \mathbb{1}_B\left(F^-(s)\right). \tag{3.26}$$

Then, as F^- is a regenerative process with finite mean and aperiodic cycle length, by e.g. part (a) of Theorem 1 from [138], the above quantity exists and may be computed in terms of the properties of a single cycle as

$$\mathrm{P}\left(F_\infty^- \in B\right) = \frac{1}{\mathrm{E}[V_1]} \mathbb{E}\left[\sum_{s=1}^{V_1} \mathbb{1}_B\left(F^-(s)\right)\right]. \tag{3.27}$$

Letting B be the event where a link is present in the G memory, we then see that

$$\mathrm{P}\left(F_\infty^- > 0\right) = \frac{1}{\mathrm{E}[V_1]} \mathbb{E}\left[\sum_{s=1}^{V_1} \mathbb{1}_{\text{l.e.}}\left(F^-(s)\right)\right] \tag{3.28}$$

$$= \frac{1}{\mathbb{E}[T_{\text{occ}}^{(1)}] + \mathbb{E}[T_{\text{gen}}^{(1)}]} \cdot \mathbb{E}\left[T_{\text{occ}}^{(1)}\right]. \tag{3.29}$$

We now show that the above expression is equal to A. Since the interarrival times of consumption requests are i.i.d. and follow a geometric distribution, we make use of the PASTA property in discrete time [137] to see that the availability from the point of view of the consumer in Definition 3.2 is equal to the time average as given above, i.e.

$$A = \lim_{t\to\infty} \frac{1}{t} \sum_{s=1}^{t} \mathbb{1}_{\text{l.e.}}\left(F^-(s)\right) = \mathrm{P}(F_\infty^- > 0), \text{ a.s.} \tag{3.30}$$

Then, (3.24) is shown by combining (3.30) with (3.29).

We now show the identity for $\overline{F}$. For this, we also use the regerative property. We define $W_0 = 0$ and W_k to be the time at which the k-th cycle ends,

$$W_k := \sum_{j=1}^{k} V_j. \tag{3.31}$$

Then, the sequence of times at which the link is removed from the G memory is

$$\{W_{k-1} + T_{\text{occ}}^{(k)}\}_{k\geq 1}. \tag{3.32}$$

We then define the subsequence

$$\{W_{i_k-1} + T_{\text{occ}}^{(i_k)}\}_{k\geq 1} \tag{3.33}$$

to be the times at which link removal is due to consumption (and not purification failure). We recall that in our model, when a consumption request arrives, it immediately removes the link from the G memory. Then, (3.33) are precisely the times at which consumption requests arrive to find a link in the G memory. In particular,

$$\{W_{i_k-1} + T_{\text{occ}}^{(i_k)}\}_{k\geq 1} = \left\{T_{\text{con}}^{(k)} : F^-\left(T_{\text{con}}^{(k)}\right) > 0\right\}_{k\geq 1}, \tag{3.34}$$

recalling that $\{T_{\text{con}}^{(k)}\}$ is the sequence of arrival times for consumption requests. Recalling Definition (3.4) of $\overline{F}$, we then see that

$$\begin{aligned}\overline{F} &= \lim_{m\to\infty} \frac{\sum_{k=1}^{m} F^-(T_{\text{con}}^{(k)}) \cdot \mathbb{1}_{\text{l.e.}}\left(F^-(T_{\text{con}}^{(k)})\right)}{\sum_{k=1}^{m} \mathbb{1}_{\text{l.e.}}\left(F^-(T_{\text{con}}^{(k)})\right)} \\ &= \lim_{m\to\infty} \frac{\sum_{k=1}^{M(m)} F^-\left(W_{i_k-1} + T_{\text{occ}}^{(i_k)}\right)}{\sum_{k=1}^{M(m)} 1},\end{aligned} \tag{3.35}$$

where we have used the identity (3.34), and defined $M(m) \leq m$ as

$$M(m) = \left|\left\{T_{\text{con}}^{(k)} : F^-\left(T_{\text{con}}^{(k)}\right) > 0,\ k \leq m\right\}\right|.$$

Then, $M(m)$ is the number of consumption requests up to time $T_{\text{con}}^{(m)}$ that arrive when a link is stored in memory. We now show that $\lim_{m\to\infty} M(m) = \infty$ a.s. so that we can apply SLLN to the above expression. To see this, recall that $\{V_k\}_{k\geq 1}$ are the i.i.d. interarrival times of a renewal process $N(t) = \sup\{k : W_k \leq t\}$. Since $|\mathbb{E}[V_1]| < \infty$, we have that $\lim_{t\to\infty} N(t) = \infty$ a.s. (see 10.1.2 of [95]). Within each of these cycles, the link is removed from memory exactly once. The probability that this is due to consumption is bounded below by $p_{\text{con}} > 0$, because for each cycle it is possible to consume directly after link generation, which occurs with probability p_{con}. Recalling the sequence of times when the link is removed due to consumption as given in (3.34), the number of these events may therefore be bounded below by a subsequence

$$\{W_{j_k-1} + T_{\text{occ}}^{(j_k)}\}_{k\geq 1} \subseteq \{W_{i_k-1} + T_{\text{occ}}^{(i_k)}\}_{k\geq 1} \tag{3.36}$$

such that the $j_k - j_{k-1}$ is geometrically distributed with parameter $\eta \geq p_{\text{con}}$. We therefore see that

$$\lim_{k\to\infty} |\{W_{j_k-1} + T_{\text{occ}}^{(j_k)}\}_{k\geq 1}| = \infty \text{ a.s.} \tag{3.37}$$

and therefore by (3.36), the total number of times when the link is consumed diverges to infinity almost surely. From (3.35), we then have

$$\begin{aligned}\overline{F} &\overset{\text{a.s.}}{=} \lim_{M\to\infty} \frac{1}{M}\sum_{k=1}^{M} F^{-}\left(W_{i_k-1} + T_{\text{occ}}^{(i_k)}\right) \\ &\overset{\text{a.s.}}{=} \mathbb{E}\left[F^{-}(T_{\text{occ}}^{(1)})|C_1\right],\end{aligned}$$

where we have used the fact that the sequence $\{F^{-}(W_{i_k-1} + T_{\text{occ}}^{(i_k)})\}_{k\geq 1}$ is i.i.d. since the process is regenerative, and the strong law of large numbers. □

In the final lemma of this section, we see that the above metrics are equal to the time averages over the whole process. This follows from a version of the well-known PASTA property (Poisson Arrivals See Time Averages) in queuing theory [137], which we can employ because the arrival of consumption requests in each time step is assumed to be a Bernoulli process.

Lemma 3.2 (Performance metrics, time average)**.** *Suppose that the 1GnB system parameters are not in edge cases (A) or (B). The performance metrics in Definition 3.4 may be computed using an average over time, i.e.*

$$A = \lim_{t\to\infty} \frac{1}{t}\sum_{s=1}^{t} \mathbb{1}_{\text{l.e.}}\left(F^{-}(s)\right), \tag{3.38}$$

and

$$\overline{F} = \lim_{t\to\infty} \frac{\sum_{s=1}^{t} F^{-}(s)\cdot \mathbb{1}_{\text{l.e.}}\left(F^{-}(s)\right)}{\sum_{s=1}^{t} \mathbb{1}_{\text{l.e.}}\left(F^{-}(s)\right)} \tag{3.39}$$

Proof. The identity for A is a direct application of the PASTA property in discrete time [137], which we also saw in the proof of Lemma 3.4.

For the second equality, from (3.21) we firstly rewrite $\overline{F}$ as

$$\overline{F} = \lim_{m\to\infty} \frac{\frac{1}{m}\sum_{j=1}^{m} F^{-}\left(T_{\text{con}}^{(j)}\right)}{\frac{1}{m}\sum_{j=1}^{m} \mathbb{1}_{\text{l.e.}}\left(F^{-}\left(T_{\text{con}}^{(j)}\right)\right)} = \frac{\overline{F}_{\text{tot}}}{A}, \tag{3.40}$$

where

$$\overline{F}_{\text{tot}} := \lim_{m\to\infty} \frac{1}{m}\sum_{j=1}^{m} F^{-}\left(T_{\text{con}}^{(j)}\right)$$

is the average fidelity seen by users, without conditioning on the fidelity being nonzero. In (3.40), we have removed the indicator function from the sum in the numerator by recalling that $F^{-}(T_{\text{con}}^{(j)}) = 0$ if the j-th consumption request does not find a link in memory. Then, since $\overline{F}_{\text{tot}} = \overline{F}\cdot A$ and by Lemma 3.4 both $\overline{F}$ and A converge, the PASTA property can be applied and we have that

$$\overline{F}_{\text{tot}} = \lim_{t\to\infty} \frac{1}{t}\sum_{s=1}^{t} F^{-}(s), \text{ a.s.} \tag{3.41}$$

Then,

$$\overline{F} = \frac{\overline{F}_{\text{tot}}}{A} = \frac{\lim_{t\to\infty} \frac{1}{t}\sum_{s=1}^{t} F^{-}(s)}{\lim_{t\to\infty} \frac{1}{t}\sum_{s=1}^{t} \mathbb{1}_{\text{l.e.}}\left(F^{-}(s)\right)} \tag{3.42}$$

$$= \lim_{t\to\infty} \frac{\sum_{s=1}^{t} F^{-}(s)}{\sum_{s=1}^{t} \mathbb{1}_{\text{l.e.}}\left(F^{-}(s)\right)} \tag{3.43}$$

$$= \lim_{t\to\infty} \frac{\sum_{s=1}^{t} F^{-}(s)\mathbb{1}_{\text{l.e.}}\left(F^{-}(s)\right)}{\sum_{s=1}^{t} \mathbb{1}_{\text{l.e.}}\left(F^{-}(s)\right)} \text{ a.s.} \tag{3.44}$$

□

In the following, we show that our performance metrics may be computed as limiting values of properties of $F(t)$. Note that this was the definition used in ref [56].

Lemma 3.3 (Performance metrics, limiting values)**.** *Suppose that the 1GnB system parameters are not in edge cases (A) or (B). Then, our performance metrics may be computed as*

$$A = \lim_{t\to\infty} \mathrm{P}(F^{-}(t) > 0) \ \textit{a.s.} \tag{3.45}$$

$$\overline{F} = \lim_{t\to\infty} \mathbb{E}\left[F^{-}(t) | F^{-}(t) > 0\right] \ \textit{a.s.} \tag{3.46}$$

Proof. Since $F^{-}(t)$ is a regenerative process with finite mean and an aperiodic cycle length, it follows that the limiting distribution is well-defined in the following sense. As in the proof of Lemma 3.4, we let F^{-}_{∞} be a random variable with distribution given by (3.26). Then, by e.g. parts (a) and (b) of Theorem 1 of [138], we have

$$\lim_{t\to\infty} \mathrm{P}\left(F^{-}(t) \in B\right) = \mathrm{P}(F^{-}_{\infty} \in B). \tag{3.47}$$

We therefore see that

$$\lim_{t\to\infty} \mathrm{P}\left(F^{-}(t) > 0\right) = \mathrm{P}(F^{-}_{\infty} > 0) = A, \tag{3.48}$$

where we have used the identity for A which we saw in (3.30) in the proof of Lemma 3.4. This shows (3.45).

To show the identity for $\overline{F}$, we make use of the renewal-reward theorem (see e.g. 10.5.1 of [95]). From the previous discussion, associated with the regenerative process $\{F(t), t \in \mathbb{N}\}$ with cycle times $\{W_k\}$, there is a renewal process $N(t) = \sup\{k : W_k \leq t\}$. We then define the reward $\tilde{R}_k$ as the sum of fidelity over the k-th cycle,

$$\tilde{R}_k = \sum_{t=W_{k-1}+1}^{W_k} F^{-}(t). \tag{3.49}$$

Then, the cumulative reward up to time t is given by

$$\tilde{C}(t) = \sum_{s=1}^{t} F^{-}(s) \tag{3.50}$$

$$= \sum_{k=1}^{N(t)} \tilde{R}_k + E(t), \tag{3.51}$$

where we have defined

$$E(t) = \sum_{s=W_{N(t)}+1}^{t} F^-(s) \tag{3.52}$$

to be the remainder of the reward that is not contained in a full cycle. Then, we see that

$$\begin{aligned}\frac{\tilde{C}(t)}{t} &\leq \frac{1}{t}\sum_{k=1}^{N(t)+1}\tilde{R}_k \\ &= \frac{\sum_{k=1}^{N(t)+1}\tilde{R}_k}{N(t)+1}\cdot\frac{N(t)+1}{t}.\end{aligned} \tag{3.53}$$

We will now the strong law of large numbers (SLLN) for both terms in the above product. In particular, the convergence of $(N(t)+1)/t$ may be seen by noticing that

$$\frac{\sum_{k=1}^{N(t)} V_k}{N(t)}\cdot\frac{N(t)}{N(t)+1} < \frac{t}{N(t)+1} \leq \frac{\sum_{k=1}^{N(t)+1} V_k}{N(t)+1} \tag{3.54}$$

and using SLLN shows that the upper and lower bound converge to $\mathbb{E}[V_1]$. From (3.53), we therefore see that

$$\lim_{t\to\infty}\frac{\tilde{C}(t)}{t} \leq \frac{\mathbb{E}\left[\tilde{R}_1\right]}{\mathbb{E}[V_1]} \tag{3.55}$$

$$= \frac{\mathbb{E}\left[\sum_{t=1}^{V_1} F^-(t)\right]}{\mathbb{E}[V_1]} \text{ a.s.} \tag{3.56}$$

Similarly,

$$\lim_{t\to\infty}\frac{\tilde{C}(t)}{t} \geq \lim_{t\to\infty}\frac{\sum_{k=1}^{N(t)}\tilde{R}_k}{N(t)+1}\cdot\frac{N(t)}{t} \tag{3.57}$$

$$= \frac{\mathbb{E}\left[\tilde{R}_1\right]}{\mathbb{E}[V_1]} \text{ a.s.} \tag{3.58}$$

Combining (3.41), (3.50), (3.56) and (3.58), we therefore see that

$$\overline{F}_{\text{tot}} = \lim_{t\to\infty}\frac{1}{t}\sum_{s=1}^{t} F^-(s) = \lim_{t\to\infty}\frac{\tilde{C}(t)}{t} = \frac{\mathbb{E}\left[\sum_{t=1}^{V_1} F^-(t)\right]}{\mathbb{E}[V_1]} \text{ a.s.} \tag{3.59}$$

Moreover, using part (b) of Theorem 1 from [138] , we see that

$$\lim_{t\to\infty}\mathbb{E}[F^-(t)] = \frac{\mathbb{E}\left[\sum_{t=1}^{V_1} F^-(t)\right]}{\mathbb{E}[V_1]}, \tag{3.60}$$

and therefore $\overline{F}_{\text{tot}} = \lim_{t\to\infty}\mathbb{E}[F^-(t)]$. Then, we have

$$\lim_{t\to\infty}\mathbb{E}[F^-(t)|F^-(t)>0] = \lim_{t\to\infty}\frac{\mathbb{E}[F^-(t)\mathbb{1}_{\text{l.e.}}(F^-(t))]}{\mathrm{P}(F^-(t)>0)} \tag{3.61}$$

$$= \lim_{t\to\infty}\frac{\mathbb{E}[F^-(t)]}{\mathrm{P}(F^-(t)>0)} = \frac{\overline{F}_{\text{tot}}}{A} = \overline{F}. \tag{3.62}$$

□

3.6.2. DERIVATION OF FORMULAE FOR PERFORMANCE METRICS

In this appendix, we prove Theorems 3.1 and 3.2, which contain the formulae for the availability and the average consumed fidelity of the 1GnB system.

For these derivations, we work with the following change of variable.

Definition 3.6 (Shifted fidelity)**.** The *shifted fidelity* H of the 1GnB system is given by

$$H := F - \frac{1}{4}, \tag{3.63}$$

where F is the fidelity of the link in the G memory.

This will simplify our calculations because under decoherence, the shifted fidelity changes due to a multiplicative exponential factor. In particular, given an initial value h of the shifted fidelity, after t time steps this reduces to

$$h \to e^{-\Gamma t} h. \tag{3.64}$$

We see that the shifted fidelity does not inherit linear terms under decoherence, in contrast to the fidelity, which decays according to (3.1). This will simplify our derivations.

After successful $(k+1)$-to-1 purification, the value h of the shifted fidelity undergoes a jump given by

$$\tilde{J}_k(h) := J_k\left(h + \frac{1}{4}\right) - \frac{1}{4} = \frac{a_k h + b_k}{c_k h + d_k} \tag{3.65}$$

where we have used (3.3). Similarly, the probability of successful purification is

$$\tilde{p}_k(h) := p_k\left(h + \frac{1}{4}\right) = c_k h + d_k. \tag{3.66}$$

Therefore, $\tilde{J}_k$ and $\tilde{p}_k$ are the jump function and success probability of the corresponding purification events for the shifted fidelity.

Finally, we notice that the range for the fidelity $F \in [0,1]$ translates to $H \in \left[-\frac{1}{4}, \frac{3}{4}\right]$. In particular, we have $H < 0$ if and only if there is no link in the G memory.

We have fully characterised the dynamics of the shifted fidelity in 1GnB (decoherence, purification, and link removal). Our two key performance metrics may then be rewritten in terms of H. Recall that with the assumption $F_{\text{new}} > 1/4$, and the depolarising decoherence model (3.1), a link exists at time t if and only if $F(t) > 1/4$, or equivalently $H(t) > 0$. Let us again denote the indicator function when acting on the shifted fidelity as

$$\mathbb{1}_{\text{link exists}}(H) \equiv \mathbb{1}_{\text{l.e.}}(H) = \begin{cases} 1 \text{ if } H \geq 0, \\ 0 \text{ if } H < 0. \end{cases}$$

Recalling Definition 3.2, the availability may then be written as

$$A = \lim_{m\to\infty} \frac{1}{m} \sum_{j=1}^{m} \mathbb{1}_{\text{l.e.}}\left(H(T_{\text{con}}^{(j)})\right). \tag{3.67}$$

Recalling Definition 3.3, the average consumed fidelity may be rewritten as

$$\begin{aligned}\overline{F} &= \lim_{m\to\infty} \frac{\sum_{j=1}^{m} F(T_{\text{con}}^{(j)}) \cdot \mathbb{1}_{\text{l.e.}}\left(F(T_{\text{con}}^{(j)})\right)}{\sum_{j=1}^{m} \mathbb{1}_{\text{l.e.}}\left(F(T_{\text{con}}^{(j)})\right)} \\ &= \lim_{m\to\infty} \frac{\sum_{j=1}^{m} \left(\frac{1}{4} + H(T_{\text{con}}^{(j)})\right) \cdot \mathbb{1}_{\text{l.e.}}\left(H(T_{\text{con}}^{(j)})\right)}{\sum_{j=1}^{m} \mathbb{1}_{\text{l.e.}}\left(H(T_{\text{con}}^{(j)})\right)} \\ &= \lim_{m\to\infty} \left[\frac{1}{4} + \frac{\sum_{j=1}^{m} H(T_{\text{con}}^{(j)}) \cdot \mathbb{1}_{\text{l.e.}}\left(H(T_{\text{con}}^{(j)})\right)}{\sum_{j=1}^{m} \mathbb{1}_{\text{l.e.}}\left(H(T_{\text{con}}^{(j)})\right)}\right] \\ &= \frac{1}{4} + \overline{H}. \end{aligned} \tag{3.68}$$

We have now written $\overline{F}$ in terms of $\overline{H}$, where

$$\overline{H} := \lim_{m\to\infty} \left[\frac{\sum_{j=1}^{m} H(T_{\text{con}}^{(j)}) \cdot \mathbb{1}_{\text{l.e.}}\left(H(T_{\text{con}}^{(j)})\right)}{\sum_{j=1}^{m} \mathbb{1}_{\text{l.e.}}\left(H(T_{\text{con}}^{(j)})\right)}\right] \tag{3.69}$$

is the average consumed *shifted* fidelity. Finding a formula for $\overline{F}$ then reduces to finding a formula for $\overline{H}$.

From now on, we will assume that the system starts with shifted fidelity $H(0) = H_{\text{new}}$, where

$$H_{\text{new}} := F_{\text{new}} - \frac{1}{4} \tag{3.70}$$

is the state of the G memory immediately after transferring a freshly generated link into memory. Note that H_{new} is a constant, as newly generated links are assumed to be identical. The subsequent dynamics of the system will then be as follows: the link may undergo decoherence followed by purification a number of times, until the link is removed. The removal is due to either consumption or purification failure. After the link is removed, entanglement generation will be attempted until success, at which point a link is transferred to the G memory with shifted fidelity H_{new}. See Figure 3.10 for an illustration of this.

Definition 3.7. We define $T_0 = 0$, $\{T_i\}_{i=1}^{\infty}$ to be the times at which H (equivalently, F) experiences a change that is due to purification, consumption or entanglement generation (or alternatively, any change that is not due to decoherence). Let $S_i := T_i - T_{i-1}$ denote the times between each jump.

We also refer to the $\{T_i\}$ as the *jump times*. See Figure 3.10 for a depiction.

Now, recall that both the time until entanglement generation and consumption are assumed to be geometrically distributed. Then, the distribution of S_i is then given by

$$S_i = \begin{cases} \min\left\{\tau_{\text{pur}}^{(i)}, \tau_{\text{con}}^{(i)}\right\} & \text{if } H(T_{i-1}) \geq 0 \\ T_{\text{gen}}^{(i)} & \text{if } H(T_{i-1}) < 0, \end{cases} \tag{3.71}$$

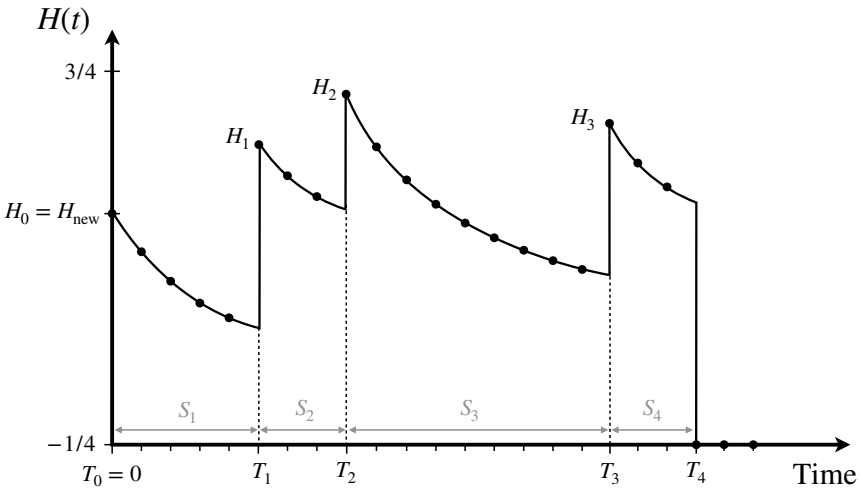


Figure 3.10: **Example dynamics of shifted fidelity in the first cycle of 1GnB.** We assume that $H(0) = H_{\text{new}}$, or equivalently that a freshly generated link is transferred to memory at time $t = 0$. The $\{T_i\}_{i\geq 0}$ are defined to be the times at which there are changes in the (shifted) fidelity that are not due to decoherence. We let T_{occ} be the first time at which the link is removed from the G memory. In the example, $T_{\text{occ}} = T_4$.

where $T_{\text{gen}}^{(i)}$, $\tau_{\text{pur}}^{(i)}$, and $\tau_{\text{con}}^{(i)}$ are independent random variables with the following distributions

$$\begin{aligned} T_{\text{gen}}^{(i)} &\sim \text{Geo}\left(1-(1-p_{\text{gen}})^n\right) \\ \tau_{\text{pur}}^{(i)} &\sim \text{Geo}\left(q(1-(1-p_{\text{gen}})^n)\right) \\ \tau_{\text{con}}^{(i)} &\sim \text{Geo}(p_{\text{con}}). \end{aligned} \tag{3.72}$$

Here, starting at jump time T_{i-1}, $T_{\text{gen}}^{(i)}$ is the time until a new link is generated and transferred to memory, $\tau_{\text{pur}}^{(i)}$ is the time until there is a successful generation and the system decides to attempt purification, and $\tau_{\text{con}}^{(i)}$ is the time until there is a consumption request.

Definition 3.8. For $i \geq 0$, we define $H_i := H(T_i)$ to be the shifted fidelity at the jump times of the process. See Figure 3.10 for an illustration.

Since we assume that the system starts with a freshly generated link in memory, we have $H_0 = H_{\text{new}}$. We note that $\{H_i\}_{i\geq 0}$ is a Markov chain.

Definition 3.9. Let T_{occ} be the first time at which the link in the G memory is removed from the system. In particular, $T_{\text{occ}} = T_N$, where

$$N = \min\{i : H_i < 0\}. \tag{3.73}$$

Note that N is finite a.s. since it is upper bounded by the time until the first consumption request arrives, which follows a geometric distribution.

In Appendix 3.6.1, we saw that $F(t)$ is a regenerative process, meaning that it can be broken down into i.i.d. cycles $V_k = T_{\text{occ}}^{(k)} + T_{\text{gen}}^{(k)}$, where $T_{\text{occ}}^{(k)}$ are the times during which the G memory is occupied and $T_{\text{gen}}^{(k)}$ are the times during which the G memory is empty. We note that in Definition 3.9, $T_{\text{occ}} = T_{\text{occ}}^{(1)}$. From now on, we also refer to $T_{\text{gen}} \equiv T_{\text{gen}}^{(1)}$.

It follows straightforwardly that $H(t) = F(t) - 1/4$ is a regenerative process with the same cycles as $F(t)$. We saw in Lemma 3.1 that the performance metrics may be rewritten in terms of the statistical properties of one cycle. This result also holds for $\overline{H}$, which we restate below. Recalling the notation introduced in Appendix 3.6.1 for F^-, we will also use the equivalent notation for H^-, i.e.

$$H^-(t) = F^-(t) - \frac{1}{4}. \tag{3.74}$$

Lemma 3.4 (Performance metrics for H, single cycle)**.** *The availability is given by*

$$A = \frac{\mathbb{E}[T_{\text{occ}}]}{\mathbb{E}[T_{\text{occ}}] + \mathbb{E}[T_{\text{gen}}]} \ \textit{a.s.} \tag{3.75}$$

and the average consumed (shifted) fidelity is given by

$$\overline{H} = \mathbb{E}\left[e^{-\Gamma S_N} H_{N-1} | \tau_{\text{con}}^{(N)} \leq \tau_{\text{pur}}^{(N)}\right] \ \textit{a.s.} \tag{3.76}$$

where $C_1 \equiv \{\tau_{\text{con}}^{(N)} \leq \tau_{\text{pur}}^{(N)}\}$ is the event that the link is consumed at time T_{occ}.

Proof. The identity (3.75) follows directly from Lemma 3.1. In the same Lemma, we saw that

$$\overline{F} = \mathbb{E}[F(T_{\text{occ}}^{(1)-}) | C_1], \tag{3.77}$$

where C_1 is the event that the first link is removed due to consumption, and we recall the notation

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

which is necessary to capture the fidelity when *consumed* at time t, since the discrete-time stochastic process is defined such that $H(T_{\text{occ}}^{(1)}) = 0$. The value of $H^-(T_{\text{occ}}^{(1)})$ is given by $e^{-\Gamma S_N} H_{N-1}$, where H_{N-1} is the value of the shifted fidelity at the previous jump time (see Definition 3.7) and S_N is the time the link spends decohering in memory from that point until the link is removed from memory (see Definition 3.9). For the conditioning, we recall from (3.71) that $C_1 \equiv \{\tau_{\text{con}}^{(N)} \leq \tau_{\text{pur}}^{(N)}\}$. □

By properties of geometric random variables, we already know that

$$\mathbb{E}[T_{\text{gen}}^{(1)}] = \frac{1}{1-(1-p_{\text{gen}})^n}.$$

To solve for our two performance metrics, it is then sufficient to find formulae for $\mathbb{E}[T_{\text{occ}}]$ and $\mathbb{E}\left[e^{-\Gamma S_N} H_{N-1} | \tau_{\text{con}}^{(N)} \leq \tau_{\text{pur}}^{(N)}\right]$. This is what we accomplish with the following results.

Definition 3.10. For $i \leq N$, let U_i denote the event that purification is attempted at the ith jump time, and $R_i \subseteq U_i$ denote the event that purification is attempted *and* succeeds at the ith jump time.

Lemma 3.5. *Let x and y be given by*

$$x := \sum_{i=1}^{\infty} \mathbb{E}\left[H_i \prod_{j=1}^{i} \mathbb{1}_{R_j}\right], \quad y := \sum_{i=1}^{\infty} \mathrm{P}(N > i). \tag{3.78}$$

Then,

$$\mathbb{E}[T_{\mathrm{occ}}] = \frac{1+y}{p_{\mathrm{con}} + q\left(1-\left(1-p_{\mathrm{gen}}\right)^n\right)(1-p_{\mathrm{con}})} \tag{3.79}$$

and

$$\mathbb{E}[e^{-\Gamma S_N} H_{N-1} | \tau_{\mathrm{con}}^{(N)} \leq \tau_{\mathrm{pur}}^{(N)}] = \frac{(H_{\mathrm{new}} + x)(p_{\mathrm{con}} + q\left(1-\left(1-p_{\mathrm{gen}}\right)^n\right)(1-p_{\mathrm{con}}))}{(1+y)\left(e^{\Gamma} - 1 + p_{\mathrm{con}} + q\left(1-\left(1-p_{\mathrm{gen}}\right)^n\right)(1-p_{\mathrm{con}})\right)}. \tag{3.80}$$

Proof. Denoting $U_N^{\mathrm{c}} = \left\{\tau_{\mathrm{con}}^{(N)} \leq \tau_{\mathrm{pur}}^{(N)}\right\}$ and using properties of the conditional expectation, we may write

$$\mathbb{E}[e^{-\Gamma S_N} H_{N-1} | \tau_{\mathrm{con}}^{(N)} \leq \tau_{\mathrm{pur}}^{(N)}] = \frac{\mathbb{E}[e^{-\Gamma S_N} H_{N-1} \mathbb{1}_{U_N^{\mathrm{c}}}]}{\mathrm{P}(U_N^{\mathrm{c}})}. \tag{3.81}$$

The denominator $\mathrm{P}(U_N^{\mathrm{c}})$ may be rewritten as

$$\begin{aligned}
\mathrm{P}(U_N^{\mathrm{c}}) &= \mathbb{E}[\mathbb{1}_{U_N^{\mathrm{c}}}] \\
&\overset{\mathrm{i}}{=} \mathbb{E}\left[\mathbb{1}_{U_1^{\mathrm{c}}} + \sum_{i=2}^{\infty} \mathbb{1}_{U_i^{\mathrm{c}}} \prod_{j=1}^{i-1} \mathbb{1}_{R_j}\right] \\
&\overset{\mathrm{ii}}{=} \mathbb{E}\left[\mathbb{1}_{U_1^{\mathrm{c}}}\right] + \sum_{i=2}^{\infty} \mathbb{E}\left[\mathbb{1}_{U_i^{\mathrm{c}}} | R_1, \ldots, R_{i-1}\right] \mathbb{E}\left[\prod_{j=1}^{i-1} \mathbb{1}_{R_j}\right] \\
&\overset{\mathrm{iii}}{=} \mathbb{E}\left[\mathbb{1}_{U_1^{\mathrm{c}}}\right]\left(1 + \sum_{i=1}^{\infty} \mathbb{E}\left[\prod_{j=1}^{i} \mathbb{1}_{R_j}\right]\right) \\
&\overset{\mathrm{iv}}{=} \mathrm{P}\left(U_1^{\mathrm{c}}\right)\left(1 + \sum_{i=1}^{\infty} \mathrm{P}\left(N > i\right)\right) \\
&= \mathrm{P}\left(U_1^{\mathrm{c}}\right)\left(1 + y\right).
\end{aligned} \tag{3.82}$$

In the above, we have used the following steps:

i. One may partition the event U_N^{c} by conditioning on the value of N as

$$U_N^{\mathrm{c}} = \bigcup_{i=1}^{\infty} \left(U_i^{\mathrm{c}} \cap \{N = i\}\right).$$

Now, notice that we have $U_i^{\mathrm{c}} \cap \{N = i\}$ exactly when successful purification occurs $i-1$ times, and the link is consumed. Therefore,

$$U_i^{\mathrm{c}} \cap \{N = i\} = U_i^{\mathrm{c}} \cap \left(\cap_{j=1}^{i-1} R_j\right). \tag{3.83}$$

Since the $U_i^c \cap \{N = i\}$ are mutually exclusive, it follows from the above that

$$\begin{aligned}\mathbb{1}_{U_N^c} &= \sum_{i=1}^{\infty} \mathbb{1}_{U_i^c \cap \{N=i\}} \\ &= \sum_{i=1}^{\infty} \mathbb{1}_{U_i^c \cap \left(\cap_{j=1}^{i-1} R_j\right)} \\ &= \mathbb{1}_{U_1^c} + \sum_{i=2}^{\infty} \mathbb{1}_{U_i^c} \prod_{j=1}^{i-1} \mathbb{1}_{R_j}. \end{aligned} \tag{3.84}$$

ii. We use linearity of taking the expectation and take the expectation inside the sum, which is possible by the monotone convergence theorem (see 5.6.12 of [95]). Then, we express the joint probability in terms of conditional probabilities.

iii. We have used the fact that

$$\mathbb{E}\left[\mathbb{1}_{U_i^c} | R_1, ..., R_{i-1}\right] = \mathrm{P}(\tau_{\mathrm{con}}^{(i)} \leq \tau_{\mathrm{pur}}^{(i)}) = \mathrm{P}(\tau_{\mathrm{con}}^{(1)} \leq \tau_{\mathrm{pur}}^{(1)}) = \mathbb{E}\left[\mathbb{1}_{U_1^c}\right].$$

iv. Notice that $N > i$ if and only if the first i jump times are due to successful purification. Therefore,

$$\{N > i\} \equiv \cap_{j=1}^{i} R_j.$$

We therefore have

$$\mathbb{E}\left[\prod_{j=1}^{i} \mathbb{1}_{R_j}\right] = \mathbb{E}\left[\mathbb{1}_{\{N>i\}}\right] = P(N > i).$$

Secondly, we rewrite the numerator of (3.81) in a similar way:

$$\begin{aligned}\mathbb{E}[e^{-\Gamma S_N} H_{N-1} \mathbb{1}_{U_N^c}] &\overset{\text{i}}{=} \mathbb{E}\left[e^{-\Gamma S_N} H_{N-1} \cdot \left(\mathbb{1}_{U_1^c} + \sum_{i=2}^{\infty} \mathbb{1}_{U_i^c} \prod_{j=1}^{i-1} \mathbb{1}_{R_j}\right)\right] \\ &\overset{\text{ii}}{=} \mathbb{E}\left[H_{\mathrm{new}} e^{-\Gamma S_1} \mathbb{1}_{U_1^c} + \sum_{i=1}^{\infty} e^{-\Gamma S_{i+1}} H_i \mathbb{1}_{U_{i+1}^c} \prod_{j=1}^{i} \mathbb{1}_{R_j}\right] \\ &\overset{\text{iii}}{=} H_{\mathrm{new}} \mathbb{E}\left[e^{-\Gamma S_1} \mathbb{1}_{U_1^c}\right] + \sum_{i=1}^{\infty} \mathbb{E}\left[e^{-\Gamma S_{i+1}} \mathbb{1}_{U_{i+1}^c} | R_1, ..., R_i\right] \mathbb{E}\left[H_i \prod_{j=1}^{i} \mathbb{1}_{R_i}\right] \\ &\overset{\text{iv}}{=} \mathbb{E}\left[e^{-\Gamma S_1} \mathbb{1}_{U_1^c}\right] \left(H_{\mathrm{new}} + \sum_{i=1}^{\infty} \mathbb{E}\left[H_i \prod_{j=1}^{i} \mathbb{1}_{R_j}\right]\right) \\ &= \mathbb{E}\left[e^{-\Gamma S_1} \mathbb{1}_{U_1^c}\right] (H_{\mathrm{new}} + x). \end{aligned} \tag{3.85}$$

In the above, we have used the following steps:

i. We have again made use of (3.84), and $H_0 := H_{\mathrm{new}}$.

ii. Again making use of (3.83), we have noticed that the indicator function selects the value of N as

$$e^{-\Gamma S_N} H_{N-1} \cdot \mathbb{1}_{U_i^c} \prod_{j=1}^{i-1} \mathbb{1}_{R_j} = e^{-\Gamma S_N} H_{N-1} \cdot \mathbb{1}_{U_i^c \cap \{N=i\}}$$
$$= e^{-\Gamma S_i} H_{i-1} \cdot \mathbb{1}_{U_i^c} \prod_{j=1}^{i-1} \mathbb{1}_{R_j}.$$

iii. We have used linearity of the expectation, and take the expectation inside the sum, which is possible by the monotone convergence theorem (see 5.6.12 of [95]). Then, we express the joint probability in terms of conditional probabilities.

iv. We have again used the fact that, conditioned on $R_1, ..., R_{i-1}$, $e^{-\Gamma S_i} \mathbb{1}_{U_i^c}$ are identically distributed, for all $i \geq 1$.

We now directly evaluate the multiplying factor in the above expressions. Using the partition $U_1^c = \bigcup_{i=1}^{\infty} \{U_1^c, S_1 = i\}$,

$$\mathbb{E}\left[e^{-\Gamma S_1} \mathbb{1}_{U_1^c}\right] = \mathbb{E}\left[e^{-\Gamma S_1} \mathbb{1}_{U_1^c} \middle| U_1\right] \mathrm{P}(U_1) + \sum_{i=1}^{\infty} \mathbb{E}\left[e^{-\Gamma S_1} \mathbb{1}_{U_1^c} \middle| U_1^c, S_1 = i\right] \mathrm{P}(U_1^c, S_1 = i)$$
$$= 0 + \sum_{i=1}^{\infty} e^{-i\Gamma} \mathrm{P}(U_1^c, S_1 = i). \tag{3.86}$$

Recalling that $U_1^c = \{\tau_{\mathrm{con}}^{(1)} \leq \tau_{\mathrm{pur}}^{(1)}\}$, we now evaluate

$$\mathrm{P}(U_1^c, S_1 = i) = \mathrm{P}(i = \tau_{\mathrm{con}}^{(1)}, i \leq \tau_{\mathrm{pur}}^{(1)})$$
$$= \mathrm{P}(i = \tau_{\mathrm{con}}^{(1)}) \cdot \mathrm{P}(i \leq \tau_{\mathrm{pur}}^{(1)})$$
$$= (1 - p_{\mathrm{con}})^{i-1} p_{\mathrm{con}} \cdot \left(1 - q(1 - (1 - p_{\mathrm{gen}})^n)\right)^{i-1}, \tag{3.87}$$

where we have used the fact that $\tau_{\mathrm{con}}^{(1)}$ and $\tau_{\mathrm{pur}}^{(1)}$ are independent, and have distributions as given in (3.72). Therefore, combining (3.86) and (3.87), it follows that

$$\mathbb{E}\left[e^{-\Gamma S_1} \mathbb{1}_{U_1^c}\right] = \sum_{i=1}^{\infty} e^{-\Gamma i} (1 - p_{\mathrm{con}})^{i-1} p_{\mathrm{con}} \cdot \left(1 - q\left(1 - \left(1 - p_{\mathrm{gen}}\right)^n\right)\right)^{i-1}$$
$$= p_{\mathrm{con}} e^{-\Gamma} \sum_{i=0}^{\infty} e^{-\Gamma i} \left(1 - q\left(1 - \left(1 - p_{\mathrm{gen}}\right)^n\right)\right)^i (1 - p_{\mathrm{con}})^i$$
$$= \frac{p_{\mathrm{con}} e^{-\Gamma}}{1 - e^{-\Gamma}\left(1 - q\left(1 - \left(1 - p_{\mathrm{gen}}\right)^n\right)\right)(1 - p_{\mathrm{con}})}, \tag{3.88}$$

where to obtain the first equality we have relabelled the summing index, and to obtain the second equality we have used the formula for a geometric series. By setting $\Gamma = 0$ in the above, we also obtain

$$\mathbb{E}\left[\mathbb{1}_{U_1^c}\right] = \mathrm{P}(U_1^c) = \frac{p_{\mathrm{con}}}{1 - \left(1 - q\left(1 - \left(1 - p_{\mathrm{gen}}\right)^n\right)\right)(1 - p_{\mathrm{con}})}$$
$$= \frac{p_{\mathrm{con}}}{p_{\mathrm{con}} + q\left(1 - \left(1 - p_{\mathrm{gen}}\right)^n\right)(1 - p_{\mathrm{con}})}. \tag{3.89}$$

Then, combining (3.82), (3.85) (3.88), (3.89) allows us to rewrite (3.81) as

$$\mathbb{E}[e^{-\Gamma S_N} H_{N-1} | \tau_{\text{con}}^{(N)} \leq \tau_{\text{pur}}^{(N)}] = \frac{(H_{\text{new}} + x)(p_{\text{con}} + q\left(1 - \left(1 - p_{\text{gen}}\right)^n\right)(1 - p_{\text{con}}))}{(1+y)\left(e^{\Gamma} - \left(1 - q\left(1 - \left(1 - p_{\text{gen}}\right)^n\right)\right)(1 - p_{\text{con}})\right)}.$$

This shows (3.80). We now show (3.79) using a similar method: firstly, we again condition on the value of N. Recalling Definitions 3.7 and 3.9, we may rewrite T_N as

$$T_N = \sum_{i=1}^{\infty} S_i \cdot \mathbb{1}_{\{N \geq i\}} = S_1 + \sum_{i=2}^{\infty} S_i \prod_{j=1}^{i-1} \mathbb{1}_{R_j},$$

where we have again used $\{N \geq i\} \equiv \cap_{j=1}^{i-1} R_j$ to obtain the second equality. Taking expectations, it follows that

$$\begin{aligned}
\mathbb{E}[T_N] &= \mathbb{E}\left[S_1 + \sum_{i=2}^{\infty} S_i \prod_{j=1}^{i-1} \mathbb{1}_{R_j}\right] \\
&= \mathbb{E}[S_1] + \sum_{i=2}^{\infty} \mathbb{E}[S_i | R_1, ..., R_{i-1}] \mathbb{E}\left[\prod_{j=1}^{i-1} \mathbb{1}_{R_j}\right] \\
&= \mathbb{E}[S_1]\left(1 + \sum_{i=1}^{\infty} \mathbb{E}\left[\prod_{j=1}^{i} \mathbb{1}_{R_j}\right]\right) \\
&= \mathbb{E}[S_1]\left(1 + y\right), \qquad (3.90)
\end{aligned}$$

where we have used the same reasoning as was used to obtain (3.82) and (3.85). It now only remains to compute $\mathbb{E}[S_1]$. Recalling that $S_1 = \min\{\tau_{\text{con}}^{(1)}, \tau_{\text{pur}}^{(1)}\}$, we see that

$$\begin{aligned}
\mathrm{P}(S_1 > i) &= \mathrm{P}(\tau_{\text{con}}^{(1)} > i, \tau_{\text{pur}}^{(1)} > i) \\
&= \mathrm{P}(\tau_{\text{con}}^{(1)} > i)(\tau_{\text{pur}}^{(1)} > i) \\
&= (1 - p_{\text{con}})^i \cdot (1 - q(1 - (1 - p_{\text{gen}})^n))^i,
\end{aligned}$$

where we have used the fact that $\tau_{\text{con}}^{(1)}$ and $\tau_{\text{pur}}^{(1)}$ are independent random variables, and their distributions which are given in (3.72). Then, we may rewrite the expectation as

$$\begin{aligned}
\mathbb{E}[S_1] = \sum_{i=0}^{\infty} \mathrm{P}(S_1 > i) &= \sum_{i=0}^{\infty} (1 - p_{\text{con}})^i \cdot (1 - q(1 - (1 - p_{\text{gen}})^n))^i \\
&= \frac{1}{1 - (1 - p_{\text{con}})(1 - q(1 - (1 - p_{\text{gen}})^n))} \\
&= \frac{1}{p_{\text{con}} + q(1 - p_{\text{con}})(1 - (1 - p_{\text{gen}})^n)},
\end{aligned}$$

where we have used the formula for a geometric series to evaluate the sum. Rearranging terms and combining the above with (3.90), we may then write this as

$$\mathbb{E}[T_N] = \frac{1 + y}{p_{\text{con}} + q(1 - p_{\text{con}})(1 - (1 - p_{\text{gen}})^n)},$$

which shows (3.79). □

Lemma 3.6. *Let x and y be defined as in (3.78). Then,*

$$x = -H_{\text{new}} + \frac{\tilde{B} - \tilde{D}H_{\text{new}} + H_{\text{new}}}{(1-\tilde{A})(1-\tilde{D}) - \tilde{B}\tilde{C}}, \quad y = -1 + \frac{1 - \tilde{A} + \tilde{C}H_{\text{new}}}{(1-\tilde{A})(1-\tilde{D}) - \tilde{B}\tilde{C}}, \tag{3.91}$$

where $\tilde{A}$, $\tilde{B}$, $\tilde{C}$, $\tilde{D}$ are defined in Theorem 3.1 in the main text.

Proof. We firstly define the quantities

$$x_i := \mathbb{E}\left[H_i \prod_{j=1}^{i} \mathbb{1}_{R_j}\right], \quad y_i := \mathrm{P}(N > i), \tag{3.92}$$

which means that, recalling (3.78), x and y may be rewritten as

$$x = \sum_{i=1}^{\infty} x_i, \quad y = \sum_{i=1}^{\infty} y_i. \tag{3.93}$$

We now show that there is a recursive relationship between the $\{x_i\}$ and the $\{y_i\}$. We firstly rewrite x_i by conditioning on the value of H_{i-1}. In particular, recalling that $\cap_{j=1}^{i} R_j = \{N > i\}$, we have

$$x_i = \mathbb{E}\left[H_i \prod_{j=1}^{i} \mathbb{1}_{R_j}\right] = \mathbb{E}\left[H_i \mathbb{1}_{\cap_{j=1}^{i} R_j}\right] = \mathbb{E}\left[H_i \mathbb{1}_{\{N>i\}}\right].$$

Then, one may partition by conditioning on the value of H_{i-1} in the following way:

$$\{N > i-1\} = \bigcup_h \{H_{i-1} = h, N > i-1\}.$$

We may then rewrite x_i as

$$\begin{aligned} x_i = {} & \mathbb{E}\left[H_i \mathbb{1}_{\{N>i\}} \middle| N \le i-1\right] \mathrm{P}(N \le i-1) \\ & + \sum_h \mathbb{E}\left[H_i \mathbb{1}_{\{N>i\}} \middle| H_{i-1} = h, N > i-1\right] \mathrm{P}(H_{i-1} = h, N > i-1), \end{aligned}$$

which simplifies to

$$x_i = 0 + \sum_h \mathbb{E}\left[H_i \mathbb{1}_{\{N>i\}} \middle| H_{i-1} = h, N > i-1\right] \mathrm{P}(H_{i-1} = h, N > i-1). \tag{3.94}$$

We now focus on evaluating $\mathbb{E}\left[H_i \mathbb{1}_{\{N>i\}} \middle| H_{i-1} = h, N > i-1\right]$. We do this for $h > 0$, as this is the only relevant range in the above formula. We firstly notice that this expression may be rewritten as

$$\begin{aligned} \mathbb{E}\left[H_i \mathbb{1}_{\{N>i\}} \middle| H_{i-1} = h, N > i-1\right] &= \mathbb{E}\left[H_i \prod_{j=1}^{i} \mathbb{1}_{R_j} \middle| H_{i-1} = h, \cap_{j=1}^{i-1} R_j\right] \\ &= \mathbb{E}\left[H_i \mathbb{1}_{R_i} \middle| H_{i-1} = h, \cap_{j=1}^{i-1} R_j\right] \\ &= \mathbb{E}\left[H_i \mathbb{1}_{R_i} \middle| H_{i-1} = h\right], \end{aligned} \tag{3.95}$$

where to obtain the final equality we have used the Markovian property of the system: given the information that $H_{i-1} = h > 0$, this is sufficient to understand the future behaviour $\{H_k\}_{k\geq i}$. This follows from the fact that $\{H_i\}$ is a Markov chain.

Recall that R_i is the event where the ith jump time is due to a purification round succeeding. Given that in the above expression we are conditioning on the value H_{i-1}, the random variables on which H_i depends are therefore the time S_i until the next round of purification, and the number of links L_i that are used for this purification (recalling that this number determines which purification protocol is used). We must therefore take the expectation over these two random variables.

Definition 3.11. For $i < N$ (the i-th successful purification round), let L_i be the number of links that were produced in the bad memories just before time T_i.

We then expand the expectation (3.95) to condition on the values taken by S_i and L_i:

$$\begin{aligned}\mathbb{E}\left[H_i \mathbb{1}_{R_i} \middle| H_{i-1} = h\right] &= \sum_{t,k} \mathbb{E}\left[H_i \mathbb{1}_{R_i} \middle| H_{i-1} = h, S_i = t, L_i = k, R_i\right] \mathrm{P}(S_i = t, L_i = k, R_i | H_{i-1} = h) \\ &= \sum_{t,k} \tilde{J}_k\left(e^{-\Gamma t} h\right) \mathrm{P}(S_i = t, L_i = k, R_i | H_{i-1} = h), \end{aligned} \tag{3.96}$$

where, recalling (3.65), $\tilde{J}_k$ is the jump function corresponding to the $(k+1)$-to-1 purification protocol from our purification policy. To evaluate (3.96), it now remains to compute the probability distribution in the weighted sum. We again condition, to find

$$\begin{aligned}&\mathrm{P}(S_i = t, L_i = k, R_i | H_{i-1} = h) \\ &\qquad = \mathrm{P}(R_i | U_i, S_i = t, L_i = k, H_{i-1} = h) \mathrm{P}(U_i, S_i = t, L_i = k | H_{i-1} = h) \\ &\qquad = \tilde{p}_k(e^{-\Gamma t} h) \mathrm{P}(U_i, S_i = t, L_i = k | H_{i-1} = h), \end{aligned} \tag{3.97}$$

where $\tilde{p}_k$ determines the probability of successful purification when employing the $(k+1)$-to-1 protocol, recalling its definition in (3.66). Now, recalling the distribution of S_i from (3.71),

$$\begin{aligned}\mathrm{P}\left(U_i,\, S_i = t,\, L_i = k | H_{i-1} = h\right) &= \mathrm{P}(\tau_{\mathrm{con}}^{(i)} > t,\, \tau_{\mathrm{pur}}^{(i)} = t,\, L_i = k) \\ &= \mathrm{P}(\tau_{\mathrm{con}}^{(i)} > t) \cdot \mathrm{P}(\tau_{\mathrm{pur}}^{(i)} = t,\, L_i = k) \\ &= (1 - p_{\mathrm{con}})^t \cdot (1 - q(1 - (1 - p_{\mathrm{gen}})^n))^{t-1} \cdot q \binom{n}{k} p_{\mathrm{gen}}^k (1 - p_{\mathrm{gen}})^{n-k}, \end{aligned} \tag{3.98}$$

where we have used the fact that $\tau_{\mathrm{pur}}^{(i)}$ and L_i are independent of $\tau_{\mathrm{con}}^{(i)}$.

Combining (3.96), (3.97) and (3.98) yields that $\mathbb{E}\left[H_i \mathbb{1}_{R_i} \middle| H_{i-1} = h\right]$ may be written as

$$\begin{aligned}&\sum_{t,k} \tilde{J}_k\left(e^{-\Gamma t} h\right) \tilde{p}_k\left(e^{-\Gamma t} h\right) (1 - p_{\mathrm{con}})^t (1 - q(1 - (1 - p_{\mathrm{gen}})^n))^{t-1} q \binom{n}{k} p_{\mathrm{gen}}^k (1 - p_{\mathrm{gen}})^{n-k} \\ &\quad = \sum_{t,k} \left(a_k e^{-\Gamma t} h + b_k\right) (1 - p_{\mathrm{con}})^t (1 - q(1 - (1 - p_{\mathrm{gen}})^n))^{t-1} q \binom{n}{k} p_{\mathrm{gen}}^k (1 - p_{\mathrm{gen}})^{n-k}, \end{aligned}$$

where we have made use of the expressions (3.65) and (3.66) that define the purification jump function and success probability for the shifted fidelity. We therefore have

$$\mathbb{E}\left[H_i \mathbb{1}_{R_i} \middle| H_{i-1} = h\right] = \sum_t \left(\tilde{a} e^{-\Gamma t} h + \tilde{b}\right)(1-p_{\text{con}})^t \cdot (1 - q(1-(1-p_{\text{gen}})^n))^{t-1} q, \tag{3.99}$$

where we have defined

$$\tilde{a} = \sum_{k=1}^{n} a_k \cdot \binom{n}{k} (1-p_{\text{gen}})^{n-k} p_{\text{gen}}^k, \quad \tilde{b} = \sum_{k=1}^{n} b_k \cdot \binom{n}{k} (1-p_{\text{gen}})^{n-k} p_{\text{gen}}^k.$$

Now, using the fact that (3.99) is a geometric series (starting from $t = 1$), we obtain

$$\mathbb{E}\left[H_i \mathbb{1}_{R_i} | H_{i-1} = h\right] = \tilde{A} h + \tilde{B}, \tag{3.100}$$

where

$$\tilde{A} = \frac{q(1-p_{\text{con}})\tilde{a}}{e^{\Gamma} - \left(1 - q + q\left(1-p_{\text{gen}}\right)^n\right)(1-p_{\text{con}})}, \quad \tilde{B} = \frac{q(1-p_{\text{con}})\tilde{b}}{p_{\text{con}} + q\left(1-\left(1-p_{\text{gen}}\right)^n\right)(1-p_{\text{con}})}. \tag{3.101}$$

Combining (3.94) and (3.100), we may then write

$$\begin{aligned} x_i &= \sum_h (\tilde{A} h + \tilde{B}) \cdot \mathrm{P}(H_{i-1} = h,\, N > i-1) \\ &= \tilde{A} \cdot \mathbb{E}[H_{i-1} \mathbb{1}_{N>i-1}] + \tilde{B} \cdot \mathrm{P}(N > i-1) \\ &= \tilde{A} x_{i-1} + \tilde{B} y_{i-1}, \end{aligned}$$

which is our first recursion relation for $\{x_i\}$ and $\{y_i\}$. We now write down an analogous recursion relation for y_i. We use the same method as for the x_i. In particular, using

$$y_i = \mathrm{P}(N > i) = \mathbb{E}\left[\mathbb{1}_{N>i}\right],$$

we again expand the expectation while conditioning on the value of H_{i-1} in a step analogous to (3.94):

$$y_i = \sum_{h>0} \mathbb{E}\left[\mathbb{1}_{\{N>i\}} | H_{i-1} = h,\, N > i-1\right] \cdot \mathrm{P}(H_{i-1} = h,\, N > i-1). \tag{3.102}$$

In a step analogous to (3.95), we rewrite the conditional expectation as

$$\mathbb{E}\left[\mathbb{1}_{\{N>i\}} | H_{i-1} = h,\, N > i-1\right] = \mathbb{E}\left[\mathbb{1}_{R_i} | H_{i-1} = h\right]. \tag{3.103}$$

We then expand the above with the distributions of S_i and L_i to obtain

$$\begin{aligned} \mathbb{E}\left[\mathbb{1}_{R_i} | H_{i-1} = h\right] &= \sum_{t,k} 1 \cdot \mathrm{P}(S_i = t, L_i = k, R_i | H_{i-1} = h) \\ &= \sum_{t,k} 1 \cdot \tilde{p}_k(e^{-\Gamma t} h) \mathrm{P}(U_i, S_i = t, L_i = k | H_{i-1} = h) \\ = \sum_{t,k} 1 \cdot \tilde{p}_k(e^{-\Gamma t} h) \cdot (1-p_{\text{con}})^t \cdot (1 - q(1-(1-p_{\text{gen}})^n))^{t-1} \cdot q \binom{n}{k} p_{\text{gen}}^k (1-p_{\text{gen}})^{n-k}, \end{aligned}$$

where in the second step we have used (3.97) and in the last step we have again used the conditional distribution in (3.98). Using again the definition for $\tilde{p}_k$ as given in (3.66), one may simplify the above to obtain

$$\mathbb{E}\left[\mathbb{1}_{R_i}|H_{i-1}=h\right]=\sum_{t>0}(\tilde{c}e^{-\Gamma t}h+\tilde{d})\cdot(1-p_{\text{con}})^t\cdot(1-q(1-(1-p_{\text{gen}})^n))^{t-1}\cdot q, \tag{3.104}$$

where we have defined

$$\tilde{c}=\sum_{k=1}^{n}c_k\cdot\binom{n}{k}(1-p_{\text{gen}})^{n-k}p_{\text{gen}}^k,\quad \tilde{d}=\sum_{k=1}^{n}d_k\cdot\binom{n}{k}(1-p_{\text{gen}})^{n-k}p_{\text{gen}}^k.$$

One may again use a geometric series to evaluate (3.104), to obtain

$$\mathbb{E}\left[\mathbb{1}_{R_i}|H_{i-1}=h\right]=\tilde{C}h+\tilde{D}, \tag{3.105}$$

where

$$\tilde{C}=\frac{q(1-p_{\text{con}})\tilde{c}}{e^{\Gamma}-\left(1-q+q\left(1-p_{\text{gen}}\right)^n\right)(1-p_{\text{con}})},\quad \tilde{D}=\frac{q(1-p_{\text{con}})\tilde{d}}{p_{\text{con}}+q\left(1-\left(1-p_{\text{gen}}\right)^n\right)(1-p_{\text{con}})}. \tag{3.106}$$

Combining (3.102), (3.103) and (3.105) then yields

$$\begin{aligned}y_i&=\sum_{h>0}(\tilde{C}h+\tilde{D})\cdot\mathrm{P}(H_{i-1}=h,\,N>i-1)\\&=\tilde{C}\cdot\mathbb{E}\left[H_{i-1}\mathbb{1}_{\{N>i-1\}}\right]+\tilde{D}\cdot P(N>i-1)\\&=\tilde{C}x_{i-1}+\tilde{D}y_{i-1},\end{aligned}$$

which completes our second recursion relation for the $\{x_i\}$ and $\{y_i\}$. We now combine these to find expressions for x and y. Given the initial values

$$x_0=\mathbb{E}\left[H_0\mathbb{1}_{N>0}\right]=\mathbb{E}\left[H_{\text{new}}\cdot 1\right]=H_{\text{new}}$$

and

$$y_0=\mathrm{P}(N>0)=1,$$

it follows that

$$\begin{aligned}x&=\sum_{i=1}^{\infty}(\tilde{A}x_{i-1}+\tilde{B}y_{i-1})=\tilde{A}(x+H_{\text{new}})+\tilde{B}(y+1)\\y&=\sum_{i=1}^{\infty}(\tilde{C}x_{i-1}+\tilde{D}y_{i-1})=\tilde{C}(x+H_{\text{new}})+\tilde{D}(y+1).\end{aligned}$$

We therefore have a linear system of equations for x and y, which may be written as

$$\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}\tilde{A}&\tilde{B}\\\tilde{C}&\tilde{D}\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}+\begin{pmatrix}\tilde{A}H_{\text{new}}+\tilde{B}\\\tilde{C}H_{\text{new}}+\tilde{D}\end{pmatrix}, \tag{3.107}$$

which has solution

$$\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{(1-\tilde{A})(1-\tilde{D}) - \tilde{B}\tilde{C}} \begin{pmatrix} 1-\tilde{D} & \tilde{B} \\ \tilde{C} & 1-\tilde{A} \end{pmatrix} \begin{pmatrix} \tilde{A}H_{\text{new}} + \tilde{B} \\ \tilde{C}H_{\text{new}} + \tilde{D} \end{pmatrix}, \tag{3.108}$$

providing us with the formulae for x and y. These may be simplified in the following way:

$$\begin{aligned} x &= \frac{(1-\tilde{D})(\tilde{A}H_{\text{new}} + \tilde{B}) + \tilde{B}(\tilde{C}H_{\text{new}} + \tilde{D})}{(1-\tilde{A})(1-\tilde{D}) - \tilde{B}\tilde{C}} = -H_{\text{new}} + \frac{\tilde{B} - \tilde{D}H_{\text{new}} + H_{\text{new}}}{(1-\tilde{A})(1-\tilde{D}) - \tilde{B}\tilde{C}} \\ y &= \frac{\tilde{C}(\tilde{A}H_{\text{new}} + \tilde{B}) + (1-\tilde{A})(\tilde{C}H_{\text{new}} + \tilde{D})}{(1-\tilde{A})(1-\tilde{D}) - \tilde{B}\tilde{C}} = -1 + \frac{1 - \tilde{A} + \tilde{C}H_{\text{new}}}{(1-\tilde{A})(1-\tilde{D}) - \tilde{B}\tilde{C}}, \end{aligned}$$

which are in the final form for x and y, as given in (3.91). □

Proof of Theorems 3.1 and 3.2. We combine Lemmas 3.4, 3.5 and 3.6. From Lemma 3.4, we recall that our performance metrics may be written in terms of properties of the first cycle. From Lemma 3.5, we recall that these may be written in terms of x and y. Finally, in Lemma 3.6 we have found formulae for x and y. In order to write down the availability, we firstly combine (3.79) and (3.91), to find

$$\begin{aligned} \mathbb{E}[T_{\text{occ}}] = \mathbb{E}[T_N] &= \frac{1+y}{p_{\text{con}} + q(1-p_{\text{con}})(1-(1-p_{\text{gen}})^n)} \\ &= \frac{1 - \tilde{A} + \tilde{C}H_{\text{new}}}{\left((1-\tilde{A})(1-\tilde{D}) - \tilde{B}\tilde{C}\right)\tilde{P}} \\ &= \frac{1 - \tilde{A} + \tilde{C}(F_{\text{new}} - \frac{1}{4})}{\left((1-\tilde{A})(1-\tilde{D}) - \tilde{B}\tilde{C}\right)\tilde{P}} \end{aligned}$$

where $\tilde{P} \coloneqq p_{\text{con}} + q(1-p_{\text{con}})(1-(1-p_{\text{gen}})^n)$. This suffices to show Theorem 3.1.

In order to write down the average consumed fidelity, we combine (3.80) and (3.91), to obtain

$$\begin{aligned} \overline{H} &\overset{\text{a.s.}}{=} \frac{\left[\tilde{B} - \tilde{D}H_{\text{new}} + H_{\text{new}}\right] \cdot \left[p_{\text{con}} + q\left(1-\left(1-p_{\text{gen}}\right)^n\right)(1-p_{\text{con}})\right]}{\left[1 - \tilde{A} + \tilde{C}H_{\text{new}}\right] \cdot \left[e^{\Gamma} - \left(1 - q\left(1-\left(1-p_{\text{gen}}\right)^n\right)\right)(1-p_{\text{con}})\right]} \\ &= \frac{q(1-p_{\text{con}})(\tilde{b} - \tilde{d}H_{\text{new}}) + H_{\text{new}}\left(p_{\text{con}} + q\left(1-\left(1-p_{\text{gen}}\right)^n\right)(1-p_{\text{con}})\right)}{q(1-p_{\text{con}})(\tilde{c}H_{\text{new}} - \tilde{a}) + e^{\Gamma} - \left(1 - q\left(1-\left(1-p_{\text{gen}}\right)^n\right)\right)(1-p_{\text{con}})} \end{aligned}$$

where we have used the formulae (3.101) and (3.106) for $\tilde{A}$, $\tilde{B}$, $\tilde{C}$, and $\tilde{D}$. The above may be rewritten as

$$\overline{H} \overset{\text{a.s.}}{=} \frac{q(1-p_{\text{con}})(\tilde{b} - \tilde{d}H_{\text{new}}) + H_{\text{new}}\left(p_{\text{con}} + qp^*_{\text{gen}}(1-p_{\text{con}})\right)}{q(1-p_{\text{con}})(\tilde{c}H_{\text{new}} - \tilde{a}) + e^{\Gamma} - \left(1 - qp^*_{\text{gen}}\right)(1-p_{\text{con}})} \tag{3.109}$$

$$= \frac{\left[p_{\text{con}} + q(1-p_{\text{con}})\left(p^*_{\text{gen}} - \tilde{d}\right)\right] \cdot H_{\text{new}} + q(1-p_{\text{con}})\tilde{b}}{\left[q(1-p_{\text{con}})\tilde{c}\right] \cdot H_{\text{new}} + e^{\Gamma} - 1 + p_{\text{con}} + q(1-p_{\text{con}})\left(p^*_{\text{gen}} - \tilde{a}\right)}, \tag{3.110}$$

where we have let $p^*_{\text{gen}} = 1-(1-p_{\text{gen}})^n$ be the effective probability of link generation. We now convert the above to $\overline{F}$. Recalling that $H_{\text{new}} = F_{\text{new}} - 1/4$, it follows that

$$\begin{aligned}\overline{F} &= \overline{H} + \frac{1}{4} \\ &\overset{\text{a.s.}}{=} \frac{\left[p_{\text{con}} + q(1-p_{\text{con}})\left(p^*_{\text{gen}} - \tilde{d}\right)\right]\cdot(F_{\text{new}} - \frac{1}{4}) + q(1-p_{\text{con}})\tilde{b}}{\left[q(1-p_{\text{con}})\tilde{c}\right]\cdot(F_{\text{new}} - \frac{1}{4}) + e^{\Gamma} - 1 + p_{\text{con}} + q(1-p_{\text{con}})\left(p^*_{\text{gen}} - \tilde{a}\right)} + \frac{1}{4} \\ &= \frac{\left[p_{\text{con}} + q(1-p_{\text{con}})\left(p^*_{\text{gen}} + \frac{\tilde{c}}{4} - \tilde{d}\right)\right]\cdot F_{\text{new}} + \frac{1}{4}\left[e^{\Gamma} - 1 + q(1-p_{\text{con}})\left(-\tilde{a} + 4\tilde{b} - \frac{\tilde{c}}{4} + \tilde{d}\right)\right]}{\left[q(1-p_{\text{con}})\tilde{c}\right]\cdot F_{\text{new}} + e^{\Gamma} - 1 + p_{\text{con}} + q(1-p_{\text{con}})\left(p^*_{\text{gen}} - \tilde{a} - \frac{\tilde{c}}{4}\right)},\end{aligned}$$

which is our formula for the average consumed fidelity in terms of the system parameters, as is given in Theorem 3.2. □

3.6.3. PURIFICATION COEFFICIENTS a_k, b_k, c_k, AND d_k

Here, we discuss the values that the coefficients a_k, b_k, c_k, and d_k of a purification protocol k are allowed to take. Note that these coefficients are in general functions of the newly generated state, ρ_{new}, although here we do not write this dependence explicitly for brevity. Then, in Subsection 3.6.3, we provide explicit expressions for the coefficients of the DEJMPS policy discussed in the main text.

The probability of success of the protocol is given by

$$p_k(F) = c_k\left(F - \frac{1}{4}\right) + d_k, \tag{3.111}$$

where the fidelity of the buffered state F can take values between $1/4$ (fully depolarised state) and 1 (perfect Bell pair). Since p_k is a probability, we must enforce $0 \le p_k \le 1$. At $F = 1/4$, this yields

$$0 \le d_k \le 1. \tag{3.112}$$

At $F = 1$, it yields

$$-\frac{4}{3}d_k \le c_k \le \frac{4}{3}(1-d_k). \tag{3.113}$$

Combining (3.112) and (3.113) yields

$$-\frac{4}{3} \le c_k \le \frac{4}{3}. \tag{3.114}$$

The jump function (output fidelity) of the protocol is given by

$$J_k(F) = \frac{1}{4} + \frac{a_k\left(F - \frac{1}{4}\right) + b_k}{c_k\left(F - \frac{1}{4}\right) + d_k}. \tag{3.115}$$

This output fidelity must also be between $1/4$ (fully depolarised state) and 1 (perfect Bell pair). This condition can be written as $0 \le a_k\left(F - \frac{1}{4}\right) + b_k \le \frac{3}{4}p_k$. At $F = 1/4$, this yields

$$0 \le b_k \le \frac{3}{4}d_k. \tag{3.116}$$

Combining (3.116) and (3.112) yields

$$0 \leq b_k \leq \frac{3}{4}. \tag{3.117}$$

Similarly, the condition on the jump function at $F = 1$ can be written as

$$-\frac{4}{3}b_k \leq a_k \leq -\frac{4}{3}b_k + \frac{3}{4}c_k + d_k. \tag{3.118}$$

Combining (3.118) with (3.112), (3.113), and (3.116), we find

$$-1 \leq a_k \leq 1. \tag{3.119}$$

DEJMPS AND CONCATENATED DEJMPS POLICIES

As explained in the main text, the DEJMPS policy applies the well-known 2-to-1 DEJMPS purification protocol [50] to the buffered link and one of the newly generated links (and discarding the rest). This policy is given by the following purification coefficients:

$$\begin{aligned} a_k &= \frac{1}{6}\left(5\rho_{00} + \rho_{11} + \rho_{22} - 3\rho_{33}\right), \\ b_k &= \frac{1}{24}\left(3\rho_{00} - 3\rho_{11} - 3\rho_{22} + 5\rho_{33}\right), \\ c_k &= \frac{2}{3}\left(\rho_{00} - \rho_{11} - \rho_{22} + \rho_{33}\right), \\ d_k &= \frac{1}{2}\left(\rho_{00} + \rho_{11} + \rho_{22} + \rho_{33}\right), \end{aligned} \tag{3.120}$$

$\forall k \in \{1, ..., n\}$, where ρ_{ii} are the diagonal elements of ρ_{new} in the Bell basis $\{|\phi^+\rangle, |\phi^-\rangle, |\psi^+\rangle, |\psi^-\rangle\}$. Note that we define the Bell states as follows:

$$|\phi^+\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}},\ |\phi^-\rangle = \frac{|00\rangle - |11\rangle}{\sqrt{2}},\ |\psi^+\rangle = \frac{|01\rangle + |10\rangle}{\sqrt{2}},\ |\psi^-\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}. \tag{3.121}$$

Regarding a concatenated or a nested DEJMPS policy, one can find the purification coefficients by applying (3.120) recursively. For each round of DEJMPS, the coefficients ρ_{ii} in (3.120) are the diagonal elements of the output state from the previous application of DEJMPS. These diagonal elements are given by [50]

$$\begin{aligned} \rho_{00} &= \frac{\sigma_{00}\sigma'_{00} + \sigma_{33}\sigma'_{33}}{P}, \\ \rho_{11} &= \frac{\sigma_{00}\sigma'_{33} + \sigma_{33}\sigma'_{00}}{P}, \\ \rho_{22} &= \frac{\sigma_{11}\sigma'_{11} + \sigma_{22}\sigma'_{22}}{P}, \\ \rho_{33} &= \frac{\sigma_{11}\sigma'_{22} + \sigma_{11}\sigma'_{22}}{P}, \end{aligned} \tag{3.122}$$

with $P = (\sigma_{00} + \sigma_{33})(\sigma'_{00} + \sigma'_{33}) + (\sigma_{11} + \sigma_{22})(\sigma'_{11} + \sigma'_{22})$, where σ_{ii} and σ'_{ii} are the Bell diagonal elements of the two input states, σ and σ'.

OPTIMAL BILOCAL CLIFFORD POLICY

In the main text, we compare the concatenated versions of the DEJMPS policy to the optimal bilocal Clifford (optimal-bC) policy. When there is a buffered link in memory and k new links are generated, the optimal-bC policy operates as follows:

- When $k = 1$, DEJMPS is applied, using the buffered link and the newly generated link as inputs.
- When $k > 1$, the optimal k-to-1 purification protocol from ref. [97] is applied to all k new links. Then, the resulting state is used for DEJMPS, together with the buffered link. This is illustrated in Fig. 3.11b.

The reason why we apply an optimal bilocal Clifford protocol followed by DEJMPS is because these bilocal Clifford protocols have been shown to be optimal when the input states are identical. Hence, they ensure that the second link used in the final DEJMPS subroutine has maximum fidelity (see ref. [97] for a comparison of the output fidelity using the optimal protocol versus concatenated DEJMPS). This combined protocol (optimal k-to-1 followed by DEJMPS, Fig. 3.11b) is not necessarily the $(k+1)$-to-1 protocol that yields the largest output fidelity. However, one would expect it to provide better buffering performance than a simple concatenation of DEJMPS subroutines (Fig. 3.11a) – nevertheless, in the main text we show that this intuition is incorrect.

Let us now show how to compute the purification coefficients a_k, b_k, c_k, and d_k of the optimal-bC policy:

- When $k = 1$ new links are generated, the purification coefficients a_1, b_1, c_1, and d_1 are given by (3.120), as in the DEJMPS policy.
- When $k > 1$, we first apply the optimal bilocal Clifford protocol, which outputs a state σ_k, with diagonal elements in the Bell basis $\sigma_{k,ii}$. The probability of success of this subroutine is θ_k. Then, the state σ_k is used as input for a final DEJMPS subroutine. Using (3.120), we obtain

$$\begin{aligned} a_k &= \frac{1}{6\theta_k}\left(5\sigma_{k,00}+\sigma_{k,11}+\sigma_{k,22}-3\sigma_{k,33}\right), \\ b_k &= \frac{1}{24\theta_k}\left(3\sigma_{k,00}-3\sigma_{k,11}-3\sigma_{k,22}+5\sigma_{k,33}\right), \\ c_k &= \frac{2}{3\theta_k}\left(\sigma_{k,00}-\sigma_{k,11}-\sigma_{k,22}+\sigma_{k,33}\right), \\ d_k &= \frac{1}{2\theta_k}\left(\sigma_{k,00}+\sigma_{k,11}+\sigma_{k,22}+\sigma_{k,33}\right). \end{aligned} \tag{3.123}$$

In the example discussed in the main text, we consider $n = 4$. We also consider the newly generated links to be Werner states [135] with fidelity F_{new}, i.e.,

$$\rho_{\text{new}} = F_{\text{new}}\left|\phi^+\right\rangle\!\left\langle\phi^+\right| + \frac{1-F_{\text{new}}}{3}\left|\phi^-\right\rangle\!\left\langle\phi^-\right| + \frac{1-F_{\text{new}}}{3}\left|\psi^+\right\rangle\!\left\langle\psi^+\right| + \frac{1-F_{\text{new}}}{3}\left|\psi^-\right\rangle\!\left\langle\psi^-\right|. \tag{3.124}$$

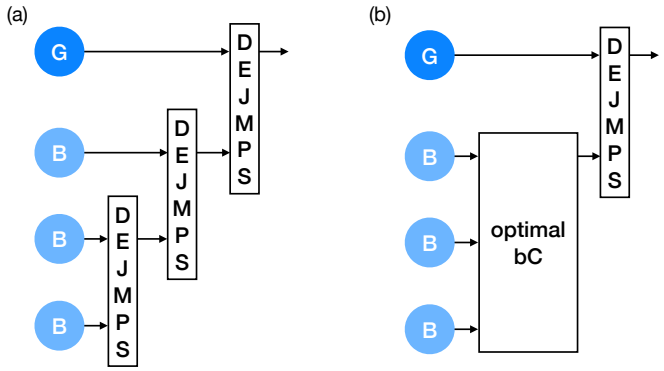


Figure 3.11: **The optimal bilocal Clifford policy applies an optimal protocol followed by DEJMPS.** Illustration of two purification policies: (a) concatenated DEJMPS and (b) optimal bilocal Clifford.

Under these assumptions, the values of $\sigma_{k,00}$ (fidelity of the output state from the optimal bilocal Clifford protocol) and θ_k are given explicitly in ref. [97]:

$$\begin{aligned}
\theta_2 &= \frac{8}{9}F_{\text{new}}^2 - \frac{4}{9}F_{\text{new}} + \frac{5}{9}, \\
\theta_3 &= \frac{32}{27}F_{\text{new}}^3 - \frac{4}{9}F_{\text{new}}^2 + \frac{7}{27}, \\
\theta_4 &= \frac{32}{27}F_{\text{new}}^4 - \frac{4}{9}F_{\text{new}}^2 + \frac{4}{27}F_{\text{new}} + \frac{1}{9},
\end{aligned} \tag{3.125}$$

$$\begin{aligned}
\sigma_{2,00} &= \frac{1}{\theta_2}\cdot\left(\frac{10}{9}F_{\text{new}}^2 - \frac{2}{9}F_{\text{new}} + \frac{1}{9}\right), \\
\sigma_{3,00} &= \frac{1}{\theta_3}\cdot\left(\frac{28}{27}F_{\text{new}}^3 - \frac{1}{9}F_{\text{new}} + \frac{2}{27}\right), \\
\sigma_{4,00} &= \frac{1}{\theta_4}\cdot\left(\frac{8}{9}F_{\text{new}}^4 + \frac{8}{27}F_{\text{new}}^3 - \frac{2}{9}F_{\text{new}}^2 + \frac{1}{27}\right),
\end{aligned} \tag{3.126}$$

where F_{new} is the fidelity of the newly generated Werner states. The rest of the diagonal elements of σ_k can be found using the code provided in our repository (https://github.com/AlvaroGI/buffering-1GnB; this code is based on the methods from ref. [97]). For $F_{\text{new}} = 0.7$, which we use in the example from the main text, we have

$$\begin{cases} \sigma_{2,11} &= 0.20589 \\ \sigma_{2,22} &= 0.02941 \\ \sigma_{2,33} &= 0.02941 \end{cases}, \quad \begin{cases} \sigma_{3,11} &= 0.14287 \\ \sigma_{3,22} &= 0.03571 \\ \sigma_{3,33} &= 0.03571 \end{cases} \text{ and } \quad \begin{cases} \sigma_{4,11} &= 0.04545 \\ \sigma_{4,22} &= 0.04545 \\ \sigma_{4,33} &= 0.04545 \end{cases}. \tag{3.127}$$

The calculations from ref. [97] can also be used to obtain θ_k and σ_k for $k > 4$, although their methods become infeasible for $k > 8$ due to the large computational cost, as discussed in their paper.

3.6.4. MONOTONICITY OF THE AVAILABILITY AND BOUNDS

In this appendix, we show that the availability of the 1GnB system (given in Theorem 3.1) is monotonically decreasing with increasing probability of purification q (Proposition 3.1). This means that the availability is maximised when no purification is performed. If any purification is performed, the availability can only decrease, until reaching its minimum value at $q = 1$. Using these ideas, we compute upper and lower bounds for the availability in Section 3.6.4.

Proof of Proposition 3.1. We start by taking the partial derivative of the availability:

$$\frac{\partial A}{\partial q} = \frac{\mathbb{E}[T_{\text{gen}}]}{\left(\mathbb{E}[T_{\text{gen}}] + \mathbb{E}[T_{\text{occ}}]\right)^2} \frac{\partial \mathbb{E}[T_{\text{occ}}]}{\partial q}, \tag{3.128}$$

where we have used (3.8), (3.9), and (3.10). Since the first term in 3.128 is always positive, the sign of $\partial A/\partial q$ is the same as the sign of $\partial \mathbb{E}[T_{\text{occ}}]/\partial q$. Hence, we only need to show that $\partial \mathbb{E}[T_{\text{occ}}]/\partial q \leq 0$. Next, we write $\mathbb{E}[T_{\text{occ}}]$ explicitly in terms of q:

$$\mathbb{E}[T_{\text{occ}}] = \frac{\varepsilon + \varepsilon' q}{\delta + \delta' q + \delta'' q^2}, \tag{3.129}$$

where

$$\begin{aligned}
\varepsilon &:= \gamma + p_{\text{con}}, \\
\varepsilon' &:= (1 - p_{\text{con}}) \left(p^*_{\text{gen}} - \tilde{a} + H_{\text{new}} \tilde{c} \right), \\
\delta &:= \varepsilon p_{\text{con}}, \\
\delta' &:= (1 - p_{\text{con}}) \left(\gamma p^*_{\text{gen}} + 2 p_{\text{con}} p^*_{\text{gen}} - p_{\text{con}} \tilde{a} - (\gamma + p_{\text{con}}) \tilde{d} \right), \\
\delta'' &:= (1 - p_{\text{con}})^2 \left((p^*_{\text{gen}})^2 - p^*_{\text{gen}} \tilde{a} - p^*_{\text{gen}} \tilde{d} + \tilde{a}\tilde{d} - \tilde{b}\tilde{c} \right),
\end{aligned} \tag{3.130}$$

with $\gamma := e^{\Gamma} - 1$, $p^*_{\text{gen}} := 1 - (1 - p_{\text{gen}})^n$ and $H_{\text{new}} := F_{\text{new}} - \frac{1}{4}$. The derivative of $\mathbb{E}[T_{\text{occ}}]$ can be written as

$$\frac{\partial \mathbb{E}[T_{\text{occ}}]}{\partial q} = \frac{\varepsilon \left(\varepsilon' p_{\text{con}} - \delta' \right) - 2\varepsilon \delta'' q - \varepsilon' \delta'' q^2}{\left(\delta + \delta' q + \delta'' q^2 \right)^2}. \tag{3.131}$$

To prove that $\partial \mathbb{E}[T_{\text{occ}}]/\partial q \leq 0$, we will now show that all three terms in the numerator are negative.

FIRST TERM FROM (3.131) - The first term can be expanded as follows:

$$\varepsilon \left(\varepsilon' p_{\text{con}} - \delta' \right) = -\varepsilon (1 - p_{\text{con}}) \left(\gamma (p^*_{\text{gen}} - \tilde{d}) + p_{\text{con}} (p^*_{\text{gen}} - \tilde{d} - H_{\text{new}} \tilde{c}) \right) \geq 0, \tag{3.132}$$

where, in the last step, we have used the following: (i) $0 \leq p_{\text{con}} \leq 1$, (ii) $\gamma := e^{\Gamma} - 1 \geq 0$,

(iii) $\tilde{d} + H_{\text{new}}\tilde{c} \leq p^*_{\text{gen}}$, and (iv) $\tilde{d} \leq p^*_{\text{gen}}$. Inequality (iii) can be shown as follows:

$$\begin{aligned}\tilde{d} + H_{\text{new}}\tilde{c} &= \sum_{k=1}^{n} (d_k + H_{\text{new}}c_k)\binom{n}{k}(1-p_{\text{gen}})^{n-k}p_{\text{gen}}^k \\ &\leq \sum_{k=1}^{n}\binom{n}{k}(1-p_{\text{gen}})^{n-k}p_{\text{gen}}^k \\ &= \sum_{k=0}^{n}\binom{n}{k}(1-p_{\text{gen}})^{n-k}p_{\text{gen}}^k - (1-p_{\text{gen}})^n \\ &= 1-(1-p_{\text{gen}})^n = p^*_{\text{gen}},\end{aligned} \tag{3.133}$$

where we have used the definition of $\tilde{c}$ and $\tilde{d}$ from Theorem 3.1 and the fact that $d_k + H_{\text{new}}c_k \leq 1$ (this is the success probability of purification protocol k when the link in memory has fidelity F_{new}). Inequality (iv) can be shown in a similar way:

$$\begin{aligned}\tilde{d} &= \sum_{k=1}^{n} d_k\binom{n}{k}(1-p_{\text{gen}})^{n-k}p_{\text{gen}}^k \\ &\leq \sum_{k=1}^{n}\binom{n}{k}(1-p_{\text{gen}})^{n-k}p_{\text{gen}}^k \\ &= \sum_{k=0}^{n}\binom{n}{k}(1-p_{\text{gen}})^{n-k}p_{\text{gen}}^k - (1-p_{\text{gen}})^n \\ &= 1-(1-p_{\text{gen}})^n = p^*_{\text{gen}},\end{aligned} \tag{3.134}$$

where we have used $d_k \leq 1$ (upper bound from (3.112)).

SECOND TERM FROM (3.131) - Regarding the second term in the numerator of (3.131), we first note that, since $p_{\text{con}} \geq 0$ and $\gamma \geq 0$, then $\varepsilon \geq 0$. Moreover, $q \geq 0$ by definition. Consequently, the second term in the numerator of (3.131) is negative if and only if $\delta'' \geq 0$, which in turn is equivalent to $(p^*_{\text{gen}})^2 - p^*_{\text{gen}}\tilde{a} - p^*_{\text{gen}}\tilde{d} + \tilde{a}\tilde{d} - \tilde{b}\tilde{c} \geq 0$. This can be shown as follows:

$$\begin{aligned}(p^*_{\text{gen}})^2 - p^*_{\text{gen}}\tilde{a} - p^*_{\text{gen}}\tilde{d} + \tilde{a}\tilde{d} - \tilde{b}\tilde{c} &\overset{\text{i}}{\geq} (p^*_{\text{gen}})^2 - p^*_{\text{gen}}\tilde{a} - p^*_{\text{gen}}\tilde{d} + \tilde{a}\tilde{d} - \tilde{b}\frac{4}{3}(p^*_{\text{gen}} - \tilde{d}) \\ &= \left(p^*_{\text{gen}} - \frac{4}{3}\tilde{b} - \tilde{a}\right)(p^*_{\text{gen}} - \tilde{d}) \\ &\overset{\text{ii}}{\geq} 0,\end{aligned} \tag{3.135}$$

with these steps:

i. We use $\tilde{b} \geq 0$ (which follows from the lower bound in (3.117)) and $\tilde{c} \leq (p^*_{\text{gen}} - \tilde{d})/H_{\text{new}}$ (shown in (3.133)). This last inequality must hold for any $H_{\text{new}} \in [0, 3/4]$, and therefore $\tilde{c} \leq 4(p^*_{\text{gen}} - \tilde{d})/3$.

ii. To show that the first factor is non-negative, we use $\tilde{a} + 4\tilde{b}/3 \leq \tilde{d} + H_{\text{new}}\tilde{c} \leq p^*_{\text{gen}}$. The first inequality can be shown using the definitions of $\tilde{a}$, $\tilde{b}$, $\tilde{c}$, and $\tilde{d}$ from Theorem 3.1 and the upper bound from (3.118); while the second inequality was shown in (3.133). The second factor $(p^*_{\text{gen}} - \tilde{d})$ is also non-negative, as shown in (3.134).

THIRD TERM FROM (3.131) - Lastly, the third term in the numerator of (3.131) is negative if and only if $\varepsilon' \geq 0$, since we just showed that $\delta'' \geq 0$. Moreover, $\varepsilon' \geq 0 \Leftrightarrow p^*_{\text{gen}} - \tilde{a} + H_{\text{new}}\tilde{c} \geq 0$. The latter can be shown as follows:

$$\begin{aligned}
p^*_{\text{gen}} - \tilde{a} + H_{\text{new}}\tilde{c} &\overset{\text{i}}{=} p^*_{\text{gen}} + \sum_{k=1}^{n} (-a_k + H_{\text{new}}c_k)\binom{n}{k}(1-p_{\text{gen}})^{n-k}p_{\text{gen}}^k \\
&\overset{\text{ii}}{\geq} p^*_{\text{gen}} + \sum_{k=1}^{n} \left(-\left(\frac{3}{4} - H_{\text{new}}\right)c_k - d_k\right)\binom{n}{k}(1-p_{\text{gen}})^{n-k}p_{\text{gen}}^k \\
&\overset{\text{iii}}{\geq} p^*_{\text{gen}} + \sum_{k=1}^{n} \left(\frac{4}{3}H_{\text{new}}(1-d_k) - 1\right)\binom{n}{k}(1-p_{\text{gen}})^{n-k}p_{\text{gen}}^k \\
&\overset{\text{iv}}{\geq} p^*_{\text{gen}} - \sum_{k=1}^{n} \binom{n}{k}(1-p_{\text{gen}})^{n-k}p_{\text{gen}}^k \\
&\overset{\text{v}}{=} 0,
\end{aligned} \tag{3.136}$$

with these steps:

i. We use the definitions of $\tilde{a}$ and $\tilde{c}$ from Theorem 3.1.

ii. We use $a_k \leq 3c_k/4 + d_k$, which can be shown using the upper bound from (3.118) in combination with the lower bound from (3.117).

iii. We use $c_k \leq 4(1-d_k)/3$ (upper bound from (3.113)).

iv. We note that $H_{\text{new}}(1-d_k) \geq 0$, since $H_{\text{new}} \geq 0$ (by definition) and $d_k \leq 1$ (as shown in (3.112)).

v. We recall the definition $p^*_{\text{gen}} := 1 - (1-p_{\text{gen}})^n = \sum_{k=0}^{n}\binom{n}{k}(1-p_{\text{gen}})^{n-k}p_{\text{gen}}^k - (1-p_{\text{gen}})^n$.

We have now shown that all three terms in the numerator of (3.131) are negative. Therefore, $\partial\mathbb{E}[T_{\text{occ}}]/\partial q \leq 0$ and, consequently, $\partial A/\partial q \leq 0$. □

UPPER AND LOWER BOUNDING THE AVAILABILITY

Since $\partial A/\partial q \leq 0$, the availability is upper bounded by the value it takes when $q = 0$. From (3.129), we have

$$\mathbb{E}[T_{\text{occ}}]\big|_{q=0} = \frac{1}{p_{\text{con}}}. \tag{3.137}$$

Combining this with (3.8), we obtain

$$A \leq A\big|_{q=0} = \frac{p^*_{\text{gen}}}{p^*_{\text{gen}} + p_{\text{con}}}, \tag{3.138}$$

with $p^*_{\text{gen}} := 1-(1-p_{\text{gen}})^n$.

To evaluate A at $q=1$, we first use (3.8) and (3.129) to write it as follows:

$$A \geq A\big|_{q=1} = \frac{p^*_{\text{gen}}\eta}{p^*_{\text{gen}}\eta + \Delta}, \tag{3.139}$$

with $\eta := \varepsilon + \varepsilon'$, $\Delta := \delta + \delta' + \delta''$, with $\varepsilon, \varepsilon', \delta, \delta', \delta''$ defined in (3.130).

The solution from (3.139) constitutes a lower bound for the availability. However, η and Δ implicitly depend on the parameters of the purification policy, a_k, b_k, c_k, and d_k, $k \in \{0,\ldots,n\}$. Next, we find a more general and meaningful lower bound that applies to any purification policy.

We start by noting that

- $\varepsilon \geq 0$ (since $p_{\text{con}} \geq 0$ and $\gamma \geq 0$),
- $\varepsilon' \geq 0$ (as shown in (3.136)),
- $\delta \geq 0$ (since $\varepsilon \geq 0$),
- $\delta' \geq 0$ (this can be shown using the fact that $\tilde{d} \leq p^*_{\text{gen}}$, shown in (3.134), and $\tilde{a} \leq p^*_{\text{gen}}$, which can be shown in a similar way as (3.134) and using (3.119)),
- and $\delta'' \geq 0$ (as shown in (3.135)).

As a consequence, none of the factors in (3.139) can be negative: $p^*_{\text{gen}} \geq 0$ (by definition), $\eta \geq 0$, and $\Delta \geq 0$. This means that we can find a lower bound for $A\big|_{q=1}$ by lower bounding η and upper bounding Δ. We first lower bound η:

$$\eta = \gamma + p_{\text{con}} + (1-p_{\text{con}})\left(p^*_{\text{gen}} - \tilde{a} + H_{\text{new}}\tilde{c}\right) \geq \gamma + p_{\text{con}}, \tag{3.140}$$

where we have used $p^*_{\text{gen}} - \tilde{a} + H_{\text{new}}\tilde{c} \geq 0$, which was shown in (3.136).

Regarding Δ, we proceed as follows:

$$\begin{aligned}
\Delta &= \delta + \delta' + \delta'' \\
&= (\gamma + p_{\text{con}})p_{\text{con}} + (1-p_{\text{con}})\left((\gamma + 2p_{\text{con}})p^*_{\text{gen}} - p_{\text{con}}(\tilde{a}+\tilde{d}) - \gamma\tilde{d}\right) \\
&\qquad + (1-p_{\text{con}})^2\left((p^*_{\text{gen}})^2 - p^*_{\text{gen}}(\tilde{a}+\tilde{d}) + \tilde{a}\tilde{d} - \tilde{b}\tilde{c}\right) \\
&\overset{\text{i}}{\leq} (\gamma + p_{\text{con}})p_{\text{con}} + (1-p_{\text{con}})\left(\gamma p^*_{\text{gen}} + 2p_{\text{con}}p^*_{\text{gen}} - p_{\text{con}}\tilde{a}\right) \\
&\qquad + (1-p_{\text{con}})^2\left((p^*_{\text{gen}})^2 - p^*_{\text{gen}}\tilde{a} + \tilde{a}\tilde{d} - \tilde{b}\tilde{c}\right) \\
&\overset{\text{ii}}{\leq} (\gamma + p_{\text{con}})p_{\text{con}} + (1-p_{\text{con}})\left(\gamma p^*_{\text{gen}} + 2p_{\text{con}}p^*_{\text{gen}} + p_{\text{con}}p^*_{\text{gen}}\right) \\
&\qquad + (1-p_{\text{con}})^2\left((p^*_{\text{gen}})^2 + (p^*_{\text{gen}})^2 + \tilde{a}\tilde{d} - \tilde{b}\tilde{c}\right) \\
&= (\gamma + p_{\text{con}})p_{\text{con}} + (1-p_{\text{con}})\left(\gamma + 3p_{\text{con}}\right)p^*_{\text{gen}} + (1-p_{\text{con}})^2\left(2(p^*_{\text{gen}})^2 + \tilde{a}\tilde{d} - \tilde{b}\tilde{c}\right) \\
&\overset{\text{iii}}{\leq} (\gamma + p_{\text{con}})p_{\text{con}} + (1-p_{\text{con}})\left(\gamma + 3p_{\text{con}}\right)p^*_{\text{gen}} + (1-p_{\text{con}})^2\left(2(p^*_{\text{gen}})^2 + p^*_{\text{gen}}\right) \\
&= (\gamma + p_{\text{con}})p_{\text{con}} + (1-p_{\text{con}})\left(1 + \gamma + 2p_{\text{con}}\right)p^*_{\text{gen}} + 2(1-p_{\text{con}})^2(p^*_{\text{gen}})^2
\end{aligned} \tag{3.141}$$

with these steps:

i. We use $-(\gamma + p_{\text{con}})\tilde{d} \leq 0$ and $-p^*_{\text{gen}}\tilde{d} \leq 0$, which follows from $\tilde{d} \geq 0$ (shown in (3.112)).

ii. Using the lower bound from (3.119) and following a similar derivation as in (3.134), one can show that $\tilde{a} \geq -p^*_{\text{gen}}$. This implies that $-p_{\text{con}}\tilde{a} \leq p_{\text{con}}p^*_{\text{gen}}$ and $-p^*_{\text{gen}}\tilde{a} \leq (p^*_{\text{gen}})^2$.

iii. We use $\tilde{a}\tilde{d} - \tilde{b}\tilde{c} \leq p^*_{\text{gen}}$. This can be shown as follows:

$$\tilde{a}\tilde{d} - \tilde{b}\tilde{c} \leq -\frac{4}{3}\tilde{b}\tilde{d} + \frac{3}{4}\tilde{c}\tilde{d} + \tilde{d}^2 - \tilde{b}\tilde{c} \leq -\frac{4}{3}\tilde{b}\tilde{d} + (1-\tilde{d})\tilde{d} + \tilde{d}^2 - \tilde{b}\tilde{c} \leq \tilde{d} \leq p^*_{\text{gen}}, \quad (3.142)$$

where we have used the upper bound from (3.118) in the first step; $\tilde{d} \geq 0$ (see (3.112)) and the upper bound from (3.113) in the second step; $\tilde{b} \geq 0$ (see (3.117)) and the lower bound from (3.113) in the third step; and (3.134) in the last step.

Lastly, combining (3.139) with the bounds from (3.140) and (3.141), we obtain

$$A \geq A\big|_{q=1} = \frac{p^*_{\text{gen}}\eta}{p^*_{\text{gen}}\eta + \Delta} \geq \frac{p^*_{\text{gen}}\left(\gamma + p_{\text{con}}\right)}{\xi + \xi' p^*_{\text{gen}} + \xi''(p^*_{\text{gen}})^2}, \quad (3.143)$$

with $\xi := \gamma p_{\text{con}} + p^2_{\text{con}}$, $\xi' := 1 + 2\gamma + (2-\gamma)p_{\text{con}} - 2p^2_{\text{con}}$, and $\xi'' := 2(1-p_{\text{con}})^2$. This lower bound is general and applies to every 1GnB system, no matter which purification policy it employs.

3.6.5. MONOTONICITY OF THE AVERAGE CONSUMED FIDELITY AND BOUNDS

In this appendix, we show that the average consumed fidelity of the 1GnB system (given in Theorem 3.2) is monotonically increasing with increasing probability of purification q (Proposition 3.2), as long as the purification policy is made of protocols that can increase the fidelity of newly generated links (i.e., $J_k(F_{\text{new}}) \geq F_{\text{new}}$, $\forall k \in \{1,\ldots,n\}$). This means that the average consumed fidelity is maximised when purification is performed every time a new link is generated ($q = 1$). Using these ideas, we compute upper and lower bounds for the average consumed fidelity in Section 3.6.5.

Proof of Proposition 3.2. Recalling from (3.68) that $\overline{F} = \overline{H} + 1/4$, showing the monotonicity of $\overline{H}$ is equivalent to showing the monotonicity of $\overline{F}$. We firstly rewrite the formula for $\overline{H}$ as given in (3.109),

$$\overline{H} = \frac{q(1-p_{\text{con}})\left[\tilde{b} - \tilde{d}H_{\text{new}} + H_{\text{new}}p^*_{\text{gen}}\right] + H_{\text{new}}p_{\text{con}}}{q(1-p_{\text{con}})\left[\tilde{c}H_{\text{new}} - \tilde{a} + p^*_{\text{gen}}\right] + e^{\Gamma} - 1 + p_{\text{con}}},$$

where $p^*_{\text{gen}} = 1 - (1-p_{\text{gen}})^n$. Now consider functions of the form $g(x) = \frac{\alpha x + \beta}{\gamma x + \delta}$. This is non-decreasing if and only if

$$\frac{\mathrm{d}g}{\mathrm{d}x} = \frac{\alpha\delta - \beta\gamma}{(\gamma x + \delta)^2} \geq 0$$
$$\Leftrightarrow \alpha\delta - \beta\gamma \geq 0.$$

We therefore see that $\overline{H}$ is non-decreasing in q if and only if

$$(1-p_{\text{con}})\left[\tilde{b}-\tilde{d}H_{\text{new}}+H_{\text{new}}p^*_{\text{gen}}\right](e^{\Gamma}-1+p_{\text{con}}) \\ -H_{\text{new}}p_{\text{con}}(1-p_{\text{con}})\left[\tilde{c}H_{\text{new}}-\tilde{a}+p^*_{\text{gen}}\right]\geq 0,$$

or equivalently

$$(e^{\Gamma}-1)\left(\tilde{b}-\tilde{d}H_{\text{new}}+H_{\text{new}}p^*_{\text{gen}}\right)+p_{\text{con}}\left(\tilde{b}-\tilde{d}H_{\text{new}}-\tilde{c}H^2_{\text{new}}+\tilde{a}H_{\text{new}}\right)\geq 0 \tag{3.144}$$

We now show this by considering the two parts of the expression:

(a) $\tilde{b}-\tilde{d}H_{\text{new}}+H_{\text{new}}p^*_{\text{gen}}\geq 0$
Recall that the jump functions $\tilde{J}_k$ map the shifted fidelity h as

$$\tilde{J}_k(h)=\frac{a_k h+b_k}{c_k h+d_k}. \tag{3.145}$$

When the input state is completely mixed ($h=0$), the probability of successful purification is

$$\tilde{p}_k(0)=d_k,$$

and so we must have $0\leq d_k\leq 1$. If $d_k>0$, the output fidelity when inputting a completely mixed state then satisfies

$$\tilde{J}_k(0)=\frac{b_k}{d_k}\geq 0$$

which implies $b\geq 0$. If $d=0$, the output fidelity as the input state approaches the completely mixed state is

$$\lim_{h\to 0}\frac{a_k h+b_k}{c_k h},$$

and since this is bounded, it must be the case that $b=0$. Therefore,

$$\tilde{b}=\sum_{k=1}^{n}b_k\cdot\binom{n}{k}(1-p_{\text{gen}})^{n-k}p^k_{\text{gen}}\geq 0$$

and

$$\tilde{d}=\sum_{k=1}^{n}d_k\cdot\binom{n}{k}(1-p_{\text{gen}})^{n-k}p^k_{\text{gen}} \\ \leq\sum_{k=1}^{n}\binom{n}{k}(1-p_{\text{gen}})^{n-k}p^k_{\text{gen}}=1-(1-p_{\text{gen}})^n=p^*_{\text{gen}}.$$

Combining the above, we obtain

$$\tilde{b}-\tilde{d}H_{\text{new}}+H_{\text{new}}p^*_{\text{gen}}\geq H_{\text{new}}(p^*_{\text{gen}}-\tilde{d})\geq 0.$$

(b) $\tilde{b} - \tilde{d}H_{\text{new}} - \tilde{c}H_{\text{new}}^2 + \tilde{a}H_{\text{new}} \geq 0$
We have that

$$\begin{aligned}
&\tilde{b} - \tilde{d}H_{\text{new}} - \tilde{c}H_{\text{new}}^2 + \tilde{a}H_{\text{new}} \\
&\qquad = \sum_{k=1}^{n} \binom{n}{k} (1-p_{\text{gen}})^{n-k} p_{\text{gen}}^k (b_k - d_k H_{\text{new}} - c_k H_{\text{new}}^2 + a_k H_{\text{new}}) \\
&\qquad = \sum_{k=1}^{n} \binom{n}{k} (1-p_{\text{gen}})^{n-k} p_{\text{gen}}^k \left(\frac{a_k H_{\text{new}} + b_k}{c_k H_{\text{new}} + d_k} - H_{\text{new}} \right) (c_k H_{\text{new}} + d_k) \\
&\qquad = \sum_{k=1}^{n} \binom{n}{k} (1-p_{\text{gen}})^{n-k} p_{\text{gen}}^k \cdot \left(\tilde{J}_k(H_{\text{new}}) - H_{\text{new}} \right) \cdot \tilde{p}_k(H_{\text{new}}),
\end{aligned}$$

which is non-negative if all jump functions $\tilde{J}_k$ satisfy

$$\tilde{J}_k(H_{\text{new}}) \geq H_{\text{new}},$$

or equivalently $J_k(F_{\text{new}}) \geq F_{\text{new}}$ Since $\Gamma \geq 0$, we therefore see that (3.144) holds. □

UPPER AND LOWER BOUNDING THE AVERAGE CONSUMED FIDELITY

Here, we only consider purification policies made of protocols that can increase the fidelity of newly generated links (i.e., $J_k(F_{\text{new}}) \geq F_{\text{new}}$, $\forall k \in \{1, ..., n\}$). For these policies, $\partial \overline{F}/\partial q \geq 0$. A tight lower bound can be found by setting $q = 0$ in (3.11):

$$\overline{F} \geq \overline{F}\big|_{q=0} = \frac{\gamma/4 + F_{\text{new}} p_{\text{con}}}{\gamma + p_{\text{con}}}, \tag{3.146}$$

where $\gamma := e^{\Gamma} - 1$.

An upper bound for $\overline{F}$ can be found by upper bounding it maximum value, which occurs at $q = 1$. Using (3.109), we can write the maximum value as

$$\overline{F}\big|_{q=1} = \frac{1}{4} + \frac{(1-p_{\text{con}})(\tilde{b} - \tilde{d}H_{\text{new}}) + H_{\text{new}}\left(p_{\text{con}} + p_{\text{gen}}^*(1-p_{\text{con}})\right)}{(1-p_{\text{con}})(\tilde{c}H_{\text{new}} - \tilde{a}) + \gamma + p_{\text{con}} + (1-p_{\text{con}})p_{\text{gen}}^*}, \tag{3.147}$$

where $p_{\text{gen}}^* = 1 - (1-p_{\text{gen}})^n$. Using (3.116) and (3.134), it can be shown that $\tilde{b} - \tilde{d}H_{\text{new}} \leq p_{\text{gen}}^*(3/4 - H_{\text{new}})$. Moreover, from (3.136), we know that $H_{\text{new}}\tilde{c} - \tilde{a} \geq -p_{\text{gen}}^*$. Applying these two inequalities to (3.147), we find the upper bound:

$$\overline{F} \leq \overline{F}\big|_{q=1} \leq \frac{1}{4} + \frac{H_{\text{new}} p_{\text{con}} + (3/4)(1-p_{\text{con}})p_{\text{gen}}^*}{\gamma + p_{\text{con}}} \tag{3.148}$$

$$= \frac{\gamma/4 + F_{\text{new}} p_{\text{con}}}{\gamma + p_{\text{con}}} + \frac{3}{4}\frac{(1-p_{\text{con}})p_{\text{gen}}^*}{\gamma + p_{\text{con}}}. \tag{3.149}$$

3.6.6. BUFFERING WITH THE 513 EC POLICY

In this appendix, we compare the performance of a 1GnB system that uses a concatenated DEJMPS policy to a system that uses the *513 EC policy*. When there is a buffered link in memory and k new links are generated, the 513 EC policy operates as follows:

- When $k = 1$, DEJMPS is applied, using the buffered link and the newly generated link as inputs.
- When $k = 5$, the purification protocol based on the $[[5,1,3]]$ quantum error-correcting code [139] from ref. [140] is applied to all k new links. Then, the output state is twirled into a Werner state (that is, it is transformed into Werner form while preserving the fidelity) and used for DEJMPS, together with the buffered link.
- Otherwise, twice-concatenated DEJMPS is applied.

This policy is heavily based on twice-concatenated DEJMPS, with the main difference being that, when $k = 5$, a different protocol is applied. Note that, when $k = 5$, we apply some twirling after the purification step to be able to use the results reported in ref. [140], where they provide the output fidelity and the success probability of the protocol but not the full density matrix of the output state.

The purification coefficients of the 513 EC policy can be computed as follows:

- When $k \neq 5$, this policy applies DEJMPS or concatenated DEJMPS. Hence, a_k, b_k, c_k, and d_k can be found as explained in Appendix 3.6.3.
- When $k = 5$, the purification coefficients are given by the output fidelity and probability of success of the 513 EC protocol reported in Figure 3 from ref. [140]. Since we apply this protocol followed by twirling and DEJMPS, we can use (3.120) to compute the purification coefficients of the whole protocol:

$$\begin{aligned} a_5 &= \frac{1}{6\theta}\left(5\sigma_{00}+\sigma_{11}+\sigma_{22}-3\sigma_{33}\right), \\ b_5 &= \frac{1}{24\theta}\left(3\sigma_{00}-3\sigma_{11}-3\sigma_{22}+5\sigma_{33}\right), \\ c_5 &= \frac{2}{3\theta}\left(\sigma_{00}-\sigma_{11}-\sigma_{22}+\sigma_{33}\right), \\ d_5 &= \frac{1}{2\theta}\left(\sigma_{00}+\sigma_{11}+\sigma_{22}+\sigma_{33}\right), \end{aligned} \tag{3.150}$$

  where σ is the output state of the 513 EC protocol after twirling: σ_{00} is the output fidelity from the 513 protocol (reported in Figure 3 from ref. [140]), and $\sigma_{11} = \sigma_{22} = \sigma_{33} = (1-\sigma_{00})/3$ (since we twirl the output state); and θ is the probability of success of the 513 EC protocol (reported in Figure 3 from ref. [140]).

Figure 3.12 shows the performance of the 513 EC policy versus DEJMPS and twice-concatenated DEJMPS. In this example, twice-concatenated DEJMPS also includes twirling before the final round of DEJMPS, to make the comparison with the 513 EC policy fairer. In Fig. 3.12a, we assume $F_{\text{new}} = 0.86$ (according to Figure 3 from ref. [140], this corresponds to $\theta = 0.869$ and $\sigma_{00} = 0.864$), and in Fig. 3.12a, we assume $F_{\text{new}} = 0.95$ (according to Figure 3 from ref. [140], this corresponds to $\theta = 0.981$ and $\sigma_{00} = 0.978$). Similar to the optimal-bC policies discussed in the main text, we observe that the 513 EC policy can be outperformed by DEJMPS, twice-concatenated DEJMPS, and replacement (Figure 3.12a). In some parameter regions, the 513 EC may provide better performance (Figure 3.12b), although this behaviour may not be achievable experimentally, as it requires

both large p_{gen} and large F_{new} – in commonly used entanglement generation protocols, there is a tradeoff between these two parameters, see e.g. ref. [141].

3.6.7. ENTANGLEMENT BUFFERING WITH CONCATENATED PURIFICATION

In this appendix, we discuss further features of 1GnB buffers that use concatenated purification policies. In 3.6.7, we consider different orderings for the purification subroutines that are being concatenated. In 3.6.7, we show that increasing the number of concatenations is beneficial when noise in memory is very strong.

DIFFERENT CONCATENATION ORDERINGS

As stated in the main text, we tested different orderings of the concatenated purification subroutines. In Figure 3.5, we showed two different orderings for a concatenated DEJMPS policy: sequentially concatenated DEJMPS and nested DEJMPS. Here, we consider a policy that applies a nested DEJMPS protocol to all the newly generated links, and then uses the output state to purify the link in memory with a final round of DEJMPS. This policy is only defined when the number of links generated is a power of 2. Hence, we assume $n = 4$ bad memories and deterministic entanglement generation ($p_{\text{gen}} = 1$) in the following example. Figure 3.13 shows the performance of this policy compared to concatenated versions of DEJMPS (in which DEJMPS is applied sequentially to all links, as shown in Figure 3.5a). The performance of all policies shown is qualitatively similar. We also observe that, in this case, nesting is better than concatenating as much as possible, but it is worse than concatenating twice.

INCREASING NUMBER OF CONCATENATIONS

In the main text, we showed that using some newly generated entangled links in the purification protocol and discarding the rest may provide a better buffering performance than implementing a more complex protocol that uses all the newly generated links. In particular, we showed that increasing the maximum number of concatenations in a concatenated DEJMPS policy does not necessarily lead to better performance. The reason was that, as we increase the number of concatenations, the overall probability of success of the protocol decreases. Nevertheless, this effect is irrelevant when noise is strong: the quality of the buffered entanglement decays so rapidly that we need a protocol that can compensate noise with large boosts in fidelity, even if the probability of failure is large. This is shown in Figure 3.14, where we display the maximum average consumed fidelity (i.e. assuming purification probability $q = 1$, see Proposition 3.2) versus the number of concatenations. When no purification is applied (zero concatenations), $\overline{F}$ is below 0.5, meaning that the good memory stores no entanglement, on average (see ref. [56]). As we increase the number of concatenations in the purification protocol, $\overline{F}$ increases, although the increase is marginal. Note that this is a consequence of the strong noise experienced by the buffered entanglement – in Figure 3.7 we showed the same plot but considering a lower noise level and the conclusions were different: increasing the number of concatenations eventually led to a decrease in average consumed fidelity.

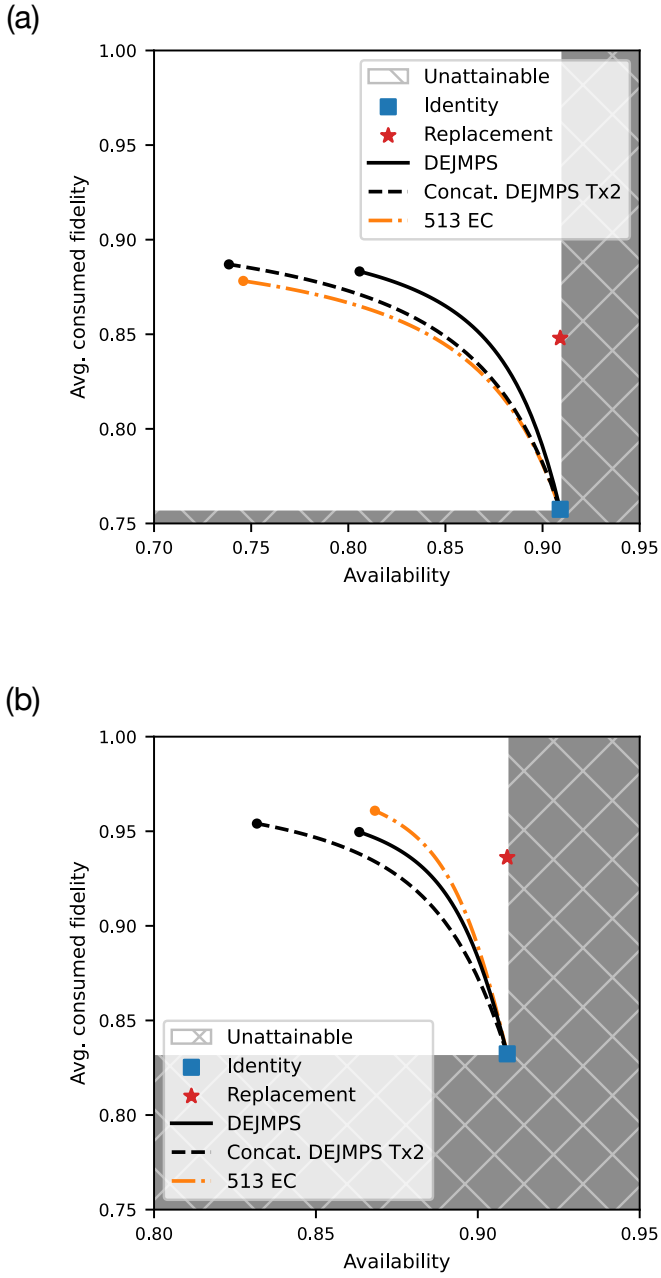


Figure 3.12: **The 513 EC policy may perform better than DEJMPS when new links are generated with a very large fidelity.** Performance of 1GnB systems with different purification policies, in terms of availability A and average consumed fidelity $\overline{F}$. In (a), newly generated links are Werner states with fidelity $F_{\text{new}} = 0.86$, while in (b) we assume $F_{\text{new}} = 0.95$. The shaded area corresponds to unattainable values of A and $\overline{F}$ (see (3.13) and (3.16)). Replacement (star marker) and identity (square marker) policies provide maximum availability. Lines represent the achievable values when using one of the following policies: DEJMPS (solid line), twice-concatenated DEJMPS with twirling (dashed line), and 513 EC (dotted line). Parameter values used in this example: $n = 5$, $p_{\text{gen}} = 1$, $p_{\text{con}} = 0.1$, and $\Gamma = 0.02$.

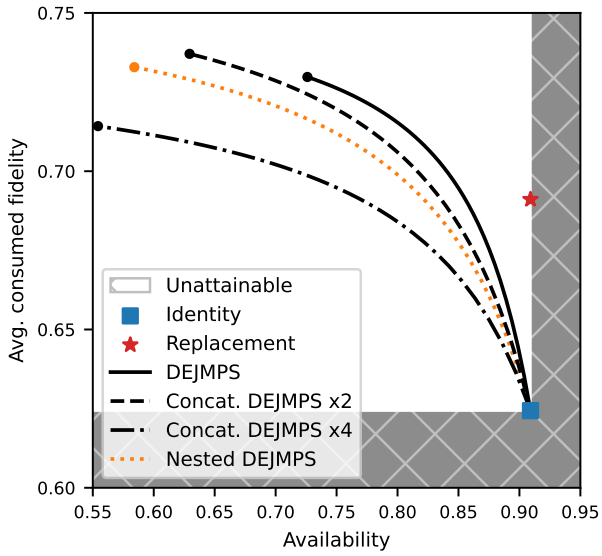


Figure 3.13: **Different concatenation orderings seem to yield qualitatively similar performance.** Performance of 1GnB systems with different purification policies, in terms of availability A and average consumed fidelity $\overline{F}$. The shaded area corresponds to unattainable values of A and $\overline{F}$ (see (3.13) and (3.16)). Lines and markers show the combinations of A and $\overline{F}$ achievable by different purification policies: identity (square marker), replacement (star marker), DEJMPS (solid line), twice-concatenated DEJMPS (dashed line), thrice-concatenated DEJMPS (dotted-dashed line), and nested DEJMPS (orange dotted line). Parameter values used in this example: $n = 4$, $p_{\text{gen}} = 1$, $F_{\text{new}} = 0.7$ (ρ_{new} is a Werner state), $p_{\text{con}} = 0.1$, and $\Gamma = 0.02$.

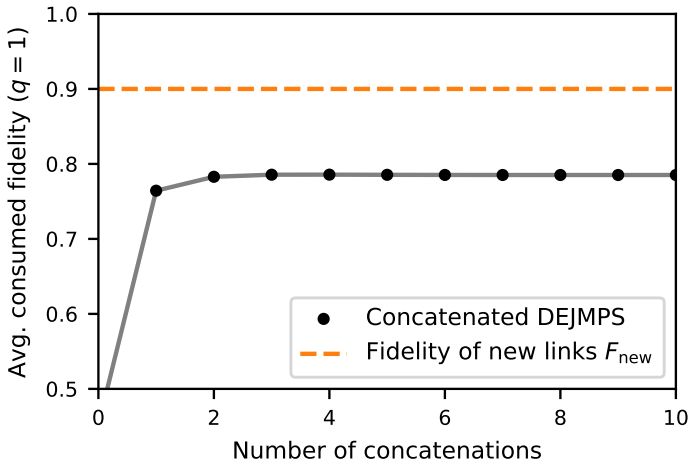


Figure 3.14: **Additional concatenation may improve the performance when noise is strong.** Maximum average consumed fidelity $\overline{F}$ achieved by a purification policy that concatenates DEJMPS a limited number of times. Zero concatenations corresponds to an identity policy (no purification is performed). One concatenation corresponds to the DEJMPS policy. Parameter values used in this example: $n = 10$, $p_{\text{gen}} = 0.5$, $F_{\text{new}} = 0.9$ (ρ_{new} is a Werner state), $p_{\text{con}} = 0.1$, and $\Gamma = 0.2$.

4

QUANTUM PROTOCOLS REQUIRING STATE GENERATION WITHIN A TIME WINDOW

Bethany Davies, Thomas Beauchamp, Gayane Vardoyan and Stephanie Wehner

Quantum protocols commonly require a certain number of quantum resource states to be available simultaneously. An important class of examples is quantum network protocols that require a certain number of entangled pairs. Here, we consider a setting in which a process generates a quantum resource state with some probability p in each time step, and stores it in a quantum memory that is subject to time-dependent noise. To maintain sufficient quality for an application, each resource state is discarded from the memory after w time steps. Let s be the number of desired resource states required by a protocol. We characterise the probability distribution $X_{(w,s)}$ of the ages of the quantum resource states, once s states have been generated in a window w. Combined with a time-dependent noise model, knowledge of this distribution allows for the calculation of fidelity statistics of the s quantum resources. We also give exact solutions for the first and second moments of the waiting time $\tau_{(w,s)}$ until s resources are produced within a window w, which provides information about the rate of the protocol. Since it is difficult to obtain general closed-form expressions for statistical quantities describing the expected waiting time $\mathbb{E}(\tau_{(w,s)})$ and the distribution $X_{(w,s)}$, we present two novel results that aid their computation in

This chapter has been published separately at Davies, Bethany, et al. "Tools for the analysis of quantum protocols requiring state generation within a time window." IEEE Transactions on Quantum Engineering 5 (2024): 1-20.

certain parameter regimes. The methods presented in this work can be used to analyse and optimise the execution of quantum protocols. Specifically, with an example of a Blind Quantum Computing (BQC) protocol, we illustrate how they may be used to infer w and p to optimise the rate of successful protocol execution.

4.1. INTRODUCTION

It is common for quantum computing and quantum network protocols to require the simultaneous availability of a certain number of high quality quantum resource states. In the domain of quantum networks, such resource states are typically entangled pairs of qubits, where the execution of protocols such as entanglement distillation and many quantum network applications require multiple entangled pairs to be available at the same time [49, 50, 12]. Another example of a resource state can be found in the domain of quantum computing, where magic state distillation relies on the presence of multiple initial magic states [142].

Here, we study the setting in which resource states are generated using a probabilistic process. In each time step, this process succeeds in generating one resource state with probability p. If the state is prepared successfully, it is immediately stored in a quantum memory that is subject to time-dependent noise. The process is repeated until all s states required by a protocol are in memory. Such a setting is ubiquitous in quantum networking, and (photonic) quantum computing. A prime example is heralded entanglement generation, which is commonly used in present-day quantum networks (see e.g. [25, 23, 143, 29, 144, 145]).

If the noise is time-dependent, this means that when a state is placed in a quantum memory its quality will degrade over time. In practice then, in order to deliver states of sufficient quality, one often imposes a window of w time steps within which all s states must be produced. If the states are produced within the desired window, the quality of the states is high enough for the application to succeed. Otherwise, the states are typically discarded (see Figure 4.1). In the context of quantum repeater protocols, such a window size is also often referred to as a 'cut-off time', and the analysis across multiple nodes is generally non-trivial [41, 42, 43, 44, 45, 35]. In the context of repeater chains, the goal is typically to deliver one state at a high rate. This is different from our case, where the goal is to deliver multiple states. If a protocol requires s quantum resource states of sufficiently high quality to exist simultaneously, this translates to a requirement of s successful generation events within the window of w time steps. The motivation of this work is to quantify the effects of noise on a quantum protocol - we consider a time window because states may be subjected to time-dependent noise in memory, and therefore must be discarded before they are too old. We remark that our methods apply to many different types of hardware, including those with long coherence times [146].

When analysing the performance of protocols that rely on such a generation of resource states, we are interested in a number of performance metrics. For example, one may be interested in the rate at which we can execute a protocol, the probability that the overall quantum protocol will be successful, or a combined metric that considers the number of successful executions of the quantum protocol per time unit. To understand and optimise such performance metrics, we are interested in understanding a number of quantities related to the system's ability to prepare the resource states required by the

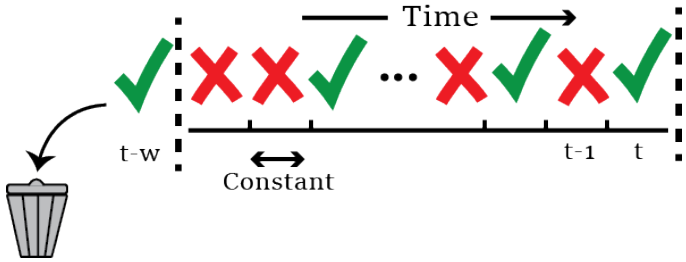


Figure 4.1: **Setup.** In each time step, a probabilistic process generates a resource state, where p is the probability of success (tick) and $1-p$ the probability of failure (cross). After generation the resource state is immediately placed into a quantum memory subject to time-dependent noise. To ensure the states have sufficient quality to enable a quantum protocol, states that are older than a specific window w of time steps are discarded (bin).

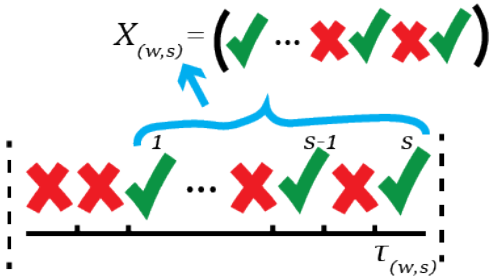


Figure 4.2: **Ending patterns.** At the first instance $\tau_{(w,s)}$ when a window of w time steps contains s successes, we are interested in how long ago each of the s links were generated. This allows one to quantify the quality of the corresponding resource states. The information of when the s resource states were produced is contained in the ending pattern $X_{(w,s)}$.

protocol.

Firstly, one may consider the *waiting time* $\tau_{(w,s)}$ until there are s successes within a window of w time steps. We remark that for fixed parameters (w,s,p) this provides us also with information about the rate at which a protocol can be carried out, when executed multiple times. Secondly, we look at the *ending pattern* $X_{(w,s)}$ (see Figure 4.2), which contains the ages of the s quantum resources at time $\tau_{(w,s)}$. Combined with a model of decoherence, this can be used to compute the quality (fidelity) of the resources immediately after the last state has been produced, which is when the quantum protocol may be executed. Obtaining the distribution of $X_{(w,s)}$ then gives the distribution of the fidelities of the resource states.

The goal of this work is to provide tools that may be used to analyse the performance of a given quantum protocol for specific choices of w, s and p, as well as to choose a combination of these parameters to optimise its performance.

4.1.1. RESULTS

Our main contributions are summarised below.

- For all values of w, s and p, we provide formulae for the first and second moments

of $\tau_{(w,s)}$ (and therefore its mean and variance), and the full distribution of $X_{(w,s)}$. For $(w,s) = (\infty, s)$, these are in a simple closed form, and similarly for $(w,s) = (w,2)$. For all other values of (w,s) we present general formulae, which are in the form of a linear system that may be solved numerically. The dimension of the system scales as w^{s-1}. For large w and s, it is therefore difficult to obtain closed-form expressions from these systems.

- We provide an efficient method to find bounds on the range of w and p for which $\mathbb{E}(\tau_{(w,s)})$ and $X_{(w,s)}$ may be approximated by $\mathbb{E}(\tau_{(\infty,s)})$ and $X_{(\infty,s)}$ to an arbitrary degree of accuracy. In a practical context, this allows one to quickly compute thresholds on the window size such that increasing w further provides no improvement for a protocol rate. Moreover, for appropriate parameter regimes, this approximation is desirable due to the fact that the dimension of the linear system to solve for the expected waiting time $\mathbb{E}(\tau_{(w,s)})$ and $X_{(w,s)}$ scales with w and s, as described above. This is in contrast to the simple closed-from expressions that can be found for the corresponding quantities in the case $w = \infty$.

- We characterise the behaviour of $\mathbb{E}(\tau_{(w,s)})$ and $X_{(w,s)}$ in the limit of a small probability of success. In particular, we show that as $p \to 0$, $\mathbb{E}(\tau_{(w,s)})$ scales as p^{-s}, and that the distribution of $X_{(w,s)}$ becomes uniform. This result may be used to gain intuition about the performance of a quantum application when the resource generation success probability is small, without needing to perform (potentially lengthy) numerical computations.

- We provide a demonstration of how these methods may be used in practice. We consider a Blind Quantum Computation protocol [147]. In our model, entanglement is consumed in the transmission of qubits from a client device to a server device. The model includes noise due to imperfect entangled links and memory decoherence. For a set-up involving a computation on four qubits, we provide an example of how the methods from the first sections may be used to choose architecture parameters that optimise the rate of the protocol.

4.1.2. RELATED WORK

To obtain our results, we draw on methods used in the mathematical literature known as scan statistics [148, 149]. Scan statistics is typically concerned with patterns and clusters in a sequence of random events. This is a field that incorporates techniques from multiple different areas of mathematics. In the quantum context, the problem is different from other areas in caring about the ending pattern distribution. In this work, we therefore use the approach that makes use of martingales, because this allows one to obtain both $\mathbb{E}(\tau_{(w,s)})$ and the distribution of $X_{(w,s)}$ [150]. It is possible to obtain the same quantities with embedded Markov chains [151], but we continue here with the martingale method; firstly because the resulting formula has a smaller dimension (it is therefore faster to compute the quantities of interest), and secondly because it has a regular structure that allows us to derive asymptotic results for small p, which is an experimentally relevant regime. One of the aims of this work is to characterise both $\tau_{(w,s)}$ and the distribution of the ending pattern $X_{(w,s)}$. However, if one is only interested in $\tau_{(w,s)}$ and not

$X_{(w,s)}$, then there exist other methods to compute $\mathbb{E}(\tau_{(w,s)})$, and also in principle the full distribution of $\tau_{(w,s)}$ - see e.g. [152] or [153] which give formulae to obtain the probability generating function of $\tau_{(w,s)}$. To our knowledge, however, these also result in a large system of equations, and not the ending pattern distribution. We therefore do not provide details of these other methods in this work. Other quantities related to the distribution of $\tau_{(w,s)}$ have been explored in great depth in the scan statistics literature, which may also have relevance to the quantum domain. For example, there exist a number of bounds and approximations for $\mathbb{P}(\tau_{(w,s)} \leq n)$ (see e.g. [148] for an overview of results), which may prove useful in allocating time for entanglement generation in a quantum network schedule. By contrast, in this work, we focus on the behaviour of $\mathbb{E}(\tau_{(w,s)})$ and $X_{(w,s)}$, and their implications for the performance of quantum protocols. To our knowledge, this work is the first to characterise the behaviour of the ending pattern distribution in certain parameter regimes, and demonstrate an explicit example of the application of results from scan statistics to a quantum protocol.

4.1.3. OUTLINE

The rest of the chapter is organised as follows. In Section 4.2, the quantities $\tau_{(w,s)}$ and $X_{(w,s)}$ are formally defined. In Sections 4.3.1 and 4.3.2, we give formulae for the first and second moments of $\tau_{(w,s)}$, and the distribution of $X_{(w,s)}$. In Sections 4.3.3 and 4.3.4, we present results that aid the understanding and approximation of these quantities. In Section 4.4.1, the behaviour and practical relevance of the results of Section 4.3 are outlined, specifically looking at $\mathbb{E}(\tau_{(w,s)})$. In Section 4.4.2, we give an example of how one may use the results of Section 4.3 to choose architecture parameters that optimise the performance of a BQC protocol. Finally, further directions are summarised in Section 4.5.

4.2. PRELIMINARIES

We view quantum resource generation attempts as a sequence of i.i.d. Bernoulli trials $(Z_i)_{i=1}^{\infty}$ with success probability $p = \mathbb{P}(Z_1 = 1) > 0$. Then, if a protocol requires $s \leq w$ quantum resources to coexist, the time taken to complete the application is dependent on the *waiting time* $\tau_{(w,s)}$ to produce s successes within a window of size w. We are also interested in the *ending pattern* $X_{(w,s)}$ which completes the process, because this contains the ages of the s quantum resources present at time $\tau_{(w,s)}$. We denote the set of possible ending patterns as $\Omega(w,s)$. This contains every possible configuration of the s successes within the scanning window, so that $X_{(w,s)} \in \Omega(w,s)$. A visualisation of how an ending pattern realises the end of the process is given in Figure 4.2. More specifically, we define

$$\Omega_l(s) := \{x \in \{0,1\}^l : x_1 = x_l = 1 \ \wedge \ \sum_{i=1}^{l} x_i = s\} \tag{4.1}$$

to be the set of all length-l binary strings $x = (x_1, ..., x_l)$ that contain s successes, two of which occur at either end of the string. Then,

$$\Omega(w,s) := \bigcup_{l=s}^{w} \Omega_l(s) \tag{4.2}$$

is the set of ending patterns. The set $\Omega(w,s)$ can be thought of as containing all clusters of s successes that were produced within a time less than or equal to w time steps. Note that the number of possible ending patterns is given by

$$|\Omega(w,s)| = \binom{w-1}{s-1}. \tag{4.3}$$

To see this, consider the fact that each ending pattern in $\Omega(w,s)$ corresponds uniquely to an ending scenario where the s successes are distributed within the window of w time steps, as may be seen from Figure 2. Since the final quantum resource must always have been prepared at the most recent time step and therefore is fixed, it remains to distribute the remaining $s-1$ successes within $w-1$ time steps, meaning that the number of possible ending patterns is restricted to (4.3). The waiting time $\tau_{(w,s)}$ is then defined by

$$\tau_{(w,s)} := \min_{x\in\Omega(w,s)} \{\tau_x\}, \tag{4.4}$$

i.e. this is the time until we see the first ending pattern in the sequence of Bernoulli trials. Here, τ_x is the time taken until one particular ending pattern x is first seen, so that for $x \in \Omega_l(s) \subset \Omega(w,s)$

$$\tau_x := \min\{t : (Z_{t-l+1}, Z_{t-l+2}, ..., Z_t) = x\}. \tag{4.5}$$

We note that $\tau_{(w,s)}$ is well-defined because it is bounded above by a geometric random variable (see Appendix 4.6.2). There is also a relationship between $\tau_{(w,s)}$ and the distribution of $X_{(w,s)}$ given by

$$\mathbb{P}(X_{(w,s)} = x) = \mathbb{P}(\tau_{(w,s)} = \tau_x), \tag{4.6}$$

recalling that $X_{(w,s)}$ takes the value of the ending pattern that completes the process. No two ending patterns can realise the end of the process at the same time since no element of $\Omega(w,s)$ contains another, and so $X_{(w,s)}$ is well-defined.

4.3. FORMULAE AND APPROXIMATIONS

In the following two sections, we provide exact solutions for the first and second moments of $\tau_{(w,s)}$, and the full distribution of $X_{(w,s)}$. Formulae are provided for all possible values of w and s. In Section 4.3.3, we look at approximating the solutions for a large w. In Section 4.3.4, we characterise the solution behaviour for small p.

4.3.1. INFINITE WINDOW

Here, we consider the case where no resource states are discarded (or equivalently when $w = \infty$) and give solutions for the first and second moments of $\tau_{(\infty,s)}$, and the distribution of $X_{(\infty,s)}$. This serves as a useful initial study of the problem, providing intuition for the case where w is large and finite.

When no states are discarded, the waiting time to see all of the successes simply becomes a sum of s i.i.d. geometric random variables with parameter p. This is known as a *negative binomial* distribution, and has an exact distribution given by

$$\mathbb{P}(\tau_{(\infty,s)} = n) = \binom{n-1}{s-1}(1-p)^{n-s}p^s \tag{4.7}$$

and expectation

$$\mathbb{E}(\tau_{(\infty,s)}) = \frac{s}{p}. \tag{4.8}$$

Note that for $w' > w$, it is always the case that $\tau_{(w',s)} \leq \tau_{(w,s)}$ (increasing the window size must always decrease the time for the process to complete), and so

$$\mathbb{E}(\tau_{(w',s)}) \leq \mathbb{E}(\tau_{(w,s)}), \text{ for } w' > w. \tag{4.9}$$

In particular, the waiting time for a finite w will always be greater than or equal to the infinite case. Then, using (4.8) and (4.9), we obtain a simple lower bound in terms of s and p

$$\mathbb{E}(\tau_{(w,s)}) \geq \frac{s}{p}. \tag{4.10}$$

The variance of $\tau_{(\infty,s)}$ is given by

$$\text{Var}(\tau_{(\infty,s)}) = \frac{s \cdot (1-p)}{p^2}, \tag{4.11}$$

from which we can see that the standard deviation is reciprocal in p.

It is also possible to derive a simple expression for the distribution of $X_{(\infty,s)}$. For a binary string $x \in \Omega_l$ that lives in the (now infinite) set of ending patterns $\Omega(\infty, s)$,

$$\mathbb{P}(X_{(\infty,s)} = x) = (1-p)^{l-s} p^{s-1}, \tag{4.12}$$

which can be seen by considering the probability of generating the remaining $l-1$ entries of B after the first success has been generated. We see that when the window size is infinite, the probability of generating ending patterns of the same length is constant.

4.3.2. FINITE WINDOW

$s = 2$

When $s = 2$, it is possible again to derive closed-form solutions for the first and second moment of $\tau_{(w,s)}$ and the distribution of $X_{(w,s)}$. We present below the formulae for $\mathbb{E}(\tau_{(w,s)})$ and the ending pattern distribution.

In this case, the ending patterns are determined by the time between the two states, i.e. $|\Omega_l(2)| = 1$. We separate the process of resource generation into two parts: generation of the first state, and generation of the second state. Generation of the first link occurs when there is no state stored in memory. This is not limited by the window, and has a generation time described by a geometric distribution with parameter p. To finish the process, the generation of the second state must happen within $w-1$ time steps of the first link being generated. When the process is finished, the time between the two states is then a geometric distribution conditional on this event, which occurs with probability $1-(1-p)^{w-1}$. Then, letting $L \in \{1, ..., w-1\}$ be the number of attempts after the first state to generate the second,

$$\mathbb{P}(L = n) = \frac{(1-p)^{n-1} p}{1-(1-p)^{w-1}}, \tag{4.13}$$

which gives the full ending pattern distribution, where $L = n$ corresponds to $X_{(w,s)} \in \Omega_{n+1}(2)$.

Now, let M be the number of times a first state must be generated until the process is finished. Since this is determined by the success of the second state within the time window, we have

$$M \sim \text{Geom}\left(1-(1-p)^{w-1}\right). \tag{4.14}$$

The total time is then given in terms of M and L by

$$\tau_{(w,2)} = \sum_{j=1}^{M} T_j + (M-1)(w-1) + L, \tag{4.15}$$

where the random variables $T_j \sim \text{Geom}(p)$ describe the number of attempts to generate the first state. Now, as shown in Appendix 4.6.1,

$$\mathbb{E}\left(\sum_{j=1}^{M} T_j\right) = \mathbb{E}(M)\mathbb{E}(T_1) = \frac{1}{\left(1-(1-p)^{w-1}\right)} \cdot \frac{1}{p}, \tag{4.16}$$

and

$$\mathbb{E}(L) = \frac{1-(1-p)^w - wp(1-p)^{w-1}}{p(1-(1-p)^{w-1})}. \tag{4.17}$$

The expected waiting time may then be computed as

$$\mathbb{E}(\tau_{(w,2)}) = \mathbb{E}(M)\mathbb{E}(T_1) + (\mathbb{E}(M)-1)(w-1) + \mathbb{E}(L), \tag{4.18}$$

from which we obtain

$$\mathbb{E}(\tau_{(w,2)}) = \frac{1}{p} + \frac{1}{p(1-(1-p)^{w-1})}. \tag{4.19}$$

The variance of $\tau_{(w,2)}$ may also be computed by making use of (4.15). The computation is given in Appendix 4.6.1.

$s > 2$

We now give a formula to exactly compute $\mathbb{E}(\tau_{(w,s)})$ and the full distribution $\{\mathbb{P}(X_{(w,s)} = x) : x \in \Omega(w,s)\}$, for a finite w. This is derived using the method from [150], which makes use of martingales. For completeness, we include an outline of the derivation in Appendix 4.6.2, where a gambling analogy is introduced to aid understanding. The resulting formula is in the form of a linear system of size $|\Omega(w,s)|+1$ that can be solved exactly. Each matrix element defining the linear system can be computed simply and efficiently. Further, in Appendix 4.6.2 we give a formula for the second moment of $\tau_{(w,s)}$, which now involves two linear systems of size $|\Omega(w,s)|$. A martingale method is also used for its derivation, and for this we refer to [153]. The second moment of $\tau_{(w,s)}$ can then be used to calculate the variance and standard deviation of $\tau_{(w,s)}$.

Before stating the first formula, we introduce some notation. We define a function $*$ that maps two binary strings $x = (x_1, ..., x_k)$ and $y = (y_1, ..., y_m)$ to a scalar value, given by

$$x * y := \sum_{j=1}^{\min(k,m)} \left(\prod_{i=1}^{j} \delta_{(x_i, y_{m-j+i})} \right), \tag{4.20}$$

where for $a, b \in \{0,1\}$, the quantity $\delta_{a,b}$ is defined as

$$\delta_{(a,b)} = \begin{cases} \frac{1}{p_a} \text{ if } a = b; \\ 0, \text{ otherwise,} \end{cases} \tag{4.21}$$

where $p_1 := p$ and $p_0 := 1 - p$. From (4.20), we see that the value of $x * y$ is obtained by comparing the overlap of successive substrings of x and y. If two substrings match exactly, then the corresponding term is included in the sum, and it is weighted by an amount that is dependent on the Bernoulli parameter p. Informally, then, $x * y$ measures how similar the structures of x and y are. A simple example of the action of $*$ is given as follows. We consider the action of $*$ on two ending patterns, recalling (4.1). Letting $s = 3$ and $w \geq 7$, suppose that $x = 1010001$ and $y = 100011$. Computing (4.20) then yields

$$x * y = \frac{1}{p_1} + 0 + 0 + 0 + \frac{1}{p_1^2} \cdot \frac{1}{p_0^3} + 0 = \frac{1}{p_1} + \frac{1}{p_1^2 p_0^3}.$$

Since all elements of $\Omega(w,s)$ start and finish with a success by their definition in (4.1), the same initial $1/p$ term will be present for any pair of ending patterns. Whether or not there are higher order terms depends on the overlap of the successive substrings. In particular, for two ending patterns $x, y \in \Omega(w,s)$, the quantity $x * y$ will be of order $1/p^s$ if and only if $x = y$.

Equipped with these definitions, we now give the formula for the expected waiting time and the ending pattern distribution.

Theorem 4.1. *Let $N := |\Omega(w,s)|$. After enumerating the ending patterns as $\Omega(w,s) \equiv \{x^{(i)} : i = 1, \ldots, N\}$, let*

$$\vec{v} = \begin{pmatrix} \mathbb{E}(\tau_{(w,s)}) \\ \mathbb{P}\left(X = x^{(1)}\right) \\ \mathbb{P}\left(X = x^{(2)}\right) \\ \vdots \\ \vdots \\ \mathbb{P}\left(X = x^{(N)}\right) \end{pmatrix}. \tag{4.22}$$

Then

$$A\vec{v} = \vec{e}_1, \tag{4.23}$$

where $\vec{e}_1 := (1, 0, \ldots, 0)^T$ is a vector of length $N+1$, and

$$A := \begin{pmatrix} 0 & 1 & 1 & \cdots & 1 \\ -1 & x^{(1)} * x^{(1)} & x^{(1)} * x^{(2)} & \cdots & x^{(1)} * x^{(N)} \\ -1 & x^{(2)} * x^{(1)} & x^{(2)} * x^{(2)} & \cdots & x^{(2)} * x^{(N)} \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ -1 & x^{(N)} * x^{(1)} & \cdots & & x^{(N)} * x^{(N)} \end{pmatrix}. \tag{4.24}$$

The matrix A is invertible since no element of $\Omega(w,s)$ contains another [150]. The solution for $\vec{v}$ therefore always exists and is unique. One may then solve the linear system (4.23) to obtain $\mathbb{E}(\tau_{(w,s)})$ and the ending pattern distribution. Code is provided in [154] that performs this task. The matrix A is fully determined by the success probability p and the parameters w and s. From (4.3), the dimension of the square matrix A is $|\Omega(w,s)|+1=\mathcal{O}(w^{s-1})$, and so the complexity of this task is increasingly difficult for large w and s. In the following sections, we derive results that aid the understanding and computation of the corresponding results in the two characteristic regimes of large w, and small p. Moreover, by the definition of the star product as given in (4.20), we see that each entry of A is a polynomial in $1/p$ and $1/(1-p)$. As discussed previously, each entry of the submatrix formed by removing the first row and column is greater than or equal to $1/p$, due to properties of the operator $*$. Entries that do not take this exact value contain higher-order terms in $1/p$, due to the fact that there is a greater overlap of the ending patterns corresponding to the row and column indices of such an entry. The entries of the highest power in $1/p$ are exactly the diagonal elements and are of order s, because a string overlaps completely with itself and contains s successes. We note that in principle, the solutions for $\vec{v}$ can be computed analytically as functions of p by inverting A directly. However, due to the scaling of the system with w and s, this is in practice computationally laborious.

4

4.3.3. APPROXIMATING WITH AN INFINITE WINDOW

Now, one might ask: how large must the window size be for the approximation $w=\infty$ to be accurate? This is desirable due to the simple analytical form of the results for the distributions of $\tau_{(\infty,s)}$ and $X_{(\infty,s)}$, as seen in Section 4.3.1. This is in contrast to the solutions presented in Section 4.3.2 for the case of a finite w, which scale with w and s. The approximation becomes valid when the window size has 'saturated' the process, so that increasing the window size does not provide any significant improvement for the rate. Alternatively, the approximation becomes accurate when $\mathbb{P}(\tau_{(w,s)}>w)$ is small. This intuition is formalised with the following theorem.

Theorem 4.2. *Let $\tau_{(w,s)}$ be the waiting time for s successes in a w-window. Let $X_{(w,s)}$ be the corresponding ending pattern. Let p denote the success probability of each trial. Let $\epsilon(w,s,p):=\mathbb{P}(\tau_{(w,s)}>w)$. Suppose that $0<p<1$ and $w<\infty$. Then*

$$\frac{\mathbb{E}(\tau_{(w,s)})-\mathbb{E}(\tau_{(\infty,s)})}{\mathbb{E}(\tau_{(w,s)})}<\epsilon(w,s,p) \tag{4.25}$$

and

$$\sum_{x\in\Omega(\infty,s)}|\mathbb{P}(X_{(w,s)}=x)-\mathbb{P}(X_{(\infty,s)}=x)|<2\epsilon(w,s,p). \tag{4.26}$$

We now look to evaluate $\epsilon(w,s,p)$. Looking back at the identity (4.7) for the distribu-

tion of $\tau_{(\infty,s)}$, we have

$$\mathbb{P}(\tau_{(w,s)} > w) = \mathbb{P}(\tau_{(\infty,s)} > w) \tag{4.27}$$

$$= \sum_{n=w+1}^{\infty} \binom{n-1}{s-1} (1-p)^{n-s} p^s. \tag{4.28}$$

To evaluate the right-hand side of (4.25) it is convenient to rewrite this as a finite sum, as provided by the following lemma. The proof of this is given in Appendix 4.6.3.

Lemma 4.1. *Let $\tau_{(w,s)}$ be the waiting time for s successes in a w-window, as defined in (4.4). Suppose that $0 < p < 1$ and $w < \infty$. Then*

$$\mathbb{P}(\tau_{(w,s)} > w) = \sum_{i=0}^{s-1} \binom{w}{i} (1-p)^{w-i} p^i. \tag{4.29}$$

This sum is simple and efficient to evaluate for constant s, and can then be used to find a range of w for which the two expectations are close. For example, if one is interested in evaluating $\mathbb{E}(\tau_{(w,s)})$, when in fact w is large enough such that it may be reasonably approximated by $\mathbb{E}(\tau_{(\infty,s)}) = s/p$, it is possible avoid solving a large system of equations with the following method. Demanding some desired error δ, one may quickly compute

$$w^* = \min\{w : \epsilon(w,s,p) < \delta\}. \tag{4.30}$$

Then, for all $w \geq w^*$, one may approximate $\mathbb{E}(\tau_{(w,s)})$ with $\mathbb{E}(\tau_{(\infty,s)}) = s/p$ with accuracy on the order of δ. The same can be done with the ending pattern distribution: if one is interested in the expectation of some fidelity quantity $F(X_{(w,s)})$, one may also approximate $\mathbb{E}(F(X_{(w,s)}))$ with $\mathbb{E}(F(X_{(\infty,s)}))$ with the same accuracy.

4.3.4. ASYMPTOTIC BEHAVIOUR OF THE EXPECTATION

From (4.9), an upper bound for $\mathbb{E}(\tau_{(w,s)})$ is given by $\mathbb{E}(\tau_{(s,s)})$, which can be written in a simple analytical form. In the case $w = s$, there is only one ending pattern x, which corresponds to the case of s consecutive successes. From (4.23), we then have

$$\mathbb{E}(\tau_{(s,s)}) = x * x = \sum_{j=1}^{s} \frac{1}{p^j} = \frac{1/p^s - 1}{1-p}, \tag{4.31}$$

which for small p satisfies $\mathbb{E}(\tau_{(s,s)}) \sim 1/p^s$. In comparison, from (4.8), the scaling of the expectation for $w = \infty$ is reciprocal in p. Further, from the form of A given in (4.24), all entries of $\vec{v}$ will be ratios of polynomials in $1/p$. Looking at the first component of $\vec{v}$, which is the waiting time expectation, this tells us that there is some integer value $\alpha_s(w)$ which dominates the scaling for small p, so that

$$\mathbb{E}(\tau_{(w,s)}) \sim \frac{c(w,s)}{p^{\alpha_s(w)}}, \tag{4.32}$$

where $c(w,s)$ is a constant. Now, recalling from (4.9) that $\mathbb{E}(\tau_{(w,s)})$ is a decreasing function of w, we therefore expect the same of $\alpha_s(w)$, which satisfies $\alpha_s(s) = s$ and $\alpha_s(\infty) = 1$. Below we show that for $w < \infty$, $\alpha_s(w)$ is always equal to s, and also derive the scaling factor $c(w,s)$.

Theorem 4.3. *Let $\tau_{(w,s)}$ be the waiting time for s successes in a w-window, as defined in (4.4). Let $X_{(w,s)}$ be the corresponding ending pattern. Let p be the success probability of the process. Then, in the limit $p \to 0$,*

$$\mathbb{E}\left(\tau_{(w,s)}\right) \sim \frac{1}{|\Omega(w,s)|p^s}, \tag{4.33}$$

and for all $x \in \Omega(w,s)$

$$\mathbb{P}(X_{(w,s)} = x) \to \frac{1}{|\Omega(w,s)|} \tag{4.34}$$

where $|\Omega(w,s)| = \binom{w-1}{s-1}$ is the number of possible ending patterns.

A proof of Theorem 4.3 is given in Appendix 4.6.3. It is interesting future work to quantify the speed of convergence of (4.33) and (4.34).

As intuition for (4.34), note that for very small p, the probability of having w failures preceding the ending pattern is high. In this case, the ending pattern distribution is equivalent to the ending pattern distribution given we succeed in w attempts, which in the limit of small p converges to the uniform distribution.

The behaviour captured by Theorems 4.2 and 4.3 may be viewed as two limiting behaviours of the problem in the regimes of small and large p, respectively. In particular, we expect that the formula provided by Theorem 4.1 becomes useful in neither regime, i.e. when p is neither too small or too large to apply either approximation. Moreover, it is important to keep in mind that such a regime will depend on the choices of w and s.

4.4. ILLUSTRATION AND APPLICATION

We expect the methods presented in the above sections to be useful in choosing the optimal window size for a quantum protocol. To this end, we firstly analyse in more detail the behaviour of $\mathbb{E}(\tau_{(w,s)})$. We then demonstrate how these methods may be used to optimise the performance of a BQC protocol.

4.4.1. ILLUSTRATION

We fix $s = 4$ as an example to showcase the characteristic behaviours of the expected waiting time. From our investigations, the solutions for other values of s display the same qualitative behaviour. To produce each figure, we compute $\mathbb{E}(\tau_{(w,s)})$ by numerically solving the linear system (4.23) for the specific choices of w, s and p. Recall that the size of this linear system scales as $\mathcal{O}(w^{s-1})$. The value $s = 4$ is small enough so that for the w values that we consider, the complexity of the problem is not too large to be solved on a laptop.

In Figure 4.3, $\mathbb{E}(\tau_{(w,4)})$ is plotted against w, with the success probability set to $p = 0.5$. We notice the convergence to the $w = \infty$ lower bound. The grey region is that given by one standard deviation above and below the expectation. Note that one also expects the standard deviation to converge to that of $\tau_{(\infty,4)}$, which is given in closed form by (4.11).

In Figure 4.4, $\mathbb{E}(\tau_{(w,4)})$ is again plotted against w, but this time for three different values of the success probability. In each case, the solution again approaches the corresponding $w = \infty$ lower bound. Each line starts at $\mathbb{E}(\tau_{(4,4)}) = (1/p^s - 1)/(1-p)$, corresponding to $w = s$, and converges to the $w = \infty$ limit. This convergence is an important

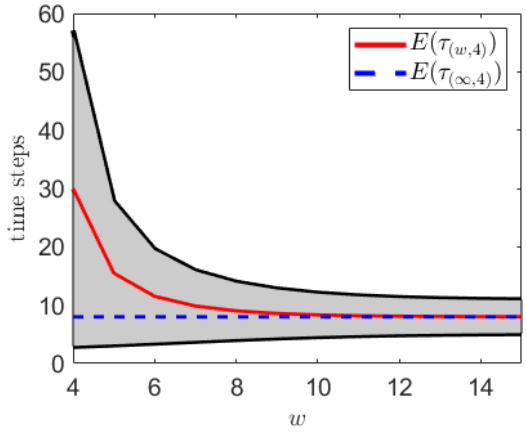


Figure 4.3: **How $\mathbb{E}(\tau_{(w,4)})$ varies with w.** We see that $\mathbb{E}(\tau_{(w,4)})$ (red line) converges to the lower bound $\mathbb{E}(\tau_{(\infty,4)}) = 4/p$ (blue line) as w becomes large. The grey region is one standard deviation of $\tau_{(w,4)}$ above and below its expectation (for $w < \infty$). All quantities are evaluated with a success probability $p = 0.5$.

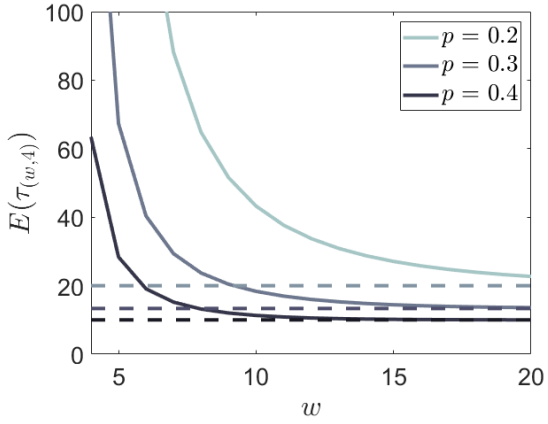


Figure 4.4: **How $\mathbb{E}(\tau_{(w,4)})$ varies with w and p.** We see that for larger p, $\mathbb{E}(\tau_{(w,4)})$ (solid line) approaches the lower bound $\mathbb{E}(\tau_{(\infty,4)}) = 4/p$ (dashed line) at a higher rate.

feature, because at some point increasing w provides no significant improvement for the protocol rate. As one would expect intuitively, the convergence occurs more quickly for a larger p, as increasing the window size effectively saturates the problem more easily. To quantify this, we can use the arguments of Section 4.3.3. For example, taking the desired margin of error to be 2%, define

$$w^* = \min\{w : \epsilon(w, s, p) < 0.02\}, \tag{4.35}$$

where $\epsilon(w, s, p)$ is given by (4.29). By Theorem 4.2 and Lemma 4.1, the approximation $\mathbb{E}(\tau_{(w,s)}) \approx \mathbb{E}(\tau_{(\infty,s)})$ is then valid to the same margin of error for all $w > w^*$. It is inter-

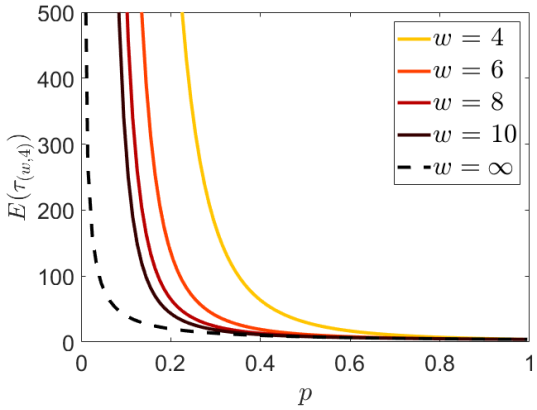


Figure 4.5: **How $\mathbb{E}(\tau_{(w,4)})$ varies with p.** We see that $\mathbb{E}(\tau_{(w,4)})$ (solid lines) demonstrates the reciprocal scaling as $p \to 0$, as encapsulated by Theorem 4.3. There is convergence to $\mathbb{E}(\tau_{(w,4)})$ (dashed line). This plot was made by discretising p into 100 points, evenly spaced in the range $(0,1)$.

esting to see how this compares to the smallest window size w^*_{true} for which the same approximation can be made, which is defined formally as

$$w^*_{\text{true}} = \min\left\{w : \frac{\mathbb{E}(\tau_{(w,s)}) - \mathbb{E}(\tau_{(\infty,s)})}{\mathbb{E}(\tau_{(w,s)})} < 0.02\right\}. \tag{4.36}$$

The value w^* is then an upper bound for w^*_{true}. For example, letting $p = 0.5$ and $s = 4$ yields $w^* = 15$, and checking with the exact solutions gives $w^*_{\text{true}} = 12$. These are plotted for more values of p in Figure 4.6. We see from the plot that as p increases, the bound appears to become tighter. The value w^*_{true} was plotted for only a few select values of p because its calculation is computationally intensive.

In Figure 4.5, $\mathbb{E}(\tau_{(w,4)})$ is plotted against p for five different values of the window size. One can see that the scaling of $\mathbb{E}(\tau_{(w,4)})$ occurs more slowly for a larger w. This indicates the reciprocal behaviour as given in (4.33), where $\mathbb{E}(\tau_{(w,4)}) \sim 1/|\Omega(w,4)|p^4$, and in the case of a larger w the constant $|\Omega(w,4)|$ suppresses the scaling. As $p \to 1$, all plots simply converge to $s = 4$, because in this case the process is deterministic. Further, we see again the convergence of the expectation to the infinite window limit. If p becomes large, we expect the problem to 'saturate' in the same sense as before, so that $\mathbb{E}(\tau_{(w,s)}) \approx \mathbb{E}(\tau_{(\infty,s)})$. The speed of this convergence can again be quantified using the results of Section 4.3.3. Demanding the same error of 2%, we take the p^* that satisfies

$$p^* = \inf\left\{p : \epsilon(w,s,p) < 0.02\right\}, \tag{4.37}$$

or equivalently, p^* is the unique value of p such that $\epsilon(w,s,p^*) = 0.02$. The value p^* is an upper bound for the true threshold p^*_{true},

$$p^*_{\text{true}} := \inf\left\{p : \left(\frac{\mathbb{E}(\tau_{(w,s)}) - \mathbb{E}(\tau_{(\infty,s)})}{\mathbb{E}(\tau_{(w,s)})}\right)\Big|_p < 0.02\right\}. \tag{4.38}$$

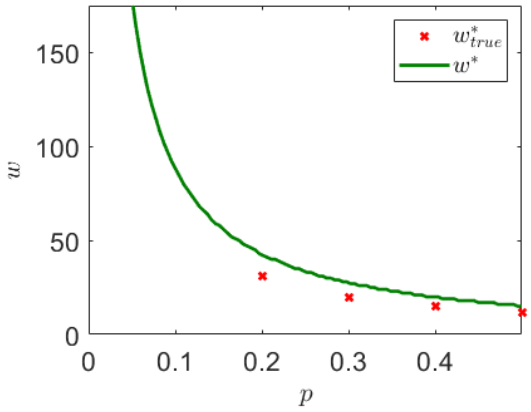


Figure 4.6: **Comparison of thresholds on w for the infinite window approximation**. One may use w^* (green line) as a threshold, which is more easily computable than w^*_{true} (red cross). See (4.35) and (4.36) for the definition of these quantities. Here we assume a desired error of 2%.

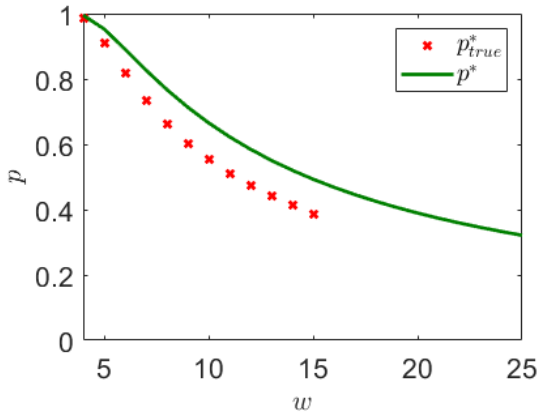


Figure 4.7: **Comparison of thresholds on p for the infinite window approximation**. One may use p^* (green line) as a threshold, which is more easily computable than p^*_{true} (red cross). See (4.37) and (4.38) for the definition of these quantities. Here we assume a desired error of 2%.

where we now include dependence of the expectations on the success probability p. In Figure 4.7, p^* and p^*_{true} are plotted against w. The value p^*_{true} is computationally intensive to find for large values of w, and has therefore only been plotted for selected small values of w. The bound p^*, however, is efficient to compute. We observe that the bound appears to be tighter for smaller w.

4.4.2. APPLICATION TO A BQC PROTOCOL

In the following, we provide an example of how the results from Section 4.3 may be used in the performance analysis of a quantum network application. We consider a verifi-

able Blind Quantum Computation (BQC) protocol [147]. This involves a client, who uses a more powerful server device to carry out a bounded-error quantum polynomial-time (BQP) computation [11], which is specified in the measurement-based formalism [20]. In this formalism, the computation is defined with respect to a graph $G = (V, E)$, where V is the set of vertices and E the set of edges. The computation is performed by firstly creating a graph state corresponding to G, and then applying a series of measurements ('measurement flow') to a subset of qubits. The BQC protocol is designed such that the server remains ignorant of the client's desired computation (blindness). Further, it ensures that the client can validate that the outcome is correct, even in the presence of some amount of noise or a malicious server (veriability). These properties are stated precisely in terms of the composable security properties of the protocol [155]. For the protocol in full detail, we refer to [147]. Here, we provide a short outline of the BQC protocol, and a simple model of how it is carried out. We then apply the results of Section 4.3 to study the performance of the protocol.

PROTOCOL FEASIBILITY

The BQC protocol involves a series of rounds. In each round, the client sends $|V|$ qubits to the server, and also a description of the measurement flow it should carry out. If the server is honest, it will then create a graph state by applying entangling gates corresponding to edges in E, carry out the corresponding measurement flow, and send the measurement outcomes back to the client.

The protocol involves interweaving two types of rounds: *computation* and *test* rounds. The computation rounds are used to carry out the client's desired computation. In these rounds, the computation measurement flow is encrypted in order to maintain blindness. The function of the test rounds is to check for deviations from the client's specified operations. Deviations could be due to noise, or the server being malicious. Each test round has the outcome of either pass or fail, and the protocol is aborted if the ratio of failed test rounds lies above a certain threshold.

Assuming the test round outcomes are i.i.d., the sufficient condition for verifiability that we will consider is given by

$$p_{\text{av}} < \frac{2\gamma - 1}{k(2\gamma - 2)}, \tag{4.39}$$

as shown in [156]. Here, p_{av} is the average probability of failure of a test round, and γ is the inherent error probability of the BQP computation. The value k is an integer and is corresponding to the k-colouring chosen by the client. This is a partition of the set of vertices into k subsets, known as colours, such that there is no edge between two vertices of the same colour. For the relevance of this to the BQC protocol, see Appendix 4.6.5 for a description of test rounds. For deterministic computations ($\gamma = 0$), (4.39) simplifies to

$$p_{\text{av}} < \frac{1}{2k}. \tag{4.40}$$

When the server is honest, the quantity p_{av} is determined on the amount by noise, which could for example arise from imperfect local operations and measurements, or imperfect memory in the server. Further, in a networked setting where the client and server are distantly separated, the client may send its qubits to the server by making use of

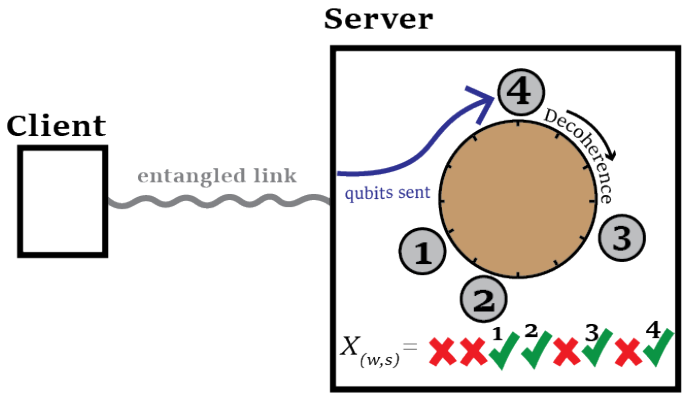


Figure 4.8: **The scenario considered for the model of BQC.** Generation of the entangled link is attempted sequentially, with success probability p. Upon success, the entangled link is immediately used to transmit qubits from the client to the server. While in the server, qubits (numbered grey circles) undergo decoherence (brown clock). Qubits are discarded from the server after they have existed for w time steps.

entanglement that has been established between the two parties. In this way, the performance of the protocol is directly dependent on properties of the quantum network architecture connecting client and server. In such architectures, however, there is in general a trade-off between rate and quality. In the case of this BQC protocol, demanding that the condition (4.39) is met then effectively places an upper bound on the rate of the protocol. In the following, we consider a simple model of the network and device architectures, and provide a demonstration of how the methods presented in Section 4.3 may be used to find architecture parameters that maximise the protocol rate, given the constraint (4.39).

MODEL OF NETWORK ARCHITECTURE

Our model of the quantum network architecture on which the BQC protocol is carried out is summarised in the bullet points below. A depiction is in Figure 4.8.

- The server is honest, meaning that it carries out all tasks specified by the client. The BQC protocol protects against malicious server activity, as well as being robust to noise. Here, we solely aim to quantify the effect of noise on the protocol.
- Entanglement generation between client and server is performed with sequential attempts. Each attempt succeeds with probability p.
- Upon entanglement success, a qubit transmission procedure takes place. We assume that each qubit comes into existence in the server memory at the end of the corresponding time step.
- Immediately after transmission, each qubit is established with fidelity $F_{\text{est}}(p)$, where $F_{\text{est}} : [0,1] \to [0,1]$ is a decreasing function. In this way, we include a trade-off between rate and fidelity that is inherent to the entanglement generation process occurring between client and server. In the following, we choose this to be $F_{\text{est}}(p) = 1 - \lambda p$. Motivation for this choice of F_{est} is given in Appendix 4.6.4.

- While they are stored in the server, qubits are subject to depolarising noise with a memory lifetime of T time steps. For a d-dimensional density matrix $\rho \in \mathscr{D}(\mathscr{H}_d)$, this has action

$$\rho \to e^{-\frac{t}{T}}\rho + \left(1 - e^{-\frac{t}{T}}\right)\frac{\mathbb{I}_d}{d}, \tag{4.41}$$

  where $\mathbb{I}_d$ is the d-dimensional identity matrix and t is the number of time steps for which ρ has existed at the server. For the case of a qubit, i.e. $d = 2$, the fidelity then decays as

$$F_{\text{est}} \to \left(F_{\text{est}} - \frac{1}{2}\right)e^{-\frac{t}{T}} + \frac{1}{2}. \tag{4.42}$$

- To reduce decoherence, the server discards a qubit once it has been in memory for w time steps.

- All local operations and measurements by the client and server devices are perfect and instantaneous. In particular, once all qubits required for the round are present in the server, it immediately and perfectly applies the measurement flow that has been specified by the client.

- Before each round, the client chooses an element of V uniformly at random. The corresponding qubit is the first one sent. The client then cycles through the qubits from V in some pre-defined order. With this added randomness, the resulting order of the qubit ages will appear completely random. We continue with this assumption because it simplifies the resulting calculation of p_{av}, by removing any dependence of the qubit ages on events that occurred beyond the last w time steps. More details of protocol test rounds are given in Appendix 4.6.5.

In our model, then, the fidelity of a qubit in the server depends only on the amount of time it has been stored there, and the entanglement generation success probability p. Notice that our set-up consists of the sequential attempted establishment of qubits at the server, and the discarding of these qubits after they have existed for a pre-defined number of time steps. We then have a situation analogous to that considered in the first part of this work, where the qubits function as the corresponding quantum resources. The methods given in Section 4.3 can then be applied to study this situation: the time taken to complete a round is $\tau_{(w,s)}$ time steps, where $s = |V|$ is the number of qubits required to produce a graph state, and $\tau_{(w,s)}$ is the waiting time as defined in Section 4.2. Furthermore, the qubit fidelities at the time when the server applies its entangling gates and measurements are determined by the ending pattern $X_{(w,s)}$ which finishes the process. More specifically, it is possible to calculate p_{av} exactly using the ending pattern distribution. We briefly describe this now.

Suppose that during a particular test round, at the time the server will carry out its local operations and measurements, the fidelities of the server qubits are $\vec{F} = (F_1, F_2, ..., F_{|V|})$. Then, given the model described in the previous section, it is possible to find a function that tells us the probability of error of a test round, $P_G(\vec{F})$. This is a polynomial in the values F_i, and has a form dependent on the graph G and the choice of k-colouring. The details of how to obtain this function are given in Appendix 4.6.5. An expression for the

average probability of error of a test round is then

$$p_{\text{av}} = \sum_{\vec{F}} \mathbb{P}(\vec{F}) P_G(\vec{F}), \tag{4.43}$$

where $\mathbb{P}(\vec{F})$ is the probability of obtaining the particular fidelity vector $\vec{F}$. Note that in the model introduced in the previous section, the qubit fidelities $\vec{F}$ are determined by the amount of time for which the qubits have been stored in the server. Moreover, recall that the ages of the links are contained exactly in the ending pattern $X_{(w,s)}$. Writing this dependence as

$$\vec{F} = \vec{F}(X_{(w,s)}), \tag{4.44}$$

we then rewrite (4.43) to obtain an expression for the average probability of test round failure,

$$p_{\text{av}} = \sum_{x \in \Omega(w,s)} \mathbb{P}(X_{(w,s)} = x) P_G\left(\vec{F}(x)\right). \tag{4.45}$$

This is a quantity that we can now evaluate using the methods introduced in Section 4.3. In this way, the tools from Section 4.3 allow for the direct connection between the feasibility of the BQC protocol, as determined by p_{av}, to its rate. Since the above formula for p_{av} is dependent on the graph structure and k-colouring, some parameter regimes may be sufficient for some calculations but not others. For example, for more complicated graphs that require a larger k, the condition (4.39) is more strict. Further, if one chooses a different graph or k-colouring for the calculation, the polynomial P_G may differ.

NUMERICAL EVALUATION

We now aim to find optimal values of the architecture parameters p and w for one round of the protocol. By *optimality*, we mean that the expected time taken to carry out a round is minimised, while ensuring that the protocol is still feasible. Note that this does not necessarily mean optimality for the full protocol, which is comprised of multiple rounds. To optimise over the full protocol, one would to do a further optimisation over more protocol parameters (for example, the ratio of computation and test rounds), which we not not consider in this work.

There is a combination of trade-offs between rate and fidelity present in our scenario: firstly due to varying the success probability, and secondly due to varying the window size. An increase in p increases the rate at which successful links are generated, but decreases the initial fidelity of qubits in the server by an amount determined by $F_{\text{est}}(p)$. We would therefore expect that a smaller value of w is required to minimise decoherence at the server, to ensure that the condition (4.39) is met. This in turn increases the expected time taken to generate all necessary entangled links within the time window. More formally, given a fixed p, we may find the minimal expected time for one round with the following procedure.

1. Find the maximum value of w such that the protocol is still feasible for this value of p,

$$w_{\max}(p) := \max\left\{ w : p_{\text{av}} < \frac{2\gamma - 1}{k(2\gamma - 2)} \right\}. \tag{4.46}$$

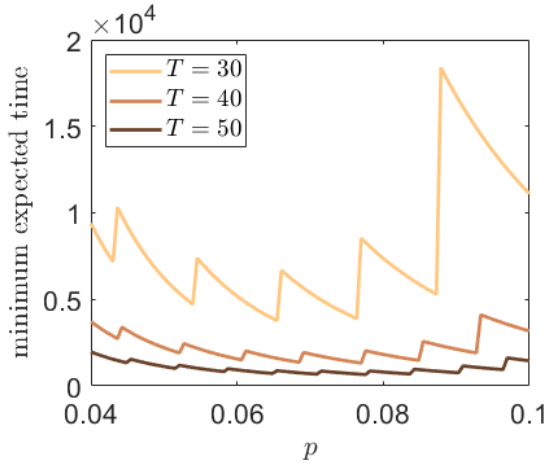


Figure 4.9: **Minimum expected time for one round of a BQC calculation vs. success probability of entanglement generation.** For a given value of p, we find the maximum window size (4.46), and then use that to compute the minimum expected waiting time. Here, p is discretised into 100 values of p that are evenly spaced in the range $[0.04, 0.1]$. We assume calculations on a square graph, which requires four entangled links to be produced within a time window.

2. Compute $\mathbb{E}\left(\tau_{(w_{\max}(p),s)}\right)\big|_p$. This is the minimum expected time for one round.

As an example of this method put into practice, we consider the case where the client would like to perform a BQP calculation on a square graph, so that $|V| = 4$. This requires $s = 4$ entangled pairs to be produced within the time window. For simplicity we will consider deterministic computations, so that the requirement on the probability of error is $p_{\text{av}} < 1/2k$. In this case, k can be chosen to be 2 (see Appendix 4.6.5 for an example of a 2-colouring of a square graph), and the sufficient condition becomes $p_{\text{av}} < 1/4$.

In Figure 4.9, the minimum expected time to carry out a round is plotted against p for three different values of the memory lifetime parameter T, and $F_0(p) = 1 - \lambda p$. The code used to produce Figures 4.9 and 4.10 is provided in [154]. We choose $\lambda = \frac{1}{2}$ in order to best display the behaviour of the solution, given our computational resources. In particular, the range of p that we plot is chosen to clearly show the region of the optimal combination of the two trade-offs. For small p, the expected waiting time is high due to the small entanglement generation probability. For large p, it is high due to the small window size required due to the decrease in F_{est}. We therefore see a region in the middle of the plot where the average waiting time is minimal, or equivalently, the rate at which rounds can be carried out is maximal. For larger T, the decoherence of qubits in memory is reduced, and so it is possible to have a larger window size without disrupting the condition on p_{av}. We thus see that the expected time for one round decreases with T. Further, there are sharp peaks in the plots for each T, which are due to the discrete nature of w. This can be explained as follows: in the middle of two peaks, it is possible to increase the value of p without disrupting the condition (4.39). However, there will come a point where this condition is in fact an equality, which is when the window cut-off must be decreased in order to maintain the minimum quality of qubits in the server.

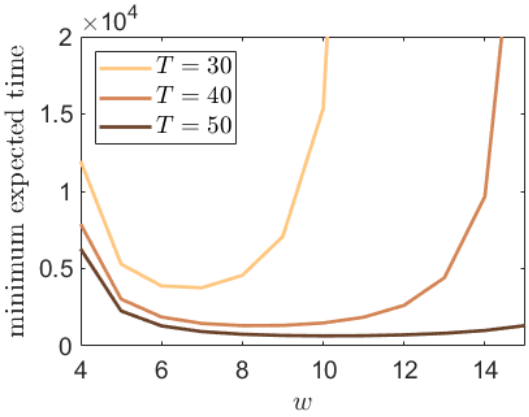


Figure 4.10: **Minimum expected time for one round of a BQC calculation vs. window size.** For each window size $4 \leq w \leq 15$, we find the maximum success probability (4.47), and then use this to compute the expected waiting time. We assume calculations on a square graph, which requires four entangled links to be produced within a time window.

Since w is a discrete parameter, this causes a jump to a higher expected time.

One can do something similar when varying the window size. Given a fixed w, we find the minimum expected time for one round with the following steps.

1. Find the maximum value of p such that the protocol is still feasible,

$$p_{\max}(w) \coloneqq \sup\left\{p : p_{\text{av}} < \frac{2\gamma - 1}{k(2\gamma - 2)}\right\}. \tag{4.47}$$

2. Compute $\mathbb{E}\left(\tau_{(w,s)}\right)\big|_{p_{\max}(w)}$.

In Figure 4.10, the minimum expected time to carry out a round is plotted against the window size. This is again in the case of a square graph, for the same three values of the memory lifetime parameter T. We see a similar behaviour as when varying the success probability: a smaller w induces a larger expected time. When s is larger, qubits are subject to more decoherence, and in order to keep condition (4.39) it is necessary to decrease the success probability. This is what induces a larger waiting time for larger w. We therefore again see an optimal region of w for which the expected time to carry out one round of the protocol is minimised.

Finally, we note that in practice, in order to optimise the full BQC protocol, one would need to consider how other aspects of the set-up, such as hardware, architecture and protocol, affect the performance. The simple scenario chosen in this work was to highlight the application of the results of Section 4.3. We see from Figures 4.9 and 4.10 that for such values of T and s, the methods from Section 4.3 enable one to make a careful choice of (w, p) that can improve the rate of rounds of the protocol by two or three times, in comparison to other non-optimal choices of (w, p) that are also sufficient for protocol feasibility.

4.5. FURTHER DIRECTIONS

With the methods presented in this work, we focus on computing both the first and second moments of $\tau_{(w,s)}$, and the full distribution of $X_{(w,s)}$. We have seen that for w finite and $s > 2$, the formulae given here to compute $\mathbb{E}(\tau_{(w,s)})$ and the distribution of $X_{(w,s)}$ are in the form of linear systems that scale as $|\Omega(w,s)|+1$. If one would like to compute the full ending pattern distribution, then this seems to be a good scaling, since the outcome is comprised of $|\Omega(w,s)|$ probabilities. However, if one is for example only interested in $\mathbb{E}(\tau_{(w,s)})$ (e.g. for computing a protocol rate), then for certain regimes of w and p it may be useful to consider a continuous approximation, where the time between successful resource generation attempt is exponentially distributed. Such a case is often considered in the scan statistics literature (for example, see [148]). However, how to study the ending pattern distribution in the continuous case is not immediately clear.

We also note that a useful tool of approximation would be to further understand the asymptotic scaling highlighted by Theorem 4.3. More specifically, it would be interesting to know exactly how fast is the approach of (4.33) and (4.34), in terms of s and w.

In the set-up of the problem, one could also consider a more realistic model of a quantum network architecture. For example, there may be parameter drift, when the success probability decreases over time due to increased noise. Further, in the more general case where the sequential attempts are not necessarily independent but Markovian, methods similar to those used in this chapter may again be applied to the problem - see [149], for example.

4.6. APPENDIX

4.6.1. IDENTITIES FOR THE CASE OF TWO RESOURCE STATES

EVALUATION OF $\mathbb{E}(\tau_{(w,2)})$

Evaluation of $\mathbb{E}(L)$. Recalling from (4.13) the distribution of L, we have

$$\begin{aligned}
\mathbb{E}(L) &= \sum_{n=1}^{w-1} \frac{n(1-p)^{n-1}p}{1-(1-p)^{w-1}} \\
&= \frac{p}{1-(1-p)^{w-1}} \sum_{n=1}^{w-1} n(1-p)^{n-1} \\
&= \frac{p}{1-(1-p)^{w-1}} \cdot -\frac{\mathrm{d}}{\mathrm{d}p} \sum_{n=1}^{w-1} (1-p)^n \\
&= \frac{p}{1-(1-p)^{w-1}} \cdot -\frac{\mathrm{d}}{\mathrm{d}p} \frac{1-(1-p)^w}{1-p} \\
&= \frac{p}{1-(1-p)^{w-1}} \cdot \frac{1-(1-p)^w - wp(1-p)^{w-1}}{p^2} \\
&= \frac{1-(1-p)^w - wp(1-p)^{w-1}}{p(1-(1-p)^{w-1})},
\end{aligned}$$

where to evaluate the sum we have used the identity for a geometric series. □

Proof that $\mathbb{E}\left(\sum_{j=1}^{M} T_j\right) = \mathbb{E}(M)\mathbb{E}(T_1)$. This is used to evaluate the expectation $\mathbb{E}(\tau_{(w,2)})$. The

random variables M and $\{T_j\}$ are independent, and since the $\{T_j\}$ are identically distributed,

$$\begin{aligned}\mathbb{E}\left(\sum_{j=1}^{M} T_j\right) &= \sum_{m=1}^{\infty} \mathbb{E}\left(\sum_{j=1}^{m} T_j\right)\mathbb{P}(M=m)\\ &= \sum_{m=1}^{\infty} \mathbb{E}(T_1)\cdot m\mathbb{P}(M=m)\\ &= \mathbb{E}(T_1)\mathbb{E}(M).\end{aligned}$$

□

EVALUATION OF VAR($\tau_{(w,2)}$)

Recall M, T_j and L, as given in Section 4.3.2. These are independent, M and T have distributions $M \sim \text{Geom}(1-(1-p)^{w-1})$, $T_j \sim \text{Geom}(p)$, and L has distribution as given in (4.13). From (4.15), we have

$$\text{Var}(\tau_{(w,2)}) = \text{Var}\left(\sum_{j=1}^{M} T_j + (M-1)(w-1)\right) + \text{Var}(L), \tag{4.48}$$

since L is independent of M and T_j. Now, letting $\text{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$ be the covariance,

$$\begin{aligned}\text{Var}\left(\sum_{j=1}^{M} T_j + (M-1)(w-1)\right) = \text{Var}\left(\sum_{j=1}^{M} T_j\right) + \text{Var}\left((M-1)(w-1)\right)\\ + 2\text{Cov}\left(\sum_{j=1}^{M} T_j, (M-1)(w-1)\right),\end{aligned} \tag{4.49}$$

where we have used the identity $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\cdot\text{Cov}(X,Y)$. We now evaluate (4.49) term by term. Firstly,

$$\begin{aligned}\mathbb{E}\left(\left(\sum_{j=1}^{M} T_j\right)^2\right) &= \sum_m \mathbb{E}\left(\left(\sum_{j=1}^{m} T_j\right)^2\right)\mathbb{P}(M=m)\\ &= \sum_m \mathbb{E}\left(\sum_{j=1}^{m} T_j^2 + \sum_{i\neq j} T_i T_j\right)\mathbb{P}(M=m)\\ &= \sum_m \left(m\mathbb{E}(T_1^2) + m(m-1)\mathbb{E}(T_1)^2\right)\mathbb{P}(M=m)\\ &= \mathbb{E}(M)\mathbb{E}(T_1^2) + \left(\mathbb{E}(M^2) - \mathbb{E}(M)\right)\mathbb{E}(T_1)^2.\end{aligned}$$

Subtracting $\mathbb{E}\left(\sum_{j=1}^{M} T_j\right)^2 = \mathbb{E}(M)^2\mathbb{E}(T_1)^2$ then yields

$$\text{Var}\left(\sum_{j=1}^{M} T_j\right) = \mathbb{E}(M)\text{Var}(T_1) + \text{Var}(M)\mathbb{E}(T_1)^2. \tag{4.50}$$

Secondly,

$$\mathrm{Var}\left((M-1)(w-1)\right) = (w-1)^2\mathrm{Var}(M). \tag{4.51}$$

Thirdly,

$$\mathrm{Cov}\left(\sum_{j=1}^{M} T_j, (M-1)(w-1)\right) = (w-1)\mathrm{Cov}\left(\sum_{j=1}^{M} T_j, M\right) \tag{4.52}$$

$$= (w-1)\left(\sum_m m^2\mathbb{E}(T_1)\mathbb{P}(M=m) - \mathbb{E}(M)^2\mathbb{E}(T_1)\right) \tag{4.53}$$

$$= (w-1)\mathrm{Var}(M)\mathbb{E}(T_1). \tag{4.54}$$

4

It now remains to evaluate $\mathrm{Var}(L)$. We firstly calculate

$$\mathbb{E}(L^2) = \sum_{n=1}^{w-1} \frac{n^2(1-p)^{n-1}p}{1-(1-p)^{w-1}} \tag{4.55}$$

$$= \mathbb{E}(L) + \frac{p}{1-(1-p)^{w-1}} \sum_{n=1}^{w-1} n(n-1)(1-p)^{n-1}. \tag{4.56}$$

Now,

$$\sum_{n=1}^{w-1} n(n-1)(1-p)^{n-1} = (1-p)\frac{\mathrm{d}^2}{\mathrm{d}p^2}\sum_{n=0}^{w-1}(1-p)^n$$

$$= (1-p)\frac{\mathrm{d}^2}{\mathrm{d}p^2}\left(\frac{1-(1-p)^w}{p}\right),$$

hence

$$\mathrm{Var}(L) = \frac{p(1-p)}{1-(1-p)^{w-1}}\frac{\mathrm{d}^2}{\mathrm{d}p^2}\left(\frac{1-(1-p)^w}{p}\right) + \mathbb{E}(L) - \mathbb{E}(L)^2. \tag{4.57}$$

We are now equipped to compute the full variance of $\tau_{(w,2)}$,

$$\mathrm{Var}(\tau_{(w,2)}) = \mathbb{E}(M)\mathrm{Var}(T_1) + \mathrm{Var}(M)\mathbb{E}(T_1)^2$$

$$+ 2(w-1)\mathrm{Var}(M)\mathbb{E}(T_1) + (w-1)^2\mathrm{Var}(M) + \mathrm{Var}(L), \tag{4.58}$$

where one may find a closed-form expression by inputting the standard identities for a geometric random variable, which are

$$\mathbb{E}(T_1) = \frac{1}{p} \tag{4.59}$$

$$\mathrm{Var}(T_1) = \frac{1-p}{p^2} \tag{4.60}$$

$$\mathbb{E}(M) = \frac{1}{1-(1-p)^{w-1}} \tag{4.61}$$

$$\mathrm{Var}(M) = \frac{(1-p)^{w-1}}{(1-(1-p)^{w-1})^2}. \tag{4.62}$$

4.6.2. ENDING PATTERN DISTRIBUTION AND WAITING TIME MOMENTS FOR A FINITE WINDOW

THE WAITING TIME IS WELL-DEFINED

We show here that τ_x can be bounded above by a geometrically distributed random variable. Using the notation $p_1 = p$, $p_0 = 1-p$ for an ending pattern $x \in \Omega_l(s)$, this exact sequence will appear in any given l consecutive trials $Z_{i-l+1},...,Z_i$ with probability $\gamma_x := p_{x_1}...p_{x_l}$. Defining a new sequence of random variables $(Y_n)_{n=1}^{\infty}$,

$$Y_n = \begin{cases} 1 \text{ if } Z_i = x_i \text{ for all } (n-1)l < i \leq nl; \\ 0, \text{ otherwise.} \end{cases} \tag{4.63}$$

Each Y_n is then Bernoulli with parameter γ_x. It takes the value 1 if the nth segment of l trials exactly matches with x. There is then an associated waiting time random variable $\tilde{\tau}_x$ that is geometric with parameter γ_x,

$$\tilde{\tau}_x := \min\{n : Y_n = 1\}. \tag{4.64}$$

Moreover, the waiting time to see x satisfies $\tau_x \leq \tilde{\tau}_x \cdot l$. Taking expectations yields

$$\mathbb{E}(\tau_{(w,s)}) \leq \mathbb{E}(\tau_x) \leq \mathbb{E}(\tilde{\tau}_x \cdot l) = \frac{l}{\gamma_x} < \infty, \tag{4.65}$$

which completes our proof. We note that the same method can be used to show that all moments of $\tau_{(w,s)}$ are finite.

THE EXPECTED WAITING TIME OF A SIMPLE PATTERN

Using the theory of martingales and a helpful gambling analogy to aid understanding, we now derive a way to numerically compute the ending pattern distribution $\{\mathbb{P}(x) : x \in \Omega(w,s)\}$, and the first and second moments of the waiting time $\tau_{(w,s)}$, in the case of a finite window size. The result of this is Theorem 4.1 in the main text. The method was introduced in [150], where they consider the more abstract case of a general sequence of discrete i.i.d random variables, and a general set of ending patterns. Here, due to its relevance to the subject of the main text, we continue with the case of i.i.d. Bernoulli trials.

It will be useful to first of all consider the case where we wait for an instance of a single pattern $x = (x_1,...,x_l) \in \{0,1\}^l$, instead of waiting for any instance of the set $\Omega(w,s)$. The former case is referred to as a *simple pattern* and the latter as a *compound pattern*. In this section we will find an exact expression for $\mathbb{E}(\tau_x)$. Here, τ_x refers to the waiting time until seeing the pattern x, and is defined in (4.5).

To provide intuition, we introduce the following scenario of gamblers in a casino. Suppose that just before the first trial is realised, a gambler, hereinafter referred to as Gambler 1, bets €1 on the outcome $\{Z_1 = x_1\}$. We also suppose that the odds are fair, so that if this is the case then she wins $€\frac{1}{p_{x_1}}$ [1]. Moreover, if she wins, then she straight away bets all of these winnings on the outcome $\{Z_2 = x_2\}$. If not, the casino keeps her €1 and

[1] i.e. the expected net gain of the gambler is zero. Calling this G, we can verify explicitly by writing $\mathbb{E}(G) = (1/p_{\lambda_1} - 1) \cdot p_{\lambda_1} + (-1) \cdot (1 - p_{\lambda_1}) = 0$.

she doesn't place any more bets. For a general n, Gambler 1 then proceeds at the nth trial in a similar way: if she has yet to lose, she bets all of her winnings on the outcome $\{Z_n = x_n\}$, and if not, she doesn't place any bet. Furthermore, at every trial we introduce a new gambler who behaves in exactly the same way, so that Gambler 2 bets €1 on the outcome $\{Z_2 = x_1\}$, and continues betting all of her winnings on the subsequent rounds being equal to the next entry of x, up until she loses a round. Gambler j bets €1 on the outcome $\{Z_j = x_1\}$ and continues with exactly the same strategy. The game stops when the sequence x first appears, which by definition is at the τ_xth trial.

Our aim now is to write down an expression for the combined net gain of the gamblers after the nth trial, for a general n. In order to do this concisely, we recall the definition (4.21) of the quantities $\delta_{(a,b)}$. Given a realisation $C_n := (c_1, ..., c_n)$ of the first n trials, the winnings of Gambler j after the nth round can then be written as

$$W^{(j)}(C_n) = \begin{cases} \delta_{(x_1,c_j)}\delta_{(x_2,c_{j+1})}...\delta_{(x_{n-j+1},c_n)} \\ \text{for } n-l+1 \leq j \leq n; \\ 0, \text{ otherwise.} \end{cases} \tag{4.66}$$

Here, we see that the functions $\delta_{(a,b)}$ allow us to elegantly write down a Gambler's winnings. With this in mind, the *combined* winnings of the gamblers after the n'th trial is

$$W(C_n) := \sum_{j=1}^{n} W^{(j)}(C_n) = \sum_{j=n-l+1}^{n} W_n^{(j)} \equiv x * C_n, \tag{4.67}$$

where we have introduced the function $*$ that was defined in (4.20), and that maps two binary strings to a scalar value. From (4.66), we see that the *net gain* of the jth gambler after the nth time-step is simply

$$G^{(j)}(C_n) := W^{(j)}(C_n) - 1, \tag{4.68}$$

and similarly, the *total net gain* of the gamblers after the nth trial is

$$G(C_n) := \sum_{j=1}^{n} G^{(j)}(C_n) = x * C_n - n. \tag{4.69}$$

We can now define a sequence of random variables $(G_n)_{n\geq 0}$,

$$G_n := G(C_n) \tag{4.70}$$

which take the value of the total net gain of the gamblers after each round. In particular, after the game ends, the total net gain is

$$G_{\tau_x} = x * x - \tau_x. \tag{4.71}$$

Note that $x * x$ is a quantity that is only dependent on the pattern x. Since the game is defined to be fair at every round, the expected total net gain when the game finishes would intuitively be equal to zero, i.e.

$$\mathbb{E}(G_{\tau_x}) = 0, \tag{4.72}$$

A neat expression for the expected waiting time to see the sequence B follows by making use of the linearity of expectation,

$$\mathbb{E}(\tau_x) = x * x. \tag{4.73}$$

To prove (4.72), we make use of the fact that $(G_n)_{n\geq 0}$ is a *martingale*, for which the following properties must hold:

(i) $\mathbb{E}(|G_n|) < \infty$.

(ii) $\mathbb{E}(G_{n+1}|G_n, ..., G_1) = G_n$

To show (i), we use the definition (4.71), and see that

$$\mathbb{E}(|G_n|) \leq x * C_n + \mathbb{E}(\tau_x) < \infty, \tag{4.74}$$

since the waiting time τ_x is well-defined, and $x * C_n$ is bounded. To show condition (ii), we use the fact that the game is fair at each round. Suppose that we have the maximum amount of information about what has happened in the first n trials, i.e. we know that they have taken the values $(c_1, ..., c_n)$. Then, the conditional expectation of G_{n+1} satisfies

$$\begin{aligned}\mathbb{E}(G_{n+1}|Z_1 = c_1, ..., Z_n = c_n) &= \sum_{c_{n+1}\in\{0,1\}} G(C_n, c_{n+1})\,\mathbb{P}(Z_{n+1} = c_{n+1}) \\ &= \sum_{j=1}^{n+1} \sum_{c_{n+1}\in\{0,1\}} G^{(j)}(C_n, c_{n+1})\, p_{c_{n+1}} \\ &= \sum_{j=1}^{n+1} G^{(j)}(C_n) = G_n,\end{aligned}$$

where to go to the final line, we have made use of the definition of $G^{(j)}$. Since the realisations of $(Z_1, ..., Z_n)$ completely determine the values of $G_1, ..., G_n$, this also shows (ii).

We now know that $(G_n)_{n\geq 0}$ is a martingale. However, this is not quite enough to show (4.72), which is what is required to obtain the final simple form for $\mathbb{E}(\tau_x)$. In particular, some extra machinery is needed, in the form of Doob's *optional stopping theorem*, a proof of which can be found in [157]. A version of this is stated below.

Theorem 4.4 (Optional stopping)**.** *Let G_n be a martingale and τ a stopping time. Suppose that there exists a constant K such that $|G_n - G_{n-1}| < K$ for all n. Suppose also that τ is a.s. finite. Then $\mathbb{E}(G_\tau) = \mathbb{E}(G_1)$.*

All that remains to be done is to show that the martingale defined in (4.71) satisfies the required properties to satisfy Theorem 4.4. Firstly, we have

$$|G_n - G_{n-1}| < x * C_n + x * C_{n-1} + 1 \leq K, \tag{4.75}$$

where

$$K = 2 \cdot \max_{C\in\{0,1\}^l}\{x * C\} + 1. \tag{4.76}$$

Secondly, we see that since τ_x is bounded above by a geometric random variable, it is a.s. finite. This gives us (4.72).

STARTING FROM ANOTHER PATTERN

We now adapt the results above in order to find the expected time to see x, given that we start already with some pattern y. We extend the gambling analogy in order to illustrate this concept, and suppose that we want to calculate the expected time until seeing y only *after* some number of rounds, m, say, have been realised. In particular, after the mth round we know the first m realisations of the i.i.d Bernoulli sequence, and we call these $y = (y_1, ..., y_m)$. At this point, the net gain of the gamblers is thus $G_m = x * y - m$. We will evaluate the net gain of the gamblers *compared* to this point after each of the n trials, which for $n \geq m$ we denote by $\tilde{G}_n$. This is simply given by

$$\tilde{G}_n = G_n - G_m = (x * C_n - n) - (x * y - m) \tag{4.77}$$

$$= x * C_n - x * y - (n - m), \tag{4.78}$$

where C_n is no longer completely general as its first m entries must correspond to y. Using the same reasoning as before, one can show that $(\tilde{G}_n)_{n\geq 0}$ is a martingale. Then, defining τ_{xy} as the waiting time to see x *given* that we have already seen pattern y, it is again possible to use Theorem 4.4 to show that

$$0 = \mathbb{E}(\tilde{G}_{\tau_{xy}}) = \mathbb{E}(x * x - x * y - \tau_{yx}),$$

and so by the linearity of expectations,

$$\mathbb{E}(\tau_{xy}) = x * x - x * y. \tag{4.79}$$

We may now use the results derived above to derive a formula for $\mathbb{E}(\tau_{(w,s)})$ and the distribution of $X_{(w,s)}$. Given $x \in \Omega(w, s)$, we write

$$\begin{aligned}\mathbb{E}(\tau_x) &= \mathbb{E}(\tau_{(w,s)}) + \mathbb{E}(\tau_x - \tau_{(w,s)}) \\ &= \mathbb{E}(\tau_{(w,s)}) + \sum_{y\in\Omega(w,s)} \mathbb{P}(X_{(w,s)} = y)\mathbb{E}(\tau_x - \tau_{(w,s)}|X_{(w,s)} = y) \\ &= \mathbb{E}(\tau_{(w,s)}) + \sum_{y\in\Omega(w,s)} \mathbb{P}(X_{(w,s)} = y)(x * x - x * y),\end{aligned}$$

Where we have noticed that $\mathbb{E}(\tau_x - \tau_{(w,s)}|X_{(w,s)} = y) = \mathbb{E}(\tau_{xy})$. Applying Theorem 4.4 and enforcing the condition that the ending pattern probabilities must sum to one then yields the formula (4.23).

FORMULA FOR THE SECOND MOMENT OF THE WAITING TIME

An extension to the gambling analogy given above can be used to derive the formula for the second moment of the waiting time, for which we refer to [149]. Here, we will only state the formula. We first of all define a new operation † that maps two elements $x, y \in \Omega(w, s)$ to a real number. If $x = (x_1, ..., x_k)$ and $y = (y_1, ..., y_m)$,

$$x \dagger y := \sum_{j=1}^{\min(k,m)} (1 - j) \prod_{i=1}^{j} \delta_{(x_i, y_{m-j+i})}. \tag{4.80}$$

Letting $\Omega \equiv \Omega(w, s)$, the second moment of $\tau_{(w,s)}$ can be found through solving the following systems

Theorem 4.5. *Let* $\{u_j\}_{1\le j\le|\Omega|}$ *and* $\{v_j\}_{1\le j\le|\Omega|}$ *solve the linear systems*

$$\sum_{j=1}^{|\Omega|} W_{ij}u_j = 1, \text{ for } 1\le i\le|\Omega|, \tag{4.81}$$

$$\sum_{j=1}^{|\Omega|} (N_{ij}u_j + W_{ij}v_j) = 1, \text{ for } 1\le i\le|\Omega| \tag{4.82}$$

with $W_{ij} := x^{(i)} * x^{(j)}$ *and* $N_{ij} := x^{(i)} \dagger x^{(j)}$. *Then,*

$$\mathbb{E}\left(\tau^2_{(w,s)}\right) = \frac{1+\left(1-\sum_j v_j - \sum_j u_j/2\right)\cdot\mathbb{E}(\tau_{(w,s)})}{\sum_j u_j/2}. \tag{4.83}$$

Code that makes use of this formula to compute $\mathbb{E}\left(\tau^2_{(w,s)}\right)$ is provided in [154].

4

4.6.3. APPROXIMATIONS

INFINITE WINDOW SIZE APPROXIMATION

Proof of Theorem 4.2. Letting $\epsilon \equiv \epsilon(w,s,p) = \mathbb{P}(\tau_{(w,s)} > w)$, the expectation of $\tau_{(w,s)}$ can be rewritten as

$$\mathbb{E}(\tau_{(w,s)}) = (1-\epsilon)\mathbb{E}\left(\tau_{(w,s)}|\tau_{(w,s)}\le w\right) + \epsilon\mathbb{E}\left(\tau_{(w,s)}|\tau_{(w,s)} > w\right). \tag{4.84}$$

Now, note that for $n \le w$

$$\mathbb{P}(\tau_{(w,s)} = n) = \mathbb{P}(\tau_{(\infty,s)} = n),$$

i.e. for this range of n the distributions of $\tau_{(w,s)}$ and $\tau_{(\infty,s)}$ exactly match. We can thus rewrite (4.84) as

$$\begin{aligned}(1-\epsilon)\mathbb{E}\left(\tau_{(\infty,s)}|\tau_{(\infty,s)}\le w\right) &+ \epsilon\mathbb{E}\left(\tau_{(w,s)}|\tau_{(w,s)} > w\right)\\ &= \mathbb{E}(\tau_{(\infty,s)}) - \epsilon\mathbb{E}\left(\tau_{(\infty,s)}|\tau_{(\infty,s)} > w\right) + \epsilon\mathbb{E}\left(\tau_{(w,s)}|\tau_{(w,s)} > w\right),\end{aligned}$$

where to obtain the last equality we have expanded $\mathbb{E}(\tau_{(\infty,s)})$ in the same way as (4.84). Now, if one considers starting the whole process again after the first w time steps, we see that $\mathbb{E}\left(\tau_{(w,s)}|\tau_{(w,s)} > w\right) \le w + \mathbb{E}(\tau_{(w,s)})$. Combining this with the fact that

$$\mathbb{E}(\tau_{(\infty,s)}|\tau_{(\infty,s)} > w) > w,$$

we find that

$$\mathbb{E}(\tau_{(w,s)}) - \mathbb{E}(\tau_{(\infty,s)}) \le \epsilon\cdot\left(w + \mathbb{E}(\tau_{(w,s)}) - w\right), \tag{4.85}$$

from which (4.25) follows.

We further bound the distance between the ending pattern distributions. Making use of

$$\mathbb{P}(X_{(w,s)} = x) = \mathbb{P}(X_{(w,s)} = x|\tau_{(w,s)}\le w)(1-\epsilon) + \mathbb{P}(X_{(w,s)} = x|\tau_{(w,s)} > w)\epsilon,$$

we have

$$\mathbb{P}(X_{(w,s)} = x) - \mathbb{P}(X_{(\infty,s)} = x) = \left(\mathbb{P}(X_{(w,s)} = x|\tau_{(w,s)} > w) - \mathbb{P}(X_{(\infty,s)} = x|\tau_{(w,s)} > w)\right)\epsilon,$$

and so, letting $\Omega \equiv \Omega(\infty, s)$,

$$\sum_{x\in\Omega} |\mathbb{P}(X_{(w,s)}=x) - \mathbb{P}(X_{(\infty,s)}=x)| < \sum_{x\in\Omega} \Big(\mathbb{P}(X_{(w,s)}=x|\tau_{(w,s)}>w) + \mathbb{P}(X_{(\infty,s)}=x|\tau_{(w,s)}>w)\Big)\epsilon$$
$$= 2\epsilon(w,s,p).$$

□

Proof of Lemma 4.1. Here, we show the identity (4.29) for $\epsilon(w,s,p)$. Letting $q = 1-p$, we have

$$\begin{aligned}
\epsilon(w,s,p) &= \sum_{n=w+1}^{\infty} \binom{n-1}{s-1} q^{n-s} p^s \\
&= \frac{p^s}{(s-1)!} \sum_{n=w+1}^{\infty} \frac{(n-1)!}{(n-s)!} q^{n-s} \\
&= \frac{p^s}{(s-1)!} \sum_{n=w+1}^{\infty} \frac{\mathrm{d}^{s-1}}{\mathrm{d}q^{s-1}} \left(q^{n-1}\right) \\
&= \frac{p^s}{(s-1)!} \frac{\mathrm{d}^{s-1}}{\mathrm{d}q^{s-1}} \left(\sum_{n=w+1}^{\infty} q^{n-1} \right) \\
&= \frac{p^s}{(s-1)!} \frac{\mathrm{d}^{s-1}}{\mathrm{d}q^{s-1}} \left(\frac{q^w}{1-q} \right) \\
&= \frac{p^s}{(s-1)!} \sum_{i=1}^{s-1} \binom{s-1}{i} \frac{\mathrm{d}^i}{\mathrm{d}q^i} \left(q^w\right) \frac{\mathrm{d}^{s-1-i}}{\mathrm{d}q^{s-1-k}} \left((1-q)^{-1}\right) \\
&= \frac{p^s}{(s-1)!} \sum_{i=1}^{s-1} \frac{(s-1)!}{i!(s-i-1)!} \frac{w!}{(w-i)!} q^{w-i} \frac{(s-i-1)!}{(1-q)^{s-i}} \\
&= \sum_{i=1}^{s-1} \binom{w}{i} q^{w-i} p^i .
\end{aligned}$$

□

ASYMPTOTIC BEHAVIOUR OF THE EXPECTED WAITING TIME AND ENDING PATTERN

Proof of Theorem 4.3. For conciseness, here we take $\Omega \equiv \Omega(w,s)$. A formula for the inverse of A is given in terms of its adjugate matrix $\operatorname{adj} A$ [158],

$$[\operatorname{adj} A]_{ij} := (-1)^{(i+j)} \det M_{ji}. \tag{4.86}$$

where M_{ij} is the $|\Omega| \times |\Omega|$ matrix obtained by removing row i and column j from A. Since A is invertible, the inverse is

$$A^{-1} = \frac{\operatorname{adj} A}{\det A}, \tag{4.87}$$

Now consider the system (4.23). If we consider solving for $\vec{v}$ by multiplying through by A^{-1}, we see that its first element is

$$\mathbb{E}(\tau_{(w,s)}) = \frac{\det B}{\det A}, \tag{4.88}$$

where B is the $|\Omega| \times |\Omega|$ matrix obtained by removing the first row and column from A, so that $B_{ij} = x^i * x^j$. Since all the entries of A are polynomials in $1/p$ and $1/q$, so are $\det B$ and $\det A$.

To proceed with showing (4.33), we characterise the scaling of $\det B$ and $\det A$ for small p. Since $q = 1 - p$ is close to 1 for small p, it suffices to only consider the powers of $1/p$ for the analysis of the asymptotic scaling as $p \to 0$ (recalling the definition of the star product (4.20)). We firstly consider $\det B$. With the observation that the higher-order terms in $1/p$ are given by the star products on the diagonal, and moreover that these each have leading order term given by $1/p^s$. The form of $\det B$, then, is a polynomial of maximum degree $1/p^{s|\Omega|}$. In fact, this is exactly the degree. One can compute this contribution by considering the matrix $\tilde{B}$ of highest powers: letting $r \equiv 1/p^s$, we have $\det B \sim \det \tilde{B}$, where

$$\tilde{B} = \begin{pmatrix} r & 0 & \dots & 0 \\ 0 & r & \dots & 0 \\ \vdots & & \ddots & 0 \\ 0 & 0 & \dots & r \end{pmatrix}, \tag{4.89}$$

and hence, $\det B \sim r^{|\Omega|} = 1/p^{s|\Omega|}$. We then do the same with A. In this case, $\det A \sim \det \tilde{A}$, where

$$\tilde{A} = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ -1 & r & 0 & \dots & 0 \\ -1 & 0 & r & \dots & 0 \\ \vdots & \vdots & & \ddots & 0 \\ -1 & 0 & 0 & \dots & r \end{pmatrix}, \tag{4.90}$$

where the existence of a 0 in the top left-hand corner now disrupts the evaluation of $\det \tilde{A}$ by multiplying along the diagonal, as we did above. Our next step is to evaluate $\det \tilde{A}$ by expanding along the top row,

$$\det \tilde{A} = \sum_{k=1}^{|\Omega|} (-1)^k \det \tilde{A}_k, \tag{4.91}$$

where $\tilde{A}_k$ is the $|\Omega| \times |\Omega|$ matrix formed by removing the first row and the kth column from $\tilde{A}$,

$$\tilde{A}_k = \begin{pmatrix} -1 & r & 0 & & \dots & & 0 \\ \vdots & 0 & \ddots & & & & \vdots \\ & & & r & 0 & & \\ & & & 0 & 0 & & \\ & & & 0 & r & & \\ \vdots & & & & & \ddots & 0 \\ -1 & 0 & \dots & & \dots & 0 & r \end{pmatrix}. \tag{4.92}$$

In $\tilde{A}_k$, the r's are placed above the diagonal in rows $1,...,k-1$, and on the diagonal in rows $k+1,...,|\Omega|$. The determinant of $\tilde{A}_1$ may be calculated by simply multiplying the diagonal elements, to obtain

$$\det \tilde{A}_1 = -r^{|\Omega|-1}. \tag{4.93}$$

We then notice that any $\tilde{A}_k$ can be transformed into $\tilde{A}_1$ by moving the kth row to the top row. This can be achieved by performing $k-1$ row operations, if it is moved by successively exchanging with the row above it $k-1$ times. Then, since each row operation incurs a factor of $(-1)^k$,

$$\det \tilde{A}_k = (-1)^{k-1} \det \tilde{A}_1 = (-1)^k r^{|\Omega|-1}. \tag{4.94}$$

With (4.91), we then see that

$$\det \tilde{A} = |\Omega| r^{|\Omega|-1}, \tag{4.95}$$

and so $\det A \sim |\Omega|/p^{s(|\Omega|-1)}$. Substituting into (4.88), we find

$$\mathbb{E}(\tau_{(w,s)}) \sim \frac{p^{s(|\Omega|-1)}}{|\Omega| p^{s|\Omega|}} = \frac{1}{|\Omega| p^s}. \tag{4.96}$$

To show (4.34), we employ a similar method. We have from (4.87) that

$$\mathbb{P}(X_{(w,s)} = x^{(k)}) = (-1)^{1+k} \frac{\det M_{k0}}{\det A} \tag{4.97}$$

$$\sim (-1)^{1+k} \frac{\det C_k}{\det X}, \tag{4.98}$$

where C_k is obtained by removing the first column and kth row from A,

$$C_k = \begin{pmatrix} 1 & \dots & & & & \dots & 1 \\ r & 0 & \dots & & & \dots & 0 \\ 0 & \ddots & & & & & \vdots \\ \vdots & & r & 0 & 0 & & \\ & & 0 & 0 & r & & \vdots \\ \vdots & & & & & \ddots & 0 \\ 0 & \dots & & & \dots & 0 & r \end{pmatrix}, \tag{4.99}$$

where in C_k the x-entries are placed below the diagonal in columns $1,...,k-1$, and on the diagonal in columns $k+1,...,|\Omega|$. We note that C_k is the transpose of B_k, with the first column multiplied by a scaling factor of -1. Therefore, $\det C_k = -\det B_k = (-1)^{k+1} r^{|\Omega|-1}$, and making use again of (4.95),

$$\mathbb{P}(X_{(w,s)} = x^{(k)}) \sim \frac{r^{|\Omega|-1}}{|\Omega| r^{|\Omega|-1}}, \tag{4.100}$$

from which the result follows. □

4.6.4. TRADE-OFF FUNCTION DUE TO ENTANGLEMENT GENERATION SCHEME

In this appendix, we motivate the linear trade-off function $F_{\text{est}} = 1 - \lambda p$ used in the BQC analysis, which describes the fidelity of qubits in the server immediately after transmission.

Let σ denote the noisy two-qubit entangled state that is produced between the client and server when there is a successful attempt. When there is a success, the client and server perform some qubit transmission procedure Λ_σ, which could for example be teleportation, or a remote state preparation protocol. We assume that this protocol establishes all qubits in the server with the same fidelity F_{est}. For example, this is the case if the noisy entangled state is depolarised

$$\sigma = \frac{4F_0 - 1}{3}|\Phi^+\rangle\langle\Phi^+| + \frac{1 - F_0}{3}\mathbb{I}_4, \tag{4.101}$$

and the standard teleportation protocol from [159] is applied. Here, $F_0 = \langle\Phi^+|\sigma|\Phi^+\rangle$ is the fidelity of σ to the target state. This involves performing a full measurement in the Bell basis $\{|\Phi_{ij}\rangle\}$ and applying the corresponding Pauli corrections. If $|\psi\rangle$ is the qubit state to be teleported, its action is given by Λ^{st},

$$\Lambda^{\text{st}}_\sigma(|\psi\rangle\langle\psi|) := \sum_{i,j} X^i Z^j \langle\Phi_{ij}|(|\psi\rangle\langle\psi| \otimes \sigma)|\Phi_{ij}\rangle Z^j X^i \tag{4.102}$$

where the Bell measurement acts on the registers containing the qubit state $|\psi\rangle$ and the first qubit of σ. Suppose that the entangled state and qubit transmission procedure are given by (4.101) and (4.102). Then after transmitting any qubit $|\psi\rangle$, the resulting fidelity is [96]

$$F_{\text{est}} = \frac{2F_0 + 1}{3}. \tag{4.103}$$

Now, one can incorporate a general rate-fidelity trade-off inherent to the entanglement generation protocol by specifying that F_{est} is a decreasing function of p. In particular, we draw here on an example from the single-photon scheme for entanglement generation. When implementing a single-photon scheme [23], the fidelity of generated states is

$$F_0(p_{\text{suc}}) = 1 - \frac{p_{\text{suc}}}{2p_{\text{det}}}, \tag{4.104}$$

where p_{suc} is the success probability of a physical entanglement attempt, and p_{det} is the probability of detecting an emitted photon. In the case of a very small p_{suc}, one might want to perform entanglement attempts in *batches* in order to minimise overhead due to communication with higher layers of the software stack (which must be notified when there is, or is not, a success). This scheme has been implemented with NV centres in diamond, where typically $p_{\text{suc}} \ll 1$ [141]. If this is the case, choosing one time step to correspond to a batch of $M \ll 1/p_{\text{suc}}$ attempts, the probability of producing at least one entangled link in a time step is

$$p = 1 - (1 - p_{\text{suc}})^M \approx M p_{\text{suc}}. \tag{4.105}$$

Substituting this into (4.103) and (4.104), we obtain a trade-off function of

$$F_{\text{est}}(p) = \frac{2\left(1 - \frac{p}{2Mp_{\text{det}}}\right) + 1}{3} = 1 - \lambda p, \tag{4.106}$$

where $\lambda := 1/(3Mp_{\text{det}})$. Since M is a freely adjustable parameter, then so is λ. The simple relationship (4.106) is also a general first-order behaviour for a decreasing function in p such that $F_{\text{est}}(0) = 1$, which justifies the choice as potentially applicable to other hardware and entanglement generation protocols.

4.6.5. COMPUTING THE ERROR PROBABILITY OF A BQC TEST ROUND

TEST ROUNDS

As mentioned previously, the protocol involves interweaving test rounds at random with computation rounds. It is the test rounds that provide verifiability of the protocol, because they allow the client to check for deviations from the ideal measurement outcomes. Recall that the goal of the client in the BQC protocol is to perform a BQP computation, which is defined in the measurement-based formalism with respect to a graph $G = (V, E)$. In one round of the protocol, the client transmits $|V|$ qubits to the server, which (if it is honest) creates a graph state by applying CZ-gates to pairs of qubits as given in the set of edges E. Before carrying out the protocol, the client chooses some *k-colouring* $\{V_j : j = 1, ..., k\}$, which is a partition of the set of vertices V into different subsets, known as *colours*, such that there is no edge between two vertices of the same colour. This k then corresponds to the k in the feasibility condition (4.39).

Before each test round, the client chooses a colour V_j uniformly at random to be the *trap colour*. A qubit corresponding to vertices from this set is then referred to as a *trap qubit*. Any other qubit is referred to as a *dummy qubit*. Each trap qubit $v \in V_j$ will be $|+_{\theta_v}\rangle := (|0\rangle + e^{i\theta_v}|1\rangle)/\sqrt{2}$, for some angle θ_v that is chosen uniformly at random from $\Theta = \{\frac{k\pi}{4} : k = 0, 1, ..., 7\}$. Each dummy qubit $v \in V \setminus V_j$ will be $|d_v\rangle$, where $d_v \in \{0, 1\}$ is chosen uniformly at random. Then, the effect of the server applying its entangling gates is to flip each trap qubit to the orthogonal basis vector a number of times that corresponds to the sum (modulo 2) of the neighbouring dummies. This is a quantity that the client can compute. After constructing the graph state, the server measures its qubits and sends the outcome to the client. The trap qubit measurement basis that is specified by the client is $\{|\pm_{\delta_v}\rangle\}$, for each trap $v \in V_j$, where $\delta_v = \theta_v + r_v\pi$, and $r_v \in \{0, 1\}$ is chosen uniformly at random. The client compares the outcomes of the trap qubits to what is expected if all states and local operations are perfect, declaring the test round to be a failure if there is at least one trap measurement that is incorrect. A depiction of a graph state, a choice of k-colouring, and a choice of qubits for a test round is given in Figure 4.11, for the case of a square graph.

ERROR PROBABILITY FOR A GENERAL GRAPH

We suppose that the client would like to know the outcome of a BQP calculation, which has corresponding graph $G = (V, E)$, and that the client has chosen a k-colouring $\{V_j\}_{j=1}^k$. Then, given that the vector of fidelities at the time the server applies its operations is $\vec{F} = (F_1, ..., F_{|V|})$, here we obtain a general form for the probability of error of the test round, $P_G(\vec{F})$. This is a generalisation of what can be found in [156], where BQC with two qubits is considered.

We firstly find the probability of error, *given* that the client has chosen trap colour V_j. Call this P_{V_j}. The client thus chooses to send the trap qubits $v \in V_j$ as states $|+_{\theta_v}\rangle$. Then, at the time when the test round is carried out, the trap qubits corresponding to vertices

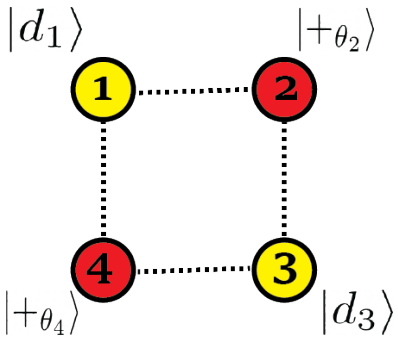


Figure 4.11: **Test rounds in the BQC protocol.** Here we choose an example where the computation is performed on a square graph. The client chooses a k-colouring: here $k = 2$, and the two sets are coloured in yellow and red. In this case, the red vertices correspond to trap qubits, and the yellow vertices correspond to dummy qubits.

$v \in V_j$ are each in a state ρ_v, where

$$\rho_v = F_v|+_{\theta_v}\rangle\langle+_{\theta_v}| + (1-F_v)|-_{\theta_v}\rangle\langle-_{\theta_v}| + (\text{o.d.1}), \tag{4.107}$$

where we use (o.d.1) to write the off-diagonal elements, with respect to the basis

$$\{|+_{\theta_v}\rangle, |-_{\theta_v}\rangle\}.$$

We don't write them out in full because these end up making no contribution to $P_{V_j}(\vec{F})$, as we will see later. Similarly, the dummy qubits $v \in V \setminus V_j$ will be in the state

$$\rho_v = F_v|d_v\rangle\langle d_v)| + (1-F_v)|d_v \oplus 1\rangle\langle d_v \oplus 1| + (\text{o.d.2}), \tag{4.108}$$

where here we use (o.d.2) to write the off-diagonal elements, this time with respect to the computational basis. The state of the server is then given by the tensor product of all of these states

$$\rho_{\text{server}} = \bigotimes_{v\in V_j} \rho_v \bigotimes_{w\in W_j} \rho_w \bigotimes_{u\in V\setminus(V_j\cup W_j)} \rho_u, \tag{4.109}$$

where we have defined $W_j \subset V \setminus V_j$ to be the set of all vertices that share an edge with a trap qubit. The server then proceeds with the next step of the BQC protocol, and applies CZ gates to all pairs of qubits corresponding to edges in E, resulting in the state

$$\rho'_{\text{server}} = U\rho_{\text{server}}U^\dagger, \tag{4.110}$$

where $U := \prod_{(w,v)\in E} CZ_{(w,v)}$. Recall that we are interested in the probability $P_{V_j}(\vec{F})$ that the this results in an error. In fact, is it simpler to find a form for the *success probability* $Q_{V_j}(\vec{F}) = 1 - P_{V_j}(\vec{F})$. An error occurs when at least one of the trap qubit measurements does not match the result that would be obtained if all states were perfect. In particular, if everything were perfect, then the client would expect the measurement outcome corresponding to the trap qubit v to be $r_v \oplus D_v$, where

$$D_v = \bigoplus_{w\in W_j:(v,w)\in E} d_w, \tag{4.111}$$

i.e. the sum (modulo 2) of all the dummy variables d_v that surround the trap qubit. The success probability is then given by

$$Q_{V_j}(\vec{F}) = \mathrm{Tr}_V\left(\left(\bigotimes_{v\in V_j} |(-1)^{r_v\oplus D_v}_{\delta_v}\rangle\langle(-1)^{r_v\oplus D_v}_{\delta_v}|\right)\rho'_{\text{server}}\right), \tag{4.112}$$

where for convenience, we are using the notation $|(+1)_\theta\rangle \equiv |+_\theta\rangle$ and $|(-1)_\theta\rangle \equiv |-_\theta\rangle$. Rewriting $\left|(-1)^{r_v+D_v}_{\delta_v}\right\rangle = \left|(-1)^{D_v}_{\theta_v}\right\rangle$, and after examining Equations (4.109) and (4.110), we see that the qubit registers corresponding to vertices $v \in V\setminus(V_j\cup W_j)$ will make no contribution to this quantity, so that

$$Q_{V_j}(\vec{F}) = \mathrm{Tr}_{V_j}\mathrm{Tr}_{W_j}\left(\left(\bigotimes_{v\in V_j} |(-1)^{D_v}_{\theta_v}\rangle\langle(-1)^{D_v}_{\theta_v}|\right)\sigma'_{\text{server}}\right), \tag{4.113}$$

with

$$\sigma'_{\text{server}} := \tilde{U}\sigma_{\text{server}}\tilde{U}^\dagger, \tag{4.114}$$

where we have defined $\sigma_{\text{server}} := \bigotimes_{v\in V_j}\rho_v\bigotimes_{w\in W_j}\rho_w$ and $\tilde{U} := \prod_{(w,v)\in E_j} CZ_{(w,v)}$, and $E_j := \{(v,w)\in E : v\in V_j\}$ to be the set of all edges between any element of V_j and any other vertex. Recalling the states of our qubits as given in (4.107) and (4.108), and defining $F^{(0)} := F$, $F^{(1)} := 1-F$ to be used as a more concise way to write some of the terms, we can then write

$$\begin{aligned}\sigma_{\text{server}} &= \bigotimes_{v\in V_j}\left(\sum_{x_v\in\{0,1\}} F_v^{(x_v)}|(-1)^{x_v}_{\theta_v}\rangle\langle(-1)^{x_v}_{\theta_v}|\right)\bigotimes_{w\in W_j}\left(\sum_{y_w\in\{0,1\}} F_w^{(y_w)}|d_w+y_w\rangle\langle d_w+y_w|\right)\\ &= \bigotimes_{v\in V_j}\left(\sum_{x_v\in\{0,1\}} F_v^{(x_v)}|(-1)^{x_v}_{\theta_v}\rangle\langle(-1)^{x_v}_{\theta_v}|\right)\bigotimes\left(\sum_{\vec{y}\in\{0,1\}^{|W_j|}}\prod_{w\in W_j} F_w^{(y_w)}|\vec{d}+\vec{y}\rangle\langle\vec{d}+\vec{y}|\right),\end{aligned} \tag{4.115}$$

where in (4.115) we have rewritten the sum to be over all length-$|W_j|$ binary strings $\vec{y}\equiv(y_w)_{w\in W_j}\in\{0,1\}^{|W_j|}$. We have also stored the dummy variables in a vector $\vec{d}$, so that $(\vec{d}+\vec{y})_w = d_w+y_w$. Again, we are not writing out the off-diagonal terms because these all disappear when we take the trace, and therefore make no contribution to the final expression. Applying the unitary operator $\tilde{U}$ yields

$$\begin{aligned}\sigma'_{\text{server}} = \sum_{\vec{y}\in\{0,1\}^{|W_j|}}\prod_{w\in W_j} F_w^{(y_w)}|\vec{d}+\vec{y}\rangle\langle\vec{d}+\vec{y}|\\ \bigotimes_{v\in V_j}\left(\sum_{x_v\in\{0,1\}} F_v^{(x_v)}|(-1)^{x_v+s_v(\vec{y})+D_v}_{\theta_v}\rangle\langle(-1)^{x_v+s_v(\vec{y})+D_v}_{\theta_v}|\right)\end{aligned}$$

where for a trap $v\in V_j$, we have defined

$$s_v(\vec{y}) := \sum_{w\in W_j:(w,v)\in E_j} y_w, \tag{4.116}$$

which is the sum of the binary variables y_w over all vertices neighbouring v. We can now start to trace out registers in order to find a final expression for $Q_{V_j}(\vec{F})$ in terms of the qubit fidelities. Taking the inner product $\langle(-1)^{D_v}_{\delta_v}|...|(-1)^{D_v}_{\delta_v}\rangle$ for all trap qubits $v \in V_j$ and tracing out W_j yields our final expression for the success probability as introduced in Equation (4.112),

$$Q_{V_j}(\vec{F}) = \sum_{\vec{y}\in\{0,1\}^{|W_j|}} \prod_{w\in W_j} F_w^{(y_w)} \prod_{v\in V_j} F_v^{(s_v(\vec{y}))}, \tag{4.117}$$

This is a polynomial in the fidelities $\vec{F} = (F_1, ..., F_{|V|})$, with a form that is completely determined by the graph structure and choice of trap colour V_j. The same thus holds for the error probability $P_{V_j}(\vec{F})$. In our model as given in Section 4.4.2, it is further necessary to incorporate the fact that the first qubit to be sent is chosen at random. The probability of error is then effectively symmetrised over the $|V|$ possible starting qubits in the following way. Without loss of generality, letting the order in which the qubits are sent to be lexicographical, the probability of error is then

$$P_{V_j}(\vec{F}) \to \tilde{P}_{V_j}(\vec{F}) := \frac{1}{|V|}\sum_j P_{V_j}(\sigma^j \vec{F}), \tag{4.118}$$

where σ is the permutation that moves the vector elements one place to the left, i.e. $\sigma(F_1, ..., F_{|V|}) = (F_2, ..., F_{|V|}, F_1)$. To obtain the final probability of error, it remains to average over the choice of trap colour, recalling that this is chosen uniformly at random. This gives us a final expression for $P_G(\vec{F})$,

$$P_G(\vec{F}) = \frac{1}{k}\sum_{j=1}^{k} \tilde{P}_{V_j}(\vec{F}). \tag{4.119}$$

ERROR PROBABILITY FOR A SQUARE GRAPH

An example of such a polynomial for the case of a square graph is as follows. Consider the k-colouring as in Figure 4.11, with red as the choice of trap colour. Suppose that when the server applies its gates and measurements, the qubits have fidelities $\vec{F} = (F_1, F_2, F_3, F_4)$. Then, according to (4.117), the success probability is given by

$$\begin{aligned} Q_{\text{red}}(\vec{F}) &= F_1F_2F_3F_4 + F_1(1-F_2)(1-F_3)(1-F_4) \\ &+(1-F_1)(1-F_2)F_3(1-F_4) + (1-F_1)F_2(1-F_3)F_4, \end{aligned}$$

and the error probability is then

$$P_{\text{red}}(\vec{F}) = 1 - Q_{\text{red}}(\vec{F}). \tag{4.120}$$

By symmetry of the square graph, the error probability $P_{\text{yellow}}(\vec{F})$ is obtained by exchanging the indices $1 \leftrightarrow 2, 3 \leftrightarrow 4$ in (4.120). In the case of the square graph, the symmetrisation maps $P_{\text{red}}(\vec{F}) \to \tilde{P}_{\text{red}}(\vec{F})$, where

$$\begin{aligned} \tilde{P}_{\text{red}}(\vec{F}) = \frac{1}{4}\big[&P_{\text{red}}(F_1, F_2, F_3, F_4) + P_{\text{red}}(F_2, F_3, F_4, F_1) \\ &+ P_{\text{red}}(F_3, F_4, F_1, F_2) + P_{\text{red}}(F_4, F_1, F_2, F_3)\big]. \end{aligned}$$

Note that the symmetries of the error functions P_{red} and P_{yellow} reflect the symmetries of the graph, i.e. they are symmetric under the interchange of $1 \leftrightarrow 3$ or $2 \leftrightarrow 4$. Then,

$$\tilde{P}_{\text{red}}(\vec{F}) = \frac{1}{2}\left(P_{\text{red}}(\vec{F}) + P_{\text{yellow}}(\vec{F})\right) \tag{4.121}$$

The other error function P_{yellow} maps to the same after the symmetrisation (4.118), i.e. $\tilde{P}_{\text{yellow}}(\vec{F}) = \tilde{P}_{\text{red}}(\vec{F})$. The probability of error, then, is given by

$$\begin{aligned} P_{\text{square}}(\vec{F}) &= \frac{1}{2}\left(\tilde{P}_{\text{yellow}}(\vec{F}) + \tilde{P}_{\text{red}}(\vec{F})\right) \\ &= \frac{1}{2}\left(P_{\text{red}}(\vec{F}) + P_{\text{yellow}}(\vec{F})\right). \end{aligned}$$

This is the function that we use with (4.45) to calculate p_{av} for our model, and compute the results for an example of a square graph in Section 4.4.2.

5

OPTIMISING ENTANGLEMENT PACKET GENERATION WITH ADAPTIVE POLICIES

Aksel Tacettin*, Tianchen Qu*, Bethany Davies*, Boris Goranov, Ioana-Lisandra Draganescu, and Gayane Vardoyan

Protocols for distributed quantum systems commonly require the simultaneous availability of n entangled states, each with a fidelity above some fixed minimum F_{app} relative to the target maximally-entangled state. However, the fidelity of entangled states degrades over time while in memory. Entangled states are therefore rendered useless when their fidelity falls below F_{app}. This is problematic when entanglement generation is probabilistic and attempted in a sequential manner, because the expected completion time until n entangled states are available can be large. Motivated by existing entanglement generation schemes, we consider a system where the entanglement generation parameters (the success probability p and fidelity F of the generated entangled state) may be adjusted at each time step. We model the system as a Markov decision process, where the policy dictates which generation parameters (p,F) to use for each attempt. We use dynamic programming to derive optimal policies that minimise the expected time until n entangled states are available with fidelity greater than F_{app}. We observe that the advantage of our optimal policies over the selected baselines increases significantly with n. In the parameter regimes explored, which are based closely on current experiments, we find that the optimal policy

*These authors contributed equally.

This chapter has been released separately at https://arxiv.org/abs/2509.17576

can provide a speed-up of as much as a factor of twenty over a constant-action policy. In addition, we propose a computationally inexpensive heuristic method to compute policies that perform either optimally or near-optimally in the parameter regimes explored. Our heuristic method can be used to find high-performing policies in parameter regimes where finding an optimal policy is intractable.

5.1. INTRODUCTION

Protocols for distributed quantum systems commonly require multiple entangled pairs of qubits, also referred to as *entangled links*, or just *links*. Examples of protocols with this requirement are applications such as verifiable blind quantum computing [147] and quantum secret sharing [160], as well as important subroutines such as entanglement purification [49, 50]. In some contexts, multiple simultaneously-existing links are collectively referred to as an *entanglement packet* [28]. The fast generation of entanglement packets is a task of fundamental importance for a functional quantum network. In this work, we find protocols that optimise the rate of entanglement packet generation, by adaptively varying a rate-fidelity trade-off mechanism available due to the entanglement generation scheme.

Here, we consider a setting with two nodes that attempt entanglement generation sequentially. In our model we assume that time is divided into discrete, uniform time steps, where in each time step, a single entanglement generation attempt is performed. This is very often the case in near-term quantum networks, where heralded entanglement generation schemes succeed probabilistically and take up a fixed amount of time, due to the transfer of classical and quantum information between distant nodes [25, 23, 24, 26, 27]. In our model, the time units are abstract and a single attempt is assumed to take up one unit of time. After an attempt, an entangled link is generated with success probability p. The link is generated with initial fidelity $F = \langle \Psi_{00} | \rho | \Psi_{00} \rangle$ to the target maximally-entangled state

$$|\Psi_{00}\rangle = \frac{1}{\sqrt{2}} (|00\rangle + |11\rangle).$$

When a link is generated successfully, it is assumed to be immediately stored in memory. While the entangled link is stored in memory, it is subject to time-dependent noise (or *decoherence*), which causes the fidelity to degrade over time. We assume a simple depolarising noise model: given a link with initial fidelity F, the fidelity of the link after t time steps is

$$F \mapsto e^{-\Gamma t} \left(F - \frac{1}{4} \right) + \frac{1}{4}, \tag{5.1}$$

where Γ is the decoherence rate. We assume that there is a fixed minimum fidelity F_{app} required by an application for each link. A link is discarded once the fidelity of the link decays below F_{app} (see Figure 5.1). The goal is to generate n simultaneously-existing links. Accordingly, we assume that each node has n memories, each with the same decoherence rate Γ.

Suppose that at time $t = 0$, there are no links stored in memory. We let T denote the first time that there are n simultaneously-existing links in memory. An important performance metric to consider is the expected time $\mathbb{E}[T]$ until completion. Depending on the system parameters, $\mathbb{E}[T]$ may be excessively large. For example, if the decoherence rate Γ is large, then links are discarded soon after they are generated. Therefore, any link in memory is likely to be discarded before enough remaining links are successfully generated for the application. The same holds if the probability of generating a link is small, or if the initial fidelity of entangled states is small.

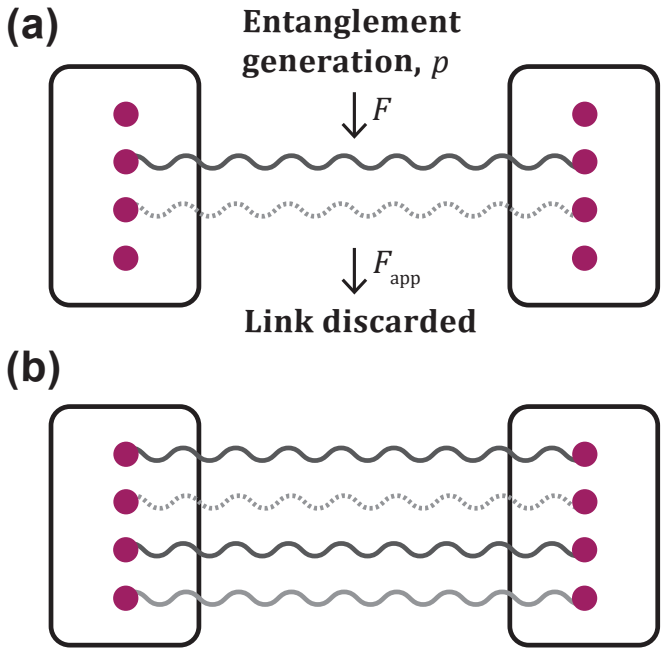


Figure 5.1: **(a) Depiction of an intermediate state with $n = 4$.** At the beginning of each time step, the system chooses generation parameters $(p, F) \in \mathscr{A}$. Then, an entanglement generation attempt is carried out, which with probability p generates a link with fidelity F. If successful, the generated link is immediately transferred to memory. While in memory, the link is subject to decoherence. If the link's fidelity falls below F_{app}, it is discarded. We are interested in the case where the system starts with no links in memory. It will then pass through a number of intermediate states, until the first time T when it reaches an absorbing state that has n links simultaneously available in memory. **(b) Depiction of an absorbing state with $n = 4$.**

We consider a setting where, before each entanglement generation attempt, the system may choose the generation parameters (p, F) from a fixed set

$$\mathscr{A} = \{(p_i, F_i) : i = 1, \ldots, |\mathscr{A}|\}, \tag{5.2}$$

where it is assumed that $F_i > F_{\text{app}}$ for all i. The choice of generation parameters at time t may depend on the current state of the system S_t, which in our model is defined by the number of links that are in memory at time t and their corresponding fidelities.

In many entanglement generation schemes, there exists a trade-off between the probability p of successful entanglement generation and the fidelity F of the generated link. In our model, this is captured by enforcing

$$F_i < F_j \Leftrightarrow p_i > p_j, \quad i \neq j. \tag{5.3}$$

A well-known example of the trade-off between p and F arises in the single-click protocol [23], where varying the bright-state parameter α results in a linear relationship between the two, $(p, F) = (\kappa\alpha, 1-\alpha)$, where κ is a constant depending on hardware parameters such as the photon loss [141]. The trade-off between p and F exists in other physical

entanglement generation schemes and could for example be due to the number of entanglement generation attempts that occur before the system is reinitialised [161], or due to the mean photon number when a weak coherent pulse is used as a single-photon source [162]. The same trade-off is also ubiquitous at a higher level: for example, if entanglement purification is employed, then in the choice of protocol there exists a fundamental trade-off between the output fidelity and the probability of the purification protocol succeeding [53]. We assume that the set of generation parameters (5.2) is finite without loss of generality (see Section 5.4.3 for details of how an infinite action space may be reduced to a finite one).

In this work, we model the system as a Markov decision process (MDP), which we subsequently solve to find optimal policies. A policy π is a map[1]

$$\pi : \mathscr{S} \to \mathscr{A}, \tag{5.4}$$

where $\mathscr{A}$ is the set of generation parameters (5.2) (also referred to as the set of *actions*) and $\mathscr{S}$ is the state space. When implementing policy π, the generation parameters chosen at time t depend on the current state S_t and are given by $\pi(S_t) \in \mathscr{A}$. The state S_t and the action $\pi(S_t)$ determine the possible values of the state in the next time step, S_{t+1}. When the policy π is employed, we write the completion time as $T \equiv T_\pi$. The policy π^* is called *optimal* if it minimises the expected completion time,

$$\mathbb{E}[T_{\pi^*}] = \min_\pi \mathbb{E}[T_\pi]. \tag{5.5}$$

Now that we have introduced the problem, we outline our main contributions:

- We use dynamic programming [163] to compute optimal policies π^* and the optimal performance $\mathbb{E}[T_{\pi^*}]$. We compare the optimal performance with the performance of the best constant-action policy π_{con} that chooses the same action in every time step. We evaluate policy performance in two parameter regimes that we call the *near-term regime* and the *far-term regime*. The parameters of the near-term regime are based on recent experiments. The far-term regime is assumed to have an improved memory lifetime, and therefore has a more complex state space and policy behaviour. Both parameter regimes have a trade-off relation (5.41) that is given by the single-click protocol [23]. In both parameter regimes, we find that the optimal policy can provide a speed-up over the constant-action policy of as much as
$$\frac{\mathbb{E}[T_{\pi^*}]}{\mathbb{E}[T_{\pi_{\text{con}}}]} \approx 0.05. \tag{5.6}$$
We conclude that, given an adjustable rate-fidelity trade-off available in the entanglement generation scheme, it can be highly advantageous to use adaptive protocols to boost the generation rate of entanglement packets. Moreover, we see that this advantage increases with n, indicating that adaptive polices will provide even more improvement as quantum networks become more sophisticated.

[1] In general, a policy can be non-deterministic [163]. However, since the state space is finite in our problem, there always exists a deterministic optimal policy [164]. For clarity, we therefore introduce the policy as deterministic.

- By computing π^* in the two parameter regimes, we gain insights about the structure of optimal policies. Based on our insights, we define an efficiently-computable heuristic policy π_{h}. Remarkably, we find that in the near-term regime, the heuristic policy is in fact optimal, i.e. $\mathbb{E}[T_{\pi_{\mathrm{h}}}] = \mathbb{E}[T_{\pi^*}]$. In the far-term regime, the heuristic policy performs close to optimally, satisfying

$$\frac{\mathbb{E}[T_{\pi_{\mathrm{h}}}] - \mathbb{E}[T_{\pi^*}]}{\mathbb{E}[T_{\pi^*}]} < 0.03 \tag{5.7}$$

  for all n for which π^* was computed. In parameter regimes where it is not possible to compute an optimal policy π^* due to scaling of the state space $|\mathscr{S}|$, one may therefore instead employ π_{h} and expect either optimal or close-to-optimal performance. For such a case in our far-term regime, the heuristic policy provides a speed-up over the constant-action policy of as much as

$$\frac{\mathbb{E}[T_{\pi_{\mathrm{h}}}]}{\mathbb{E}[T_{\pi_{\mathrm{con}}}]} \approx 1.05 \cdot 10^{-6}. \tag{5.8}$$

5

The remainder of this work is structured as follows. In Section 5.2, we summarise related work. Then, in Section 5.3, we formally define the MDP and briefly introduce dynamic programming. In Section 5.4, we present our results: in Section 5.4.1 we firstly present an analytical solution for the optimal policy and its performance for $n = 2$. In Section 5.4.2, we present our efficiently-computable heuristic policy. In Section 5.4.3, we consider the example of a single-click entanglement generation scheme, and compare the performance of the optimal policy, heuristic policy and baselines in the two parameter regimes of interest. We also extract general conclusions about the properties one can expect of optimal policies. We conclude and suggest possible future extensions of our work in Section 5.5.

5.2. RELATED WORK

The scenario studied in this work is a generalisation of the one studied in [58]. In [58], the generation parameters (p, F) were assumed to be the same in each time step. In this work, we allow the system to choose instead from the set of generation parameters (5.2). The system studied in [58] is therefore equivalent to the constant-action policy π_{con} that chooses the same generation parameters in every state. We note that the methods used in this work are very different from [58]: here we formulate the problem as an MDP and perform optimisation with dynamic programming, whereas in [58] analytical solutions were derived for $\mathbb{E}[T_{\pi_{\mathrm{con}}}]$.

MDP-based techniques have previously shown their value for the optimisation of a range of quantum network protocols. For example, they have been used to find optimal entanglement swapping policies in repeater networks [42, 45, 165]. Approximate reinforcement learning approaches, which efficiently find approximate solutions to problems formulated as MDPs, have also been utilised in the context of quantum networks, for example for designing entanglement routing schemes [166], optimising quantum repeater chains for secret key distribution [167] or entanglement distribution [62], and designing new and improved communication protocols, particularly in networks with

asymmetric features [168]. We also note that other works have optimised the performance of quantum network protocols by varying the trade-off between the success probability and output fidelity, most often optimising over the bright-state parameter [75, 156, 169]. However, other than [58], the aforementioned studies all optimise the delivery of a single entangled link. In our work, we optimise the delivery of multiple links in the form of an entanglement packet, which is a fundamentally different problem.

5.3. METHODS

5.3.1. CONSTRUCTING THE MARKOV DECISION PROCESS

An MDP is defined as a 4-tuple $(\mathcal{S}, \mathcal{A}, \mathrm{P}, R)$ where $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space, P is the transition function, and R is the reward function. In the following, we elaborate on each component of the MDP for our system.

STATE SPACE

The fidelity of each link in memory is fully characterised by the time it will survive before being discarded, which we refer to as the *time-to-live* (TTL) of a link. Suppose that a link has fidelity $F > F_{\text{app}}$. Recalling our decoherence model (5.1), the TTL of the link is given by

$$t_{\text{TTL}}(F) = \left\lceil \frac{1}{\Gamma} \ln\left(\frac{F - \frac{1}{4}}{F_{\text{app}} - \frac{1}{4}} \right) \right\rceil. \tag{5.9}$$

The ceiling function is taken because we work in discrete time. Letting the maximum fidelity of a newly generated link be denoted by

$$F_{\max} = \max\{F : (p, F) \in \mathcal{A}\},$$

the maximum TTL is given by

$$t_{\max} = t_{\text{TTL}}(F_{\max}). \tag{5.10}$$

The state S_t of the system at time t characterises the relevant system information. For our system, the state is the number of links stored in memory and their TTLs (corresponding to link fidelities). If there are m links in memory, the state s is given by

$$s = \{t_1, \ldots, t_m\}, \tag{5.11}$$

where $t_i \in [t_{\max}]$ is the TTL of the ith link, and $[t_{\max}] = \{1, \ldots, t_{\max}\}$. Formally, s is a multiset that may contain multiple elements of the same value, because it is possible that $t_i = t_j$ for $i \neq j$. Since the n memories are assumed to be identical, the ordering of the TTLs in any given state is assumed to be decreasing without loss of generality, i.e. $t_i \geq t_j$ for $i < j$.

As an example, if $S_t = \{3, 2\}$ then at time t there are two links in memory, one of which will be discarded after three time steps, and the other after two time steps. The first link has a higher fidelity than the second link, because it will be discarded later.

We denote the set of all states with m links in memory as

$$\mathcal{S}_m = \left\{ \{t_1, \ldots, t_m\} : t_i \in [t_{\max}], t_1 \geq \cdots \geq t_m \right\}. \tag{5.12}$$

We denote the state with no links in memory as the empty set $\varnothing \equiv \{\}$. The process is completed when there are n links in memory, or equivalently, when it reaches a state $s \in \mathscr{S}_n$. We also refer to $\mathscr{S}_n$ as the set of *absorbing states*. See Figure 5.1 for an illustration. The full state space is given by

$$\mathscr{S} = \bigcup_{m=1}^{n-1} \mathscr{S}_m \cup \{\varnothing\}. \tag{5.13}$$

We also denote the combined state space with the absorbing states as

$$\mathscr{S}^+ = \mathscr{S} \cup \mathscr{S}_n. \tag{5.14}$$

The time T_π until completion (when policy π is employed) may then be written explicitly as

$$T_\pi = \min\{t : S_t \in \mathscr{S}_n\}. \tag{5.15}$$

5

ACTION SPACE

The action space $\mathscr{A}$ is the set of generation parameters (5.2). The possible generation parameters depend on the specific entanglement generation scheme used. We note that $\mathscr{A}$ is without loss of generality finite (see Section 5.4.3). In Section 5.4.3 we give an example of $\mathscr{A}$ for the single-click protocol.

TRANSITION FUNCTION

Suppose that at time t, the system is in state s, and that action $a \in \mathscr{A}$ is chosen. Then, the transition function $\mathrm{P}(s'|s,a)$ determines the probability of transitioning to state s' in time step $t+1$. We now write down the transition function P explicitly for our system. We suppose that $s = \{t_1, \ldots, t_m\}$ and that the action chosen is $a = (p, F)$. There are two transitions that can occur. The first is when the entanglement generation attempt succeeds, which occurs with probability p. In this case, a new link is generated with TTL $t_{\mathrm{TTL}}(F)$, as given by (5.9). All other links decohere by one time step and are discarded if the fidelity falls below F_{app}, or equivalently when the TTL becomes zero. Since the TTL (5.9) is the inverse of the decoherence map (5.1), decoherence over a single time step simply causes all TTLs to reduce by one. In the event of entanglement generation success, the state in the next time step is therefore given by

$$s'_{\mathrm{succ}} = \{t_j - 1 : t_j \in s, t_j > 1\} \cup \{t_{\mathrm{TTL}}(F)\}. \tag{5.16}$$

The second possible transition is when entanglement generation fails, which occurs with probability $1-p$. The state in the next time step is given by

$$s'_{\mathrm{fail}} = \{t_j - 1 : t_j \in s, t_j > 1\}. \tag{5.17}$$

We therefore have

$$\mathrm{P}(s'|s,a) = \begin{cases} p, & \text{if } s' = s'_{\mathrm{succ}} \\ 1-p, & \text{if } s' = s'_{\mathrm{fail}} \\ 0, & \text{otherwise.} \end{cases} \tag{5.18}$$

Similarly, the transitions from the state $\varnothing$ are given by

$$\mathrm{P}(s'|\varnothing, a) = \begin{cases} p, & \text{if } s' = \{t_{\mathrm{TTL}}(F)\} \\ 1-p, & \text{if } s' = \varnothing \\ 0, & \text{otherwise.} \end{cases} \tag{5.19}$$

The transition function is fully defined by (5.18) and (5.19).

REWARD

Since our objective is to minimise the expected completion time $\mathbb{E}[T_\pi]$, the reward is $R(s,a) = -1$ for all $a \in \mathscr{A}$ and $s \in \mathscr{S}^+$.

5.3.2. DYNAMIC PROGRAMMING

We use a dynamic programming algorithm known as *policy iteration* to compute optimal policies. We let $R_t := R(S_t, a)$ be the reward at time step t, given that the state is S_t and action a is taken. Then, the *value* of a state $s \in \mathscr{S}$ under policy π is defined as

$$v_\pi(s) := \mathbb{E}\left[\sum_{k=1}^{T_\pi - t} R_{t+k} \middle| S_t = s\right]. \tag{5.20}$$

This is the definition value for episodic problems with no discounting factor – see e.g. Chapter 3 of [163] for more details of how the value is defined for other systems. In our system,

$$\begin{aligned} v_\pi(s) &= \mathbb{E}\left[\sum_{k=1}^{T_\pi - t} -1 | S_t = s\right] && (5.21)\\ &= -\mathbb{E}[T_\pi - t | S_t = s] && (5.22)\\ &= -\mathbb{E}[T_\pi | S_0 = s]. && (5.23) \end{aligned}$$

Then, $v_\pi(s)$ is simply the expected completion time, given that the process starts in state s and policy π is employed.

For all $s \in \mathscr{S}$ and any policy π, we can calculate $v_\pi(s)$, which is a step known as *policy evaluation*. We do this with *iterative policy evaluation*, where the update rule

$$v_k(s) := -1 + \sum_{s' \in \mathscr{S}} \mathrm{P}(s'|s, \pi(s)) v_{k-1}(s') \tag{5.24}$$

is recursively applied. Letting $v_k = v_{k-1} = v_\pi$, (5.24) is known as the *Bellman equation* for v_π. The sequence $\{v_k\}$ obtained can be shown to converge to v_π [163].

Policy iteration computes an optimal policy π^* by starting with an arbitrary policy π_0, and then iteratively updating it until the value $v_\pi(s)$ of each state s has been maximised. Let π_k be the policy at the k-th iteration. The policy is then updated by maximising the value with respect to all $a \in \mathscr{A}$ for each state, which again can be achieved by considering the Bellman equation for $v_{\pi_{k-1}}$,

$$\begin{aligned} \pi_k(s) &= \operatorname*{argmax}_{a \in \mathscr{A}} \sum_{s' \in \mathscr{S}} \mathrm{P}(s'|s,a)(-1 + v_{\pi_{k-1}}(s')) && (5.25)\\ &= \operatorname*{argmax}_{a \in \mathscr{A}} \sum_{s' \in \mathscr{S}} \mathrm{P}(s'|s,a) v_{\pi_{k-1}}(s'). && (5.26) \end{aligned}$$

Since both $|\mathscr{S}|$ and $|\mathscr{A}|$ are finite in our MDP, policy iteration converges to an optimal policy π^* after a finite number of iterations [163].

5.4. RESULTS

5.4.1. ANALYTICAL SOLUTION FOR $n = 2$

We now analyse the case $n = 2$, for which we are able to find a closed-form expression for the optimal policy π^* and its expected waiting time. This will help to provide intuition for the properties we expect from optimal policies for larger n.

We recall the definition of the state space (5.13). For $n = 2$, the state space is written explicitly as

$$\mathscr{S} = \{\varnothing, \{1\}, \{2\}, \ldots, \{t_{\max}\}\}. \tag{5.27}$$

For now, we write $\pi^*(\varnothing) = (p_\varnothing, F_\varnothing) \in \mathscr{A}$. The precise choice of $\pi^*(\varnothing)$ will be fixed later. We then notice that, if a single link is present in memory, one must maximise the success probability. This is because the process will be completed if the second link is generated before the first is discarded. Then, although maximising the success probability also minimises the fidelity of the generated link, this would not matter because the second link does not impact future behaviour. In particular, we define the action

$$(p_{\max}, F_{\min}) = \underset{(p,F)\in\mathscr{A}}{\operatorname{argmax}}\{p\}. \tag{5.28}$$

Note that by the rate-fidelity trade-off (5.3) assumed in $\mathscr{A}$, we also have $F_{\min} = \min\{F : (p, F) \in \mathscr{A}\}$. We then set

$$\pi^*(s) \coloneqq (p_{\max}, F_{\min}), \text{ for } s \in \mathscr{S}_1 \setminus \{\{1\}\}. \tag{5.29}$$

In (5.29), we have not assigned the same action to $\pi^*(\{1\})$ because, when there is a link with TTL of one, the link will be discarded in the next time step. Thus, there is no chance of completing the process in the next time step. One should therefore not maximise the success probability, as we do for the other states in (5.29). In particular, $\varnothing$ and $\{1\}$ both contain zero *viable links*, which are the number of links that have a non-zero probability of surviving until the system reaches an absorbing state $s \in \mathscr{S}_n$. We therefore view $\varnothing$ and $\{1\}$ as equivalent states, and they are assigned the same action

$$\pi^*(\{1\}) = \pi^*(\varnothing) = (p_\varnothing, F_\varnothing). \tag{5.30}$$

Equivalent states may also be identified when $n > 2$ to perform a reduction of the state space $\mathscr{S}$ (see Appendix 5.6.1), which can speed up the computation of optimal policies with dynamic programming.

Given the structure of the optimal policy from (5.29) and (5.30), we now explicitly compute the expected completion time $\mathbb{E}[T_{\pi^*}]$. Assuming that the system starts in the state $S_0 = \varnothing$, it will alternate between two phases until completion:

(A) There are no viable links in memory (the state is either $\varnothing$ or $\{1\}$). The system attempts entanglement generation with parameters $(p_\varnothing, F_\varnothing)$ until a link is successfully generated.

(B) The newly generated link survives in memory for $t_{\mathrm{TTL}}(F_\varnothing)$ time steps, until it is discarded. During the first $t_{\mathrm{TTL}}(F_\varnothing)-1$ time steps when the link is in memory, the system attempts entanglement generation with parameters $(p_{\max}, F_{\min})$. If at least one of these attempts is successful, the process is completed. If none of these attempts are successful, then in the final time step, the state is $\{1\}$ and the system returns to phase (A).

The expected completion time may then be computed using exactly the same method as the one used in Section III-B of [58]. The method makes use of properties of the geometric distribution. We obtain

$$\mathbb{E}[T_{\pi^*}] = \frac{1}{p_{\max}} + \frac{1}{p_\varnothing\left(1-(1-p_{\max})^{t_{\mathrm{TTL}}(F_\varnothing)-1}\right)}, \tag{5.31}$$

and the solution for $\pi^*(\varnothing)$ and $\pi^*(\{1\})$ is therefore given by

$$(p_\varnothing, F_\varnothing) = \underset{(p,F)\in\mathscr{A}}{\operatorname{argmin}}\left\{\frac{1}{p\left(1-(1-p_{\max})^{t_{\mathrm{TTL}}(F)-1}\right)}\right\}. \tag{5.32}$$

We have now fully defined the optimal policy $\pi^*(s)$ for all $s\in\mathscr{S}$, and this completes the analytical solution. We have also derived a formula for its performance in (5.31).

The analysis for $n=2$ offers valuable insights into the expected behaviour of the optimal policy for $n>2$. As we extend to cases with larger n, we anticipate that the optimal policy will continue to select the action $(p_{\max}, F_{\min})$ for states with $n-1$ viable links in memory. For other states, the action chosen by the optimal policy must correctly balance the probability of generation p with the time-to-live $t_{\mathrm{TTL}}(F)$. This is because a link must be generated quickly, but there must also be sufficient time for the remaining links to be generated while that link is in memory. For $n=2$, the correct balance is captured by (5.32).

5.4.2. HEURISTIC POLICY

The size of the state space $|\mathscr{S}|$ scales exponentially with n, meaning that for large n computing optimal policies with policy iteration (Section 5.3.2) is intractable. See Appendix 5.6.1 for a detailed analysis of the scaling. With our setup, we are only able to compute optimal policies for $n\leq 7$ in the parameter regimes that we explore in Section 5.4.3. Here, we propose a heuristic policy π_{h} that is efficient to compute for large n and that performs close to optimally for $n\leq 7$ (see Section 5.4.3). One can therefore expect π_{h} to exhibit high performance for $n>7$.

Recalling the notion of a viable link that was introduced in Section 5.4.1, we define the function

$$N_v : \mathscr{S} \to \{0,1,\ldots,n-1\}$$

such that $N_v(s)$ outputs the number of viable links in state s. For example, recalling the discussion in Section 5.4.1, we have $N_v(\varnothing)=N_v(\{1\})=0$. More generally, given the state $s=\{t_1,\ldots,t_m\}$, the m-th link with TTL t_m is viable if there is a chance of completing the process (i.e. generating $n-m$ remaining links) before the link expires. Therefore, the link is viable if $t_m>n-m$. Recall from Section 5.3.1 that the labelling of the TTLs is assumed

to be decreasing, i.e. $t_i \geq t_j$ for $i < j$. If the m-th link is viable, then for all $k \in \{1,\dots,m-1\}$, we therefore have $t_k > n - m$ and all other links are viable. The total number of viable links in the state s is given by

$$N_v(s) \coloneqq \max\{j : t_j > n - j\}. \tag{5.33}$$

Given the state $s = \{t_1,\dots,t_m\}$, we define a second function $v : \mathscr{S} \to \mathscr{S}$ such that

$$v(s) = \begin{cases} \{t_1,\dots,t_{N_v(s)}\} & \text{if } N_v(s) \geq 1 \\ \varnothing, & \text{if } N_v(s) = 0. \end{cases} \tag{5.34}$$

The state $v(s)$ contains the viable links of s. Identifying $s \equiv v(s)$ can be used to reduce the size of the state space — see Appendix 5.6.1 for more details.

We now define the heuristic policy π_{h}. As we did in Section 5.4.1 for the case $n = 2$, we set the actions for states with zero viable links to an arbitary value,

$$\pi_{\mathrm{h}}(s) \coloneqq (p_\varnothing, F_\varnothing) \in \mathscr{A}, \text{ if } N_v(s) = 0. \tag{5.35}$$

Further, we use the intuition established in Section 5.4.1 and enforce that the policy must choose the maximum-probability action (5.28) when there are $n-1$ viable links in memory,

$$\pi_{\mathrm{h}}(s) \coloneqq (p_{\max}, F_{\min}), \text{ if } N_v(s) = n-1. \tag{5.36}$$

We now consider states s such that $0 < N_v(s) < n-1$. For such states, the heuristic policy chooses the highest-probability action that generates a link TTL greater than or equal to the smallest TTL of a viable link in memory. Explicitly, we set

$$\pi_{\mathrm{h}}(s) \coloneqq \underset{(p,F)\in\mathscr{A}}{\operatorname{argmax}}\left\{p :\ t_{\mathrm{TTL}}(F) \geq t_{N_v(s)} - 1\right\}, \quad \text{if } 0 < N_v(s) < n-1. \tag{5.37}$$

In particular, we see that the action taken in a given state s is only dependent on its viable links $v(s)$,

$$\pi_{\mathrm{h}}(s) = \pi_{\mathrm{h}}(v(s)) \text{ for all } s \in \mathscr{S}. \tag{5.38}$$

In (5.36) and (5.37), we have now fixed all actions apart from $(p_\varnothing, F_\varnothing)$, from (5.35). The choice of $(p_\varnothing, F_\varnothing)$ may be fixed by either policy evaluation or performing a Monte Carlo simulation of the system while employing π_{h} with each $(p_\varnothing, F_\varnothing) \in \mathscr{A}$. Then, one may select the value that minimises $\mathbb{E}[T_{\pi_{\mathrm{h}}}]$. This step involves a maximum of $|\mathscr{A}|$ policy evaluation steps (either carried out exactly by solving the Bellman equations (5.24), or approximately by performing a Monte Carlo simulation). The complexity of policy evaluation scales exponentially with n, but in practice the possibility to evaluate performance with simulation allows for the computation of the heuristic and its performance for regimes with larger state spaces than the optimal policy (see Section 5.4).

The policy π_{h} is particularly simple in the case where, for any $t \in \{1,\dots,t_{\max}\}$, there exists $(p,F) \in \mathscr{A}$ such that $t_{\mathrm{TTL}}(F) = t$. In other words, in $\mathscr{A}$ there are generation parameters that can produce a link with any TTL. In such a case, (5.37) is simply a *matching heuristic*, which ensures that all viable links have the same TTL, while maximising the success probability as much as possible.

The intuition behind our heuristic policy is as follows. With (5.37) the heuristic policy ensures that, when at least one viable link in memory will soon expire (small TTL), the heuristic policy tries to quickly generate all links within the remaining time. On the other hand, if all links in memory have a large TTL, the heuristic policy tries to generate similarly long-lasting links. This is a property we will also see in Section 5.4 in the structure of optimal policies: it is a trend that, as the links in memory decohere, the optimal policies prioritise higher-probability generation parameters. In fact, for one of the two parameter regimes considered, we see that the heuristic policy is in fact optimal.

5.4.3. PERFORMANCE COMPARISON

BASELINES

In the following, we compare our optimal and heuristic policies with two baseline policies. Firstly, we consider the *best constant-action policy* π_{con}, where

$$\pi_{\text{con}}(s) \coloneqq (p, F) \text{ for all } s \in \mathscr{S} \tag{5.39}$$

and $(p, F) \in \mathscr{A}$ is the value that minimises the performance $\mathbb{E}[T_{\pi_{\text{con}}}]$. The performance of this policy is well-studied with analytical methods in [58].

As a second baseline, we also consider the policy π_{ran} with uniformly random actions. Unlike all policies mentioned previously, this policy is non-deterministic. At each time step, the action is chosen from $\mathscr{A}$ uniformly at random:

$$\pi_{\text{ran}}(S_t) = (p, F) \text{ with prob. } \frac{1}{|\mathscr{A}|}, \quad \text{for all } (p, F) \in \mathscr{A}, \ t \geq 0. \tag{5.40}$$

PARAMETER REGIMES

In the performance evaluation, we consider generation parameters that correspond to batched executions of the single-click protocol. When photon loss is high, one execution of the single-click protocol has a very low success probability of generating entanglement [141]. Therefore, it can be beneficial for a single entanglement generation attempt to consist of M executions of the single-click protocol, which increases the probability of success while minimising overhead due to communication with higher layers of the software stack [29]. We also refer to these as *batched attempts*. The attempt is declared successful if at least one of the M single-click executions succeeds. In Appendix 5.6.2, we show that for batched single-click attempts, the trade-off between p and F is accurately approximated as

$$F = \lambda \ln\big(1 - p\big) + 1. \tag{5.41}$$

Here, $\lambda = 1/(2p_{\text{det}}M)$ is a fixed parameter that depends on the probability of detecting an emitted photon p_{det} and the batch size M. The relation (5.41) holds when $M \sim p_{\text{det}}^{-1}$ is large. The trade-off (5.41) is valid for a sufficiently small range $p \in (0, q]$. Given that all newly generated links must have fidelity F such that $F > F_{\text{app}}$, by (5.41) the maximum success probability q satisfies

$$q < 1 - e^{\frac{F_{\text{app}} - 1}{\lambda}}. \tag{5.42}$$

Noting that (5.41) enables a continuous choice of (p, F), we discretise the action space as follows. The maximum TTL of a newly generated link given the trade-off (5.41) is given by

$$t_{\max} = \lim_{p \to 0} t_{\mathrm{TTL}}(\lambda \ln(1-p) + 1) = t_{\mathrm{TTL}}(1). \tag{5.43}$$

Similarly, we let

$$t_{\min} = t_{\mathrm{TTL}}(\lambda \ln(1-q) + 1) \tag{5.44}$$

be the minimum TTL of a newly generated link. Then, for each $i \in \{t_{\min}, \ldots, t_{\max}\}$, we define p_i as the maximum probability with which one can generate a link with TTL i,

$$p_i = \max\{p : t_{\mathrm{TTL}}(\lambda \ln(1-p) + 1) = i\}. \tag{5.45}$$

We then work with the finite action space

$$\mathscr{A} = \left\{(p_i, \lambda \ln(1-p_i) + 1) : i \in \{t_{\min}, \ldots, t_{\max}\}\right\}. \tag{5.46}$$

We now compare the performance of our optimal and heuristic policies to our baselines in two parameter regimes. The first parameter regime has a high decoherence rate $\Gamma = 0.19$, which we refer to as the *near-term regime*. The second parameter regime has a low decoherence rate $\Gamma = 0.1$, which we refer to as the *far-term regime*. In both cases, we set the minimum required fidelity to be $F_{\mathrm{app}} = 1/2$.

The parameter regimes are chosen as follows. Suppose that the memory lifetime of each memory qubit consists of N executions of the single-click protocol. In recent experiments, $N \approx 5300$ [170]. Since an entangled link is stored in two qubits and decoherence acts on both of them, the total memory lifetime is then $N/2$ executions (for an explanation, see e.g. Supplementary Note 1 of [45]). Recalling that, in our model, M executions of the single-click protocol are batched into a single unit of time, the decoherence rate of the link is then $\Gamma = 2M/N$. The batch size in current experiments is $M \approx 500$–1000 [29, 15], but since this is not a hardware parameter we regard it as freely adjustable. See Table 5.1 for the specific choices of parameters for both the near-term and far-term regimes.

We note that the parameter choices put a restriction on the number of links n it is possible to have simultaneously in memory because necessarily $n \leq t_{\max}$, where $t_{\max}$ is the maximum TTL from (5.43). Then, a reduced Γ will increase $t_{\max}$, thereby allowing for more links to be present in memory. We note that, if n is increased, correspondingly each node must also contain at least n high-quality memories, which is experimentally challenging.

We ran our policy iteration experiments on a 2020 MacBook Air with an Apple M1 chip (8-core CPU), 16 GB unified memory, and a 256 GB SSD. Policy evaluation and simulation experiments were performed on a Asus ROG G14-GA401QM with an AMD Ryzen 9 5900HS (8-core CPU), 16 GB unified memory, and a 1 TB SSD. The limited memory and CPU resources restricted the size and complexity of our workloads. All code used in the experiments is publicly available [171].

NEAR-TERM REGIME

The parameters for this regime are $\Gamma = 0.19$, $\lambda = 2$, $F_{\mathrm{app}} = 1/2$ (see also Table 5.1). The action space $\mathscr{A}$ is given by (5.46). For the near-term regime, from (5.43) we have $t_{\max} = 6$, which is the maximum value of n.

Table 5.1: Parameters for near-term and far-term regimes

Parameter	Near-term	Far-term
N (memory lifetime in number of single-click executions)	5263.15 ≈ 5300 [170]	20000
p_{det} (photon detection probability)	5×10^{-4} $\approx 4.4 \times 10^{-4}$ [141]	5×10^{-4}
M (batch size)	500	1000
$\lambda = 1/(2p_{\text{det}}M)$ (batched single-click trade-off parameter (5.41))	2	1
$\Gamma = 2M/N$ (decoherence rate, (5.1))	0.19	0.1
$t_{\max}$ (maximum number of required links, (5.43))	6	11
F_{app} (minimum fidelity of links for application)	1/2	1/2

5.2 presents the expected completion times for the optimal policy, heuristic policy, and baselines. The optimal policy is only computed for $n \leq 5$, because for $n = 6$ our solver takes too long to converge due to the size of the state space becoming too large. See Appendix 5.6.1 for a detailed analysis of the state space scaling. The heuristic policy and the random-action policy performance are computed exactly by solving the corresponding Bellman equations (5.24). We note that, for the random-action policy, the Bellman equations currently written in (5.24) require a small generalisation to non-deterministic policies, which can e.g. be found in Chapter 3 of [163]. The optimal policy and its performance are computed with policy iteration (see Section 5.3.2). The performance of the constant-action policy is computed for all values of n with analytical methods from [58].

Remarkably, for all values of n for which the optimal policy can be computed ($n \leq 5$) we find that the heuristic policy is optimal, i.e. $\mathbb{E}[T_{\pi_h}] = \mathbb{E}[T_{\pi^*}]$. This is already expected for $n = 2$, because in that regime our heuristic is identical to the analytical solution for the optimal policy as presented in Section 5.4.1. However, for $n > 2$, the state space and transitions are more complex and one must use dynamic programming to show that a policy is optimal. In fact, the optimal policy turns out to match exactly with the heuristic policy (because multiple optimal policies can exist, this was not necessarily the case). Given the optimal performance of the heuristic for smaller n, one might reasonably expect that the performance of the heuristic policy is close-to-optimal for larger n, such as for $n = 6$ as is shown in Figure 5.2. From the figure, we see that the heuristic policy

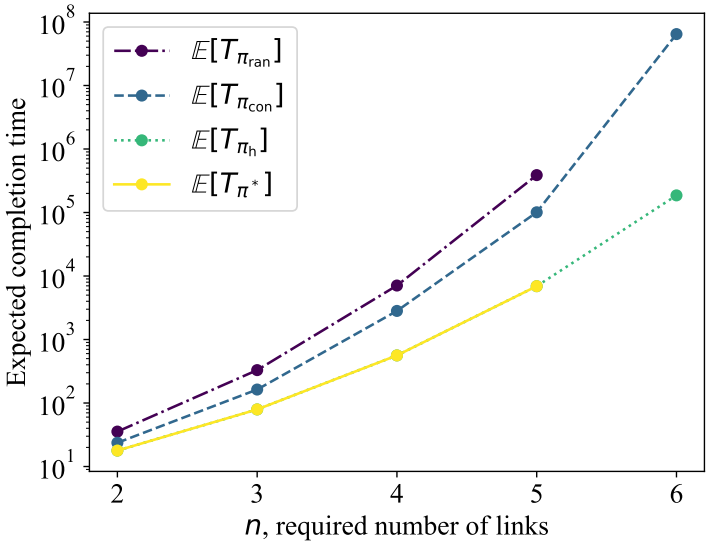


Figure 5.2: **Policy performance for the near-term regime.** The parameters for the near-term regime are $\Gamma = 0.19$, $\lambda = 2$, and $F_{\text{app}} = 1/2$ (see Table 5.1 and explanation in the main text). Plotted are the expected completion time for the optimal policy π^*, the heuristic policy π_{h}, the uniformly random policy π_{ran} and the constant-action policy π_{con}. There are no error bars as all points were computed either analytically or with policy evaluation.

provides an improvement over the constant-action policy by approximately two orders of magnitude. As discussed in Section 5.4.2, the complexity of computing the heuristic policy is in practice more efficient than finding the optimal policy with dynamic programming. For a parameter regime with a large state space where the optimal policy cannot be computed with dynamic programming, it is thus highly valuable to have this efficiently-computable heuristic policy that shows a close-to-optimal performance.

Figure 5.3 shows the ratio in performance of the optimal policy with the two baselines and the heuristic policy. As previously discussed, the heuristic policy provides optimal performance for the values of n investigated, meaning that the ratio is one. The advantage in performance increases with n, with the optimal (and heuristic) policies providing a performance increase of up to a factor of $\mathbb{E}[T_{\pi_{\text{con}}}]/\mathbb{E}[T_{\pi^*}] \approx 14$ for the constant-action policy and $\mathbb{E}[T_{\pi_{\text{ran}}}]/\mathbb{E}[T_{\pi^*}] \approx 56$ for the random-action policy.

In Figure 5.4, the structure of the optimal policy is shown as a heat map for $n = 5$. Specifically, it is shown how the action $\pi^*(s)$ depends on the number of viable links in the state $N_v(s)$ and the minimum TTL of a viable link $\min\{t : t \in v(s)\}$. We see that the optimal policy (and the heuristic policy) usually chooses high-probability (low-fidelity) actions for states containing more viable links and states containing viable links that will expire more quickly. By contrast, when there are zero viable links, the optimal policy increases the fidelity as much as possible at the expense of the success probability, choosing the action that generates a link with the maximum TTL $t_{\max} = 6$.

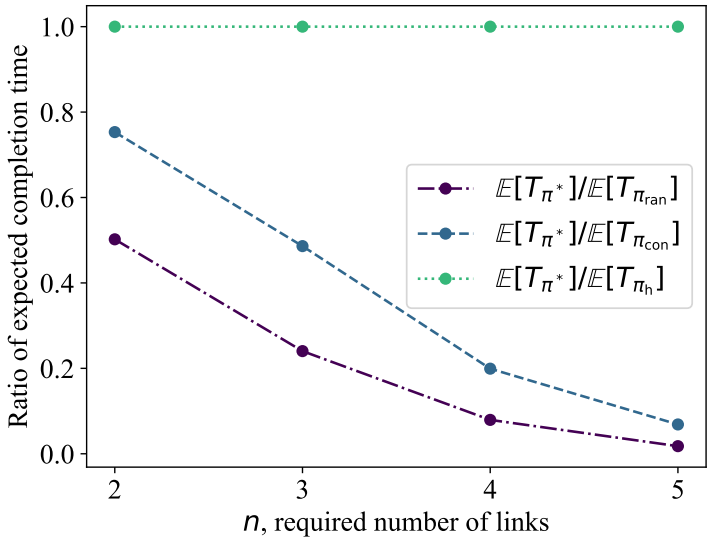


Figure 5.3: **Relative policy performance for the near-term regime.** Plotted are the ratio of the performance of the optimal policy π^* to the performance of the heuristic policy π_{h} and baselines π_{ran}, π_{con}. There are no error bars as all points were computed either analytically or with policy evaluation.

5

FAR-TERM REGIME

The parameters for this regime are $\Gamma = 0.1$, $\lambda = 1$, $F_{\text{app}} = 1/2$ (see also Table 5.1). The action space $\mathscr{A}$ is given by (5.46). For the far-term regime, from (5.43) we have $t_{\max} = 11$, which is the maximum value of the required number of links n. In the near-term regime, the maximum number of required links was $n = 6$. We thus see that the far-term regime has a larger state space for the same n (see Appendix 5.6.1).

Figure 5.5 presents the expected completion time for the optimal policy, heuristic and baselines in the far-term regime. As was the case for the near-term regime, due to the scaling of the state space, we are only able to compute optimal policies for $n \leq 7$. The random-action policy is computed with policy evaluation for $n \leq 6$ and Monte Carlo simulation for $n = 7$. The performance of the constant-action policy is computed analytically for all n with methods from [58]. The performance of the heuristic policy is computed with policy evaluation for $n \leq 8$ and Monte Carlo simulation for $n = 9, 10, 11$.

From Figures 5.5 and 5.6, we again observe that the heuristic policy maintains a remarkably close performance to the optimal policy for all values of n. As n increases, we again see a greater advantage provided by the optimal policy and heuristic policy over the baselines. For $n = 7$ we see the maximum improvement, where the optimal policy improves on the performance of both the constant-action policy by a factor of $\mathbb{E}[T_{\pi_{\text{con}}}]/\mathbb{E}[T_{\pi^*}] \approx 19$ and random-action policy by a factor of $\mathbb{E}[T_{\pi_{\text{ran}}}]/\mathbb{E}[T_{\pi^*}] \approx 139$.

We saw in the near-term regime that the heuristic policy is in fact optimal. In the far-term regime, we find that the heuristic is optimal for $n = 2$, which again is expected because for $n = 2$ the heuristic matches the analytical optimal policy from Section 5.4.1. For $n > 2$, we find that the heuristic policy is no longer optimal but still exhibits strong

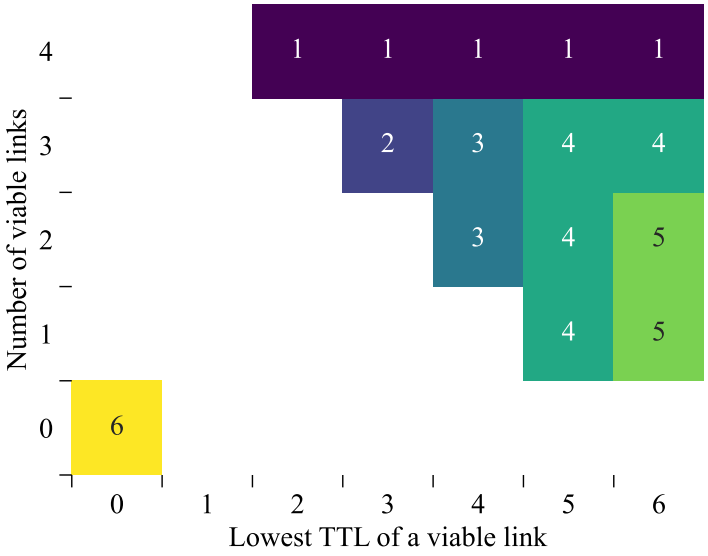


Figure 5.4: **Actions chosen by the optimal policy π^* for the near-term regime and $n = 5$ required links.** The x-axis indicates the lowest TTL of a link in a state s, $\min\{t : t \in v(s)\}$ from (5.34). The y-axis indicates the number of viable links $N_v(s)$ from (5.33). We note that there can be multiple states that take the same value on both the y-axis and x-axis. In the heat map, the most commonly-chosen action is shown for all states taking those values. A darker colour (lower number) means that the optimal policy prioritises success probability instead of fidelity. The number in each box is the TTL of the generated link corresponding to the most-commonly chosen action. We see that the optimal policy prioritises the success probability in states with more viable links and states with viable links that will expire more quickly. We note that certain states in the heat map are inaccessible, such as states with more than one viable link with the lowest TTL being six, since only a single link can be generated at a time. Nevertheless, they are displayed for clarity.

performance. Although the lines for the optimal policy and heuristic policy appear to overlap in Figure 5.5, we see in Figure 5.6 that for large n, the heuristic policy exhibits slightly worse performance than the optimal value. However, the deviation is not significant, and for all $n \leq 7$ the ratio of the two performances satisfies

$$\frac{\mathbb{E}[T_{\pi_h}] - \mathbb{E}[T_{\pi^*}]}{\mathbb{E}[T_{\pi^*}]} < 0.03. \tag{5.47}$$

Figure 5.7 visualises the optimal policy as a heat map for the case where the required number of links is $n = 7$. We again see the same patterns also seen from Figure 5.4: in the far-term regime, the optimal policy usually chooses high-probability (low-fidelity) actions for states containing more viable links and states containing viable links that will expire more quickly. However, in the far-term regime we also see these rules broken in certain circumstances. For example, consider a state s with one viable link $N_v(s) = 1$ with TTL $t_{N_v(s)} = 7$. In the state s, we see from Figure 5.7 that the optimal policy most commonly chooses the same generation parameters as it would for states with no viable links. A potential reason for this is that even though the single viable link may still survive until the remaining links are generated, the probability that it does so is very low. Thus,

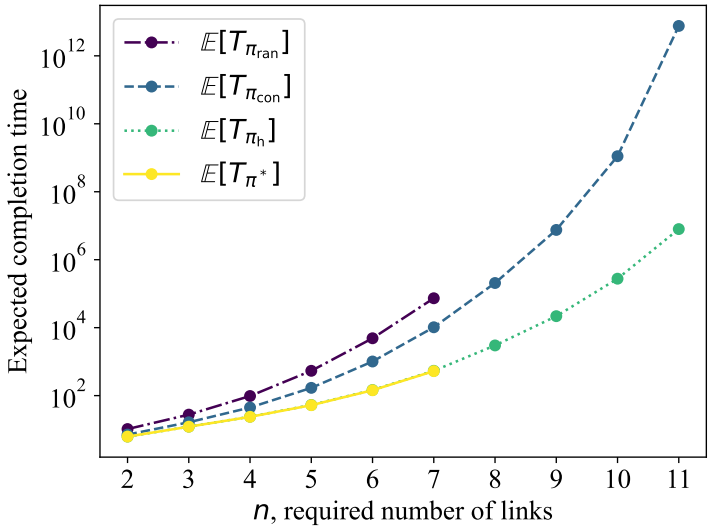


Figure 5.5: **Policy performance for the far-term regime.** The parameters for the far-term regime are $\Gamma = 0.19$, $\lambda = 2$, and $F_{\text{app}} = 1/2$ (see Table 5.1 and explanation in the main text). Plotted are the expected completion time for the optimal policy π^*, the heuristic policy π_{h}, the uniformly random policy π_{ran} and the constant-action policy π_{con}. Error bars are included for simulated values of $\mathbb{E}[T_{\pi_{\text{ran}}}]$ at $n = 7$ and $\mathbb{E}[T_{\pi_{\text{h}}}]$ at $n = 9, 10, 11$ with a confidence interval of three standard deviations, but are too small to be visible.

we learn that sometimes it is worth abandoning viable links that have a low TTL and directly start generating new links with a high TTL. Extending our heuristic to account for this is a potential avenue for improvement. We also notice that, for states with no viable links, the optimal policy chooses to generate a link with a TTL of ten, even though it is in principle possible to generate a link with a TTL of 11. We therefore see that the TTL (fidelity) should not necessarily be maximised when generating the first link, because the generation probability might be sacrificed too much. We already account for this in our heuristic policy, by making the initial action arbitrary in the step (5.35) and later optimising over this parameter.

5.5. CONCLUSION AND FUTURE WORK

We have considered a scenario where entanglement generation attempts are sequential in time and the generation parameters may be chosen in each time step. By formulating the problem as an MDP, we have found policies that minimise the expected time to generate multiple simultaneously-existing links. In the parameter regimes explored, we have seen that our optimal policies provide a significant improvement over the constant-action and random-action baselines. We have also found a heuristic method to compute policies that exhibit either optimal or close-to-optimal performance in all parameter regimes explored. The heuristic method is more efficient than finding optimal policies with dynamic programming, and we therefore expect it to be useful in situations where the optimal policy cannot be computed due to the scaling of the state space (e.g. when

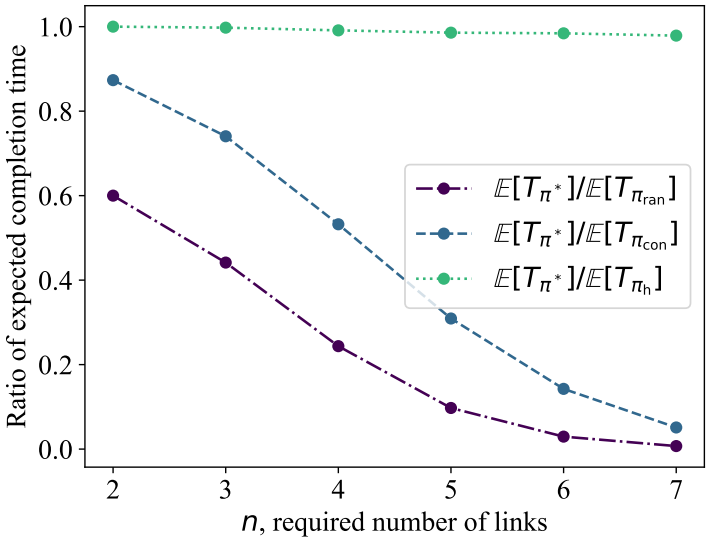


Figure 5.6: **Relative policy performance for the far-term regime.** Plotted are the ratio of the performance of the optimal policy π^* to the performance of the heuristic policy π_{h} and baselines π_{ran}, π_{con}. Error bars are included for $\mathbb{E}[T_{\pi^*}]/\mathbb{E}[T_{\pi_{\text{ran}}}]$ at $n = 7$ with a confidence interval of three standard deviations, but are too small to be visible.

the number of required links n is large). Our work highlights that adaptive protocols leveraging the rate-fidelity trade-off inherent to the entanglement generation scheme can be extremely helpful in improving quantum network performance. Moreover, in certain important entanglement generation schemes, the rate-fidelity trade-off is easily tunable (for example, by varying the bright-state parameter or mean photon number [141, 162]), and so our adaptive protocols are readily implementable.

Our model does not assume a specific entanglement generation protocol. In our results, we have used the example of a batched single-click protocol. It would also be interesting to study optimal policies for different entanglement generation schemes. A different entanglement generation scheme may result in a different action space, and therefore optimal policies with distinct properties to the batched single-click case.

For parameter regimes in which finding the optimal policy is not feasible with dynamic programming, a fruitful direction of research may be deep reinforcement learning, which finds approximate solutions.

5.6. APPENDIX

5.6.1. STATE SPACE SIZE

Here, we quantify exactly the size of the state space, $|\mathcal{S}|$. We recall from Section 5.3.1 that $\mathcal{S} = \bigcup_{m=1}^{n-1} \mathcal{S}_m \cup \{\varnothing\}$, where

$$\mathcal{S}_m = \left\{ \{t_1, \ldots, t_m\} : t_i \in [t_{\max}], t_1 \geq \cdots \geq t_m \right\}. \tag{5.48}$$

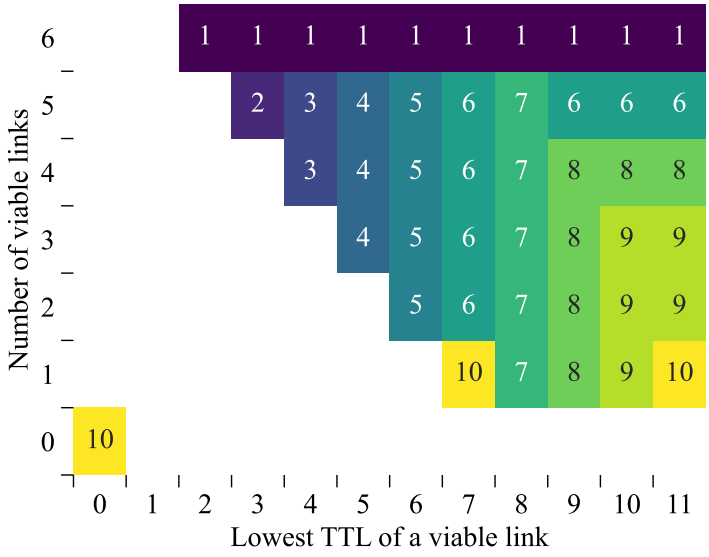


Figure 5.7: **Actions chosen by the optimal policy π^* for the far-term regime and $n = 7$ required links.** The x-axis indicates the lowest TTL of a link in the state s, $\min\{t : t \in v(s)\}$ from (5.34). The y-axis indicates the number of viable links $N_v(s)$ from (5.33). We note that there can be multiple states that take the same value on both the y-axis and x-axis. In the heat map, the most commonly-chosen action is shown for all states taking those values. A darker colour means that the optimal policy prioritises success probability instead of fidelity. The number in each box is the TTL of the generated link corresponding to the most-commonly chosen action. We see that the optimal policy prioritises the success probability in states with more viable links and states with viable links that will expire more quickly. However, there are some outliers, such as states with one viable link with a TTL of seven. We note that certain states in the heat map are inaccessible, such as states with more than one viable link with the lowest TTL being 11, since only a single link can be generated at a time. Nevertheless, they are displayed for clarity.

We firstly quantify $|\mathscr{S}_m|$. We note that each state in $\mathscr{S}_m$ corresponds to a unique outcome when choosing an element of $[t_{\max}]$ exactly m times, with replacement but without different permutations. Therefore, we have

$$|\mathscr{S}_m| = \left(\!\!\binom{t_{\max}}{m}\!\!\right), \tag{5.49}$$

where $\left(\!\binom{n}{k}\!\right)$ denotes the k-combination with repetitions. Then,

$$|\mathscr{S}| = |\{\varnothing\}| + \sum_{m=1}^{n-1} |\mathscr{S}_m| \tag{5.50}$$

$$= \sum_{m=0}^{n-1} \left(\!\!\binom{t_{\max}}{m}\!\!\right) \tag{5.51}$$

$$= \sum_{m=0}^{n-1} \binom{t_{\max}+m-1}{m}, \tag{5.52}$$

where in the first step we have used the fact that $\mathscr{S}_i$ and $\mathscr{S}_j$ are mutually exclusive sets for $i \neq j$, in the second step we have used (5.49), and in the last step we have used the formula

$$\left(\!\!\binom{n}{k}\!\!\right) = \binom{n+k-1}{k}. \tag{5.53}$$

We can further simplify (5.52) using the hockey-stick identity [172],

$$\sum_{j=0}^{w-r} \binom{j+r}{r} = \binom{w+1}{w-r} \tag{5.54}$$

for $w \geq r$. Then,

$$|\mathscr{S}| = \sum_{m=0}^{n-1} \binom{t_{\max}-1+m}{m} \tag{5.55}$$

$$= \sum_{m=0}^{n-1} \binom{t_{\max}-1+m}{t_{\max}-1} \tag{5.56}$$

$$= \binom{t_{\max}+n-1}{n-1}. \tag{5.57}$$

We can then bound the state space size as follows,

$$\binom{t_{\max}+n-1}{n-1} = \prod_{i=0}^{n-2} \frac{t_{\max}+n-1-i}{n-1-i} \tag{5.58}$$

$$\geq \prod_{i=0}^{n-2} \frac{t_{\max}+(n-1)}{n-1} \tag{5.59}$$

$$= \left(1+\frac{t_{\max}}{n-1}\right)^{n-1}. \tag{5.60}$$

To obtain (5.59), we have used the fact that

$$y = \frac{A-x}{B-x} \tag{5.61}$$

is an increasing function of x, for $A > B$.

By definition of our problem, we have $t_{\max} \geq n-1$. Then, by (5.60) we see that $|\mathscr{S}| \geq 2^{n-1}$, i.e. the size of the state space scales exponentially with n (within the feasible region $n \leq t_{\max}$).

We now quantify the reduction in the size of the state space that may be obtained by identifying states only with their viable links. Recall the function N_v that counts the number of viable links, defined in (5.33). We also recall the function v, such that $v(s)$ only contains the viable links of s, defined in (5.34). We define the reduced state space $\mathscr{S}_{\text{red}} \subset \mathscr{S}$ by identifying $s \equiv v(s)$ for all $s \in \mathscr{S}$. Then,

$$\mathscr{S}_{\text{red}} = \{s \in \mathscr{S} : v(s) = s\}. \tag{5.62}$$

We now compute $|\mathscr{S}_{\text{red}}|$. For $m = 0, \ldots, n-1$, we define

$$\mathscr{S}_{m,\text{red}} := \{s \in \mathscr{S}_{\text{red}} : N_v(s) = m\} \tag{5.63}$$

where we note that for $s \in \mathscr{S}_{\text{red}}$, we simply have $|s| = N_v(s)$, or equivalently all links are viable. The set $\mathscr{S}_{m,\text{red}}$ is interpreted as the set of states containing exactly m viable links. We note that $\mathscr{S}_{0,\text{red}} = \{\varnothing\}$. We also note that $\mathscr{S}_{m,\text{red}} \subset \mathscr{S}_m$. We then have

$$|\mathscr{S}_{\text{red}}| = \left|\bigcup_{m=0}^{n-1} \mathscr{S}_{m,\text{red}}\right| \tag{5.64}$$

$$= 1 + \sum_{m=1}^{n-1} \left|\mathscr{S}_{m,\text{red}}\right|, \tag{5.65}$$

where in the second step we have used the fact that $\mathscr{S}_{i,\text{red}}$ and $\mathscr{S}_{j,\text{red}}$ are mutually exclusive for $i \neq j$. We now compute $|\mathscr{S}_{m,\text{red}}|$. Recalling from Section 5.4.2 that the state $s = \{t_1, \ldots, t_m\}$ contains m viable links when $t_m > n - m$, it follows that

$$\mathscr{S}_{m,\text{red}} = \{\{t_1, \ldots, t_m\} : t_i \in [t_{\max}] \setminus [n-m], t_1 \geq \cdots \geq t_m\}. \tag{5.66}$$

Then, we deduce $|\mathscr{S}_{m,\text{red}}|$ by the same counting argument as was used above for $|\mathscr{S}_m|$, to find

$$|\mathscr{S}_{m,\text{red}}| = \left(\!\!\binom{t_{\max} + m - n}{m}\!\!\right). \tag{5.67}$$

Combining with (5.65), we then have

$$|\mathscr{S}_{\text{red}}| = 1 + \sum_{m=1}^{n-1} \left(\!\!\binom{t_{\max} + m - n}{m}\!\!\right) \tag{5.68}$$

$$= 1 + \sum_{m=1}^{n-1} \binom{t_{\max} + 2m - n - 1}{m}, \tag{5.69}$$

where we again made use of (5.53).

5.6.2. TRADE-OFF RELATION FOR BATCHED SINGLE-CLICK SCHEME

Here, we derive a trade-off relation between p and F for batched attempts of the single-click entanglement generation protocol [23]. In the limit of high photon losses ($\eta \ll 1$), if the success probability of a single execution of the single-click protocol is p_{succ}, the fidelity of a generated link is

$$F = 1 - \frac{p_{\text{succ}}}{2p_{\text{det}}}, \tag{5.70}$$

where p_{det} is the probability that an emitted photon is detected [14]. When $p_{\text{succ}} \ll 1$, as is typically the case for NV centres, it can be beneficial to perform attempts in batches [29] in order to avoid overhead in the software stack (the outcome of each attempt must be communicated to higher layers of the software stack). We now consider the case

where one entanglement generation attempt corresponds to M executions of the single-click protocol. If at least one of the M execution succeeds, then the entanglement generation attempt is successful. The success probability of an individual attempt is therefore given by

$$p = 1 - (1 - p_{\text{succ}})^M. \tag{5.71}$$

Letting $\lambda := 1/(2p_{\text{det}}M)$, from (5.70) we have that

$$p = 1 - \left(1 - \frac{(1-F)}{\lambda M}\right)^M. \tag{5.72}$$

Then, using the fact that $\lim_{K\to\infty}(1+\frac{x}{K})^K = e^x$, when $M \sim p_{\text{det}}^{-1}$ is large the above is approximated as

$$p \approx 1 - e^{-\frac{1-F}{\lambda}}, \tag{5.73}$$

which results in the trade-off relation

$$F \approx 1 + \lambda \ln(1-p). \tag{5.74}$$

The relation (5.74) is used in the illustration of our results in Section 5.4.3. The trade-off relation $F \approx 1 - \lambda p$ that was found in [58] can be seen as the first-order approximation of (5.74) at $p = 0$.

We now show that (5.74) is an accurate approximation in the regime where $M \sim p_{\text{det}}^{-1}$ is large. To bound the error of the approximation, we use the following inequality:

$$\left(1 - \frac{1}{y}\right)^{y-1} > \frac{1}{e} > \left(1 - \frac{1}{y}\right)^{y}, \quad \text{for } y > 1. \tag{5.75}$$

The proof of (5.75) is given in Appendix 5.6.3. We recall that the true trade-off is given by

$$\tilde{p} = 1 - \left(1 - \frac{(1-F)}{\lambda M}\right)^M \tag{5.76}$$

and the approximate trade-off

$$p = 1 - e^{-\frac{1-F}{\lambda}}. \tag{5.77}$$

We firstly claim that $p < \tilde{p}$. Letting $r = (1-F)/\lambda$ and $y = M/r$, we have $r > 0$. Then,

$$\tilde{p} - p = e^{-r} - \left(1 - \frac{r}{M}\right)^M \tag{5.78}$$

$$= e^{-r} - \left(1 - \frac{1}{y}\right)^{yr} > 0, \tag{5.79}$$

where in the final step we have made use of the upper bound from (5.75). We now bound

the difference $|p-\tilde{p}|$. We have

$$|\tilde{p}-p| = e^{-r} - \left(1-\frac{1}{y}\right)^{yr} \tag{5.80}$$

$$< e^{-r} - \left(e^{-1}\left(1-\frac{1}{y}\right)\right)^{r} \tag{5.81}$$

$$= e^{-r}\left(1-\left(1-\frac{1}{y}\right)^{r}\right) \tag{5.82}$$

$$< 1-\left(1-\frac{1}{y}\right)^{r}, \tag{5.83}$$

where in the second inequality we have used the lower bound from (5.75), and in the final inequality we have used $e^{-r}<1$ (since $r>0$). In particular, making use of the power series expansion, we see that

$$|\tilde{p}-p| \approx \frac{r}{y} = \mathcal{O}\left(\frac{1}{M}\right), \tag{5.84}$$

i.e. as long as $M \sim p_{\text{det}}^{-1}$ is large, our approximation remains accurate.

5

5.6.3. PROOF OF (5.75)

To prove the left inequality of (5.75), we firstly take the logarithm of both sides, to obtain

$$(y-1)\ln\left(1-\frac{1}{y}\right) > -1, \text{ for } y>1. \tag{5.85}$$

For ease of notation, we let

$$f(y) \coloneqq (y-1)\ln\left(1-\frac{1}{y}\right). \tag{5.86}$$

To prove (5.85), we firstly claim that

$$\lim_{y\to\infty} f(y) = -1. \tag{5.87}$$

Letting $z \coloneqq 1/y$, (5.87) is equivalent to

$$\lim_{z\to 0}\left(\frac{1-z}{z}\right)\ln(1-z) = -1, \tag{5.88}$$

which one may verify with L'Hôpital's rule: differentiating both the numerator and denominator,

$$\lim_{z\to 0}\frac{(1-z)\ln(1-z)}{z} = \lim_{z\to 0}\frac{-\ln(1-z)-1}{1} \tag{5.89}$$

$$= -1. \tag{5.90}$$

Having shown (5.87), to prove (5.85) it suffices to show that the function $f(y)$ is decreasing. We have

$$\frac{\mathrm{d}f}{\mathrm{d}y} = \ln\left(1-\frac{1}{y}\right) + (y-1)\frac{\frac{1}{y^2}}{1-\frac{1}{y}} \tag{5.91}$$

$$= \ln\left(1-\frac{1}{y}\right) + \frac{1}{y}. \tag{5.92}$$

Since

$$\ln(1+z) < z \text{ for } z > -1, \tag{5.93}$$

by letting $y = -1/z$ it follows from the above that $\frac{\mathrm{d}f}{\mathrm{d}y} < 0$, which suffices to prove the left inequality of (5.75).

We now show the right inequality of (5.75), which taking the logarithm of both sides is equivalent to

$$-\frac{1}{y} > \ln\left(1-\frac{1}{y}\right), \text{ for } y > 1. \tag{5.94}$$

Again, letting $y = -1/z$, by (5.93) this holds.

6

ON THE ACCURACY OF TWIRLED APPROXIMATIONS IN REPEATER CHAINS

Bethany Davies, Guus Avis and Stephanie Wehner

In the performance analysis of quantum networks, it is common to approximate bipartite entangled states as either being Bell-diagonal or Werner states. We refer to these as twirled approximations because it is possible to bring any state to such a form with a twirling map. Although twirled approximations can simplify calculations, they can lead to an inaccuracy in performance estimates. The goal of this work is to quantify this inaccuracy. We consider repeater chains where end-to-end entanglement is achieved by performing an entanglement swap at each repeater in the chain. We consider two scenarios: postselected and non-postselected entanglement swapping, where postselection is performed based on the Bell-state measurement outcomes at the repeaters. We show that, for non-postselected swapping, the Bell-diagonal approximation is exact for the computation of the Bell-diagonal elements of the end-to-end state. We find that the Werner approximation accurately approximates the end-to-end fidelity when the infidelity of each initial state is small with respect to the number of repeaters in the chain. For postselected swapping, we find bounds on the difference in end-to-end fidelity from what is obtained with the twirled approximation, for initial states with a general noisy form.

This chapter has been released separately at https://arxiv.org/abs/2509.16689

6.1. INTRODUCTION

A common form of quantum repeater utilises *entanglement swapping* [32, 12, 173]. Consider a simple scenario with two end nodes, each equipped with a single qubit, and an intermediate (repeater) node equipped with two qubits. Suppose that the repeater node shares the entangled two-qubit states ρ_1 and ρ_2 with each end node, such that the total initial state of the chain is $\rho_1 \otimes \rho_2$. An entanglement swap transforms $\rho_1 \otimes \rho_2$ into a two-qubit state ρ' shared between the end nodes. An entanglement swap consists of a Bell-state measurement (BSM) at the repeater node and classical communication of the BSM outcome to the end nodes, followed by local Pauli corrections at the end nodes. If the level of noise in the initial states, the BSM and the Pauli corrections is low enough, then the end-to-end state ρ' will be entangled (see Figure 6.1a).

In a *repeater chain*, $N-1$ repeater nodes are placed between the end nodes. Entanglement is firstly shared between adjacent nodes, in the form of N entangled two-qubit states $\bigotimes_{k=1}^{N} \rho_k$. End-to-end entanglement is achieved by performing an entanglement swap at each repeater node.

Swapping-based repeaters are the form of repeater most within experimental reach, and present-day demonstrations of quantum networks have distributed entanglement in such a way — see e.g. [14, 170]. For this reason, theoretical work on the design and performance analysis of large-scale quantum networks typically uses swapping-based repeaters as a basic assumption in the network model [80, 63, 75, 174, 175, 156, 89, 169, 78].

By the *Bell-diagonal* and *Werner* approximations of the two-qubit state ρ, we respectively refer to the states

$$\mathscr{B}(\rho) = \sum_{i,j=0}^{1} \lambda_{ij} \left|\Psi_{ij}\right\rangle\!\left\langle\Psi_{ij}\right|, \tag{6.1}$$

$$\mathscr{W}(\rho) = \frac{4F-1}{3} \left|\Psi_{00}\right\rangle\!\left\langle\Psi_{00}\right| + \frac{(1-F)}{3} I_4 \tag{6.2}$$

such that $\lambda_{ij} = \left\langle\Psi_{ij}\right| \rho \left|\Psi_{ij}\right\rangle$, $F = \lambda_{00}$ is the fidelity with respect to $|\Psi_{00}\rangle$, and I_4 is the identity matrix. Both approximations are equivalent to applying the symmetrising map $\mathscr{B}$ ($\mathscr{W}$) to ρ [135, 47], which is also known as *twirling* the state ρ. We therefore refer to the Bell-diagonal and Werner approximations as *twirled approximations*.

When modelling states in a repeater chain, it can be convenient to use twirled approximations for the initial states of the chain. See Figure 6.1b for an illustration of how twirled approximations are used in a repeater chain. Without the the approximation, we input a pair of two-qubit initial states $\rho_1 \otimes \rho_2$, and after the entanglement swap we have end-to-end state ρ'. With the approximation, each initial state is twirled with the map $\mathscr{B}$ ($\mathscr{W}$), and after the entanglement swap the end-to-end state is $\rho'_{\mathscr{B}}$ ($\rho'_{\mathscr{W}}$).

Twirled approximations have a symmetrised form, which requires only a few parameters to be specified and has a direct operational interpretation. For the Bell-diagonal approximation (6.1), there are three parameters λ_{01}, λ_{11}, and λ_{10}, which are interpreted as the probabilities of X, Y and Z errors when applying a Pauli channel to the state $|\Psi_{00}\rangle$ to obtain the noisy state $\mathscr{B}(\rho)$. For the Werner approximation (6.2), only a single parameter is required, which is the fidelity F [135]. Another important property is that the

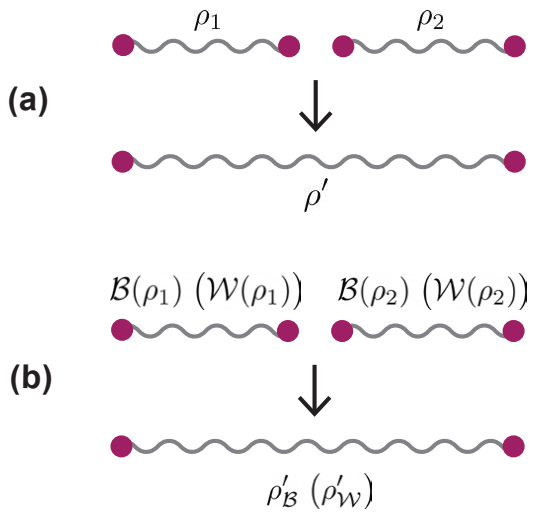


Figure 6.1: **Twirled approximations in a repeater chain with $N = 2$ initial states.** (a) The entanglement swapping of a pair of two-qubit entangled states $\rho_1 \otimes \rho_2$ results in the end-to-end state ρ'. (b) With the Bell-diagonal (Werner) approximation $\mathcal{B}(\rho_k)$ ($\mathcal{W}(\rho_k)$) of the initial states, the end-to-end state $\rho'_{\mathcal{B}}$ ($\rho'_{\mathcal{W}}$) is simpler to compute than in case (a).

symmetric form is preserved after entanglement swapping, i.e. $\rho'_{\mathcal{B}}$ is Bell-diagonal and $\rho'_{\mathcal{W}}$ is Werner. Consequently, $\rho'_{\mathcal{B}}$ ($\rho'_{\mathcal{W}}$) is often simpler to compute than ρ', which has many advantages in the performance analysis of large-scale quantum networks, potentially with complex topologies. In particular, twirled approximations enable the analytical study of high-level network performance metrics, because the metrics may then be more easily understood as a function of the initial states and therefore low-level properties of the hardware [75, 64, 56, 176, 174, 175]. Twirled approximations enable a more efficient numerical simulation and optimisation of large-scale quantum networks [89, 169, 78, 62, 177, 178]. They are also used to avoid making overly specific assumptions when modelling quantum hardware [73, 79, 178], because any noise model can in principle be transformed into such a case with twirling.

Despite their advantages, twirled approximations can cause inaccuracies in performance estimates. For example, suppose that the initial states $\rho_1 \otimes \rho_2$ have the same fidelity $\langle\Psi_{00}|\rho_k|\Psi_{00}\rangle = F$, for $k = 1,2$. Using the Werner approximation for both states in this scenario, the initial state is $\mathcal{W}(\rho_1) \otimes \mathcal{W}(\rho_2)$, and the end-to-end state $\rho'_{\mathcal{W}}$ is therefore also Werner with fidelity $F'_{\mathcal{W}} = \langle\Psi_{00}|\rho'_{\mathcal{W}}|\Psi_{00}\rangle = F^2 + (1-F)^2/3$ [173]. However, depending on the exact form of the initial states, the true fidelity $F' = \langle\Psi_{00}|\rho'|\Psi_{00}\rangle$ of the end-to-end state can lie (potentially significantly) above or below the value of $F'_{\mathcal{W}}$. The same holds if we use the Bell-diagonal approximations to obtain end-to-end fidelity $F'_{\mathcal{B}}$. The principle question that we address in our work is: *what is the maximum difference* $|F' - F'_{\mathcal{B}}|$ ($|F' - F'_{\mathcal{W}}|$) *between the true end-to-end fidelity* F' *and the end-to-end fidelity with the twirled approximation* $F'_{\mathcal{B}}$ ($F'_{\mathcal{W}}$)*?*

We consider two different scenarios: *postselected* and *non-postselected* swapping. In postselected swapping, the end-to-end state $\rho'_{\vec{s}}$ is postselected on the BSM outcomes $\vec{s} = (s_1, \ldots, s_{N-1})$ obtained at the $N-1$ repeater nodes. In non-postselected swapping, the end-to-end state is a weighted average of the postselected outcomes. Let $p'_{\vec{s}}$ be the probability of measuring $\vec{s}$. Then, the end-to-state after non-postselected swapping is $\sum_{\vec{s}} p'_{\vec{s}} \rho'_{\vec{s}}$. We note that this is equivalent to not having knowledge of the specific BSM outcomes $\vec{s}$ obtained (although we note that, in the swapping protocol, the BSM outcomes are still commicated and Pauli corrections are still applied).

Having introduced the problem, we now outline our main contributions. In Section 6.4, we study the case of **non-postselected swapping** on a chain with N initial states $\bigotimes_{k=1}^{N} \rho_k$ and $N-1$ repeaters, we consider a general class of entanglement-swapping protocols that we term *swap-and-correct* protocols (see Definition 6.3). Swap-and-correct protocols consist of BSMs and Pauli corrections that can be applied at any node in the chain. For all such protocols, we show that:

(i) $\mathscr{B}(\rho') = \rho'_{\mathscr{B}}$, i.e. the Bell-diagonal approximation is exact for the computation of the Bell-diagonal components of the end-to-end state (Theorem 6.1). The Bell-diagonal components include the fidelity, and so $F' = F'_{\mathscr{B}}$.

(ii) If the initial fidelities $F_k = \langle \Psi_{00} | \rho_k | \Psi_{00} \rangle$ satisfy $1 - F_k \ll 1/N$ for all $k = 1, \ldots, N$, then $F' \approx F'_{\mathscr{W}}$, i.e. the Werner approximation accurately approximates the end-to-end fidelity. More precisely, we have $|F' - F'_{\mathscr{W}}| = \mathcal{O}((1-F)^2 N^2)$ (Theorem 6.2).

A key insight is that, in many important cases, non-postselected swapping and the use of the Bell-diagonal approximation are equivalent. Such cases include protocols whose performance may be expressed solely in terms of the Bell-diagonal elements $\mathscr{B}(\rho')$ of the end-to-end state ρ'. For example, consider the channel $\Lambda^{\text{tel}}_{\rho'}$ induced by standard quantum teleportation with resource state ρ' [159]. In Proposition 6.1, we show that $\Lambda^{\text{tel}}_{\rho'} = \Lambda^{\text{tel}}_{\mathscr{B}(\rho')}$. By result (i), we thus have $\Lambda^{\text{tel}}_{\mathscr{B}(\rho')} = \Lambda^{\text{tel}}_{\rho'_{\mathscr{B}}}$. In particular, the Bell-diagonal approximation for each initial state in the chain is exact when subsequently performing teleportation over the end-to-end state. However, exactness does not hold for all applications: we also see that for quantum key distribution, for certain input states using the twirled approximation can lead to a large reduction in performance (Section 6.5.2).

In Section 6.4, we study the case of **postselected swapping** on a repeater chain with two initial states $\rho_1 \otimes \rho_2$ and one repeater, we restrict to swapping initial states of the form $\rho_k = p |\Psi_{00}\rangle\langle\Psi_{00}| + (1-p)\sigma_k$ such that $F = \langle \Psi_{00} | \rho_k | \Psi_{00} \rangle$. The density matrix σ_k is interpreted as an arbitrary noise term. We fix F to perform a comparison with the twirled approximation, and also fix $p \in [0, F]$ to provide meaningful bounds. (It is necessary to fix a second parameter because for any F, there exists a state ω with $F = \langle \Psi_{00} | \omega | \Psi_{00} \rangle$ such that there is a probability $p'_s > 0$ of obtaining $\rho'_s = |\Psi_{00}\rangle\langle\Psi_{00}|$ when swapping the initial states $\omega^{\otimes 2}$ [179]. Thus, if F is the only fixed parameter, it is always possible to obtain unit end-to-end fidelity with some non-zero probability.) Letting F'_s denote the end-to-end fidelity postselected on BSM outcome s, we find:

(iii) A tight, analytical upper bound for the achievable end-to-end fidelity, $F'_s \leq 1 - 2p(1-F)$ (Theorem 6.3). We show tightness by finding an example of a state ρ_{opt}

such that $\rho_{\text{opt}}^{\otimes 2}$ achieves this upper bound. The state ρ_{opt} has a physical interpretation: it corresponds to applying a Y-rotation (of a specified angle) with probability $1-p$ to a qubit of $|\Psi_{00}\rangle$.

(iv) A simple, analytical lower bound for F_s' in terms of p and F. Unlike the upper bound, this is not tight. Moreover, we find a tighter lower bound for F_s' by formulating the problem as a semi-definite program. We perform a symmetry reduction of the problem, enabling efficient computation of the bound.

Our simple formulation with the parameters p and F, as well as the efficiently computable bounds, allows for direct interpretation and comparison with twirled approximations (see Section 6.5.1). For example, let us consider swapping the initial states $\rho_R^{\otimes 2}$, where

$$\rho_R = p\,|\Psi_{00}\rangle\langle\Psi_{00}| + (1-p)\,|01\rangle\langle 01|\,.$$

The state ρ_R (up to a local unitary rotation) closely approximates states generated in certain physical entanglement generation schemes [127, 23]. It has

$$p = \langle\Psi_{00}|\,\rho_R\,|\Psi_{00}\rangle = F.$$

By (iii), we see that

$$F_s'(\rho_R^{\otimes 2}) \leq 1 - 2F(1-F) = F^2 + (1-F)^2.$$

Therefore, $F_s'(\rho_R^{\otimes 2}) - F_{\mathscr{W}}' \leq 2(1-F)^2/3$. For large F, we see that the Werner approximation does not cause a large reduction in the maximum end-to-end fidelity.

Building on this example, in Section 6.5.1, we provide further discussion of how our bounds may be used to assess whether the twirled approximation is accurate for given input states. In Section 6.5.2, we further discuss the implications of our results for an explicit example application: specifically, we look at the impact of twirled approximations for the performance of quantum key distribution over a repeater chain.

6.2. RELATED WORK

The idea that entanglement can increase (or decrease) after a postselected entanglement swap has existed for many years [179]. Much work has since focused on a fundamental investigation of how much the entanglement can change after the entanglement swap, when compared to the initial states [180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190]. To answer this question, in many studies, the concurrence is used as a measure of entanglement [180, 181, 183, 184, 185, 186, 187, 189, 190] because for this there exists a computable formula in the two-qubit case [70]. Other work has used the negativity [188], or instead of measures of entanglement, used measures of quantum correlation [182].

In the simplest case, entanglement swapping can be seen as applying teleportation to one end of an entangled state [159]. Much work has focused on the analysis and optimisation of quantum teleportation with a noisy resource state – see e.g. [191, 192, 193, 194]. However, to our knowledge, no systematic comparison has been performed with the twirled approximation. Moreover, the idea of a postselected swap is related to that of probabilistic teleportation [195, 196, 197], where a qubit may be teleported with maximum fidelity, even if the resource state not maximally entangled. This is often made

possible by measuring in a non-maximally-entangled basis and postselecting based on the measurement outcome. Other work has focused on the effect of noise in the resource state on probabilistic teleportation [198]. Again, to our knowledge, no systematic comparison of (probabilistic) teleportation has been carried out with the twirled approximation of the resource state.

Finally, we note that twirling is a technique that is used widely outside of the context of repeater chains. For example, twirling is used in quantum error correction to reduce a general noise channel to a Pauli channel [46]. There has been work on quantifying the accuracy of such an approximation for the calculation of the error correction threshold [199, 200]. Twirling is also used as a simplifying step in security proofs [4]. Twirling is also used in randomised benchmarking [201, 202], not as an approximation to the noise model, but as a tool that can be applied to extract important information about a noisy gate set.

6.3. NON-POSTSELECTED SWAPPING

6.3.1. PRELIMINARIES

Let X, Y and Z denote the usual Pauli gates, given by

$$X \equiv \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y \equiv \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z \equiv \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{6.3}$$

In what follows, we will denote the Bell basis vectors as

$$\left|\Psi_{ij}\right\rangle \coloneqq I_2 \otimes X^i Z^j \left[\frac{1}{\sqrt{2}}\left(|00\rangle + |11\rangle\right)\right], \tag{6.4}$$

where $i, j \in \{0, 1\}$.

In the following lemmas, we restate the well-known results for the Bell-diagonal and Werner twirling of a two-qubit state.

Lemma 6.1 (Bell-diagonal twirl)**.** *Suppose that Alice and Bob share the two-qubit state ρ and each apply the same gate chosen from $\{I_2, X, Y, Z\}$ uniformly at random, creating the channel*

$$\rho \mapsto \mathscr{B}(\rho) = \frac{1}{4}\Big[\rho + (X\otimes X)\rho(X^\dagger \otimes X^\dagger) + (Y\otimes Y)\rho(Y^\dagger \otimes Y^\dagger) + (Z\otimes Z)\rho(Z^\dagger \otimes Z^\dagger)\Big]. \tag{6.5}$$

Then, $\mathscr{B}(\rho)$ is diagonal in the Bell basis and $\left\langle\Psi_{ij}\right|\mathscr{B}(\rho)\left|\Psi_{ij}\right\rangle = \left\langle\Psi_{ij}\right|\rho\left|\Psi_{ij}\right\rangle$, for all $i, j \in \{0, 1\}$. In other words, the eigenvalues of $\mathscr{B}(\rho)$ are given by the diagonal elements of ρ when written in the Bell basis.

Proof. See Appendix A of [47]. □

Lemma 6.2 (Werner twirl)**.** *Suppose that Alice and Bob share the two-qubit state ρ. Alice applies the unitary U to register A and Bob applies U^* to register B. The unitary U is chosen uniformly at random from the Haar measure, creating the channel*

$$\rho \mapsto \mathscr{W}(\rho) = \int \left(U \otimes U^*\right)\rho\left(U \otimes U^*\right)^\dagger \mathrm{d}U. \tag{6.6}$$

Then, the resultant state is of the form

$$\mathscr{W}(\rho) = \frac{4F-1}{3} |\Psi_{00}\rangle\langle\Psi_{00}| + \frac{1-F}{3} I_4, \tag{6.7}$$

where $F = \langle\Psi_{00}|\rho|\Psi_{00}\rangle$ *is the fidelity of* ρ *to* $|\Psi_{00}\rangle$*, and* I_4 *is the identity matrix.*

Proof. See Section V of [48]. □

In what follows, we refer to states of the form (6.7) as *Werner states* [135], which is standard terminology in the field of quantum networks (see e.g. [80, 173, 19]). States of the form (6.7) are also sometimes referred to as isotropic states, which are the states that are invariant under the application of $U \otimes U^*$, for any unitary U. However, note that the term *Werner state* is also sometimes used to refer to the states which have $U \otimes U$ symmetry, which were originally studied in [135]. These are equivalent to the states (6.7) up to a Pauli Y rotation on one of the two qubits. We note that in order to avoid sampling unitaries uniformly from the Haar measure, which can be computationally expensive and difficult to realise experimentally, it is possible to implement the map (6.6) instead by sampling from a finite set of correlated Pauli gates [47].

Definition 6.1. Given a two-qubit state ρ, we refer to $\mathscr{B}(\rho)$ as the *Bell-diagonal approximation* of ρ. We refer to $\mathscr{W}(\rho)$ as the *Werner approximation* of ρ.

In this work, we refer to the Bell-diagonal (Werner) approximation *in a repeater chain* as when the approximation is used for all initial states in the chain (see Figure 6.1).

6.3.2. REPEATER CHAINS WITH $N = 2$ INITIAL STATES

In this section, we consider non-postselected swapping on repeater chains with $N = 2$ initial states. This section can be seen as a warm-up for our main results, which are presented in Sections 6.3.3 and 6.4.

The entanglement swapping protocol for $N = 2$ initial states that we consider in this work is implemented using the standard teleportation protocol from [159]. Given that two parties initially share entanglement, the standard teleportation protocol uses a BSM, classical communication and Pauli corrections to transport a quantum state between the two parties. The protocol is given in Algorithm 1 and illustrated in Figure 6.2a. If all states and measurements involved are perfect, then the protocol teleports a quantum state perfectly and with probability one. If the initial shared entanglement is noisy, the teleportation protocol effectively sends the quantum state down a noisy channel. Properties of such channels, such as the teleportation fidelity, have been widely studied [203]. The shared entangled state used for teleportation is also referred to as a *resource state*, because this is consumed in the protocol in order to teleport the target state.

Algorithm 1 The standard teleportation protocol [159]

1: Input qubit state σ_C, two-qubit resource state ρ_{AB}.
2: Perform measurement on registers CA in Bell basis (6.4), to obtain outcome ij.
3: Apply correction $Z^j X^i$ on register B.

Postselected on BSM outcome ij, the output state after standard teleportation is

$$\sigma'_{ij} = \frac{1}{p'_{ij}} (Z^j X^i)_B \operatorname{Tr}_{CA} \left[\left|\Psi_{ij}\right\rangle\!\left\langle\Psi_{ij}\right|_{CA} \sigma_C \otimes \rho_{AB} \right] (Z^j X^i)_B^\dagger, \tag{6.8}$$

where

$$p'_{ij} = \operatorname{Tr} \left[\left|\Psi_{ij}\right\rangle\!\left\langle\Psi_{ij}\right|_{CA} \sigma_C \otimes \rho_{AB} \right] \tag{6.9}$$

is the associated probability of obtaining measurement outcome ij. The weighted average of the postselected outcomes is given by

$$\sigma' = \sum_{i,j=0}^{1} \sigma'_{ij} p'_{ij} \tag{6.10}$$

$$= \sum_{i,j=0}^{1} (Z^j X^i)_B \left\langle\Psi_{ij}\right| \sigma_C \otimes \rho_{AB} \left|\Psi_{ij}\right\rangle_{CA} (Z^j X^i)_B^\dagger \tag{6.11}$$

$$=: \Lambda_\rho^{\text{tel}}(\sigma), \tag{6.12}$$

where $\Lambda_\rho^{\text{tel}}$ is the channel induced by standard teleportation with resource state ρ.

Proposition 6.1 (Exactness of Bell-diagonal approximation for teleportation)**.** *Let $\Lambda_\rho^{\text{tel}}(\sigma)$ denote the result of teleporting a qubit state σ (register C) with a two-qubit resource state ρ (registers AB) using the non-postselected standard teleportation protocol, as defined in (6.12). Let $\mathscr{B}(\rho)$ denote the Bell-diagonal approximation of ρ. Then,*

$$\Lambda_\rho^{\text{tel}}(\sigma) = \Lambda_{\mathscr{B}(\rho)}^{\text{tel}}(\sigma), \tag{6.13}$$

i.e. the channel $\Lambda_\rho^{\text{tel}}$ is invariant under the Bell-diagonal twirling of ρ.

Proof. Recalling that $\left|\Psi_{ij}\right\rangle_{CA} = (X^i Z^j)_A \left|\Psi_{00}\right\rangle_{CA}$, we may rewrite (6.11) in the following way: bringing the sum inside the inner product and using the identity (6.5) for Bell-diagonal twirling yields

$$\Lambda_\rho^{\text{tel}}(\sigma) = 4 \left\langle\Psi_{00}\right| \sigma_C \otimes \mathscr{B}(\rho_{AB}) \left|\Psi_{00}\right\rangle_{CA}. \tag{6.14}$$

Now,

$$\Lambda_{\mathscr{B}(\rho)}^{\text{tel}}(\sigma) = 4 \left\langle\Psi_{00}\right| \sigma_C \otimes \mathscr{B}^2(\rho_{AB}) \left|\Psi_{00}\right\rangle_{CA} \tag{6.15}$$

$$= 4 \left\langle\Psi_{00}\right| \sigma_C \otimes \mathscr{B}(\rho_{AB}) \left|\Psi_{00}\right\rangle_{CA} = \Lambda_\rho^{\text{tel}}(\sigma), \tag{6.16}$$

where in the second line we have used the fact that the Bell-diagonal twirling of Bell-diagonal states leaves them invariant. □

Interestingly, Proposition 6.1 allows one to derive a simple form for the standard teleportation channel.

Corollary 6.1 (Standard teleportation is a Pauli channel)**.** *Let $\Lambda_\rho^{\text{tel}}(\sigma)$ denote the result of teleporting a qubit state σ (register C) with a two-qubit resource state ρ (registers AB) using the non-postselected standard teleportation protocol, as defined in (6.12). Then, $\Lambda_\rho^{\text{tel}}(\sigma)$ is a Pauli channel.*

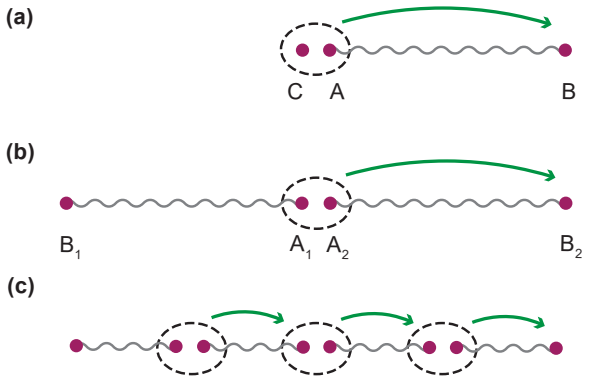


Figure 6.2: **Teleportation and entanglement swapping.** (a) The standard teleportation protocol involves a BSM on the target qubit and one qubit of an entangled resource state (wavy line), classical communication and corrections on the other half of the resource state (green arrow). (b) The standard entanglement swapping protocol involves teleporting one half of an entangled state. (c) An example of a swap-and-correct protocol for a repeater chain is applying teleportation sequentially.

Proof. See Appendix 6.7.1. □

We now define the entanglement swapping protocol for $N = 2$ initial states. In the following, we let $\mathscr{D}(\mathscr{H})$ denote the set of density operators acting on Hilbert space $\mathscr{H}$.

Definition 6.2. For $k = 1, 2$, let $\mathscr{H}_{A_k}$ and $\mathscr{H}_{B_k}$ be qubit Hilbert spaces. Given a pair of two-qubit initial states $\rho_1 \otimes \rho_2$ such that $\rho_k \in \mathscr{D}(\mathscr{H}_{A_k} \otimes \mathscr{H}_{B_k})$, the *standard entanglement swapping* protocol (or just *swapping*) is defined by applying the standard teleportation protocol to teleport register A_1 to register B_2, using ρ_2 as the resource state.

See Figures 6.2a and 6.2b for an illustration of the entanglement swapping protocol for $N = 2$.

The end-to-end state after a *postselected swap* with BSM outcome ij is then given by applying the map (6.8) to the appropriate qubits.

The end-to-end state ρ' after a *non-postselected swap* is given by applying the standard teleportation channel (6.12) to the appropriate qubits:

$$\rho' = (I_2 \otimes \Lambda^{\text{tel}}_{\rho_2})(\rho_1). \tag{6.17}$$

A simple extension of Proposition 6.1 means that the Bell-diagonal approximation of the state ρ_2 is exact for non-postselected entanglement swapping. However, this is not the case for postselected entanglement swapping. We will study postselected swapping in Section 6.4.

We now compute the result of swapping Bell-diagonal states. For convenience, we denote a Bell-diagonal state as a length-four vector,

$$\mathscr{B}(\rho) = \sum_{i,j} \lambda_{ij} \left|\Psi_{ij}\right\rangle\!\left\langle\Psi_{ij}\right| \equiv (\lambda_{00}, \lambda_{01}, \lambda_{10}, \lambda_{11})^T. \tag{6.18}$$

Lemma 6.3 (Postselected swapping of Bell-diagonal states)**.** *Let ρ'_{ij} be the end-to-end state after performing a postselected swap on a pair of Bell-diagonal states $\mathscr{B}(\rho_1)\otimes\mathscr{B}(\rho_2)$ such that $\mathscr{B}(\rho_1)\equiv(\lambda_0,\ldots,\lambda_3)^T$ and $\mathscr{B}(\rho_2)\equiv(\mu_0,\ldots,\mu_3)^T$, with BSM outcome ij. Then, $\rho'_{ij}=\rho'_{\mathscr{B}}$ for all i and j, where*

$$\rho'_{\mathscr{B}}\equiv(\lambda'_0,\ldots,\lambda'_3)^T, \tag{6.19}$$

is Bell-diagonal, and

$$\begin{pmatrix}\lambda'_0\\ \lambda'_1\\ \lambda'_2\\ \lambda'_3\end{pmatrix}=\begin{pmatrix}\lambda_0\mu_0+\lambda_1\mu_1+\lambda_2\mu_2+\lambda_3\mu_3\\ \lambda_0\mu_1+\lambda_1\mu_0+\lambda_2\mu_3+\lambda_3\mu_2\\ \lambda_0\mu_2+\lambda_2\mu_0+\lambda_3\mu_1+\lambda_1\mu_3\\ \lambda_0\mu_3+\lambda_3\mu_0+\lambda_1\mu_2+\lambda_2\mu_1\end{pmatrix}. \tag{6.20}$$

Moreover, the probability of this BSM outcome is

$$p'_{ij}=\frac{1}{4} \tag{6.21}$$

for all i and j.

Proof. See Appendix 6.7.1. □

We see from Lemma 6.3 that, when the initial states are Bell-diagonal, the end-to-end state after a non-postselected swap is given by

$$\rho'=\sum_{i,j=0}^{1}\frac{1}{4}\rho'_{ij}=\rho'_{\mathscr{B}}. \tag{6.22}$$

Therefore, when the initial states of the chain are Bell-diagonal, we see that postselected and non-postselected swapping give the same end-to-end state.

Corollary 6.2. *Let ρ' be the end-to-end state after performing a non-postselected entanglement swap on a pair of two-qubit initial states $\rho_1\otimes\rho_2$, as defined in (6.17). Then,*

$$\mathscr{B}(\rho')=\rho'_{\mathscr{B}} \tag{6.23}$$

where $\rho'_{\mathscr{B}}$ is the output state (6.19) given by swapping the Bell-diagonal approximations $\mathscr{B}(\rho_1)$ and $\mathscr{B}(\rho_2)$.

Proof. See Appendix 6.7.1. □

From Corollary 6.2, we see that the Bell-diagonal approximation is exact for the computation of the Bell-diagonal elements of the end-to-end state ρ'. This greatly simplifies the calculation of many important properties of the end-to-end state. For quantum network protocols and applications that use the end-to-end state as a resource, the Bell-diagonal approximation $\mathscr{B}(\rho')$ contains important information. For example, if one is performing QKD, the secret-key fraction of certain well-known protocols is invariant under Bell-diagonal twirling of the resource state [204, 205, 206]. Furthermore, as we have seen in Proposition 6.1, if standard teleportation is performed over the end-to-end link,

then the only contributing components are again the Bell-diagonal elements of ρ'. Even if the protocol performance is not only dependent on the Bell-diagonal elements, these elements may still contain important information about application feasibility. For example, $\mathscr{B}(\rho')$ contains the information of the fidelity to $|\Psi_{00}\rangle$, which is an important metric to deduce the feasibility and performance of entanglement purification protocols. It is a common assumption in purification protocols that the initial states are Bell-diagonal twirled [50, 47, 97, 52].

From Lemma 6.3, we obtain the well-known formula for the end-to-end state when the initial states are Werner [173].

Corollary 6.3. *Let $\rho'_{\mathscr{W}}$ be the end-to-end state after swapping a pair of two-qubit Werner states $\mathscr{W}(\rho_1)\otimes\mathscr{W}(\rho_2)$ with fidelities F_1, F_2. Then, $\rho'_{\mathscr{W}}$ is a Werner state with fidelity $F'_{\mathscr{W}} = F_1F_2 + (1-F_1)(1-F_2)/3$.*

Proof. A Werner state with fidelity F is Bell-diagonal, with the final three eigenvalues equal to one another:

$$\left(F, \frac{1-F}{3}, \frac{1-F}{3}, \frac{1-F}{3}\right)^T. \tag{6.24}$$

Then, by applying Lemma 6.3, swapping two Werner states with fidelities F_1 and F_2 results in a Werner state with fidelity $F'_{\mathscr{W}} = F_1F_2 + (1-F_1)(1-F_2)/3$. As for Lemma 6.3, this result holds for both postselected and non-postselected swapping. □

Following from Corollary 6.3, defining the *Werner parameter*

$$w := \frac{4F-1}{3}, \tag{6.25}$$

we see that the Werner parameter of the end-to-end state is $w' = w_1w_2$. Then, to compute the Werner parameter for the output state, one only needs to multiply the Werner parameters of the initial states. This is a well-used result in the performance analysis of quantum networks: if one is studying a large network, which could be a repeater chain or a more complex graph topology, it simplifies the analysis greatly to only consider the quality of each link to be described by one parameter F, which evolves under an entanglement swap according to the simple multiplicative relation – see e.g. [80, 75, 45]. We note that similar multiplicative relations have also been found for general Bell-diagonal states [64].

6.3.3. REPEATER CHAINS WITH $N > 2$ INITIAL STATES

Here, we consider non-postselected swapping over repeater chains with N initial states and $N-1$ repeaters. Due to the freedom of the order in which to perform entanglement swapping and Pauli corrections, we present a generalised class of swapping protocols on chains with N initial states that we term *swap-and-correct* protocols. For non-postselected swapping, we then go on to generalise the results of Section 6.3.2, presenting an exactness result for the Bell-diagonal approximation (Theorem 6.1) and an accuracy result for the Werner approximation (Theorem 6.2).

Suppose that non-postselected entanglement swapping is applied *sequentially*. By sequentially, we mean that the standard swapping protocol (Definition 6.2) is performed

$N-1$ times, moving from one side of the chain to the other (Figure 6.2c). Then, by Proposition 6.1, the Bell-diagonal approximation is again exact for each of the $N-1$ states that were treated as the resource state.

In practice, though, entanglement swapping is not likely to be implemented sequentially, because it requires classical communication and Pauli corrections after every BSM before the next BSM can be applied. With this strategy there is excessive classical communication time, and each swapped state has to spend an increasingly large amount of time waiting in memory before a BSM is applied to its qubit(s). This is problematic if qubits are subject to time-dependent noise while stored in memory, since added noise on the initial states can be detrimental to the quality of the final end-to-end state. For example, it may be beneficial to, instead of applying Pauli corrections sequentially after each BSM, apply them at the end nodes after all $N-1$ BSMs have been carried out. In this way, all $N-1$ BSMs at each node and classical communication of the outcomes may be carried out simultaneously, reducing the total amount of time the initial states must spend waiting in memory. We therefore look to generalise Corollary 6.2 to *all* possible strategies of performing BSMs and Pauli corrections for a repeater chain of arbitrary length.

We firstly present a definition of the class of entanglement swapping protocols under consideration. We term these *swap-and-correct* protocols. The outcomes of the $N-1$ BSMs define the *syndrome* $\vec{s}$. The syndrome is a length-$(N-1)$ list of Pauli operators such that $s_i \equiv X^m Z^n$ means that outcome mn was measured on node i. The Pauli correction at each stage of a swap-and-correct protocol depends on the result of the syndrome up to that point.

Definition 6.3 (Swap-and-correct protocol, informal)**.** For a length-N repeater chain, a *swap-and-correct* protocol $\mathscr{P}$ dictates where to apply Pauli corrections, given the $N-1$ BSM outcomes that form the syndrome $\vec{s}$. More specifically, $\mathscr{P}$ is a map

$$\mathscr{P} : \{I, X, Z, XZ\}^{N-1} \to \{I, X, Z, XZ\}^{N+1} \tag{6.26}$$

such that $\mathscr{P}_k(\vec{s})$ is the correction applied to node k for $k = 0, \ldots, N$. Moreover, for any syndrome $\vec{s}$, $\mathscr{P}$ transforms $|\Psi_{00}\rangle\langle\Psi_{00}|^{\otimes N}$ into $|\Psi_{00}\rangle\langle\Psi_{00}|$. We refer to this as the *correctness property*.

We note that for clarity, some details have been omitted from the above. For example, there must be some associated ordering of the BSMs, so that corrections always depend on past outcomes. This is an important property that imposes more restrictions on $\mathscr{P}$. We refer to Appendix 6.7.1 for the full technical definition. We also note that, for the $N-1$ repeater nodes, the protocol does not specify which of the two qubits in the node the correction is applied to. This is because a BSM projection will be applied following any correction, which means that both choices are equivalent (see Appendix 6.7.1).

Some examples of swap-and-correct protocols are:

- *Sequential teleportation.* Here, $\mathscr{P}_0(\vec{s}) = \mathscr{P}_1(\vec{s}) = I_2$ and $\mathscr{P}_k(\vec{s}) = s_{k-1}$ for $k = 2, \ldots, N$.
- *Correct at end.* Here, $\mathscr{P}_k(\vec{s}) = I_2$ for all $k = 0, \ldots, N-1$, and $\mathscr{P}_N(\vec{s}) = \prod_{k=1}^{N-1} s_k$.

Given a swap-and-correct protocol with syndrome $\mathscr{P}$, we denote the outcome state of a postselected swap with syndrome $\vec{s}$ as $\rho'_{\vec{s}}$, and the probability of measuring $\vec{s}$ as $p'_{\vec{s}}$. Then, the outcome after non-postselected swapping is given by

$$\rho' = \sum_{\vec{s}} \rho'_{\vec{s}} p'_{\vec{s}} \tag{6.27}$$

$$=: \Lambda_{\mathscr{P}}(\rho_{\text{in}}), \tag{6.28}$$

where $\rho_{\text{in}} = \otimes_{k=1}^{N} \rho_k$ is the initial state comparised of N two-qubit entangled states, and $\Lambda_{\mathscr{P}}$ is the channel induced by non-postselected swapping with $\mathscr{P}$.

We now present a generalisation of Corollary 6.2 for swap-and-correct protocols.

Theorem 6.1 (Exactness of Bell-diagonal approximation)**.** *Let $\mathscr{P}$ be a swap-and-correct protocol for repeater chains with N initial states. Let $\rho_{\text{in}} = \otimes_{k=1}^{N} \rho_k$ denote the N initial two-qubit states. Let $\Lambda_{\mathscr{P}}$ be the channel induced by non-postselected swapping with $\mathscr{P}$. Let $\mathscr{B}_{[N]}$ denote the Bell-diagonal twirling of states $1,\dots N$, such that*

$$\mathscr{B}_{[N]}\left(\rho_{\text{in}}\right) = \otimes_{k=1}^{N} \mathscr{B}(\rho_k) \tag{6.29}$$

is the Bell-diagonal approximation of the initial states. Then,

$$\mathscr{B}\left(\Lambda_{\mathscr{P}}(\rho_{\text{in}})\right) = \Lambda_{\mathscr{P}}(\mathscr{B}_{[N]}\left(\rho_{\text{in}}\right)). \tag{6.30}$$

Moreover, the above quantity is independent of the swap-and-correct protocol $\mathscr{P}$, i.e.

$$\mathscr{B}\left(\Lambda_{\mathscr{P}}(\rho_{\text{in}})\right) = \Lambda_{\text{seq}}\left(\mathscr{B}_{[N]}\left(\rho_{\text{in}}\right)\right) \tag{6.31}$$

where seq *is the protocol where standard teleportation is applied sequentially on each repeater.*

Proof. See Appendix 6.7.1. □

We see from Theorem 6.1 that for non-postselected swapping with any swap-and-correct protocol, the Bell-diagonal approximation is exact for the computation of the Bell-diagonal components of the end-to-end state. As for sequential swapping, one may then simply recursively apply the map (6.20) $N-1$ times to compute the end-to-end state.

We now turn to study the Werner approximation. With the following results, we quantify the error incurred by using the Werner approximation to compute the end-to-end fidelity in a repeater chain.

Lemma 6.4. *Consider applying the (non-postselected or postselected) sequential swapping protocol to a repeater chain with N Bell-diagonal states $\otimes_{k=1}^{N} \mathscr{B}(\rho_k)$, where $F_k = \langle\Psi_{00}|\,\rho_k\,|\Psi_{00}\rangle$ is the fidelity of the k-th initial state. Let $F' = \langle\Psi_{00}|\,\rho'\,|\Psi_{00}\rangle$ be the fidelity of the end-to-end state ρ'. Then,*

$$\prod_{k=1}^{N} F_k \leq F' \leq \frac{1}{2}\prod_{k=1}^{N}(2F_k - 1) + \frac{1}{2}. \tag{6.32}$$

Proof. See Appendix 6.7.1. □

One may combine Lemma 6.4 and Theorem 6.1 to obtain the following result for when swapping of N general two-qubit states.

Corollary 6.4. *Let $\mathscr{P}$ be a swap-and-correct protocol for repeater chains with N initial states. Let $\rho_{\text{in}} = \otimes_{k=1}^{N} \rho_k$ denote the N initial two-qubit states, where ρ_k has fidelity F_k. Let $\Lambda_{\mathscr{P}}$ be the channel induced by non-postselected swapping with $\mathscr{P}$. Then, the end-to-end fidelity after non-postselected swapping with $\mathscr{P}$ satisfies*

$$\prod_{k=1}^{N} F_k \leq \langle \Psi_{00} | \Lambda_{\mathscr{P}}(\rho_{\text{in}}) | \Psi_{00} \rangle \leq \frac{1}{2} \prod_{k=1}^{N} (2F_k - 1) + \frac{1}{2}. \tag{6.33}$$

We make the following remarks. The upper bound from (6.33) is tight: for example, this is saturated when the twirled initial states are of the form $\mathscr{B}(\rho_k) = (F_k, 1-F_k, 0, 0)$. From (6.20), it can be seen that swapping two Bell-diagonal states of rank two (in the same subspace) outputs another Bell-diagonal state of rank two, in the identical subspace. Given that the rank-two ansatz is preserved, one may find a simple rule for the fidelity decay after a swap: the parameter $x := 2F - 1$ evolves multiplicatively under swapping as $x' = x_1 x_2$, which is analogous to the evolution of the Werner parameter as we saw in (6.25). We therefore see that the upper bound is tight. The lower bound is tight for $N = 2$. For example, this is saturated when $\mathscr{B}(\rho_1) \equiv (F_1, 1-F_1, 0, 0)$ and $\mathscr{B}(\rho_2) \equiv (F_2, 0, 1-F_2, 0)$. We do not believe that the lower bound is tight for $N > 2$, but we leave further investigation of this to future work.

Theorem 6.2 (Accuracy of Werner approximation)**.** *Let $\mathscr{P}$ be a swap-and-correct protocol for repeater chains with N initial states. Let $\rho_{\text{in}} = \otimes_{k=1}^{N} \rho_k$ denote the N initial two-qubit states. Let $\Lambda_{\mathscr{P}}$ be the map induced by non-postselected swapping with $\mathscr{P}$. Let*

$$F' = \langle \Psi_{00} | \Lambda_{\mathscr{P}}(\rho_{\text{in}}) | \Psi_{00} \rangle \tag{6.34}$$

be the true end-to-end fidelity and $F'_{\mathscr{W}}$ be the end-to-end fidelity with the Werner approximation. Let F_k be the fidelity of ρ_k. If $\epsilon_k = 1 - F_k < \epsilon$ for all k, then

$$|F' - F'_{\mathscr{W}}| \leq \binom{N}{2} \epsilon^2 + \mathscr{O}(N^3 \epsilon^3). \tag{6.35}$$

In particular, if $N\epsilon \ll 1$, then

$$F' \approx F'_{\mathscr{W}}. \tag{6.36}$$

Proof. See Appendix 6.7.1. □

We note that if we do *not* have $N\epsilon \ll 1$, we do not expect the Werner approximation to be accurate: for example, swapping N identical Werner states with fidelity F results in a Werner state with fidelity

$$F_{\text{out}} = \frac{3}{4} \left(\frac{4F-1}{3} \right)^N + \frac{1}{4}, \tag{6.37}$$

which with F fixed goes to $\frac{1}{4}$ as $N \to \infty$. By contrast, for identical initial states the lower and upper bounds in (6.32) go to 0 and $\frac{1}{2}$ respectively.

6.4. POSTSELECTED SWAPPING

6.4.1. PARAMETERISATION OF INITIAL STATES

In the previous section, we studied the accuracy of twirled approximations for non-postselected swapping. We now study the accuracy of approximations for postselected swapping, for repeater chains with $N = 2$ initial states. Specifically, we address the following question: *how large (or small) can the fidelity of the outcome state become, postselected on a specific BSM outcome?* We will see that for certain initial states, after a postselected swap, the end-to-end state can exhibit a large variation in fidelity from what is obtained with a twirled approximation.

To illustrate the potential effect of postselecting on the BSM outcome, we consider an example that was introduced in [179]. Consider swapping the initial states $|\Psi_\theta\rangle\langle\Psi_\theta|^{\otimes 2}$, where

$$|\Psi_\theta\rangle = \cos(\theta)|00\rangle + \sin(\theta)|11\rangle \tag{6.38}$$

with the standard swapping protocol (Definition 6.2). This state has fidelity to $|\Psi_{00}\rangle$ given by

$$\left|\langle\Psi_{00}|\Psi_\theta\rangle\right|^2 = \cos^2(\theta - \frac{\pi}{4}), \tag{6.39}$$

which can take any value between 0 and 1, depending on the value of θ. The possible outcomes for the end-to-end state after the swap are

$$\begin{cases} |\Psi_{00}\rangle, & \text{with prob. } 2\sin^2(\theta)\cos^2(\theta), \\ \frac{1}{C}\left(\cos^2(\theta)|00\rangle + \sin^2(\theta)|11\rangle\right), & \\ & \text{with prob. } C^2 = \cos^4(\theta) + \sin^4(\theta). \end{cases} \tag{6.40}$$

In the above, the first outcome occurs when obtaining a measurement outcome corresponding to the odd-parity Bell states $|\Psi_{1j}\rangle$, and the second outcome occurs when obtaining a measurement outcome corresponding to even-parity Bell states, $|\Psi_{0j}\rangle$. We see from the above that for any $\theta \notin \{0, \pi/2, \pi, 3\pi/2\}$, there is a non-zero probability of obtaining an outcome state that is maximally entangled. This is in contrast to calculating the outcome of a non-postselected swap: the Bell-diagonal approximation of each initial state is

$$\mathscr{B}\left(|\Psi_\theta\rangle\langle\Psi_\theta|^{\otimes 2}\right) = F\left|\Phi^+\right\rangle\left\langle\Phi^+\right| + (1-F)|\Phi^-\rangle\langle\Phi^-| \tag{6.41}$$

where $F = \cos^2(\theta - \frac{\pi}{4})$. By Corollary 6.2 and Lemma 6.3, the end-to-end fidelity of a non-postselected swap is $F^2 + (1-F)^2$. One may also check that this is the weighted average of the outcomes in (6.40). We therefore see that for certain states, after a postselected entanglement swap, there is a non-zero probability of obtaining a significantly higher (or lower) fidelity outcome than the non-postselected case. Recalling Lemma 6.3, this variation may be attributed to off-Bell-diagonal terms in the initial states (in this example, $|\Psi_\theta\rangle\langle\Psi_\theta|$).

The variation in end-to-end fidelity is useful to characterise because some applications benefit from further information about the quality of the state. Postselecting on the Bell-state outcome can make certain tasks feasible: typically quantum applications are only feasible if on average, the level of noise in the resource state is below a certain

threshold [147, 207, 4]. Then, if a protocol is carried out by consuming many copies of the resource state, then by postselecting on certain swap outcomes a protocol can be made feasible, when for the non-postselected case it may not be. Furthermore, if the protocol is already feasible in the non-postselected case, knowledge of the full distribution of end-to-end fidelity can improve performance by making use of postselection. In Section 6.5.2 we provide an example of this for the case of quantum key distribution.

In the remainder of the section, we will find bounds on the variation in the end-to-end fidelity after a postselected swap. Specifically, for this problem we consider states of the form

$$\rho = p|\Psi_{00}\rangle\langle\Psi_{00}| + (1-p)\sigma, \tag{6.42}$$

where $F = \langle\Psi_{00}|\rho|\Psi_{00}\rangle$ is the fidelity, σ is a density matrix, and necessarily $p \leq F$. The last condition follows from the fact that ρ and σ are density matrices (positive semi-definite operators with unit trace).

The form (6.42) is relevant for two reasons. Firstly, every state may be written in this form, which can be interpreted as an ensemble of the pure Bell state (probability p), and the state σ (probability $1-p$). The state σ can be interpreted as a noise component, which is not necessarily orthogonal to $|\Psi_{00}\rangle$. The parameters p and F may be computed efficiently given the state ρ (see Section 6.4.2), if not just directly deducible by inspection of the form of ρ. Then, understanding the limits of the end-to-end fidelity of states of the form (6.42) has direct applications in a practical scenario. Secondly, fixing the parameter p as well as F is more restrictive than only fixing F, which makes it possible to find meaningful bounds. In order to formalise this idea, we firstly define the set of states of interest for fixed p and F.

Definition 6.4. Let $F \in [0,1]$ and $p \leq F$. We denote

$$\begin{aligned} S_{p,F} := \Big\{\rho : \exists\, \sigma \text{ s.t. } & \rho = p|\Psi_{00}\rangle\langle\Psi_{00}| + (1-p)\sigma, \\ & \langle\Psi_{00}|\rho|\Psi_{00}\rangle = F, \\ & \sigma \text{ density matrix}\Big\}. \end{aligned} \tag{6.43}$$

to be the set of all states of the form (6.42).

Proposition 6.2. *Let $F \in [0,1]$ and $p_2 < p_1 \leq F$. Then, $S_{p_1,F} \subset S_{p_2,F}$, but $S_{p_2,F} \nsubseteq S_{p_1,F}$.*

Proof. See Appendix 6.7.2. □

From Proposition 6.2, we see that increasing p (keeping F fixed) provides a more restrictive form for the state (6.42). For example, the set $S_{0,F}$ contains all valid two-qubit states with fidelity F, and the set $S_{F,F}$ contains only the states that have σ orthogonal to $|\Psi_{00}\rangle\langle\Psi_{00}|$. The set $S_{p,1}$ has only one element, which is $|\Psi_{00}\rangle\langle\Psi_{00}|$.

In the following, we will study the limits of the end-to-end fidelity for states $\rho \in S_{p,F}$. We firstly introduce simplified notation for the end-to-end fidelity, which will be helpful in the following sections.

Definition 6.5. Consider performing a postselected swap on a pair of two-qubit states $\rho_1 \otimes \rho_2$. If the BSM outcome ij is obtained, we denote the end-to-end fidelity by $F'_{ij}(\rho_1 \otimes$

ρ_2), and the probability of this BSM outcome as $p'_{ij}(\rho_1 \otimes \rho_2)$. These quantities are written explicitly as

$$F'_{ij} = \frac{1}{p'_{ij}} \mathrm{Tr}\left[|\Psi_{ij}\rangle\langle\Psi_{ij}|_{B_1B_2} |\Psi_{ij}\rangle\langle\Psi_{ij}|_{A_1A_2} \rho_1 \otimes \rho_2 \right] \tag{6.44}$$

$$p'_{ij} = \mathrm{Tr}\left[|\Psi_{ij}\rangle\langle\Psi_{ij}|_{A_1A_2} \rho_1 \otimes \rho_2 \right], \tag{6.45}$$

where we refer to Figure 6.2b for a depiction of the qubit registers B_1, A_1, A_2 and B_2.

We now wish to find bounds on the end-to-end fidelity F'_{ij}, for initial states $\rho_k \in S_{p,F}$. For clarity, we take the initial states to have the same parameters p and F. We note that the results from this section also hold for the more general case ($\rho_k \in S_{p_k,F_k}$), for which the proofs are carried out in the Appendix.

We firstly show that it is enough to consider just a single BSM outcome.

Proposition 6.3. *Let $F \in [0,1]$ and $p \leq F$, and*

$$F'_{ij,\max} := \max_{\rho_k} \left\{ F'_{ij}(\rho_1 \otimes \rho_2) \text{ s.t. } \rho_k \in S_{p,F} \right\}$$

$$F'_{ij,\min} := \min_{\rho_k} \left\{ F'_{ij}(\rho_1 \otimes \rho_2) \text{ s.t. } \rho_k \in S_{p,F} \right\}$$

where F'_{ij} is the postselected end-to-end fidelity (Definition 6.5). Then, the above quantities are independent of i and j, or alternatively

$$F'_{ij,\max} = F'_{00,\max} \equiv F'_{\max}(p,F) \tag{6.46}$$

$$F'_{ij,\min} = F'_{00,\min} \equiv F'_{\min}(p,F) \tag{6.47}$$

for all i, j.

Proof. See Appendix 6.7.2. We use the idea that one may always rotate ρ_2 by a suitable Pauli to find a $\omega_2 \in S_{p,F}$ such that $F'_{ij}(\rho_1 \otimes \rho_2) = F'_{00}(\rho_1 \otimes \omega_2)$. □

6.4.2. ANALYTICAL BOUNDS

Here, we use analytical methods to find an exact expression for $F'_{\max}(p,F)$ and a lower bound for $F'_{\min}(p,F)$. In Section 6.4.3, we will find tighter lower bounds on $F'_{\min}(p,F)$ using semidefinite programming (SDP).

To study $F'_{\max}(p,F)$ and $F'_{\min}(p,F)$ we firstly establish a simplified formula for the end-to-end fidelity F'_{00}.

Lemma 6.5. *Consider performing a postselected swap on a pair of two-qubit states $\rho_1 \otimes \rho_2$ such that $\rho_k \in S_{p,F}$ and*

$$\rho_k = p|\Psi_{00}\rangle\langle\Psi_{00}| + (1-p)\sigma_k, \tag{6.48}$$

where σ_k is a density matrix, for $k = 1,2$. Let F'_{ij} and p'_{ij} denote the postselected end-to-end fidelity and probability (Definition 6.5). Then,

$$F'_{ij}(\rho_1 \otimes \rho_2) = \frac{2pF - p^2 + 4(1-p)^2 \tilde{p}'_{ij} \tilde{F}'_{ij}}{2p - p^2 + 4(1-p)^2 \tilde{p}'_{ij}}, \tag{6.49}$$

and

$$p'_{ij}(\rho_1\otimes\rho_2) = \frac{p}{2} - \frac{p^2}{4} + (1-p)^2\tilde{p}'_{ij} \tag{6.50}$$

where $\tilde{F}'_{ij} \coloneqq F'_{ij}(\sigma_1\otimes\sigma_2)$ *and* $\tilde{p}'_{ij} \coloneqq p'_{ij}(\sigma_1\otimes\sigma_2)$ *are the swap statistics of the noisy components (Definition 6.5).*

Proof. See Appendix 6.7.2. □

From Lemma 6.5, we see that the swap statistics of states in $S_{p,F}$ may be understood only in terms of the corresponding swap statistics of the noisy components σ_k. Lemma 6.5 is used in the proof of Theorem 6.3, where we find an exact expression for $F'_{\max}(p,F)$.

Theorem 6.3. *Consider performing a postselected swap on a pair of two-qubit states* $\rho_1\otimes\rho_2$*. Let* $F'_{\max}(p,F)$ *be the maximum achievable end-to-end fidelity for* $\rho_k \in S_{p,F}$ *with* $k = 1,2$*, as defined in (6.46). Then,*

$$F'_{\max}(p,F) = 1-2p(1-F). \tag{6.51}$$

In particular, the initial states $\rho_{\text{opt}}^{\otimes 2}$ *satisfy* $F'_{00}(\rho_{\text{opt}}^{\otimes 2}) = F'_{\max}(p,F)$*, where*

$$\rho_{\text{opt}} = p\,|\Psi_{00}\rangle\langle\Psi_{00}| + (1-p)\,|\psi\rangle\langle\psi|, \tag{6.52}$$

and

$$|\psi\rangle = \sqrt{\tilde{F}}\,|\Psi_{00}\rangle + \sqrt{1-\tilde{F}}\,|\Psi_{11}\rangle \tag{6.53}$$

with $\tilde{F} = (F-p)/(1-p)$

Proof. See Appendix 6.7.2. □

We note that the saturating state (6.52) may be interpreted as $|\Psi_{00}\rangle$ having undergone a Y-rotation of a specified angle with probability $1-p$. We note that in the more general case $\rho_k \in S_{p_k,F_k}$, the equality (6.51) instead becomes an upper bound (see Appendix 6.7.2).

We notice that Theorem 6.3 implies a similar result for the swapping of identical states. More specifically, for any $\rho \in S_{p,F}$ it follows that

$$F'_{ij}(\rho^{\otimes 2}) \leq 1-2p(1-F) \tag{6.54}$$

for any BSM outcome ij.

We see from Theorem 6.3 that $F'_{\max}(p,F)$ is decreasing in p. The decreasing behaviour is expected, because from Proposition 6.2, the set $S_{p,F}$ shrinks as p increases and F is fixed. One may tighten the bound (6.54) by finding the largest possible q such that $\rho \in S_{q,F}$, which can be achieved by solving the optimisation problem

$$\begin{aligned} p^* = \max_{q} \quad & q \\ \text{s.t.} \quad & \rho - q\,|\Psi_{00}\rangle\langle\Psi_{00}| \geq 0. \end{aligned} \tag{6.55}$$

The problem (6.55) may be solved efficiently using a simple SDP solver, which we provide in our repository [208]. The tightened bound is then given by

$$F'_{ij}(\rho^{\otimes 2}) \leq 1 - 2p^*(1-F). \tag{6.56}$$

A simple demonstration of this procedure is with that of the Werner state (6.7). We rewrite this as

$$\mathscr{W}(\rho) = p\,|\Psi_{00}\rangle\langle\Psi_{00}| + (1-p)\frac{I_4}{4}, \tag{6.57}$$

where $p = (4F-1)/3$ and $F = \langle\Psi_{00}|\,\rho\,|\Psi_{00}\rangle$. Since $I_4/4$ is a density matrix, it follows that $\mathscr{W}(\rho) \in S_{p,F}$. However, we notice that $\mathscr{W}(\rho)$ may also be rewritten in Bell-diagonal form with the coefficients as given in (6.24), and therefore $\rho_W \in S_{F,F}$. The second case gives the tighter upper bound for the end-to-end fidelity,

$$F'_{ij}\big(\mathscr{W}(\rho)^{\otimes 2}\big) \leq F'_{\max}(F,F) = 1 - 2F(1-F), \tag{6.58}$$

which can be easily validated with Corollary 6.3.

We make the following further observations about Theorem 6.3. At $p = 0$, the expression simplifies to $F'_{\max}(0,F) = 1$. This is expected, because recalling the state $|\Psi_\theta\rangle$ from (6.38), we have $|\Psi_\theta\rangle\langle\Psi_\theta| \in S_{0,F}$ such that $F = \cos^2(\theta - \frac{\pi}{4})$. Then,

$$1 = F'_{ij}\left(|\Psi_\theta\rangle\langle\Psi_\theta|^{\otimes 2}\right) \leq F'_{\max}(0,F), \tag{6.59}$$

which implies the same. If $\rho \in S_{F,F}$, the noisy component σ is orthogonal to $|\Psi_{00}\rangle\langle\Psi_{00}|$. In such a case, we see that (6.51) simplifies to

$$\begin{aligned} F'_{ij}(\rho^{\otimes 2}) \leq F'_{\max}(F,F) &= 1 - 2F(1-F) && (6.60)\\ &= F^2 + (1-F)^2 && (6.61)\\ &= F'_{ij}(\rho_{\mathrm{R2}}^{\otimes 2}). && (6.62) \end{aligned}$$

where ρ_{R2} is any Bell-diagonal state of rank two with fidelity F. In the final step, we have recalled the formula for the postselected swapping of Bell-diagonal states from Lemma 6.3. In particular, the state $\rho_{\mathrm{R2}} \in S_{F,F}$ provides optimal end-to-end fidelity for initial states in $S_{F,F}$.

Now that we have characterised $F'_{\max}(p,F)$, we turn to studying $F'_{\min}(p,F)$. In the following Proposition, we derive an analytical lower bound for this quantity.

Proposition 6.4. *Let $F'_{\min}(p,F)$ be the minimum end-to-end fidelity for $\rho_k \in S_{p,F}$, as defined in (6.47). Then,*

$$F'_{\min}(p,F) \geq \frac{p(2F-p)}{1+(1-p)^2}. \tag{6.63}$$

Proof. See Appendix 6.7.2. □

6.4.3. LOWER BOUND WITH SDP

Unlike the upper bound in Theorem 6.3, the lower bound (6.63) is not tight. In this section, we find a tighter lower bound for $F'_{\text{min}}(p,F)$ using semi-definite programming (SDP).

Recalling its definition in (6.47), $F'_{\text{min}}(p,F)$ is the solution to the optimisation problem

$$\begin{aligned} \min_{\rho_1\otimes\rho_2} \quad & F'_{ij}(\rho_1\otimes\rho_2) \\ \text{s.t.} \quad & \text{Tr}\left[|\Psi_{00}\rangle\langle\Psi_{00}|\rho_k\right] = F, \\ & \text{Tr}\left[\rho_k\right] = 1, \\ & \rho_k - p|\Psi_{00}\rangle\langle\Psi_{00}| \geq 0, \text{ for } k=1,2. \end{aligned} \tag{6.64}$$

Here, the constraints ensure that we are optimising over $\rho_k \in S_{p,F}$. The first constraint enforces ρ_k has fixed fidelity F, and the final two constraints ensure that the noisy component σ_k is a valid density matrix. Recalling from Lemma 6.5 the formula for swapping two noisy states, (6.64) may be written as

$$\begin{aligned} \min_{\sigma_1\otimes\sigma_2} \quad & \frac{2pF - p^2 + 4(1-p)^2\cdot\tilde{p}'_{00}\cdot\tilde{F}'_{00}}{2p - p^2 + 4(1-p^2)\cdot\tilde{p}'_{00}} \\ \text{s.t.} \quad & p + (1-p)\text{Tr}\left[|\Psi_{00}\rangle\langle\Psi_{00}|\sigma_k\right] = F, \\ & \text{Tr}[\sigma_k] = 1, \\ & \sigma_k \geq 0, \qquad \text{for } k=1,2. \end{aligned} \tag{6.65}$$

To obtain (6.65), we have reparameterised the problem to optimise over the noisy components σ_k. The quantities $\tilde{F}'_{ij} := F'_{ij}(\sigma_1\otimes\sigma_2)$ and $\tilde{p}'_{ij} := p'_{ij}(\sigma_1\otimes\sigma_2)$ are the corresponding swap statistics when only swapping the noisy components σ_k.

The domain in (6.65) is the set of product states $\sigma_1\otimes\sigma_2$, where σ_k is a two-qubit density matrix. The domain is non-convex. Moreover, the objective function is rational, and not manifestly convex. These two details make (6.65) difficult to approach using numerical methods. We will therefore perform a relaxation of the domain, which transforms this the problem into one that is solvable with SDP. SDP is a commonly-used technique in quantum information [209]. The SDP formulation opens up the possibility of using several well-studied and efficient solvers, and moreover has an important feature that, under certain conditions, the solver converges to a global optimum.

In order to study (6.65) with SDP, we perform two steps. Firstly, we linearise the objective function. Since the objective function of (6.65) is rational, we fix its denominator and introduce the new constraint

$$\begin{aligned} \delta &= \frac{p}{2} - \frac{p^2}{4} + (1-p)^2\tilde{p}'_{00} \\ &= \frac{p}{2} - \frac{p^2}{4} + (1-p)^2\,\text{Tr}\left[|\Psi_{00}\rangle\langle\Psi_{00}|_{A_1A_2}\,\sigma_1\otimes\sigma_2\right]. \end{aligned} \tag{6.66}$$

For conciseness, we rewrite the above as

$$\text{Tr}\left[|\Psi_{00}\rangle\langle\Psi_{00}|_{A_1A_2}\,\sigma_1\otimes\sigma_2\right] = \tilde{\delta}(p,\delta) \tag{6.67}$$

where

$$\tilde{\delta}(p,\delta) := \frac{4\delta - 2p + p^2}{4(1-p)^2}. \tag{6.68}$$

Recalling Lemma 6.5, this is fixing the total probability to be δ. Since p is fixed, δ and $\tilde{\delta}$ are interchangeable via the linear relation (6.68).

Similarly, we rewrite both fidelity constraints as

$$\mathrm{Tr}\big[\,|\Psi_{00}\rangle\langle\Psi_{00}|_{A_1A_2}\,\sigma_k\big] = \tilde{F}, \tag{6.69}$$

where

$$\tilde{F}(p,F) := \frac{F-p}{1-p} \tag{6.70}$$

is the fidelity of the noisy component. Moreover, given that δ is fixed, we notice that the objective function is given by

$$\frac{1}{4\delta}\left(2pF - p^2 + 4(1-p)^2 \cdot \tilde{p}'_{00}\tilde{F}'_{00}\right),$$

and so it suffices to optimise over

$$\tilde{p}'_{00}\tilde{F}'_{00} = \mathrm{Tr}\big[\,|\Psi_{00}\rangle\langle\Psi_{00}|_{B_1B_2}|\Psi_{00}\rangle\langle\Psi_{00}|_{A_1A_2}\,\sigma_1\otimes\sigma_2\big] \tag{6.71}$$

which is a linear function of $\sigma_1\otimes\sigma_2$. With our constraints and objective function reformulated as (6.67), (6.69) and (6.71), we are now interested in the solution to

$$\begin{aligned}
\min_{\sigma_1\otimes\sigma_2}\quad & \mathrm{Tr}\big[\,|\Psi_{00}\rangle\langle\Psi_{00}|_{B_1B_2}\,|\Psi_{00}\rangle\langle\Psi_{00}|_{A_1A_2}\,\sigma_1\otimes\sigma_2\big] \\
\text{s.t.}\quad & \mathrm{Tr}\big[\,|\Psi_{00}\rangle\langle\Psi_{00}|_{A_1A_2}\,\sigma_1\otimes\sigma_2\big] = \tilde{\delta}(p,\delta), \\
& \mathrm{Tr}\big[\,|\Psi_{00}\rangle\langle\Psi_{00}|_{A_kB_k}\,\sigma_k\big] = \tilde{F}(p,F), \\
& \mathrm{Tr}[\sigma_k] = 1, \\
& \sigma_k \geq 0, \quad \text{for } k = 1,2.
\end{aligned} \tag{6.72}$$

Letting $H^*(p,F,\delta)$ be the solution to (6.72), we have quantity

$$F'_{\min}(F,p) = \min_{\delta}\frac{1}{\delta}\left(\frac{Fp}{2} - \frac{p^2}{4} + (1-p)^2 H^*(p,F,\delta)\right). \tag{6.73}$$

As well as linearising the objective function, fixing δ allows one to study the rate-fidelity trade-off in the entanglement swapping protocol. This is useful because, in the performance analysis of quantum networks, it is important to understand both fidelity metrics and rate metrics in entanglement distribution protocols. For example, if a state provides a high fidelity with an excessively low probability of success, then this may no longer be very useful or relevant. Notice that the definitions of $F'_{\max}(p,F)$ and $F'_{\min}(p,F)$ in (6.46) and (6.47) are currently agnostic to the probability of obtaining the BSM outcome with minimum and maximum fidelity. Fixing the swap probability is a mechanism to study this: with such a constraint, for a given probability δ of a given swap outcome, one may study the limits of the fidelity. The same study was carried out in [53], where the authors

use SDP to study the maximum fidelity that can be achieved with practical purification protocols, given a fixed success probability of purification. In Appendix 6.7.4, we provide further discussion and analysis of the rate-fidelity trade-off.

Recalling that the domain over which we optimise in (6.72) is not convex (product states), we perform a relaxation of the domain. In particular, we use

$$\sigma_1 \otimes \sigma_2 \in \text{SEP} \subset \text{PPT}, \tag{6.74}$$

where SEP is the set of separable states, and PPT is the set of four-qubit states that are still positive after taking the partial transpose with respect to the registers A_2 and B_2 [102]. Relaxing the domain of (6.72) results in the following:

$$\begin{aligned} \min_{\sigma} \quad & \text{Tr}\big[\,|\Psi_{00}\rangle\langle\Psi_{00}|_{B_1B_2}\,|\Psi_{00}\rangle\langle\Psi_{00}|_{A_1A_2}\,\sigma\big] \\ \text{s.t.} \quad & \text{Tr}\big[\,|\Psi_{00}\rangle\langle\Psi_{00}|_{A_1A_2}\,\sigma\big] = \tilde{\delta}(p,\delta), \\ & \text{Tr}\big[\,|\Psi_{00}\rangle\langle\Psi_{00}|_{B_1A_1}\,\sigma\big] = \tilde{F}(p,F), \\ & \text{Tr}\big[\,|\Psi_{00}\rangle\langle\Psi_{00}|_{B_2A_2}\,\sigma\big] = \tilde{F}(p,F), \\ & \text{Tr}[\sigma] = 1, \\ & \sigma \geq 0, \;\; \sigma^{\Gamma} \geq 0. \end{aligned} \tag{6.75}$$

where M^{Γ} denotes taking the partial transpose of M on the registers A_2 and B_2. The optimisation problem (6.75) may now be solved with SDP. One may greatly reduce the number of parameters in the optimisation by using the fact that the objective function and all constraints are invariant under the application of correlated unitaries. See Appendix 6.7.3 for the full details of the symmetrisation procedure. After symmetrisation, the number of free parameters in the optimisation is reduced from 256 to fewer than 48.

Letting $H^*_{\text{rel}}(p,F,\delta)$ be the solution to (6.75), by the relaxation (6.74) it follows that

$$H^*(p,F,\delta) \geq H^*_{\text{rel}}(p,F,\delta). \tag{6.76}$$

Recalling (6.73), F'_{min} is then bounded below by

$$F'_{\text{min}}(p,F) \geq \min_{\delta} \frac{1}{\delta}\left(\frac{Fp}{2} - \frac{p^2}{4} + (1-p)^2 H^*_{\text{rel}}(p,F,\delta)\right), \tag{6.77}$$

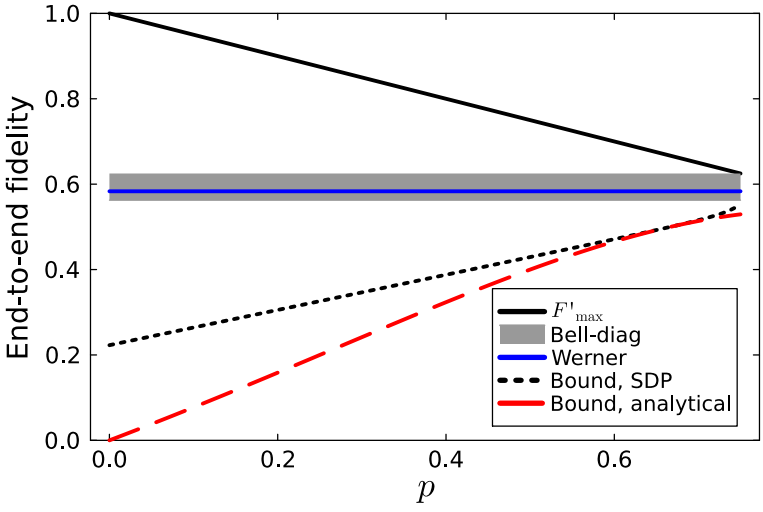
where the optimisation is performed in the feasible region of δ (see Appendix 6.7.3 for the calculation of the feasible region). After symmetrisation of (6.75), since the numerical optimisation over δ is over a single parameter in a bounded domain, (6.77) is efficient to compute (on the order of a few seconds).

6.5. DISCUSSION

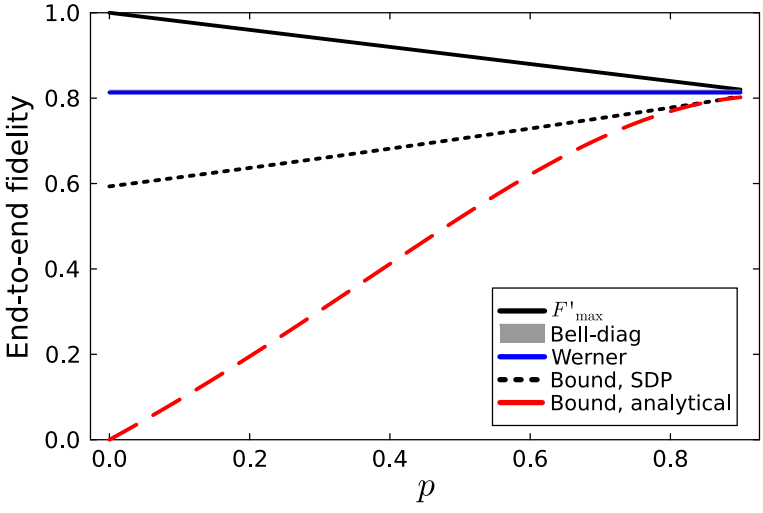
6.5.1. BOUNDS COMPARISON

Here, we illustrate the results from Sections 6.3 and 6.4 with examples. In particular, we will see how the parameters p and F affect the accuracy of twirled approximations.

For fixed fidelity F, plotted in Figures 6.3a and 6.3b is $F'_{\text{max}}(p,F)$ as found in Theorem 6.3, and the lower bounds for $F'_{\text{min}}(p,F)$. In Figure 6.4, $p = 0$ is fixed, and the same



(a)



(b)

Figure 6.3: **Bounds for the end-to-end fidelity when swapping states $\rho_1 \otimes \rho_2$ with $\rho_k = p|\Psi_{00}\rangle\langle\Psi_{00}| + (1-p)\sigma_k$, for $p \in [0, F]$ ($\rho_k \in S_{0,F}$ for $k = 1, 2$), with (a) $F = 0.75$ and (b) $F = 0.9$.** The black solid line is the tight upper bound on the postselected end-to-end fidelity, $F'_{\max}(p, F)$. The red dashed line is the analytical lower bound on the postselected end-to-end fidelity, $F'_{\min}(p, F)$. The black dotted line is the SDP lower bound for $F'_{\min}(p, F)$, given in (6.77). With the Bell-diagonal approximation, the end-to-end fidelity $F'_{\mathscr{B}}$ will lie in the grey region. With the Werner approximation, the end-to-end fidelity $F'_{\mathscr{W}}$ will lie on the blue line. The plot is made for 100 values p uniformly spaced within this interval.

quantities are plotted. In all cases we have tested, the SDP lower bound for $F'_{\min}(p, F)$ is tighter than the analytical lower bound. The grey region is where the end-to-end fidelity

will lie if the Bell-diagonal approximation is used for the initial states. In particular, the grey region $[F^2, F^2+(1-F)^2]$ is the region between the best and worst end-to-end fidelity for Bell-diagonal states of fidelity F, from Lemma 6.4. The grey region depends only on F, and hence is constant in Figures 6.3a and 6.3b. By Theorem 6.1, the grey region is also where the end-to-end fidelity will lie after a non-postselected swap.

In Theorem 6.2, we saw that for $1-F \ll 1/N$, the Werner approximation is accurate for non-postselected swapping. Then, given that $N=2$ is fixed in Figure 6.4, for large F the Bell-diagonal region is concentrated tightly around the Werner line.

In Figure 6.4, because $p=0$ is fixed, the maximum end-to-end fidelity is constant at $F'_{\max}(0,F)=1$. This is expected from the discussion at the beginning of Section 6.4 where we saw that, when only fixing the fidelity of the input states, one may always find states that swap to unit fidelity. As well as the lower bounds for $F'_{\min}(0,F)$, we have plotted the lowest-fidelity outcome of the state $|\psi\rangle$ that was given in (6.53) as an example of a state giving output fidelity $F'_{\max}(p,F)$. The state $|\psi\rangle$ provides very good postselected swap statistics for the output fidelity of certain BSM outcomes. Since the end-to-end fidelity for a non-postselected swap must lie within the grey region and this is the weighted average of the postselected outcomes (Corllary 6.2), the low-fidelity outcomes lie significantly below the grey region. In particular, the state $|\psi\rangle$ can also give an exceptionally low end-to-end fidelity. We plot this line in order to give an upper bound for the tightness of the SDP lower bound for $F'_{\min}(p,F)$.

In Figures 6.3a and 6.3b, we see that for $p=F$, $F'_{\max}$ meets the upper limit of the grey region. The reason is what was seen in (6.62): when the noisy component σ_k is orthogonal to $|\Psi_{00}\rangle$, any rank-two Bell-diagonal state ρ_{R2} provides an optimal end-to-end fidelity, but also lies in the grey region due to being Bell-diagonal. We outline the practical relevance in the following way. Let consider swapping the initial states $\rho_1 \otimes \rho_2$ with $\rho_k \in S_{F,F}$. Let $F'_{\mathscr{B}}$ ($F'_{\mathscr{W}}$) denote the end-to-end fidelity with the Bell-diagonal (Werner) approximation, such that

$$F'_{\mathscr{B}} = F'_{ij}(\mathscr{B}(\rho_1)\otimes\mathscr{B}(\rho_2)), \tag{6.78}$$

$$F'_{\mathscr{W}} = F'_{ij}(\mathscr{W}(\rho_1)\otimes\mathscr{W}(\rho_2)). \tag{6.79}$$

Let $(ij)^*$ denote the highest-fidelity BSM outcome after swapping $\rho_1\otimes\rho_2$, such that

$$F'_{(ij)^*}(\rho_1\otimes\rho_2) = \max_{i,j} F'_{ij}(\rho_1\otimes\rho_2). \tag{6.80}$$

Then, the corresponding output fidelity necessarily satisfies $F'_{(ij)^*}(\rho_1\otimes\rho_2) \geq F'_{\mathscr{B}}$, since by Corollary 6.2,

$$F'_{\mathscr{B}} = \sum_{ij} p'_{ij} F'_{ij}. \tag{6.81}$$

Recalling from Lemma 6.4 that $F'_{\mathscr{B}} \geq F^2$, the maximum deviation above the Bell-diagonal approximation is bounded as

$$F'_{(ij)^*}(\rho_1\otimes\rho_2) - F'_{\mathscr{B}} \leq F'_{\max}(F,F) - F^2 \tag{6.82}$$

$$= (1-F)^2. \tag{6.83}$$

Recalling from Corollary 6.3 that $F'_{\mathscr{W}} = F^2 + (1-F)^2/3$, the maximum deviation above the Werner approximation is therefore

$$F'_{(ij)^*}(\rho_1 \otimes \rho_2) - F'_{\mathscr{W}} \leq F'_{\max}(F,F) - F'_{\mathscr{W}} \tag{6.84}$$

$$= \frac{2}{3}(1-F)^2. \tag{6.85}$$

Then, by (6.83) and (6.85) we see that for large F, a large deviation above the twirled approximation is not possible when the input states $\rho_k \in S_{F,F}$ have an orthogonal component, i.e.

$$F'_{(ij)^*}(\rho_1 \otimes \rho_2) \approx F'_{\mathscr{B}} \tag{6.86}$$

$$F'_{(ij)^*}(\rho_1 \otimes \rho_2) \approx F'_{\mathscr{W}}. \tag{6.87}$$

For example, consider swapping the initial states $\rho_R^{\otimes 2}$ with

$$\rho_R = p\,|\Psi_{00}\rangle\langle\Psi_{00}| + (1-p)\,|01\rangle\langle 01|, \tag{6.88}$$

which in some contexts is referred to as the R state. Up to a local unitary rotation, such a state closely approximates states generated in certain physical entanglement generation schemes [127, 23]. It has an orthogonal, non-Bell-diagonal noisy component $|01\rangle\langle 01|$. We see from our analysis that, for large F (p), twirled approximations will not cause a large decrease in end-to-end fidelity because of the orthogonal noisy component.

As another example of a direct application of our bounds, let us consider the S state [53],

$$\rho_S = p\,|\Psi_{00}\rangle\langle\Psi_{00}| + (1-p)\,|11\rangle\langle 11|. \tag{6.89}$$

The state ρ_S has a non-orthogonal noisy component $|11\rangle\langle 11|$, with fidelity $|\langle\Psi_{00}|11\rangle|^2 = 1/2$. By direct inspection of ρ_S, we see that $\rho_S \in S_{p,F}$, where $F = (1+p)/2$. By Theorem 6.3,

$$F'_{ij}(\rho_S^{\otimes 2}) \leq F'_{\max}\left(p, \frac{1+p}{2}\right) \tag{6.90}$$

$$= 1 - 2p(1-(1+p)/2) \tag{6.91}$$

$$= (1-p)^2 + p. \tag{6.92}$$

Moreover, we have

$$F'_{\mathscr{B}} \geq F^2 = \left(\frac{1+p}{2}\right)^2 = p + \frac{1}{4}(1-p)^2, \tag{6.93}$$

and

$$F'_{\mathscr{W}} = F^2 + \frac{(1-F)^2}{3} \tag{6.94}$$

$$= \left(\frac{1+p}{2}\right)^2 + \left(\frac{1-p}{2}\right)^2 \tag{6.95}$$

$$= p + \frac{1}{2}(1-p)^2. \tag{6.96}$$

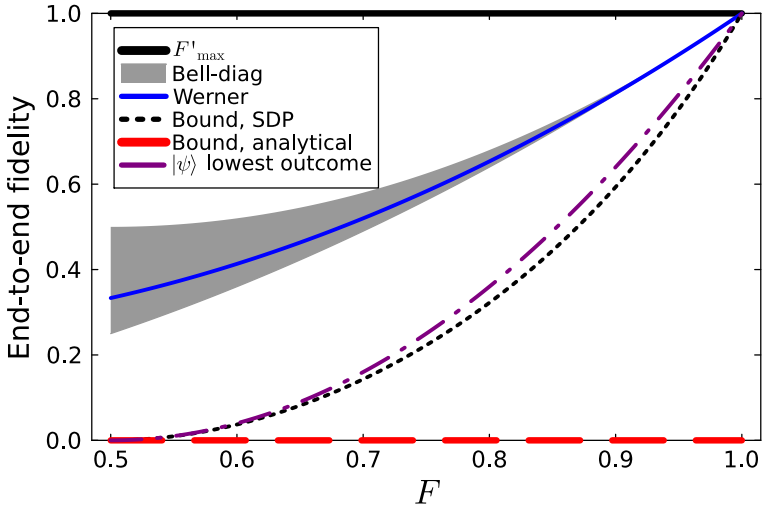


Figure 6.4: **Bounds for the end-to-end fidelity when swapping states $\rho_1 \otimes \rho_2$ with initial fidelity $F \in [0.5, 1]$ ($\rho_k \in S_{0,F}$ for $k = 1, 2$).** The black solid line is the tight upper bound on the postselected end-to-end fidelity, $F'_{\max}(0, F)$. The red dashed line is the analytical lower bound for the postselected end-to-end fidelity, $F'_{\min}(0, F)$. The black dotted line is the SDP lower bound for $F'_{\min}(0, F)$. The purple dot-dash line is the lowest-fidelity outcome of $|\psi\rangle$, as defined in (6.53) with $p = 0$. With the Bell-diagonal approximation, the end-to-end fidelity $F'_{\mathscr{B}}$ will lie in the grey region. With the Werner approximation, the end-to-end fidelity $F'_{\mathscr{W}}$ will lie on the blue line. The plot is made for 100 values of F uniformly spaced within the interval.

Therefore, combining (6.92), (6.93) and (6.96), we see that the maximum deviation *above* the Bell-diagonal (Werner) approximations when swapping the initial states $\rho_S^{\otimes 2}$ is bounded above by

$$F'_{\max}\left(p, \frac{1+p}{2}\right) - F'_{\mathscr{B}} = \frac{3(1-p)^2}{4}. \tag{6.97}$$

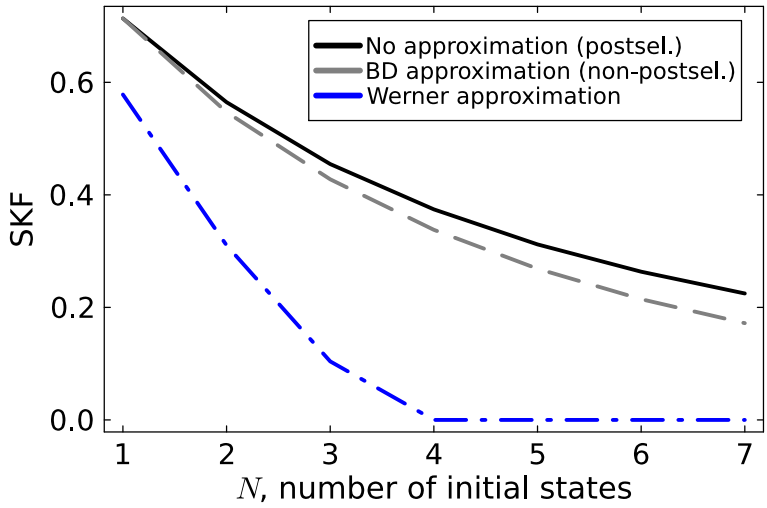
$$F'_{\max}\left(p, \frac{1+p}{2}\right) - F'_{\mathscr{W}} = \frac{(1-p)^2}{2}. \tag{6.98}$$

Consequently, for large p (equivalently, large F), we conclude that twirled approximations do not cause large inaccuracies in estimating the output fidelity when the initial states are $\rho_S^{\otimes 2}$.

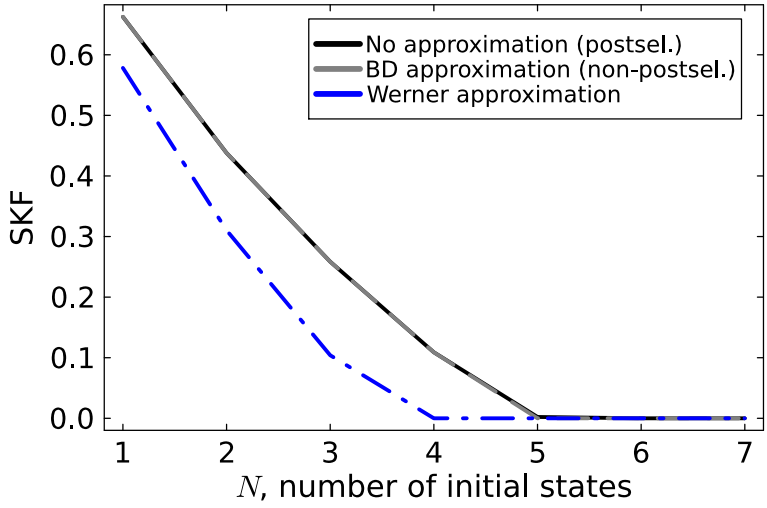
We note that one may also perform a similar study for the deviation *below* the twirled approximations by computing the difference with the analytical or SDP lower bounds for $F'_{\min}(p, F)$.

6.5.2. EXAMPLE: QUANTUM KEY DISTRIBUTION

We now carry out a numerical study of the accuracy of twirled state approximations when quantum key distribution (QKD) is performed. It has been shown previously that postselecting on the syndrome when using error correction in a repeater chain can give an advantage [210, 211, 212]. Here, we extend these results by pointing out that an advantage can also be obtained in the absence of error correction by postselecting on the



(a)



(b)

Figure 6.5: **Secret-key fraction when performing quantum key distribution over a repeater chain with N initial states** $\rho^{\otimes N}$ when (a) $\rho = \rho_{\text{opt}}$ with $F = 0.95$ and $p = 0.5$ from (6.52), and (b) $\rho = \rho_R$ from (6.88) with fidelity $F = 0.95$. The black line is the secret-key fraction of the postselected protocol, the grey dashed line is the secret-key fraction when the Bell-diagonal approximation is used for the initial states $\mathscr{B}(\rho)^{\otimes N}$ (or equivalently, the secret-key fraction that is obtained with the non-postselected protocol), and the blue dot-dashed line is the secret-key fraction when the Werner approximation is used for the initial states $\mathscr{W}(\rho)^{\otimes N}$.

swap outcomes, and moreover that this advantage is the exact loss in performance when using the Bell-diagonal approximation.

Let us consider performing QKD over the end-to-end state of a repeater chain that initially has N identical two-qubit states, $\rho^{\otimes N}$. We assume that entanglement swapping is performed with the *correct-at-end* protocol, where all BSMs may be performed simultaneously and a single Pauli correction is performed at one of the end nodes. Recalling Definition 6.3, this is a swap-and-correct protocol, and we denote it as $\mathscr{P}$. Let $\vec{s}$ be the swap syndrome, which holds the information of the $N-1$ swap outcomes. We let $\rho'_{\vec{s}}$ be the final two-qubit state held by the end nodes, postselected on the syndrome being $\vec{s}$, and $p'_{\vec{s}}$ the probability of measuring $\vec{s}$. The end nodes use the resulting end-to-end states to perform the BBM92 protocol for QKD [104], which is also known as entanglement-based BB84 [204]. The quantum bit error rate (QBER) of this protocol in the X (Z) basis is the probability that when both end nodes measure their state in the X (Z) basis, they obtain different outcomes. In the protocol, the end nodes randomly perform such measurements and then use their outcomes to distil a secret key between them. The number of secret bits that can be obtained per measurement of a state σ in the asymptotic limit is the secret-key fraction, which is given by [213]

$$\mathrm{SKF}(\sigma) = \max(0, 1 - h(Q_X(\sigma)) - h(Q_Z(\sigma))), \tag{6.99}$$

where $h(x) = -x\log_2(x) - (1-x)\log_2(1-x)$ is the binary entropy function and $Q_X(\sigma)$ and $Q_Z(\sigma)$ are the QBER of the state σ in the X and Z basis respectively. While the secret-key fraction is one in the perfect case when both QBERs are zero, it will become zero when the error rates are too large. We note that, the two Pauli bases used throughout the protocol (in this case X and Z) may be chosen from the Pauli bases. For example, if the X and Y bases are chosen instead, then the secret-key fraction (6.99) will instead depend on the QBER in the Y-basis, $Q_Y(\sigma)$. The secret-key fraction is invariant under the Bell-diagonal approximation,

$$\mathrm{SKF}(\sigma) = \mathrm{SKF}(\mathscr{B}(\sigma)). \tag{6.100}$$

For a proof of this, we refer to Appendix 6.7.5. Let λ_{ij} be the Bell-diagonal elements of σ, such that

$$\mathscr{B}(\sigma) = \sum_{i,j=0}^{1} \lambda_{ij} \left|\Psi_{ij}\right\rangle\!\left\langle\Psi_{ij}\right|. \tag{6.101}$$

Then, the QBER in each measurement basis is given by

$$Q_X(\sigma) = \lambda_{01} + \lambda_{11} \tag{6.102}$$
$$Q_Y(\sigma) = \lambda_{10} + \lambda_{01} \tag{6.103}$$
$$Q_Z(\sigma) = \lambda_{10} + \lambda_{11}. \tag{6.104}$$

Given our setup, we compare two different ways in which the end nodes can distil a secret key. The first option is to process measurement outcomes without keeping track of the syndrome. We call this the *non-postselected protocol*, and has secret-key fraction $\mathrm{SKF}(\rho')$, where

$$\rho' = \sum_{\vec{s}} p'_{\vec{s}} \rho'_{\vec{s}} = \Lambda_{\mathscr{P}}(\rho^{\otimes N}), \tag{6.105}$$

and $\Lambda_{\mathscr{P}}(\rho^{\otimes N})$ is the channel induced by non-postselected swapping with $\mathscr{P}$. By (6.100) and Theorem 6.1, we have

$$\mathrm{SKF}(\rho') = \mathrm{SKF}(\mathscr{B}(\rho')) = \mathrm{SKF}(\rho'_{\mathscr{B}}), \tag{6.106}$$

where $\rho'_{\mathscr{B}}$ is the output state with the Bell-diagonal approximation. In particular, the secret-key fraction with the non-postselected protocol is exactly what is obtained with the Bell-diagonal approximation.

The second option for the distillation of secret key is to divide all measurement outcomes into different blocks based on their corresponding syndromes and process each block separately. We call this the *postselected protocol* and write its secret-key fraction as $\mathrm{SKF}_{\mathrm{con}}$. It can be calculated as

$$\sum_{\vec{s}} p'_{\vec{s}} \mathrm{SKF}(\rho'_{\vec{s}}). \tag{6.107}$$

Because the SKF function is convex within the domain where it is nonzero, we have

$$\sum_{\vec{s}} p'_{\vec{s}} \mathrm{SKF}(\rho'_{\vec{s}}) \geq \mathrm{SKF}(\rho'). \tag{6.108}$$

We consider two types of initial states: firstly, we consider the state ρ_{opt} from (6.52) that achieves the highest end-to-end fidelity. Secondly, we consider the R state ρ_R from (6.88). The QKD measurement basis for each state was the one found to provide maximum secret-key fraction. The results can be seen in Figure 6.5. We see from Figure 6.5a that, when the initial states are $\rho_{\mathrm{opt}}^{\otimes N}$, the Bell-diagonal approximation (equivalently, the non-postselected protocol) causes a significant reduction in the secret-key fraction, especially for repeater chains with a larger number of initial states N. By contrast, from Figure 6.5b we see that when the initial states are $\rho_R^{\otimes N}$, the Bell-diagonal approximation (non-postselected protocol) causes a negligible reduction in the resulting secret-key fraction. This behaviour reflects the discussion in Section 6.5.1, where we saw that there is not a significant difference in end-to-end fidelity from the Bell-diagonal approximation when swapping $\rho_R^{\otimes 2}$ (see (6.86)). This is in contrast to swapping $\rho_{\mathrm{opt}}^{\otimes 2}$, which admits the greatest possible variation in end-to-end fidelity above the Bell-diagonal approximation.

We see in Figure 6.5a that the secret-key fraction is reduced drastically when the Werner approximation is used for the initial states $\mathscr{W}(\rho_{\mathrm{opt}})^{\otimes N}$, and we see that the length of the chain over which it is possible to distil key is limited to $N = 3$ initial states. By contrast, with the Bell-diagonal approximation, one can distil key for any length of chain. The reason for this is as follows: we note that $\mathscr{B}(\rho_{\mathrm{opt}}) = F|\Psi_{00}\rangle\langle\Psi_{00}| + (1-F)|\Psi_{11}\rangle\langle\Psi_{11}|$ is a rank-two Bell-diagonal state. Then, recalling Lemma 6.4 and surrounding discussion, the resulting state ρ'_{opt} after non-postselected swapping $\rho_{\mathrm{opt}}^{\otimes N}$ has Bell-diagonal components

$$\mathscr{B}\left(\rho'_{\mathrm{opt}}\right) = \left(\frac{1}{2} + \frac{1}{2}(2F-1)^N\right)|\Psi_{00}\rangle\langle\Psi_{00}| + \left(\frac{1}{2} - \frac{1}{2}(2F-1)^N\right)|\Psi_{11}\rangle\langle\Psi_{11}|. \tag{6.109}$$

By (6.102)-(6.104), the resulting QBER in each basis is then

$$Q_X(\rho'_{\text{opt}}) = \frac{1}{2} - \frac{1}{2}(2F-1)^N \tag{6.110}$$

$$Q_Y(\rho'_{\text{opt}}) = 0 \tag{6.111}$$

$$Q_Z(\rho'_{\text{opt}}) = \frac{1}{2} - \frac{1}{2}(2F-1)^N. \tag{6.112}$$

Choosing to measure in the X and Y bases then provides the highest secret-key fraction, given by

$$\text{SKF}\left(\rho'_{\text{opt}}\right) = 1 - h\left(\frac{1}{2} - \frac{1}{2}(2F-1)^N\right) \tag{6.113}$$

$$> 1 - h\left(\frac{1}{2}\right) = 0. \tag{6.114}$$

In particular, the fact that $\mathscr{B}(\rho_{\text{opt}})$ is of rank two means that the secret-key fraction is greater than zero for any number of initial states N in the chain. By contrast, with the Werner approximation for the initial states $\mathscr{W}(\rho_{\text{opt}})^{\otimes N}$, by Corollary 6.3 the end-to-end state $\rho'_{\mathscr{W}}$ is also Werner with fidelity given in (6.37). The corresponding QBER for each basis is then

$$\begin{aligned} Q_X(\rho'_{\mathscr{W}}) = Q_Y(\rho'_{\mathscr{W}}) &= Q_Z(\rho'_{\mathscr{W}}) \\ &= \frac{1}{2} - \frac{1}{2}\left(\frac{4F-1}{3}\right)^N. \end{aligned} \tag{6.115}$$

We therefore see that

$$\text{SKF}(\rho'_{\mathscr{W}}) = \max\left(0, 1 - 2h\left(\frac{1}{2} - \frac{1}{2}\left(\frac{4F-1}{3}\right)^N\right)\right), \tag{6.116}$$

which will eventually decrease to zero as N increases. Since in Figure 6.5b, the initial states ρ_R are each set to have the same fidelity $\langle\Psi_{00}|\rho_R|\Psi_{00}\rangle = \langle\Psi_{00}|\rho_{\text{opt}}|\Psi_{00}\rangle = 0.95$, we have $\mathscr{W}(\rho_R) = \mathscr{W}(\rho_{\text{opt}})$, and the Werner approximation gives the same result in both cases.

When the initial states are instead R states $\rho_R^{\otimes N}$, in contrast to the case of the optimal states, the secret-key fraction with the Bell-diagonal approximation will eventually reach zero. This is because the Bell-diagonal approximation of an R state is given by

$$\mathscr{B}(\rho_R) = F|\Psi_{00}\rangle\langle\Psi_{00}| + \frac{1}{2}(1-F)|\Psi_{10}\rangle\langle\Psi_{10}| + \frac{1}{2}(1-F)|\Psi_{11}\rangle\langle\Psi_{11}|,$$

and this has rank three. From the map (6.20), it can be seen that swapping identical rank-three Bell-diagonal states results in a rank-four state. Therefore, the non-postselected outcome of swapping the states $\rho_R^{\otimes N}$ will result in an end-to-end state ρ'_R such that $\mathscr{B}(\rho'_R)$ has rank four. In particular, the secret-key fraction will eventually decrease to zero as the number of swaps N increases, unlike the behaviour we saw for ρ_{opt} in (6.114), where the secret-key fraction was always positive since the end-to-end state was always within the rank-two subspace.

6.6. CONCLUSION

We have seen that, for non-postselected swapping, using twirled approximations in a repeater chain can be exact or highly accurate in certain important scenarios. In particular, the Bell-diagonal approximation is exact for evaluating the Bell-diagonal components of the end-to-end state, and in many scenarios, non-postselected swapping and Bell-diagonal twirling are therefore equivalent. Moreover, for non-postselected swapping the Werner approximation is accurate in a high-fidelity regime compared to the number of initial states in the chain. The disadvantages of twirled approximations mostly arise when postselecting on the BSM measurement outcome. For postselected swapping, we have presented bounds on the end-to-end fidelity, given a general noisy form for the initial states when there are $N = 2$ initial states in the chain. With an example of evaluating the secret-key fraction when performing QKD, we demonstrated how the insights from our work may be used to determine whether the twirled approximation is accurate in a given scenario.

6.7. APPENDIX

6.7.1. NON-POSTSELECTED SWAPPING

REPEATER CHAINS WITH $N = 2$

Proof of Corollary 6.1. Let the Bell-diagonal elements of ρ be given by λ_{ij}, as in (6.1). Then, by (6.14), we have

$$\Lambda_\rho^{\text{tel}}(\sigma) = 4\,\langle\Psi_{00}|\sigma_C \otimes \mathscr{B}(\rho_{AB})\,|\Psi_{00}\rangle_{CA} \tag{6.117}$$

$$= 4\sum_{i,j=0}^{1} \lambda_{ij}\,\langle\Psi_{00}|\sigma_C \otimes \left|\Psi_{ij}\right\rangle\!\left\langle\Psi_{ij}\right|_{AB}|\Psi_{00}\rangle \tag{6.118}$$

$$= \sum_{i,j=0}^{1} \lambda_{ij} X^i Z^j \sigma (X^i Z^j)^\dagger, \tag{6.119}$$

where in the final step we have used

$$\langle\Psi_{00}|\sigma_C \otimes \left|\Psi_{ij}\right\rangle\!\left\langle\Psi_{ij}\right|_{AB}|\Psi_{00}\rangle_{CA} = (X^i Z^j)_B\,\langle\Psi_{00}|\sigma_C \otimes |\Psi_{00}\rangle\langle\Psi_{00}|_{AB}|\Psi_{00}\rangle_{CA}\,(X^i Z^j)_B^\dagger, \tag{6.120}$$

and the noticed that

$$\langle\Psi_{00}|\sigma_C \otimes |\Psi_{00}\rangle\langle\Psi_{00}|_{AB}|\Psi_{00}\rangle_{CA} = \frac{1}{4} X^i Z^j \sigma (X^i Z^j)^\dagger$$

is the (non-normalised) result after the perfect teleportation of σ_C. □

Proof of Lemma 6.3. In this proof, we make use of the flip-flop trick, which is that for any linear operator M, we have

$$M \otimes I\,|\Psi_{00}\rangle = I \otimes M^T\,|\Psi_{00}\rangle\,. \tag{6.121}$$

This is also known as the flip-flop trick.

Suppose that in the entanglement swap, the BSM outcome is mn. The output state is then given by

$$\rho'_{mn} = \frac{L_{mn}}{\mathrm{Tr}\,[L_{mn}]}, \tag{6.122}$$

where

$$L_{mn} = (Z^n X^m)_{B_2} \langle \Psi_{mn}| \mathscr{B}(\rho_1) \otimes \mathscr{B}(\rho_2) |\Psi_{mn}\rangle_{A_1A_2} (X^m Z^n)_{B_2}. \tag{6.123}$$

We now compute L_{mn}. We firstly consider the impact of each diagonal element:

$$\begin{aligned}
&(Z^n X^m)_{B_2} \langle \Psi_{mn}|_{A_1A_2} \left[\left|\Psi_{i_1 j_1}\right\rangle_{B_1A_1} \otimes \left|\Psi_{i_2 j_2}\right\rangle_{A_2B_2} \right] \\
&\quad = \langle \Psi_{00}|_{A_1A_2} (Z^n X^m)_{A_2} \left[(X^{i_1} Z^{j_1})_{A_1} (X^{i_2} Z^{j_2})_{A_2} (Z^m X^m)_{B_2} |\Psi_{00}\rangle_{B_1A_1} \otimes |\Psi_{00}\rangle_{A_2B_2} \right] \\
&\quad \overset{a}{=} (Z^n X^m Z^{j_2} X^{i_2} X^m Z^n X^{i_1} Z^{j_1})_{B_2} \langle \Psi_{00}|_{A_1A_2} \left[|\Psi_{00}\rangle_{B_1A_1} \otimes |\Psi_{00}\rangle_{A_2B_2} \right] \\
&\quad \overset{b}{=} \pm 1 \cdot \left(X^{2m+i_1+i_2} Z^{2n+j_1+j_2} \right)_{B_2} \cdot \frac{1}{2} |\Psi_{00}\rangle_{B_1B_2} \\
&\quad \overset{c}{=} (\pm 1) \cdot \frac{1}{2} \left|\Psi_{i_1+i_2, j_1+j_2}\right\rangle_{B_1B_2},
\end{aligned} \tag{6.124}$$

where the addition in the subscript is modulo 2. In step (a), we have made use of the flip-flop trick multiple times to move all Pauli operators onto register B_2. In step (b), we have used the fact that

$$\langle \Psi_{00}|_{A_1A_2} \left[|\Psi_{00}\rangle_{B_1A_1} \otimes |\Psi_{00}\rangle_{A_2B_2} \right] = \frac{1}{2} |\Psi_{00}\rangle_{B_1B_2} \tag{6.125}$$

and that reordering Pauli operators may sometimes incur a factor of -1. In step (c), we have used the definition (6.4) of the Bell basis. Relabelling the eigenvalues as $\mathscr{B}(\rho_1) = \sum_{i,j} \lambda_{ij} |\Psi_{ij}\rangle\langle\Psi_{ij}|$ and $\mathscr{B}(\rho_1) = \sum_{i,j} \mu_{ij} |\Psi_{ij}\rangle\langle\Psi_{ij}|$, from (6.123) we see that L_{mn} is given by

$$\begin{aligned}
&\sum_{i_1,j_1,i_2,j_2} \lambda_{i_1 j_1} \mu_{i_2 j_2} Z^m X^n \langle \Psi_{mn}|_{A_1A_2} \left[\left|\Psi_{i_1 j_1}\right\rangle\!\left\langle\Psi_{i_1 j_1}\right|_{B_1A_1} \otimes \left|\Psi_{i_2 j_2}\right\rangle\!\left\langle\Psi_{i_2 j_2}\right|_{A_2B_2} \right] |\Psi_{mn}\rangle_{A_2B_2} X^n Z^m \\
&\quad = \sum_{i_1,j_1,i_2,j_2} \lambda_{i_1 j_1} \mu_{i_2 j_2} \cdot (\pm 1)^2 \cdot \frac{1}{4} \left|\Psi_{i_1+i_2, j_1+j_2}\right\rangle\!\left\langle\Psi_{i_1+i_2, j_1+j_2}\right|_{B_1A_1},
\end{aligned}$$

and so

$$\mathrm{Tr}\,[L_{mn}] = \sum_{i_1,j_1,i_2,j_2} \lambda_{i_1 j_1} \mu_{i_2 j_2} \cdot \frac{1}{4} = \frac{1}{4} = p'_{mn},$$

due to normalisation of the initial states. In the above, p'_{mn} is the probability of obtaining outcome mn in the BSM. From (6.122), we therefore see that the full swap outcome after measuring mn is

$$\rho'_{B_1B_1} = \sum_{i_1,j_1,i_2,j_2} \lambda_{i_1 j_1} \mu_{i_2 j_2} \left|\Psi_{i_1+i_2, j_1+j_2}\right\rangle\!\left\langle\Psi_{i_1+i_2, j_1+j_2}\right|_{B_1B_2}.$$

In particular, this is Bell-diagonal and independent of the measurement outcome. In the four-vector notation from (6.18), for $\mathscr{B}(\rho_1) \equiv (\lambda_0,\ldots,\lambda_3)^T$ and $\mathscr{B}(\rho_2) \equiv (\mu_0,\ldots,\mu_3)^T$, the end-to-end state is $\rho'_{mn} \equiv (\lambda'_0,\ldots,\lambda'_3)^T$, where

$$\begin{pmatrix} \lambda'_0 \\ \lambda'_1 \\ \lambda'_2 \\ \lambda'_3 \end{pmatrix} = \begin{pmatrix} \lambda_0\mu_0 + \lambda_1\mu_1 + \lambda_2\mu_2 + \lambda_3\mu_3 \\ \lambda_0\mu_1 + \lambda_1\mu_0 + \lambda_2\mu_3 + \lambda_3\mu_2 \\ \lambda_0\mu_2 + \lambda_2\mu_0 + \lambda_3\mu_1 + \lambda_1\mu_3 \\ \lambda_0\mu_3 + \lambda_3\mu_0 + \lambda_1\mu_2 + \lambda_2\mu_1 \end{pmatrix} \equiv \rho'_{\mathscr{B}}. \tag{6.126}$$

□

NON-POSTSELECTED SWAPPING ON REPEATER CHAINS WITH $N > 2$

Definition 6.6 (Swap-and-correct protocol, technical)**.** For a length-N repeater chain, a *swap-and-correct* protocol $\mathscr{P}$ dictates where to apply Pauli corrections. Given the $N-1$ BSM outcomes that form the syndrome $\vec{s}$, $\mathscr{P}$ is a map

$$\mathscr{P} : \{I, X, Z, XZ\}^{N-1} \to \{I, X, Z, XZ\}^{N+1} \tag{6.127}$$

such that $\mathscr{P}_k(\vec{s})$ is the correction applied to node k, given the syndrome $\vec{s}$. The syndrome is denoted such that $s_i \equiv X^m Z^n$ means that outcome $|\Psi_{mn}\rangle$ was measured on node i. Moreover, $\mathscr{P}$ satisfies the following two properties:

(A) $\mathscr{P}$ **is physically implementable.** For any swap-and-correct protocol $\mathscr{P}$, there exists an associated permutation $\alpha \in \mathrm{Sym}(N-1)$ in which the $N-1$ BSMs are carried out, where Sym denotes the symmetric group. For $\mathscr{P}$ to be physical, then before the kth BSM, the correction $\mathscr{P}_{\alpha(k)}$ must only depend on outcomes of BSMs that have already been carried out, which are given by $(\alpha(1),\ldots,\alpha(k-1))$.

(B) $\mathscr{P}$ **is correct.** For any syndrome $\vec{s}$, $\mathscr{P}$ transforms $|\Psi_{00}\rangle\langle\Psi_{00}|^{\otimes N}$ into $|\Psi_{00}\rangle\langle\Psi_{00}|$.

By the assumption (A), we slightly abuse notation to write $\mathscr{P}_{\alpha(k)}(s_{\alpha(1)},\ldots,s_{\alpha(k-1)}) \equiv \mathscr{P}_{\alpha(k)}(\vec{s})$. Note that the first correction, $\mathscr{P}_{\alpha(1)}$, is independent of $\vec{s}$.

In particular, given a swap-and-correct protocol $\mathscr{P}$, it may be executed as follows. Given syndrome $\vec{s}$,

(1) Apply correction $\mathscr{P}_{\alpha(1)} \equiv \mathscr{P}_{\alpha(1)}(\vec{s})$ to node $\alpha(1)$. Apply BSM at node $\alpha(1)$ to get outcome $s_{\alpha(1)}$.

(2) Apply correction $\mathscr{P}_{\alpha(2)}(s_{\alpha(1)}) \equiv \mathscr{P}_{\alpha(2)}(\vec{s})$ to node $\alpha(2)$. Apply BSM at node $\alpha(2)$ to get outcome $s_{\alpha(2)}$.

⋮

($N-1$) Apply correction $\mathscr{P}_{\alpha(N-1)}(s_{\alpha(1)},\ldots,s_{\alpha(N)}) \equiv \mathscr{P}_{\alpha(N-1)}(\vec{s})$ to node $\alpha(N-1)$. Apply BSM at node $\alpha(N-1)$ to get outcome $s_{\alpha(N-1)}$.

(N) Apply corrections $\mathscr{P}_0(\vec{s})$ and $\mathscr{P}_N(\vec{s})$ to nodes 0 and N.

We note that in principle, corrections may be applied at any point in the protocol up to the BSM on that node. Similarly, corrections may be applied at the end nodes at any point in the protocol. However, both strategies are captured by the above formalism by simply combining all corrections and applying them just before the BSM (for the repeater nodes), and after all BSMs (for the end nodes).

In the following, we consider swap-and-correct protocols $\mathscr{P}$. As discussed above, we note that such a protocol consists of Bell-state measurement

For the end nodes, which have indices 0 and N, there is only one qubit that this can be applied to. However, the repeater nodes $1,\dots N-1$ each hold two qubits, and so there is a choice of which qubit to apply the correction. We now claim that applying the correction to either qubit will give the same result.

Let the qubit registers in the kth node be denoted as k_1 and k_2. Given a swap-and-correct protocol with associated permutation $\alpha \in \mathrm{Sym}$, the correction at that node is $\mathscr{P}_{\alpha^{-1}(k)}(\vec{s})$. Directly afterwards, the BSM on node k will be applied. Suppose that the correction is applied to register k_1, and the BSM outcome is $s \in A$ (recalling from Definition 6.6 that we use of A to denote the result of the BSM). Letting $c = \mathscr{P}_{\alpha^{-1}(k)}(\vec{s}) \in A$ denote the correction, the resulting projection on the total state ρ of the chain is then

$$\mathrm{Tr}_{k_1k_2}\left[s_{k_2}|\Psi_{00}\rangle\langle\Psi_{00}|_{k_1k_2}\, s_{k_2}^\dagger c_{k_1}\rho c_{k_1}^\dagger\right] \overset{\text{i}}{=} \mathrm{Tr}_{k_1k_2}\left[s_{k_2}c_{k_2}|\Psi_{00}\rangle\langle\Psi_{00}|_{k_1k_2}\, c_{k_2}^\dagger s_{k_2}^\dagger \rho\right] \tag{6.128}$$

$$\overset{\text{ii}}{=} \mathrm{Tr}_{k_1k_2}\left[c_{k_2}s_{k_2}|\Psi_{00}\rangle\langle\Psi_{00}|_{k_1k_2}\, s_{k_2}^\dagger c_{k_2}^\dagger \rho\right] \tag{6.129}$$

$$\overset{\text{iii}}{=} \mathrm{Tr}_{k_1k_2}\left[s_{k_2}|\Psi_{00}\rangle\langle\Psi_{00}|_{k_1k_2}\, s_{k_2}^\dagger c_{k_2}\rho c_{k_2}^\dagger\right] \tag{6.130}$$

where we have (i) used cyclicity of the trace and the flip-flop trick (6.121), (ii) used the fact that $sc = \pm cs$ for $c, s \in A$, and (iii) used cyclicity of the trace and $s^\dagger = \pm s$, for $s \in A$. In particular, we notice that (6.130) corresponds to applying the Pauli correction to register k_2. We therefore see that, as long as a BSM is applied after the correction, it does not matter to which qubit the correction is applied. The same holds for Pauli operators arising from Bell-diagonal twirling (see Definition 6.9).

Definition 6.7. For a unitary U we denote the corresponding channel as

$$U(\rho) = U\rho U^\dagger. \tag{6.131}$$

We will use the above notation principally for $U \in A$, where $A = \{I, X, Z, XZ\}$ is the set of possible Pauli corrections. We will let $(s)_k$ denote the Pauli correction s applied to node k.

Definition 6.8. For any map Λ acting on two-qubit states, we denote $(\Lambda)_k$ to be this channel applied to the two qubits in node k.

Definition 6.9 (Twirling of the kth state)**.** For $k = 1,\dots N-1$, we denote the twirling map of the kth state in the chain (shared between nodes $k-1$ and k) as

$$\mathscr{B}_{k-1,k} = \frac{1}{4}\sum_{s\in A}(s)_{k-1} \circ (s)_k. \tag{6.132}$$

In (6.132), we have recalled the Bell-diagonal twirling map from Lemma 6.1. To simplify notation, in (6.132) we have not specified the specific qubit of each repeater nodes to which twirling is applied, because we will always be interested in the case where the map $\mathscr{B}_{k-1,k}$ is applied *before* the BSMs. By the same argument used to obtain (6.130), applying the correction to either qubit of the repeater node is equivalent.

Lemma 6.6 (Properties of Pauli operators and Bell-diagonal twirling)**.** *The following properties hold:*

(i) For $s_1, s_2 \in A$,
$$(s_1)_k \circ (s_2)_k = (s_1 s_2)_k = (s_2 s_1)_k = (s_2)_k \circ (s_1)_k. \tag{6.133}$$

(ii) For $s \in A$, $(s^\dagger)_k = (s)_k$.

(iii) For $s \in A$, *we have* $(s)_{k-1} \circ \mathscr{B}_{k-1,k} = \mathscr{B}_{k-1,k} \circ (s)_k$.

Proof. (i) We use the fact that, although interchanging the order of the Pauli operators may incur a sign difference $s_1 s_2 = \pm s_2 s_1$, this will not affect the channel (6.131) because the sign is global.

(ii) We use that for any $s \in A$ we have $s^\dagger = \pm s$, and the incurred sign does not affect the channel.

(iii) Recall the channel $\mathscr{B}_k$ from (6.132). For $s \in A$, we have
$$(s)_{k-1} \circ \mathscr{B}_{k-1,k} = \frac{1}{4} \sum_{s' \in A} (ss')_{k-1} \circ (s')_k \tag{6.134}$$
$$= \frac{1}{4} \sum_{r \in A} (r)_{k-1} \circ (s^\dagger r)_k \tag{6.135}$$
$$= \mathscr{B}_{k-1,k} \circ (s)_k, \tag{6.136}$$
where in the second line we have made the change of variable $r = ss'$, and used properties (i) and (ii).

□

Definition 6.10 (Bell-state projection, single repeater node)**.** For $k = 1, \ldots, N-1$, suppose that BSM outcome mn is obtained at node k. Let $s = X^m Z^n \in A$. We denote the unnormalised map corresponding to projection onto this Bell state as
$$(M_s)_k(\rho) = \mathrm{Tr}_k \left[s |\Psi_{00}\rangle\langle\Psi_{00}|_k s^\dagger \cdot \rho \right]. \tag{6.137}$$

From (6.131), we note that (6.137) may be written as
$$(M_s)_k = (M_I \circ s)_k. \tag{6.138}$$

Definition 6.11 (Postselected projection, swap-and-correct protocol)**.** Let $\mathscr{P}$ be a swap-and-correct protocol for repeater chains with N initial states and $N-1$ repeaters. Then, Let $\vec{s} \in A^{N-1}$ be the swap syndrome. Wwe define
$$\Lambda_{\mathscr{P},\vec{s}} = \bigcirc_{k=1}^{N-1} (M_{s_k})_k \bigcirc_{j=0}^{N} (\mathscr{P}_j(\vec{s}))_j. \tag{6.139}$$
to be the non-normalised map where syndrome $\vec{s}$ is measured and $\mathscr{P}$ is applied.

Lemma 6.7 (Correctness property)**.** *For any swap-and-correct protocol $\mathscr{P}$ and syndrome $\vec{s}$, the correctness condition (property B in Definition 6.6) implies that*

$$\mathscr{P}_0(\vec{s})\cdot\mathscr{P}_N(\vec{s})\cdot\prod_{j=1}^{N-1}s_j\mathscr{P}_j(\vec{s})=\pm I. \tag{6.140}$$

Proof. Recalling (6.139), (6.138) and (6.133), we may rewrite

$$\Lambda_{\mathscr{P},\vec{s}}=\bigcirc_{k=1}^{N-1}(M_I)_k\circ\bigcirc_{j=1}^{N-1}(s_j\mathscr{P}_j(\vec{s}))_j\circ(\mathscr{P}_0(\vec{s}))_0\circ(\mathscr{P}_N(\vec{s}))_N. \tag{6.141}$$

We firstly note that, given $s\in A$ and the pure state $|\Psi_{00}\rangle\langle\Psi_{00}|_{i,j}$ shared between registers i and j, we have

$$(s)_i\left(|\Psi_{00}\rangle\langle\Psi_{00}|_{i,j}\right)=(s)_j\left(|\Psi_{00}\rangle\langle\Psi_{00}|_{i,j}\right), \tag{6.142}$$

where we have used the flip-flop trick (6.121).

Letting $\rho_{\text{ideal}}=|\Psi_{00}\rangle\langle\Psi_{00}|^{\otimes N}$, it follows from (6.141) and (6.142) that one may move all Pauli operators to act on node N,

$$\begin{aligned}\Lambda_{\mathscr{P},\vec{s}}\left(\rho_{\text{ideal}}\right)&=\bigcirc_{k=1}^{N-1}(M_I)_k\circ\bigcirc_{j=1}^{N-1}(s_j\mathscr{P}_j(\vec{s}))_N\circ(\mathscr{P}_0(\vec{s}))_N\circ(\mathscr{P}_N(\vec{s}))_N(\rho_{\text{ideal}})\\&=\bigcirc_{k=1}^{N-1}(M_I)_k\circ\left(\mathscr{P}_0(\vec{s})\cdot\mathscr{P}_N(\vec{s})\cdot\prod_{j=1}^{N-1}s_j\mathscr{P}_j(\vec{s})\right)_N(\rho_{\text{ideal}}).\end{aligned} \tag{6.143}$$

Now, by recursively applying (6.125), we see that

$$\bigcirc_{k=1}^{N-1}(M_I)_k(\rho_{\text{ideal}})=\frac{1}{4^{N-1}}|\Psi_{00}\rangle\langle\Psi_{00}|_{0,N}. \tag{6.144}$$

Then, (6.143) simplifies to

$$\Lambda_{\mathscr{P},\vec{s}}\left(\rho_{\text{ideal}}\right)=\frac{1}{4^{N-1}}\left(\mathscr{P}_0(\vec{s})\cdot\mathscr{P}_N(\vec{s})\cdot\prod_{j=1}^{N-1}s_j\mathscr{P}_j(\vec{s})\right)_N\left(\rho_{\text{ideal}}\right), \tag{6.145}$$

from which we see that $\text{Tr}\left[\Lambda_{\mathscr{P},\vec{s}}\left(\rho_{\text{ideal}}\right)\right]=1/4^{N-1}$. The output state after measuring syndrome $\vec{s}$ and applying protocol $\mathscr{P}$ is then

$$\frac{\Lambda_{\mathscr{P},\vec{s}}\left(\rho_{\text{ideal}}\right)}{\text{Tr}\left[\Lambda_{\mathscr{P},\vec{s}}\left(\rho_{\text{ideal}}\right)\right]}=\left(\mathscr{P}_0(\vec{s})\cdot\mathscr{P}_N(\vec{s})\cdot\prod_{j=1}^{N-1}s_j\mathscr{P}_j(\vec{s})\right)_N\left(\rho_{\text{ideal}}\right). \tag{6.146}$$

For the correctness condition $\Lambda_{\mathscr{P},\vec{s}}\left(\rho_{\text{ideal}}\right)=\rho_{\text{ideal}}$ to hold, we therefore require

$$\mathscr{P}_0(\vec{s})\cdot\mathscr{P}_N(\vec{s})\cdot\prod_{j=1}^{N-1}s_j\mathscr{P}_j(\vec{s})=\pm I. \tag{6.147}$$

□

Lemma 6.8. *For any swap-and-correct protocol $\mathscr{P}$ and syndrome $\vec{s}$, we have*

$$\Lambda_{\mathscr{P},\vec{s}} \circ \bigcirc_{k=1}^{N} \mathscr{B}_{k-1,1} = \Lambda_{\mathrm{seq},\vec{I}} \circ \bigcirc_{k=1}^{N} \mathscr{B}_{k-1,1}, \tag{6.148}$$

where $\vec{I} = (I,\dots,I)$ is the syndrome with outcome 00 *at every repeater node, and* seq *is the sequential swapping protocol. In particular, if Bell-diagonal twirling is firstly applied to all initial states, then the end-to-end state is independent of $\vec{s}$ and $\mathscr{P}$.*

Proof. Recalling the identity (6.139) for $\Lambda_{\mathscr{P},\vec{s}}$ and property (iii) in Lemma 6.6, we see that

$$\Lambda_{\mathscr{P},\vec{s}} \circ \bigcirc_{l=1}^{N} \mathscr{B}_{l-1,l} = \bigcirc_{k=1}^{N-1} (M_{s_k})_k \circ \bigcirc_{l=1}^{N} \mathscr{B}_{l-1,1} \circ \bigcirc_{j=0}^{N} (\mathscr{P}_j(\vec{s}))_N. \tag{6.149}$$

In particular, we have moved all Pauli corrections to act at register N (the end node). Recalling (6.138), we can do the same with all Pauli operators arising from the syndrome, to obtain

$$\begin{aligned}\Lambda_{\mathscr{P},\vec{s}} \circ \bigcirc_{l=1}^{N} \mathscr{B}_{l} &= \bigcirc_{k=1}^{N-1} (M_I)_k \circ \bigcirc_{l=1}^{N} \mathscr{B}_{l-1,1} \circ \bigcirc_{j=1}^{N-1} (s_j \mathscr{P}_j(\vec{s}))_N \circ (\mathscr{P}_0(\vec{s}))_N \circ (\mathscr{P}_N(\vec{s}))_N \\ &= \bigcirc_{k=1}^{N-1} (M_I)_k \circ \bigcirc_{l=1}^{N} \mathscr{B}_{l-1,1} \circ \left(\mathscr{P}_0(\vec{s}) \cdot \mathscr{P}_N(\vec{s}) \cdot \prod_{j=1}^{N-1} s_j \mathscr{P}_j(\vec{s}) \right)_N. \end{aligned} \tag{6.150}$$

By Lemma 6.7, due to the correctness property of $\mathscr{P}$, the term inside the brackets is simply the identity. Then,

$$\Lambda_{\mathscr{P},\vec{s}} \circ \bigcirc_{l=1}^{N} \mathscr{B}_{l-1,1} = \bigcirc_{k=1}^{N-1} (M_I)_k \circ \bigcirc_{l=1}^{N} \mathscr{B}_{l-1,1}. \tag{6.151}$$

We now note that $\Lambda_{\mathrm{seq},\vec{I}} = \bigcirc_{k=1}^{N-1} (M_I)_k$, where seq denotes the sequential swapping protocol. This is because, given syndrome $\vec{I}$, all corrections specified by this protocol are the identity. Therefore,

$$\Lambda_{\mathscr{P},\vec{s}} \circ \bigcirc_{l=1}^{N} \mathscr{B}_{l-1,1} = \Lambda_{\mathrm{seq},\vec{I}} \circ \bigcirc_{l=1}^{N} \mathscr{B}_{l}. \tag{6.152}$$

□

Proof of Theorem 6.1. Recalling Definition 6.11, we may write down the map corresponding to *non-postselected* swapping with swap-and-correct protocol $\mathscr{P}$ as

$$\Lambda_{\mathscr{P}} \coloneqq \sum_{\vec{s} \in A^{N-1}} \Lambda_{\mathscr{P},\vec{s}}. \tag{6.153}$$

where the sum is over $\vec{s} \in A^{N-1}$, and $\Lambda_{\mathscr{P},\vec{s}}$ is given in (6.139). We now claim that

$$\mathscr{B}_{0,N} \circ \Lambda_{\mathscr{P}} = \Lambda_{\mathrm{seq},\vec{I}} \circ \bigcirc_{k=1}^{N} \mathscr{B}_{k-1,k}, \tag{6.154}$$

where

$$\mathscr{B}_{0,N} = \frac{1}{4} \sum_{s \in A} (s)_0 \circ (s)_N \tag{6.155}$$

denotes the Bell-diagonal twirling of the end-to-end state. We firstly make use of (6.153) to write

$$\mathscr{B}_{0,N} \circ \Lambda_{\mathscr{P}} = \sum_{\vec{s}} \mathscr{B}_{0,N} \circ \Lambda_{\mathscr{P},\vec{s}}, \tag{6.156}$$

Recalling (6.141), we have

$$\mathscr{B}_{0,N} \circ \Lambda_{\mathscr{P},\vec{s}} = \mathscr{B}_{0,N} \circ \bigcirc_{k=1}^{N-1} (M_I)_k \circ \bigcirc_{j=1}^{N-1} (s_j \mathscr{P}_j(\vec{s}))_j \circ (\mathscr{P}_0(\vec{s})\mathscr{P}_N(\vec{s}))_N \tag{6.157}$$

$$= \frac{1}{4} \sum_{t \in A} \bigcirc_{k=1}^{N-1} (M_I)_k \circ (t)_0 \circ \bigcirc_{j=1}^{N-1} (s_j \mathscr{P}_j(\vec{s}))_j \circ (t\mathscr{P}_0(\vec{s})\mathscr{P}_N(\vec{s}))_N. \tag{6.158}$$

In the first step, we have used (6.141), and then moved the correction $(\mathscr{P}_0(\vec{s}))_0$ to act on register N. This is possible by applying property (iii) of Lemma 6.6 with the twirling operator $\mathscr{B}_{0,N}$. In the second step, we have expanded $\mathscr{B}_{0,N}$ according to (6.155).

We now perform a change of variables: for $j = 1, \ldots, N-1$, we let

$$r_0 := t, \quad r_j := s_{\alpha(j)} \mathscr{P}_{\alpha(j)}(\vec{s}) \text{ for } j = 1, \ldots, N-1, \tag{6.159}$$

where α is the permutation associated with $\mathscr{P}$ in which BSMs are carried out (see Definition 6.6). We now claim that (6.159) defines a bijective map $(t, \vec{s}) \leftrightarrow \vec{r}$, for $\vec{r} = (r_0, \ldots, r_{N-1}) \in A^N$ and $(t, \vec{s}) \in A^N$. It suffices to show that the map (6.159) is invertible. We show invertibility by explicitly writing down the inverse of (6.159). Given $\vec{r}$, one may recursively deduce $(t, \vec{s})$ with the map

$$\begin{aligned} t &= r_0 \\ s_{\alpha(1)} &= \mathscr{P}_{\alpha(1)}(\vec{s}) r_1 \\ s_{\alpha(2)} &= \mathscr{P}_{\alpha(2)}(\vec{s}) r_2 \\ &\vdots \\ s_{\alpha(N-1)} &= \mathscr{P}_{\alpha(N-1)}(\vec{s}) r_{N-1}, \end{aligned}$$

where at each step, the multiplier $\mathscr{P}_{\alpha(k)}(\vec{s})$ may be computed using the values $(s_{\alpha(1)}, \ldots, s_{\alpha(k-1)})$ which have been found in the previous steps. This is due physicality property of the swap-and-correct protocol (see property A in Definition 6.6). This enforces that corrections at a given node must only depend on the BSM outcomes that have previously been obtained. Recalling the identity (6.140), we see that

$$t\mathscr{P}_0(\vec{s})\mathscr{P}_N(\vec{s}) = \pm \prod_{j=0}^{N-1} r_j. \tag{6.160}$$

Given the bijective map $(t, \vec{s}) \leftrightarrow \vec{r}$, we may therefore combine (6.156) with (6.158) and (6.160) to write

$$\mathscr{B}_{0,N} \circ \Lambda_{\mathscr{P}} = \frac{1}{4} \sum_{\vec{r} \in A^{N-1}} \left[\bigcirc_{k=1}^{N-1} (M_I)_k \circ (r_0)_0 \circ \bigcirc_{j=1}^{N-1} (r_j)_j \circ \left(\prod_{i=0}^{N-1} r_i \right)_N \right]. \tag{6.161}$$

We now claim that the above is equal to $\Lambda_{\text{seq},\vec{I}} \circ \bigcirc_{k=1}^{N} \mathscr{B}_{k-1,k}$. to see this, we perform another change of variables to $\vec{u} = (u_0, \ldots, u_{N-1}) \in A^{N-1}$, given by

$$u_0 = r_0, \quad u_k = \prod_{j=0}^{k} r_j, \quad k = 1, \ldots, N-1 \tag{6.162}$$

The above map has inverse

$$r_0 = u_0, \quad r_k = u_{k-1} u_k, \quad k = 1, \ldots, N-1. \tag{6.163}$$

and is therefore bijective. Performing this change of variable on (6.161), we obtain

$$\mathscr{B}_{0,N} \circ \Lambda_{\mathscr{P}} = \frac{1}{4} \sum_{\vec{u} \in A^{N-1}} \left[\bigcirc_{k=1}^{N-1} (M_I)_k \circ (u_0)_0 \circ \bigcirc_{j=1}^{N-1} (u_{j-1} u_j)_j \circ (u_{N-1})_N \right] \tag{6.164}$$

$$= \frac{1}{4} \sum_{\vec{u} \in A^{N-1}} \left[\bigcirc_{k=1}^{N-1} (M_I)_k \circ \bigcirc_{j=0}^{N-1} (u_j)_j \circ (u_j)_{j+1} \right]. \tag{6.165}$$

Recalling (6.132), this may be rewritten as

$$\mathscr{B}_{0,N} \circ \Lambda_{\mathscr{P}} = \frac{4^N}{4} \bigcirc_{k=1}^{N-1} (M_I)_k \circ \bigcirc_{j=1}^{N} \mathscr{B}_{j-1,j} \tag{6.166}$$

$$= 4^{N-1} \Lambda_{\text{seq},\vec{I}} \circ \bigcirc_{j=1}^{N} \mathscr{B}_{j-1,j} \tag{6.167}$$

where have have recalled that $\Lambda_{\text{seq},\vec{I}} = \bigcirc_{k=1}^{N-1} (M_I)_k$. In particular, from Lemma (6.8) we recall that, for any syndrome $\vec{s}$, we have

$$\Lambda_{\mathscr{P},\vec{s}} \circ \bigcirc_{k=1}^{N} \mathscr{B}_{k-1,1} = \Lambda_{\text{seq},\vec{I}} \circ \bigcirc_{k=1}^{N} \mathscr{B}_{k-1,1}.$$

Then, noting that $|A^{N-1}| = 4^{N-1}$, from (6.167) we have

$$\mathscr{B}_{0,N} \circ \Lambda_{\mathscr{P}} = \sum_{\vec{s} \in A^{N-1}} \Lambda_{\mathscr{P},\vec{s}} \circ \bigcirc_{k=1}^{N} \mathscr{B}_{k-1,1} \tag{6.168}$$

$$= \Lambda_{\mathscr{P}} \circ \bigcirc_{k=1}^{N} \mathscr{B}_{k-1,1}, \tag{6.169}$$

where we have used the identity (6.153) for $\Lambda_{\mathscr{P}}$. □

Proof of Lemma 6.4. We then write the eigenvalues of $\mathscr{B}(\rho_1)$ and $\mathscr{B}(\rho_2)$ as

$$\mathscr{B}(\rho_1) \equiv (F_1, (1-F_1)\vec{u})^T,$$
$$\mathscr{B}(\rho_2) \equiv (F_2, (1-F_2)\vec{v})^T,$$

where $\vec{u}$ and $\vec{v}$ are length-3 vectors such that $u_1 + u_2 + u_3 = v_1 + v_2 + v_3 = 1$ and $u_i, v_i \geq 0$. By Lemma 6.3, The fidelity of the end-to-end state is

$$F' = F_1 F_2 + (1-F_1)(1-F_2)(\vec{u} \cdot \vec{v}),$$

where $\vec{u} \cdot \vec{v} = u_1 v_1 + u_2 v_2 + u_3 v_3$. The end-to-end fidelity F' satisfies

$$F_1 F_2 \leq F' \leq F_1 F_2 + (1-F_1)(1-F_2). \tag{6.170}$$

For the lower bound, we have used $\vec{u}\cdot\vec{v}\geq 0$. Note that this is saturated whenever $\vec{u}$ and $\vec{v}$ are orthogonal. For the upper bound, we have used

$$\vec{u}\cdot\vec{v}\leq(u_1+u_2+u_3)(v_1+v_2+v_3)=1,$$

which is saturated exactly when $\vec{u}=\vec{v}$ and $\vec{u}$ is one of $(1,0,0)^T$, $(0,1,0)^T$, or $(0,0,1)^T$. In particular, this is when the initial states $(B)(\rho_k)$ are of rank two and have the same non-zero eigenvectors.

To show the bounds for a length-N repeater chain, we apply the bounds (6.170) inductively. For the upper bound, we notice that

$$F_1F_2+(1-F_1)(1-F_2)=\frac{1}{2}(2F_1-1)(2F_2-1)+\frac{1}{2}, \tag{6.171}$$

and applying the upper and lower bounds from (6.170) inductively gives

$$\prod_{k=1}^{N}F_k\leq F'\leq\frac{1}{2}\prod_{k=1}^{N}(2F_k-1)+\frac{1}{2}. \tag{6.172}$$

By Lemma 6.3, the end-to-end state is the same for both postselected and non-postselected swapping. □

Proof of Theorem 6.2. Let $\mathscr{W}_{[N]}$ denote the Bell-diagonal twirling of all initial states $\rho_{\text{in}}=\bigotimes_{k=1}^{N}\rho_k$ such that

$$\mathscr{W}_{[N]}\left(\rho_{\text{in}}\right)=\otimes_{k=1}^{N}\mathscr{W}(\rho_k). \tag{6.173}$$

Then,

$$F'=\langle\Psi_{00}|\,\Lambda_{\mathscr{P}}(\rho_{\text{in}})\,|\Psi_{00}\rangle \tag{6.174}$$

$$F'_{\mathscr{W}}=\langle\Psi_{00}|\,\Lambda_{\mathscr{P}}\left(\mathscr{W}_{[N]}\left(\rho_{\text{in}}\right)\right)|\Psi_{00}\rangle \tag{6.175}$$

are the true end-to-end fidelity and the end-to-end fidelity with the Werner approximation.

$$F'=\langle\Psi_{00}|\,\Lambda_{\mathscr{P}}(\rho_{\text{in}})\,|\Psi_{00}\rangle=\langle\Psi_{00}|\,\mathscr{B}\left(\Lambda_{\mathscr{P}}(\rho_{\text{in}})\right)|\Psi_{00}\rangle \tag{6.176}$$

$$=\langle\Psi_{00}|\,\Lambda_{\text{seq}}(\mathscr{B}_{[N]}\left(\rho_{\text{in}}\right))\,|\Psi_{00}\rangle\,, \tag{6.177}$$

where in the first step we have used the fact that the fidelity is invariant under Bell-diagonal twirling, and in the second step we have applied Theorem 6.1. Now, since Werner states are Bell-diagonal, we have $\mathscr{B}\left(\mathscr{W}(\rho_k)\right)=\mathscr{W}(\rho_k)$ and $\mathscr{B}_{[N]}\left(\mathscr{W}_{[N]}(\rho_{\text{in}})\right)=\mathscr{W}_{[N]}(\rho_{\text{in}})$. In particular, we see that

$$\Lambda_{\mathscr{P}}\left(\mathscr{W}_{[N]}\left(\rho_{\text{in}}\right)\right)=\Lambda_{\mathscr{P}}\left(\mathscr{B}_{[N]}\left(\mathscr{W}_{[N]}\left(\rho_{\text{in}}\right)\right)\right)=\mathscr{B}\left(\Lambda_{\text{seq}}\left(\mathscr{W}_{[N]}\left(\rho_{\text{in}}\right)\right)\right) \tag{6.178}$$

$$=\Lambda_{\text{seq}}\left(\mathscr{W}_{[N]}\left(\rho_{\text{in}}\right)\right), \tag{6.179}$$

where in the first step we have applied Theorem 6.1, and in the second step we have used the fact that the sequential swapping of Bell-diagonal states results in a Bell-diagonal state (see Lemma 6.3). Then,

$$F'_{\mathscr{W}}=\langle\Psi_{00}|\,\Lambda_{\mathscr{P}}\left(\mathscr{W}_{[N]}\left(\rho_{\text{in}}\right)\right)|\Psi_{00}\rangle=\langle\Psi_{00}|\,\Lambda_{\text{seq}}\left(\mathscr{W}_{[N]}\left(\rho_{\text{in}}\right)\right)|\Psi_{00}\rangle\,. \tag{6.180}$$

In particular, both (6.177) and (6.180) are the end-to-end fidelity after swapping N Bell-diagonal states. For (6.177), the fidelity of the kth state is $\langle\Psi_{00}|\mathscr{B}(\rho_k)|\Psi_{00}\rangle = F_k$. For (6.180), the fidelity of the kth state is again $\langle\Psi_{00}|\mathscr{W}(\rho_k)|\Psi_{00}\rangle = F_k$. In particular, the bounds from Lemma 6.4 apply to both F' and $F'_{\mathscr{W}}$. Letting $\epsilon_k = 1 - F_k$, we then have

$$|F' - F'_{\mathscr{W}}| \leq \frac{1}{2}\prod_{k=1}^{N}(2F_k - 1) + \frac{1}{2} - \prod_{k=1}^{N} F_k \tag{6.181}$$

$$= \frac{1}{2}\prod_{k=1}^{N}(1-2\epsilon_k) + \frac{1}{2} - \prod_{k=1}^{N}(1-\epsilon_k) \tag{6.182}$$

$$\overset{\text{i}}{=} \frac{1}{2}\left(1 - 2\sum_k \epsilon_k + 4\sum_{k,l}\epsilon_k\epsilon_l + \mathscr{O}(N^3\epsilon^3)\right) + \frac{1}{2} - \left(1 - \sum_k \epsilon_k + \sum_{k,l}\epsilon_k\epsilon_l + \mathscr{O}(N^3\epsilon^3)\right) \tag{6.183}$$

$$\overset{\text{ii}}{=} \sum_{k,l}\epsilon_k\epsilon_l + \mathscr{O}(N^3\epsilon^3) \tag{6.184}$$

$$\overset{\text{iii}}{\leq} \binom{N}{2}\epsilon^2 + \mathscr{O}(N^3\epsilon^3), \tag{6.185}$$

where in (i) we have performed a series expansion to second order in the infidelity, in (ii) we have noticed that the first- and second-order terms cancel, and in (iii) we have used $\epsilon_k = 1 - F_k < \epsilon$ for all $k = 1, \ldots, N$ and the fact that there are $\binom{N}{2}$ second-order terms in the sum. □

6.7.2. POSTSELECTED SWAPPING

Proof of Proposition 6.2. Consider $\rho \in S_{p_1}$. By definition, this can be written as

$$\rho = p_1 |\Psi_{00}\rangle\langle\Psi_{00}| + (1-p_1)\sigma, \tag{6.186}$$

where σ is a valid density matrix. Letting $p_1 = p_2 + (p_1 - p_2)$, we may rewrite ρ as

$$\rho_1 = p_2 |\Psi_{00}\rangle\langle\Psi_{00}| + (1-p_2)\left[\frac{p_1 - p_2}{1-p_2}|\Psi_{00}\rangle\langle\Psi_{00}| + \frac{1-p_1}{1-p_2}\sigma\right].$$

Since $p_1 > p_2$, the term in the brackets is a valid ensemble of states and therefore also a density matrix. For the converse, consider for example the states

$$\rho = p_2 |\Psi_{00}\rangle\langle\Psi_{00}| + (1-p_2)|\Psi_{11}\rangle\langle\Psi_{11}|.$$

and suppose that this can be written in the form (6.186). Then, one may show that

$$|\Psi_{11}\rangle\langle\Psi_{11}| = \frac{p_1 - p_2}{1-p_2}|\Psi_{00}\rangle\langle\Psi_{00}| + \frac{1-p_1}{1-p_2}\sigma$$

computing the fidelity to $|\Psi_{00}\rangle$ then yields

$$0 = \frac{p_1 - p_2}{1-p_2} + \frac{1-p_1}{1-p_2}\langle\Psi_{00}|\sigma|\Psi_{00}\rangle$$

which results in a contradiction as the RHS is positive (σ is a density matrix). □

Proposition 6.5 (Proposition 6.2, general version)**.** *For fixed p_1, p_2, F_1, F_2, let*

$$F'_{ij,\max} \coloneqq \max\left\{F'_{ij}(\rho_1,\rho_2) \text{ s.t. } \rho_k \in S_{p_k,F_k}\right\}$$
$$F'_{ij,\min} \coloneqq \min\left\{F'_{ij}(\rho_1,\rho_2) \text{ s.t. } \rho_k \in S_{p_k,F_k}\right\},$$

where F'_{ij} is given in Lemma 6.5. Then, the above quantities are independent of (i,j), or alternatively

$$F'_{ij,\max} = F'_{00,\max} \equiv F'_{\max}$$
$$F'_{ij,\min} = F'_{00,\min} \equiv F'_{\min}.$$

Proof of Proposition 6.5. Consider $\rho_k \in S_{p_k,F_k}$. We now show that for any (i,j) there is $\omega_k \in S_{p_k,F_k}$ such that

$$F'_{ij}(\rho_1,\rho_2) = F'_{00}(\omega_1,\omega_2). \tag{6.187}$$

We claim that this is the case for $\omega_1 = \rho_1$ and $\omega_2 = (Z^jX^i \otimes Z^jX^i)\rho_2(X^iZ^j \otimes X^iZ^j)$. Firstly, it is clear that $\omega_1 \in S_{p_1,F_1}$. Also,

$$\begin{aligned}\omega_2 &= (Z^jX^i \otimes Z^jX^i)\rho_2(X^iZ^j \otimes X^iZ^j)\\ &= (Z^jX^i \otimes Z^jX^i)\left(p_2|\Psi_{00}\rangle\langle\Psi_{00}| + (1-p_2)\sigma_2\right)(X^iZ^j \otimes X^iZ^j)\\ &= p_2|\Psi_{00}\rangle\langle\Psi_{00}| + (1-p_2)\eta_2 \end{aligned}\tag{6.188}$$

where

$$\eta_2 = (X^iZ^j \otimes X^iZ^j)\sigma_2(Z^jX^i \otimes Z^jX^i)$$

is the noisy component of ω_2. In (6.188), we have used the flip-flop trick, which means that the Paulis applied to both registers have no effect on the pure $|\Psi_{00}\rangle$ component. One may use the same trick to show that

$$\begin{aligned}\langle\Psi_{00}|\omega_2|\Psi_{00}\rangle &= \langle\Psi_{00}|(Z^jX^i \otimes Z^jX^i)\rho_2(X^iZ^j \otimes X^iZ^j)|\Psi_{00}\rangle\\ &= \langle\Psi_{00}|\rho_2|\Psi_{00}\rangle = F_2,\end{aligned}\tag{6.189}$$

and that therefore by (6.188) and (6.189), $\omega_2 \in S_{p_2,F_2}$. We now show (6.187). Recalling Definition 6.5, we have

$$p'_{00}(\omega_1\otimes\omega_2) = \mathrm{Tr}\left[|\Psi_{00}\rangle\langle\Psi_{00}|_{A_1A_2}\,\rho_1\otimes\omega_2\right] \tag{6.190}$$
$$= \mathrm{Tr}\left[|\Psi_{00}\rangle\langle\Psi_{00}|_{A_1A_2}\,\omega_1\otimes(Z^jX^i)_{A_2}(Z^jX^i)_{B_2}\rho_2(X^iZ^j)_{A_2}(X^iZ^j)_{B_2}\right] \tag{6.191}$$
$$= \mathrm{Tr}\left[\left|\Psi_{ij}\right\rangle\!\left\langle\Psi_{ij}\right|_{A_1A_2}\rho_1\otimes\rho_2\right] \tag{6.192}$$
$$= p'_{ij}(\rho_1\otimes\rho_2). \tag{6.193}$$

One may similarly show that

$$\mathrm{Tr}\left[|\Psi_{00}\rangle\langle\Psi_{00}|_{B_1B_2}|\Psi_{00}\rangle\langle\Psi_{00}|_{A_1A_2}\,\omega_1\otimes\omega_2\right] = \mathrm{Tr}\left[\left|\Psi_{ij}\right\rangle\!\left\langle\Psi_{ij}\right|_{B_1B_2}\left|\Psi_{ij}\right\rangle\!\left\langle\Psi_{ij}\right|_{A_1A_2}\rho_1\otimes\rho_2\right], \tag{6.194}$$

from which we see that

$$F'_{00}(\omega_1 \otimes \omega_2) = \frac{1}{p'_{00}(\omega_1 \otimes \omega_2)} \operatorname{Tr}\left[|\Psi_{00}\rangle\langle\Psi_{00}|_{B_1B_2} |\Psi_{00}\rangle\langle\Psi_{00}|_{A_1A_2} \omega_1 \otimes \omega_2\right] \tag{6.195}$$

$$= \frac{1}{p'_{ij}(\rho_1 \otimes \rho_2)} \operatorname{Tr}\left[|\Psi_{ij}\rangle\langle\Psi_{ij}|_{B_1B_2} |\Psi_{ij}\rangle\langle\Psi_{ij}|_{A_1A_2} \rho_1 \otimes \rho_2\right] \tag{6.196}$$

$$= F'_{ij}(\rho_1 \otimes \rho_2). \tag{6.197}$$

□

We now derive the formula for the end-to-end fidelity as contained in Lemma 6.5. In the main text, the result is stated for the swapping of identical states for simplicity. Here, for the completeness we state and prove the same Lemma for non-identical states.

Lemma 6.9 (Lemma 6.5, general version)**.** *Consider performing a postselected swap on a pair of two-qubit states $\rho_1 \otimes \rho_2$ such that $\rho_k \in S_{p_k,F_k}$ for $k = 1,2$. Let $F'_{ij} \equiv F'_{ij}(\rho_1 \otimes \rho_2)$ denote the end-to-end fidelity after measuring outcome ij in the BSM, and p'_{ij} the probability of measuring that outcome (Definition 6.5). Then,*

$$p'_{ij} = \frac{p_1 + p_2 - p_1 p_2}{4} + (1-p_1)(1-p_2)\tilde{p}'_{ij} \tag{6.198}$$

and

$$F'_{ij} = \frac{p_1 F_2 + p_2 F_1 - p_1 p_2 + 4(1-p_1)(1-p_2)\tilde{p}'_{ij}\tilde{F}'_{ij}}{p_1 + p_2 - p_1 p_2 + 4(1-p_1)(1-p_2)\tilde{p}'_{ij}}, \tag{6.199}$$

where $\tilde{F}'_{ij} := F'_{ij}(\sigma_1 \otimes \sigma_2)$ and $\tilde{p}'_{ij} := p'_{ij}(\sigma_1 \otimes \sigma_2)$ are the corresponding swap statistics of the noisy components.

Proof of Lemma 6.9. We start with two states of the form

$$\rho_1 = p_1 |\Psi_{00}\rangle\langle\Psi_{00}| + (1-p_1)\sigma_1$$
$$\rho_2 = p_2 |\Psi_{00}\rangle\langle\Psi_{00}| + (1-p_2)\sigma_2,$$

where $F_1 = \langle\Psi_{00}|\rho_1|\Psi_{00}\rangle$ and $F_2 = \langle\Psi_{00}|\rho_2|\Psi_{00}\rangle$. We carry out the usual swapping protocol: we start with the state $\rho_1 \otimes \rho_2$, which may be expanded as

$$\begin{aligned}\rho_1 \otimes \rho_2 = {} & p_1 p_2 \cdot |\Psi_{00}\rangle\langle\Psi_{00}| \otimes |\Psi_{00}\rangle\langle\Psi_{00}| + p_1(1-p_2) \cdot |\Psi_{00}\rangle\langle\Psi_{00}| \otimes \sigma_2 \\ & + (1-p_1)p_2 \cdot \sigma_1 \otimes |\Psi_{00}\rangle\langle\Psi_{00}| + (1-p_1)(1-p_2) \cdot \sigma_1 \otimes \sigma_2.\end{aligned} \tag{6.200}$$

Supposing that the middle station measures the BSM outcome ij, the output state is

$$\rho'_{ij} = \frac{1}{p'_{ij}} \langle\Psi_{ij}|\rho_1 \otimes \rho_2|\Psi_{ij}\rangle_{A_1A_2}, \tag{6.201}$$

where p'_{ij} is the probability of obtaining the BSM outcome ij, given by

$$p'_{ij} = \operatorname{Tr}_{B_1B_2}\left[\langle\Psi_{ij}|\rho_1 \otimes \rho_2|\Psi_{ij}\rangle_{A_1A_2}\right]. \tag{6.202}$$

We now obtain expressions for p'_{ij} and ρ'_{ij} in terms of p_i, F_i and σ_i. From (6.4), we have

$$\begin{aligned}\langle\Psi_{ij}|\rho_1\otimes\rho_2|\Psi_{ij}\rangle_{A_1A_2} &= \frac{p_1p_2}{4}|\Psi_{ij}\rangle\langle\Psi_{ij}| + \frac{p_1(1-p_2)}{4}(X^iZ^j\otimes I_2)\sigma_2(Z^jX^i\otimes I_2)\\ &+ \frac{p_2(1-p_1)}{4}(I_2\otimes X^iZ^j)\sigma_1(I_2\otimes Z^jX^i) + (1-p_1)(1-p_2)\langle\Psi_{ij}|\sigma_1\otimes\sigma_2|\Psi_{ij}\rangle_{A_1A_2},\end{aligned} \tag{6.203}$$

where the first three terms correspond to perfect teleportation (without Pauli corrections). From (6.202), we now calculate p'_{ij} by taking the trace of the above. Notice that the first three terms are proportional to valid density matrices. Therefore,

$$p'_{ij} = \frac{p_1p_2}{4} + \frac{p_1(1-p_2)}{4} + \frac{p_2(1-p_1)}{4} + (1-p_1)(1-p_2)\tilde{p}'_{ij}, \tag{6.204}$$

where $\tilde{p}'_{ij} = p'_{ij}(\sigma_1\otimes\sigma_2)$. Then,

$$\begin{aligned}F'_{ij} &= \langle\Psi_{ij}|\rho'_{ij}|\Psi_{ij}\rangle_{B_1B_2}\\ &= \frac{1}{p_{ij}}\left(\frac{p_1p_2}{4} + \frac{p_1(1-p_2)}{4}\tilde{F}_2 + \frac{p_2(1-p_1)}{4}\tilde{F}_1 + (1-p_1)(1-p_2)\tilde{F}'_{ij}\tilde{p}'_{ij}\right)\end{aligned}$$

where $\tilde{F}_k = \langle\Psi_{00}|\sigma_k|\Psi_{00}\rangle$ is the fidelity of the noisy components, and $\tilde{F}'_{ij} = F'_{ij}(\sigma_1\otimes\sigma_2)$. Recalling the fidelity constraint on our initial states ρ_k, we may rewrite

$$F_k = p_k + (1-p_k)\tilde{F}_k$$

and simplify our formula for the end-to-end fidelity to

$$F'_{ij} = \frac{1}{p'_{ij}}\left(\frac{p_1p_2}{4} + \frac{p_1(F_2-p_2)}{4} + \frac{p_2(F_1-p_1)}{4} + (1-p_1)(1-p_2)\tilde{F}'_{ij}\tilde{p}'_{ij}\right). \tag{6.205}$$

□

Theorem 6.4 (Theorem 6.3, general version)**.** *Consider performing a postselected swap on a pair of two-qubit states $\rho_1\otimes\rho_2$ such that $\rho_k\in S_{p_k,F_k}$ for $k=1,2$. Then,*

$$F'_{00}(\rho_1\otimes\rho_2) \le 1 - p_1(1-F_2) - p_2(1-F_1). \tag{6.206}$$

In particular, for the case $p_1=p_2$, $F_1=F_2$, the above bound is tight and therefore

$$F'_{\max}(p,F) = 1-2p(1-F), \tag{6.207}$$

where $F'_{\max}(p,F)$ is defined in (6.46).

Proof of Theorem 6.4. From Lemma 6.9, after measuring we obtain an outcome fidelity of

$$F'_{00} = \frac{p_1F_2 + p_2F_1 - p_1p_2 + 4(1-p_1)(1-p_2)\tilde{F}'_{00}\tilde{p}'_{00}}{p_1+p_2-p_1p_2+4(1-p_1)(1-p_2)\tilde{p}'_{00}}. \tag{6.208}$$

Now, since $\tilde{F}_{00} \leq 1$, we have

$$F'_{00} \leq \frac{p_1F_2 + p_2F_1 - p_1p_2 + 4(1-p_1)(1-p_2)\tilde{F}'_{00}\tilde{p}'_{00}}{p_1 + p_2 - p_1p_2 + 4(1-p_1)(1-p_2)\tilde{F}'_{00}\tilde{p}'_{00}}. \tag{6.209}$$

We notice that the RHS of (6.209) is of the form

$$\frac{a+x}{b+x} = 1 - \frac{b-a}{b+x}, \tag{6.210}$$

for $b-a = p_1(1-F_2) + p_2(1-F_1) \geq 0$. Then, (6.210) is a non-decreasing function of $x = \tilde{F}'_{00}\tilde{p}'_{00}$. We recall that

$$\tilde{F}'_{00}\tilde{p}'_{00} = \langle\phi|\,\sigma_1 \otimes \sigma_2\,|\phi\rangle, \tag{6.211}$$

where

$$|\phi\rangle = |\Psi_{00}\rangle_{A_1A_2} \otimes |\Psi_{00}\rangle_{B_1B_2}. \tag{6.212}$$

Since the state $|\phi\rangle$ is a product of two maximally entangled qubit states, this is a maximally entangled state between two registers of dimension $d = 4$, shared between the four-dimensional registers B_1A_1 and A_2B_2. Therefore, since $\sigma_1 \otimes \sigma_2$ is unentangled with repect to these registers, from [96], the fidelity to $|\phi\rangle$ is bounded as

$$\tilde{F}'_{00}\tilde{p}'_{00} = \langle\phi|\,\sigma_1 \otimes \sigma_2\,|\phi\rangle \leq \frac{1}{d} = \frac{1}{4}. \tag{6.213}$$

Combining the above results, it follows that

$$\begin{aligned} F'_{00} &\leq \frac{p_1F_2 + p_2F_1 - p_1p_2 + 4(1-p_1)(1-p_2)\cdot\frac{1}{4}}{p_1 + p_2 - p_1p_2 + 4(1-p_1)(1-p_2)\cdot\frac{1}{4}} \\ &= 1 - p_1(1-F_2) - p_2(1-F_1). \end{aligned} \tag{6.214}$$

For the case $p_1 = p_2 = p$, $F_1 = F_2 = F$, we now show that the above bound is tight. Letting $\rho_k \in S_{p,F}$ such that

$$\rho_1 = \rho_2 = p\,|\Psi_{00}\rangle\langle\Psi_{00}| + (1-p)\,|\psi\rangle\langle\psi|, \tag{6.215}$$

with

$$|\psi\rangle = \sqrt{\tilde{F}}\,|\Psi_{00}\rangle + \sqrt{1-\tilde{F}}\,|\Psi_{11}\rangle \tag{6.216}$$

where $\tilde{F} = (F-p)/(1-p)$, we now compute the values of $\tilde{F}'_{00}$ and $\tilde{p}'_{00}$ for $\sigma_1 = \sigma_2 = |\psi\rangle$.

We firstly expand the initial state as

$$|\psi\rangle^{\otimes 2} = \tilde{F}\,|\Psi_{00}\rangle^{\otimes 2} + (1-\tilde{F})\,|\Psi_{11}\rangle^{\otimes 2} + \sqrt{\tilde{F}(1-\tilde{F})}\,|\Psi_{00}\rangle \otimes |\Psi_{11}\rangle + \sqrt{\tilde{F}(1-\tilde{F})}\,|\Psi_{11}\rangle \otimes |\Psi_{00}\rangle. \tag{6.217}$$

We compute the action of a Bell-state measurement on each component as the following. Recalling that $|\Psi_{11}\rangle_{AB} = I \otimes XZ\,|\Psi_{00}\rangle_{AB}$ and that

$$\langle\Psi_{00}|_{A_1A_2}\Big[\,|\Psi_{00}\rangle \otimes |\Psi_{00}\rangle\,\Big] = \frac{1}{2}\,|\Psi_{00}\rangle_{B_1B_2}, \tag{6.218}$$

we have

$$\langle\Psi_{00}|_{A_1A_2}\left[|\Psi_{11}\rangle\otimes|\Psi_{11}\rangle\right] = (XZ)_{B_1}(XZ)_{B_2}\langle\Psi_{00}|_{A_1A_2}\left[|\Psi_{00}\rangle\otimes|\Psi_{00}\rangle\right] \tag{6.219}$$

$$= \frac{1}{2}(XZ)_{B_1}(XZ)_{B_2}|\Psi_{00}\rangle_{B_1B_2} \tag{6.220}$$

$$= \frac{1}{2}|\Psi_{00}\rangle_{B_1B_2}, \tag{6.221}$$

where we have used the flip-flop trick (6.121) to move the Pauli gates from one register to the other. Using the same method, it may be shown that

$$\langle\Psi_{00}|_{A_1A_2}\left[|\Psi_{00}\rangle\otimes|\Psi_{11}\rangle\right] = (XZ)_{B_2}\frac{1}{2}|\Psi_{00}\rangle_{B_1B_2}$$

$$= \frac{1}{2}|\Psi_{11}\rangle_{B_1B_2} \tag{6.222}$$

and

$$\langle\Psi_{00}|_{A_1A_2}\left[|\Psi_{11}\rangle\otimes|\Psi_{00}\rangle\right] = (XZ)_{B_1}\cdot\frac{1}{2}|\Psi_{00}\rangle_{B_1B_2} \tag{6.223}$$

$$= (ZX)_{B_2}\frac{1}{2}|\Psi_{00}\rangle_{B_1B_2} \tag{6.224}$$

$$= -\frac{1}{2}|\Psi_{11}\rangle_{B_1B_2}. \tag{6.225}$$

Then, recalling (6.217), we see that

$$\langle\Psi_{00}|_{A_1A_2}\left[|\psi\rangle^{\otimes 2}\right] = \left(\frac{\tilde{F}}{4}+\frac{1-\tilde{F}}{4}\right)|\Psi_{00}\rangle_{B_1B_2} + \sqrt{\tilde{F}(1-\tilde{F})}\left(|\Psi_{11}\rangle-|\Psi_{11}\rangle\right) \tag{6.226}$$

$$= \frac{1}{4}|\Psi_{00}\rangle_{B_1B_2}. \tag{6.227}$$

We therefore see that

$$\tilde{p}'_{00} = \frac{1}{4}, \quad \tilde{F}'_{00} = 1, \tag{6.228}$$

and substituting these values into (6.208), we see that the bound is saturated. □

Proposition 6.6 (Proposition 6.4, general version)**.** *Consider performing a postselected swap on a pair of two-qubit states $\rho_1\otimes\rho_2$ such that $\rho_k\in S_{p_k,F_k}$ for $k=1,2$. Then,*

$$F'_{00}(\rho_1\otimes\rho_2) \geq \frac{p_1F_2+p_2F_1-p_1p_2}{1+(1-p_1)(1-p_2)}. \tag{6.229}$$

Proof of Proposition 6.6. From Lemma 6.9, after measuring we obtain an outcome fidelity of

$$F'_{00} = \frac{p_1F_2+p_2F_1-p_1p_2+4(1-p_1)(1-p_2)\tilde{F}'_{00}\tilde{p}'_{00}}{p_1+p_2-p_1p_2+4(1-p_1)(1-p_2)\tilde{p}'_{00}}. \tag{6.230}$$

Since $\tilde{F}'_{00}\tilde{p}'_{00} \geq 0$, we have

$$F'_{00} \geq \frac{p_1F_2 + p_2F_1 - p_1p_2}{p_1 + p_2 - p_1p_2 + 4(1-p_1)(1-p_2)\tilde{p}'_{00}}. \tag{6.231}$$

Now, recalling that

$$\tilde{p}'_{00} = \mathrm{Tr}_{B_1B_2}\left[|\Psi_{00}\rangle\langle\Psi_{00}|_{A_1A_2}\,\sigma_1 \otimes \sigma_2\right], \tag{6.232}$$

we see that $\tilde{p}'_{00} \leq 1/2$ as this is the fidelity between a separable state and a maximally entangled state with $d = 2$ (see proof of Theorem 6.4). Therefore,

$$F'_{00} \geq \frac{p_1F_2 + p_2F_1 - p_1p_2}{p_1 + p_2 - p_1p_2 + 2(1-p_1)(1-p_2)} \tag{6.233}$$

$$= \frac{p_1F_2 + p_2F_1 - p_1p_2}{1 + (1-p_1)(1-p_2)}. \tag{6.234}$$

Recalling (6.47), for the case $p_1 = p_2 = p$ and $F_1 = F_2 = F$, the result

$$F'_{\min}(p,F) \geq \frac{2pF - p^2}{2p - p^2 + 2(1-p)^2} \tag{6.235}$$

follows directly. □

6.7.3. SDP SYMMETRISATION

Here, we perform a symmetry reduction of the optimisation problem (6.75), which we restate below for convenience:

$$\begin{aligned}
\min_{\sigma} \quad & \mathrm{Tr}\big[\,|\Psi_{00}\rangle\langle\Psi_{00}|_{B_1B_2}\,|\Psi_{00}\rangle\langle\Psi_{00}|_{A_1A_2}\,\sigma\big] \\
\text{s.t.} \quad & \mathrm{Tr}\big[\,|\Psi_{00}\rangle\langle\Psi_{00}|_{A_1A_2}\,\sigma\big] = \tilde{\delta}, \\
& \mathrm{Tr}\big[\,|\Psi_{00}\rangle\langle\Psi_{00}|_{A_1B_1}\,\sigma\big] = \tilde{F}, \\
& \mathrm{Tr}\big[\,|\Psi_{00}\rangle\langle\Psi_{00}|_{A_2B_2}\,\sigma\big] = \tilde{F}, \\
& \mathrm{Tr}[\sigma] = 1, \\
& \sigma \geq 0, \quad \sigma^{\Gamma} \geq 0.
\end{aligned} \tag{6.236}$$

In the above, the number of parameters involved in the optimisation is the number required to parameterise a quantum state over four qubits, which is of the order of $16^2 = 256$. In the following, we will reduce this number by identifying symmetries of the above optimisation problem, which will then enable us to find the solution more efficiently.

Firstly, we rewrite the above in terms of a rotated target state. In particular, we notice that since the set of PPT states is invariant under the application of local unitaries, it follows that the above is equivalent to the following. Applying the ZX operator to registers A_1 and B_2, the objective function of the above transforms to

$$\begin{aligned}
&\mathrm{Tr}\big[\,|\Psi_{00}\rangle\langle\Psi_{00}|_{B_1B_2}\,|\Psi_{00}\rangle\langle\Psi_{00}|_{A_1A_2}\,(ZX)_{B_1}(ZX)_{A_2}\sigma(XZ)_{B_1}(XZ)_{A_2}\big] \\
&\qquad = \mathrm{Tr}\big[\,|\Psi_{11}\rangle\langle\Psi_{11}|_{B_1B_2}\,|\Psi_{11}\rangle\langle\Psi_{11}|_{A_1A_2}\,\sigma\big],
\end{aligned} \tag{6.237}$$

and the other constraints transform similarly, and so (6.236) is equivalent to

$$\begin{aligned} \min_{\sigma} \quad & \mathrm{Tr}\big[\,|\Psi_{11}\rangle\langle\Psi_{11}|_{B_1B_2}\,|\Psi_{11}\rangle\langle\Psi_{11}|_{A_1A_2}\,\sigma\big] \\ \text{s.t.} \quad & \mathrm{Tr}\big[\,|\Psi_{11}\rangle\langle\Psi_{11}|_{A_1A_2}\,\sigma\big] = \tilde{\delta}, \\ & \mathrm{Tr}\big[\,|\Psi_{11}\rangle\langle\Psi_{11}|_{A_1B_1}\,\sigma\big] = \tilde{F}, \\ & \mathrm{Tr}\big[\,|\Psi_{11}\rangle\langle\Psi_{11}|_{A_2B_2}\,\sigma\big] = \tilde{F}, \\ & \mathrm{Tr}[\sigma] = 1, \\ & \sigma \geq 0, \;\; \sigma^{\Gamma} \geq 0, \end{aligned} \tag{6.238}$$

The reason for studying (6.238) instead of (6.236) is due to the symmetry properties of the $|\Psi_{11}\rangle$ state. In particular, for any one-qubit unitary U, the state $|\Psi_{11}\rangle$ satisfies

$$(U\otimes U)\,|\Psi_{11}\rangle\langle\Psi_{11}|\,(U\otimes U)^{\dagger} = |\Psi_{11}\rangle\langle\Psi_{11}|\,. \tag{6.239}$$

Therefore, under the transformation

$$\sigma \mapsto (U^{\otimes 4})\sigma(U^{\otimes 4})^{\dagger}, \tag{6.240}$$

it can be seen that the objective function and constraints of (6.238) are invariant. If σ_{opt} is an optimal solution of (6.238), then

$$\overline{\sigma}_{\mathrm{opt}} = \int (U^{\otimes 4})\sigma_{\mathrm{opt}}(U^{\otimes 4})^{\dagger}\mathrm{d}U \tag{6.241}$$

is also optimal, where the integration is over the Haar measure. The state $\overline{\sigma}_{\mathrm{opt}}$ is invariant under the map (6.240). In order to solve (6.238) it therefore suffices to optimise over the set of operators that are invariant under the symmetry (6.240). These states are given by

$$\sum_{\tau\in S_4} r_\tau M_\tau : r_\tau \in \mathbb{C}, \tag{6.242}$$

where S_4 is the symmetric group. In the above,

$$M_\tau = \sum_{\vec{k}\in\{0,1\}^4} \left|\tau(\vec{k})\right\rangle\left\langle\vec{k}\right| \tag{6.243}$$

where $\tau(\vec{k}) = (k_{\tau^{-1}(1)}, k_{\tau^{-1}(2)}, k_{\tau^{-1}(3)}, k_{\tau^{-1}(4)})$. Then, M_τ is the operator that permutes the four registers according to the permutation τ. The unitary operators M_τ form a representation of the symmetric group S_4, and in particular satisfy

$$M_{\tau_1}M_{\tau_2} = M_{\tau_1\tau_2}. \tag{6.244}$$

We therefore see from (6.242) that in order to parameterise a state that is invariant under the map (6.240), one requires a maximum of $2\cdot|S_4| = 48$ parameters. Now, we make use of the identity

$$|\Psi_{11}\rangle\langle\Psi_{11}| = \frac{1}{2}\,(I_2 - \mathrm{SWAP})\,, \tag{6.245}$$

where SWAP is the operation that swaps the two qubits, and we will rewrite the constraints and objective function of (6.238). In the following, we will denote a permutation by its decomposition into cycles [214]. For example, the permutation that swaps registers A_1 and A_2 is denoted by $(A_1 A_2)$. Then, given the symmetrised form (6.242), the first constraint of (6.238) becomes

$$\tilde{\delta} = \mathrm{Tr}\left[|\Psi_{11}\rangle\langle\Psi_{11}|_{A_1A_2}\,\sigma\right] \tag{6.246}$$

$$= \mathrm{Tr}\left[\frac{1}{2}\left(M_e - M_{(A_1A_2)}\right)\sum_{\tau\in S_4} r_\tau M_\tau\right] \tag{6.247}$$

$$= \frac{1}{2}\mathrm{Tr}\left[\sum_{\tau\in S_4} r_\tau\left(M_\tau - M_{(A_1A_2)\tau}\right)\right] \tag{6.248}$$

$$= \frac{1}{2}\sum_{\tau\in S_4} r_\tau\left(\mathrm{Tr}\left[M_\tau\right] - \mathrm{Tr}\left[M_{(A_1A_2)\tau}\right]\right) \tag{6.249}$$

$$= v^T r, \tag{6.250}$$

where v is a vector indexed by elements of S_4, with $v_\tau := \left(\mathrm{Tr}\left[M_\tau\right] - \mathrm{Tr}\left[M_{(A_1A_2)\tau}\right]\right)/2$. This is a linear constraint on the vector r. We may perform the same procedure for all constraints and the objective function in (6.238), transforming this into

$$\begin{aligned}
\min_{r\in\mathbb{C}^{|S_4|}} \quad & u^T r \\
\text{s.t.} \quad & v^T r = \tilde{\delta}, \\
& w_1^T r = \tilde{F}, \quad w_2^T r = \tilde{F}, \\
& x^T r = 1, \\
& \sum_{\tau\in S_4} r_\tau M_\tau = \Big(\sum_{\tau\in S_4} r_\tau M_\tau\Big)^\dagger, \\
& \sum_{\tau\in S_4} r_\tau M_\tau \geq 0, \quad \sum_{\tau\in S_4} r_\tau (M_\tau)^\Gamma \geq 0.
\end{aligned} \tag{6.251}$$

In the above, u, v w and x are all vectors indexed by elements of S_4, with

$$u_\tau := \frac{1}{4}\left(\mathrm{Tr}\left[M_\tau\right] - \mathrm{Tr}\left[M_{(A_1A_2)\tau}\right] - \mathrm{Tr}\left[M_{(A_1A_2)\tau}\right] + \mathrm{Tr}\left[M_{(A_1A_2)(B_1B_2)\tau}\right]\right) \tag{6.252}$$

$$v_\tau := \frac{1}{2}\left(\mathrm{Tr}\left[M_\tau\right] - \mathrm{Tr}\left[M_{(A_1A_2)\tau}\right]\right) \tag{6.253}$$

$$(w_k)_\tau := \frac{1}{2}\left(\mathrm{Tr}\left[M_\tau\right] - \mathrm{Tr}\left[M_{(A_kB_k)\tau}\right]\right) \tag{6.254}$$

$$x_\tau := \mathrm{Tr}\left[M_\tau\right]. \tag{6.255}$$

The values $\mathrm{Tr}\left[M_\tau\right]$, and therefore the vectors u, v, w and x, may be computed using the following proposition.

Proposition 6.7. *Suppose that $\tau, \tau' \in S_4$ are conjugate, i.e. there is a ν such that $\tau' = \nu^{-1}\tau\nu$. Then,*

$$\mathrm{Tr}(M_{\tau'}) = \mathrm{Tr}(M_\tau).$$

In particular, $\mathrm{Tr}[M_\tau]$ *is determined by the conjugacy class (cycle type) of* τ*, of which there are the following five possibilities:*

$$\mathrm{Tr}[M_\tau] = \begin{cases} 16 \text{ if cycle type 1+1+1+1} \\ 8 \text{ if 2+1+1} \\ 4 \text{ if 2+2} \\ 4 \text{ if 3+1} \\ 2 \text{ if 4.} \end{cases} \tag{6.256}$$

Proof. Since $\tau' = \nu^{-1}\tau\nu$, we have $M_{\tau'} = M_{\nu^{-1}\tau\nu} = M_{\nu^{-1}} M_\tau M_\nu = M_\nu^{-1} M_\tau M_\nu$. Then,

$$\mathrm{Tr}[M_{\tau'}] = \mathrm{Tr}\left[M_\nu^{-1} M_\tau M_\nu\right] = \mathrm{Tr}[M_\tau].$$

In particular, $\mathrm{Tr}[M_{\tau'}] = \mathrm{Tr}[M_\tau]$ for all $\tau' \in \mathrm{Cl}(\tau)$, where

$$\mathrm{Cl}(\tau) = \{\nu^{-1}\tau\nu : \nu \in \mathrm{Sym}(4)\}$$

is the conjugacy class of τ'. Now, for the symmetric group, the conjugacy class is determined by the cycle type [214]. The group $\mathrm{Sym}(4)$ has five cycle types. One may then compute the values (6.256) by computing $\mathrm{Tr}[M_\tau]$ for a given example τ of each cycle type. □

6

6.7.4. FURTHER ANALYSIS OF SDP LOWER BOUND

In this appendix, we further investigate the behaviour of the SDP lower bound for $F'_{\min}(p,F)$, which was presented in Section 6.4.3. We firstly note that one may formulate an upper bound for $F'_{\max}(p,F)$ with the same method (i.e. performing a maximisation of the objective functions of (6.75) and (6.77) instead of a minimisaton). Although this is not necessary, because in Theorem 6.3 we have an explicit solution for $F'_{\max}(p,F)$, we computed this solution in order to better understand the range of the end-to-end fidelity after the PPT relaxation. Interestingly, the result of this was always $F'_{\max}(p,F)$: in all cases tested, the corresponding SDP upper bound was tight. More specifically, it has a simple linear form in terms of p and F, and the corresponding probability of measuring the outcome (δ^*) takes a constant value of $1/4$. This matches the example of the optimal state given in Theorem 6.3.

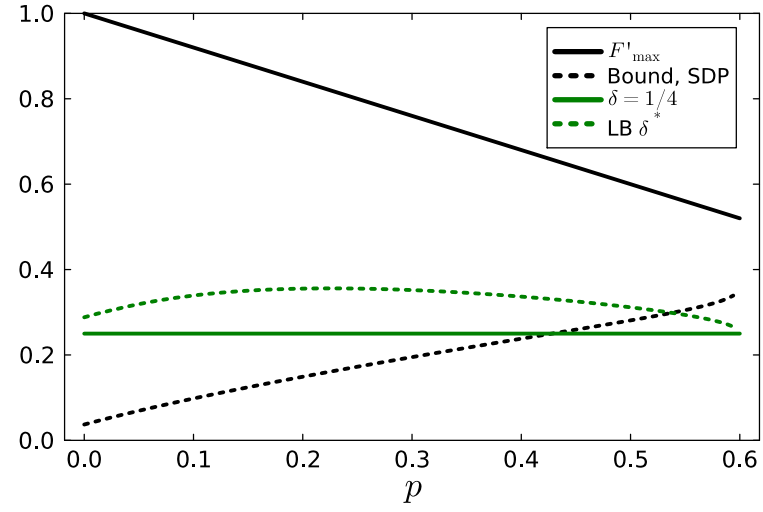
In the following, we will see that the SDP lower bound does not have these characteristics.

OPTIMAL VALUE OF δ

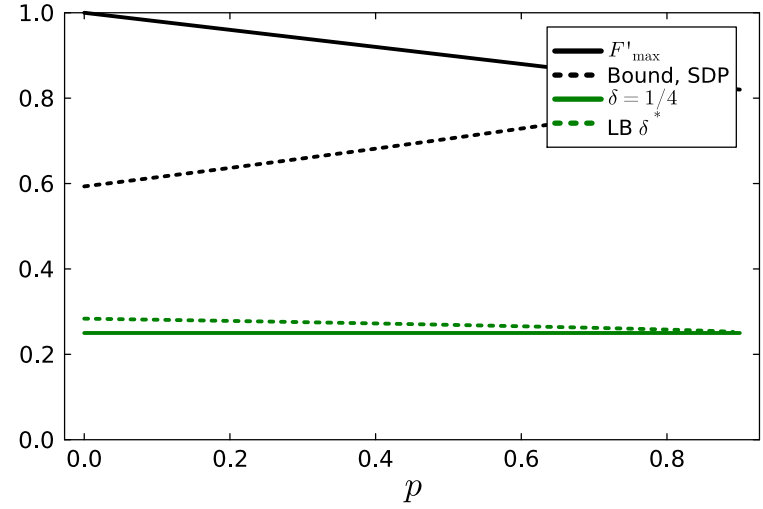
Despite the fact that the SDP upper bound is tight and has a simple analytical form, in the case of no permutation symmetry we did not observe the same for the SDP lower bound. In order to further understand its behaviour, one may analyse the value of the postselected swap probability δ^* that minimises the expression of the lower bound (6.77), i.e.

$$F'_{\min}(p,F) \geq \min_\delta \frac{1}{\delta}\left(\frac{Fp}{2} - \frac{p^2}{4} + (1-p)^2 H^*_{\mathrm{rel}}(p,F,\delta)\right) \tag{6.257}$$

$$= \frac{1}{\delta^*}\left(\frac{Fp}{2} - \frac{p^2}{4} + (1-p)^2 H^*_{\mathrm{rel}}(p,F,\delta^*)\right). \tag{6.258}$$



(a)



(b)

Figure 6.6: **Probability of the optimal postselected swap outcome plotted against** p, for (a) $F = 0.6$ and (b) $F = 0.9$. Each is plotted for 100 values of p, uniformly spaced in the interval $[0, F]$. The black solid line is $F'_{\max}(p, F)$, and the green solid line is the postselected swap probability for the states saturating this value ($\delta = 1/4$). The black dotted line is the lower bound found with SDP, and the green dotted line is the postselected swap probability of the optimal state for the lower bound.

Recalling that the states saturating the upper bound have constant postselected swap probability of 1/4, we see from Figure 6.6 that δ^* usually lies above this value, and is not constant. We were not able to find a good functional fit in terms of p and F for δ^*.

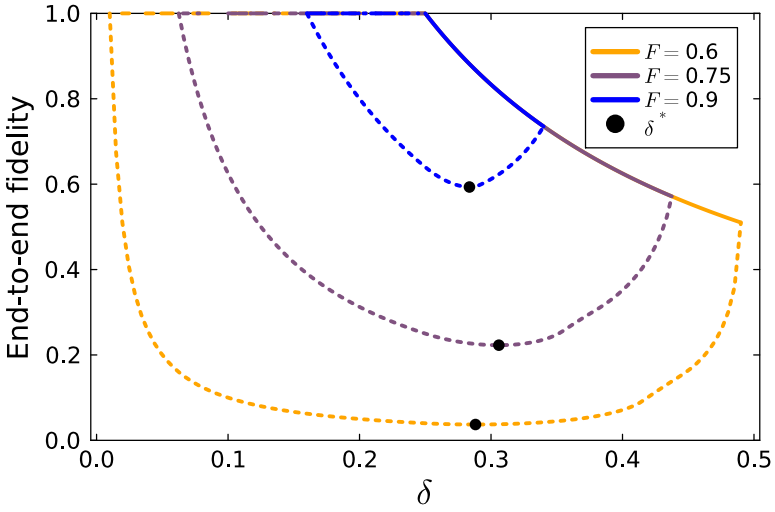


Figure 6.7: **Bounds on the end-to-end fidelity, given a postselected swap probability δ.** For three different values of F and $p = 0$, each is plotted for 100 values of δ, uniformly spaced in the feasible region $[\delta_{\min}, \delta_{\max}]$ for each value of F. The dotted lines are the lower bounds, bound by solving (6.259). The solid lines are the upper bounds, found by solving the corresponding maximisation problem of (6.259). Also shown are the values δ^* that form the SDP lower bound for $F'_{\min}(p, F)$ from (6.258).

Similarly, we were not able to find a functional fit for the SDP lower bound: although for large values of F this appears to be linear in some range of p (see Figures 6.3a and 6.3b in the main text), we see from Figure 6.6 that for $F = 0.6$ this is not the case. We leave further analysis of the behaviour of δ^* to future work.

DEPENDENCE ON δ

It was mentioned in the main text that fixing the parameter δ while optimising the end-to-end fidelity can aid to further understand the trade-off between rate and fidelity inherent to the entanglement swapping process. In particular, solving the problem (6.75) provides an answer to the question 'given that the probability of the postselected swap is δ, how small (large) can my fidelity become after swapping'? In order to answer how small it can become, one may solve the following semi-definite program

$$\begin{aligned} \min_{\sigma} \quad & \mathrm{Tr}\big[\,|\Psi_{00}\rangle\langle\Psi_{00}|_{B_1B_2}\,|\Psi_{00}\rangle\langle\Psi_{00}|_{A_1A_2}\,\sigma\big] \\ \text{s.t.} \quad & \mathrm{Tr}\big[\,|\Psi_{00}\rangle\langle\Psi_{00}|_{A_1A_2}\,\sigma\big] = \tilde{\delta}(p,\delta), \\ & \mathrm{Tr}\big[\,|\Psi_{00}\rangle\langle\Psi_{00}|_{B_1A_1}\,\sigma\big] = \tilde{F}(p,F), \\ & \mathrm{Tr}\big[\,|\Psi_{00}\rangle\langle\Psi_{00}|_{B_2A_2}\,\sigma\big] = \tilde{F}(p,F), \\ & \mathrm{Tr}[\sigma] = 1, \\ & \sigma \geq 0, \quad \sigma^{\Gamma} \geq 0. \end{aligned} \tag{6.259}$$

This was also given in (6.75) in the main text. To find bounds on how large the fidelity can become, one may perform instead a maximisation of (6.259), or simply replace the ob-

jective function by a minus sign. These bounds are shown in Figure 6.7. These are plotted for $p = 0$ and three different values of F, in the feasible range of δ. Interestingly, the upper bounds for each of the three fidelity values always coincide in their corresponding feasible region. This extends the observation from the beginning of Section 6.4, where we saw that no matter the value of the initial fidelity, it is always possible to obtain a unit end-to-end fidelity (from Figure 6.7, we observe this for values of δ below $1/4$, where all upper bounds are equal to one). Recalling the state $|\psi\rangle = \sqrt{F}|\Psi_{00}\rangle + \sqrt{1-F}|\Psi_{11}\rangle$ that swaps to unit fidelity with probability $1/4$, we therefore conclude that at the point $\delta = 1/4$ the upper bound is tight. Then, $|\psi\rangle$ is optimal in the sense that it has the maximum probability of swapping to perfect fidelity. Indeed, beyond $\delta = 1/4$, we see there is necessarily a decrease in the end-to-end fidelity if we demand that δ is larger than this value. It can also be seen from the figure that at the extremal values of the feasible region of δ, the upper and lower bounds meet. In particular, when δ becomes close to $\delta_{\min}$, the lower bound goes to one. We conclude that if the postselected swap probability is made as small as possible, the end-to-end fidelity will necessarily increase to one. Despite this, the lower bound is not monotonic in δ: for large values of δ, we see from the figure that it increases again before joining up with the upper bound at $\delta_{\max}$. From this behaviour we may conclude that, for the values of F tested, the numerical optimiser over δ that is employed in (6.257) to find the SDP lower bound is indeed finding the global minimum. This is highlighted by the black circles in Figure 6.7, which are the values δ^* of the postselected swap probability that minimise the lower bound. The value of the objective function at each point is the SDP lower bound for $F_{\min}(0, F)$.

6.7.5. INVARIANCE OF SECRET-KEY FRACTION

As discussed in Section 6.5.2, the secret-key fraction depends on the QBER according to (6.99). The QBER is defined as the probability that, when both nodes measure their state in the X (Z) basis, they obtain different outcomes. Letting σ be the entangled state shared between the two nodes, the probability of obtaining an error when measuring in the Z basis is given by

$$Q_Z = \langle 01|\sigma|01\rangle + \langle 10|\sigma|10\rangle \tag{6.260}$$

Noting that

$$|01\rangle\langle 01| + |10\rangle\langle 10| = \frac{1}{2}(I_4 - Z \otimes Z)\sigma,$$

we have

$$\begin{aligned} Q_Z(\sigma) &= \mathrm{Tr}\left[(|01\rangle\langle 01| + |10\rangle\langle 10|)\sigma\right] && (6.261)\\ &= \frac{1}{2}\mathrm{Tr}\left[(I_4 - Z \otimes Z)\sigma\right] && (6.262)\\ &= \frac{1}{2}(1 - \mathrm{Tr}\left[(Z \otimes Z)\sigma\right]). && (6.263) \end{aligned}$$

Now, since we have $(Z \otimes Z)|\Psi_{ij}\rangle = \pm|\Psi_{ij}\rangle$ for all Bell states $|\Psi_{ij}\rangle$, we see that $\mathrm{Tr}\left[(Z \otimes Z)\sigma\right]$ only depends on the Bell-diagonal elements of σ, and therefore $Q_Z(\sigma) = Q_Z(\mathscr{B}(\sigma))$. The

same holds for measuring in the X-basis:

$$Q_X(\sigma) = \frac{1}{2}(1 - \mathrm{Tr}\left[(X \otimes X)\sigma\right]) \tag{6.264}$$

$$= \frac{1}{2}(1 - \mathrm{Tr}\left[(X \otimes X)\mathscr{B}(\sigma)\right]) \tag{6.265}$$

$$= Q_X(\mathscr{B}(\sigma)). \tag{6.266}$$

By (6.99), we therefore have $\mathrm{SKF}(\sigma) = \mathrm{SKF}(\mathscr{B}(\sigma))$.

BIBLIOGRAPHY

[1] H Jeff Kimble. "The quantum internet". In: *Nature* 453.7198 (2008), pp. 1023–1030.

[2] Benjamin Schumacher. "Sending entanglement through noisy quantum channels". In: *Physical Review A* 54.4 (1996), p. 2614.

[3] Artur K Ekert. "Quantum cryptography based on Bell's theorem". In: *Phys. Rev. Lett.* 67.6 (1991), p. 661.

[4] Valerio Scarani et al. "The security of practical quantum key distribution". In: *Reviews of modern physics* 81.3 (2009), pp. 1301–1350.

[5] Anne Broadbent, Joseph Fitzsimons, and Elham Kashefi. "Universal blind quantum computation". In: *2009 50th annual IEEE symposium on foundations of computer science*. IEEE. 2009, pp. 517–526.

[6] Joseph F Fitzsimons. "Private quantum computation: an introduction to blind quantum computing and related protocols". In: *npj Quantum Information* 3.1 (2017), p. 23.

[7] Richard Jozsa et al. "Quantum clock synchronization based on shared prior entanglement". In: *Physical Review Letters* 85.9 (2000), p. 2010.

[8] Peter Komar et al. "A quantum network of clocks". In: *Nature Physics* 10.8 (2014), pp. 582–587.

[9] Daniel Gottesman, Thomas Jennewein, and Sarah Croke. "Longer-baseline telescopes using quantum repeaters". In: *Physical review letters* 109.7 (2012), p. 070503.

[10] Koji Azuma et al. "Quantum repeaters: From quantum networks to the quantum internet". In: *Reviews of Modern Physics* 95.4 (2023), p. 045006.

[11] Michael A Nielsen and Isaac Chuang. *Quantum computation and quantum information*. 2002.

[12] Stephanie Wehner, David Elkouss, and Ronald Hanson. "Quantum internet: A vision for the road ahead". In: *Science* 362.6412 (2018), eaam9288.

[13] Sreraman Muralidharan et al. "Optimal architectures for long distance quantum communication". In: *Scientific reports* 6.1 (2016), p. 20463.

[14] Matteo Pompili et al. "Realization of a multinode quantum network of remote solid-state qubits". In: *Science* 372.6539 (2021), pp. 259–264.

[15] Arian J Stolk et al. "Metropolitan-scale heralded entanglement of solid-state qubits". In: *Science advances* 10.44 (2024), eadp6442.

[16] Eric Bersin et al. "Development of a Boston-area 50-km fiber quantum network testbed". In: *Physical Review Applied* 21.1 (2024), p. 014024.

[17] Ji-Gang Ren et al. “Ground-to-satellite quantum teleportation”. In: *Nature* 549.7670 (2017), pp. 70–73.

[18] Peter Drmota et al. “Verifiable blind quantum computing with trapped ions and single photons”. In: *Physical Review Letters* 132.15 (2024), p. 150604.

[19] Koji Azuma et al. “Tools for quantum network design”. In: *AVS Quantum Science* 3.1 (2021).

[20] Robert Raussendorf and Hans J Briegel. “A one-way quantum computer”. In: *Physical review letters* 86.22 (2001), p. 5188.

[21] Johannes Borregaard et al. “One-way quantum repeater based on near-deterministic photon-emitter interfaces”. In: *Physical Review X* 10.2 (2020), p. 021071.

[22] Paul Hilaire et al. “Error-correcting entanglement swapping using a practical logical photon encoding”. In: *Physical Review A* 104.5 (2021), p. 052623.

[23] Carlos Cabrillo et al. “Creation of entangled states of distant atoms by interference”. In: *Physical Review A* 59.2 (1999), p. 1025.

[24] L-M Duan et al. “Long-distance quantum communication with atomic ensembles and linear optics”. In: *Nature* 414.6862 (2001), pp. 413–418.

[25] Sean D Barrett and Pieter Kok. “Efficient high-fidelity quantum computation using matter qubits and linear optics”. In: *Physical Review A* 71.6 (2005), p. 060310.

[26] Nicolas Sangouard et al. “Long-distance entanglement distribution with single-photon sources”. In: *Physical Review A—Atomic, Molecular, and Optical Physics* 76.5 (2007), p. 050301.

[27] Hans KC Beukers et al. “Remote-entanglement protocols for stationary qubits with photonic interfaces”. In: *PRX Quantum* 5.1 (2024), p. 010202.

[28] Thomas R Beauchamp et al. “A Modular Quantum Network Architecture for Integrating Network Scheduling with Local Program Execution”. In: *arXiv preprint arXiv:2503.12582* (2025).

[29] Matteo Pompili et al. “Experimental demonstration of entanglement delivery using a quantum network stack”. In: *npj Quantum Information* 8.1 (2022), p. 121.

[30] H-J Briegel et al. “Quantum repeaters: the role of imperfect local operations in quantum communication”. In: *Phys. Rev. Lett.* 81.26 (1998), p. 5932.

[31] Severin Daiss et al. “A quantum-logic gate between distant quantum-network modules”. In: *Science* 371.6529 (2021), pp. 614–617.

[32] Marek Zukowski et al. “" Event-ready-detectors" Bell experiment via entanglement swapping.” In: *Physical review letters* 71.26 (1993).

[33] Norbert Lütkenhaus, John Calsamiglia, and K-A Suominen. “Bell measurements for teleportation”. In: *Physical Review A* 59.5 (1999), p. 3295.

[34] John Calsamiglia and Norbert Lütkenhaus. “Maximum efficiency of a linear-optical Bell-state analyzer”. In: *Applied Physics B* 72 (2001), pp. 67–71.

[35] Filip Rozpędek et al. “Parameter regimes for a single sequential quantum repeater”. In: *Quantum Science and Technology* 3.3 (2018), p. 034002.

[36] R Zhao et al. “Long-lived quantum memory”. In: *Nature Physics* 5.2 (2009), pp. 100–104.

[37] Mihir K Bhaskar et al. “Experimental demonstration of memory-enhanced quantum communication”. In: *Nature* 580.7801 (2020), pp. 60–64.

[38] Ye Wang et al. “Single-qubit quantum memory exceeding ten-minute coherence time”. In: *Nature Photonics* 11.10 (2017), pp. 646–650.

[39] Denis D Sukachev et al. “Silicon-vacancy spin qubit in diamond: a quantum memory exceeding 10 ms with single-shot state readout”. In: *Physical review letters* 119.22 (2017), p. 223602.

[40] Peter C Humphreys et al. “Deterministic delivery of remote entanglement on a quantum network”. In: *Nature* 558.7709 (2018), pp. 268–273.

[41] 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)*. IEEE. 2020, pp. 158–168.

[42] Sumeet Khatri. “Policies for elementary links in a quantum network”. In: *Quantum* 5 (2021), p. 537.

[43] Ludmiła Praxmeyer. “Reposition time in probabilistic imperfect memories”. In: *arXiv preprint arXiv:1309.3407* (2013).

[44] Evgeny Shchukin, Ferdinand Schmidt, and Peter van Loock. “Waiting time in quantum repeaters with probabilistic entanglement swapping”. In: *Physical Review A* 100.3 (2019), p. 032322.

[45] Álvaro G Iñesta et al. “Optimal entanglement distribution policies in homogeneous repeater chains with cutoffs”. In: *npj Quantum Information* 9.1 (2023), p. 46.

[46] Wolfgang Dür et al. “Standard forms of noisy quantum operations via depolarization”. In: *Physical Review A—Atomic, Molecular, and Optical Physics* 72.5 (2005), p. 052326.

[47] Charles H Bennett et al. “Mixed-state entanglement and quantum error correction”. In: *Physical Review A* 54.5 (1996), p. 3824.

[48] Michał Horodecki and Paweł Horodecki. “Reduction criterion of separability and limits for a class of distillation protocols”. In: *Physical Review A* 59.6 (1999), p. 4206.

[49] Charles H Bennett et al. “Purification of noisy entanglement and faithful teleportation via noisy channels”. In: *Physical review letters* 76.5 (1996), p. 722.

[50] David Deutsch et al. “Quantum privacy amplification and the security of quantum cryptography over noisy channels”. In: *Physical review letters* 77.13 (1996), p. 2818.

[51] Naomi H Nickerson, Joseph F Fitzsimons, and Simon C Benjamin. “Freely scalable quantum technologies using cells of 5-to-50 qubits with very lossy and noisy photonic links”. In: *Physical Review X* 4.4 (2014), p. 041041.

[52] Stefan Krastanov, Victor V Albert, and Liang Jiang. “Optimized entanglement purification”. In: *Quantum* 3 (2019), p. 123.

[53] Filip Rozpędek et al. “Optimizing practical entanglement distillation”. In: *Physical Review A* 97.6 (2018), p. 062333.

[54] Liang Jiang et al. “Quantum repeater with encoding”. In: *Physical Review A—Atomic, Molecular, and Optical Physics* 79.3 (2009), p. 032325.

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

[56] Bethany Davies, Álvaro G Iñesta, and Stephanie Wehner. “Entanglement buffering with two quantum memories”. In: *Quantum* 8 (2024), p. 1458.

[57] Álvaro G Iñesta et al. “Entanglement buffering with multiple quantum memories”. In: *arXiv preprint arXiv:2502.20240* (2025).

[58] Bethany Davies et al. “Tools for the analysis of quantum protocols requiring state generation within a time window”. In: *IEEE Transactions on Quantum Engineering* (2024).

[59] Álvaro G Iñesta et al. “Quantum circuit switching with one-way repeaters in star networks”. In: *2024 IEEE International Conference on Quantum Computing and Engineering (QCE)*. Vol. 1. IEEE. 2024, pp. 1857–1867.

[60] Ian Tillman et al. “Calculating the capacity region of a quantum switch”. In: *2024 IEEE International Conference on Quantum Computing and Engineering (QCE)*. Vol. 1. IEEE. 2024, pp. 1868–1878.

[61] Gayane Vardoyan et al. “On the stochastic analysis of a quantum entanglement switch”. In: *ACM SIGMETRICS Performance Evaluation Review* 47.2 (2019), pp. 27–29.

[62] Stav Haldar et al. “Fast and reliable entanglement distribution with quantum repeaters: principles for improving protocols using reinforcement learning”. In: *Physical Review Applied* 21.2 (2024), p. 024041.

[63] Liang Jiang et al. “Optimal approach to quantum communication using dynamic programming”. In: *Proceedings of the National Academy of Sciences* 104.44 (2007), pp. 17291–17296.

[64] Kenneth Goodenough, Tim Coopmans, and Don Towsley. “On noise in swap ASAP repeater chains: exact analytics, distributions and tight approximations”. In: *arXiv preprint arXiv:2404.07146* (2024).

[65] Karim Elsayed, Wasiur R KhudaBukhsh, and Amr Rizk. “On the Fidelity Distribution of Link-level Entanglements under Purification”. In: *arXiv preprint arXiv:2310.18198* (2023).

[66] Tim Coopmans et al. “Netsquid, a network simulator for quantum information using discrete events”. In: *Communications Physics* 4.1 (2021), p. 164.

[67] Xiaoliang Wu et al. “SeQUeNCe: a customizable discrete-event simulator of quantum networks”. In: *Quantum Science and Technology* 6.4 (2021), p. 045027.

[68] Julius Wallnöfer et al. “Faithfully simulating near-term quantum repeaters”. In: *PRX Quantum* 5.1 (2024), p. 010351.

[69] Ryosuke Satoh et al. “Quisp: a quantum internet simulation package”. In: *2022 IEEE international conference on quantum computing and engineering (QCE)*. IEEE. 2022, pp. 353–364.

[70] William K Wootters. “Entanglement of formation of an arbitrary state of two qubits”. In: *Physical Review Letters* 80.10 (1998), p. 2245.

[71] Guifré Vidal and Reinhard F Werner. “Computable measure of entanglement”. In: *Physical Review A* 65.3 (2002), p. 032314.

[72] Daniel Gottesman et al. “Security of quantum key distribution with imperfect devices”. In: *International Symposium onInformation Theory, 2004. ISIT 2004. Proceedings.* IEEE. 2004, p. 136.

[73] Francisco Ferreira da Silva et al. “Requirements for upgrading trusted nodes to a repeater chain over 900 km of optical fiber”. In: *Quantum Science and Technology* 9.4 (2024), p. 045041.

[74] Samuel Oslovich, Bart van der Vecht, and Stephanie Wehner. “Compilation strategies for quantum network programs using Qoala”. In: *arXiv preprint arXiv:2505.06162* (2025).

[75] Gayane Vardoyan and Stephanie Wehner. “Quantum network utility maximization”. In: *2023 IEEE International Conference on Quantum Computing and Engineering (QCE)*. Vol. 1. IEEE. 2023, pp. 1238–1248.

[76] Antonio Acín, J Ignacio Cirac, and Maciej Lewenstein. “Entanglement percolation in quantum networks”. In: *Nature Physics* 3.4 (2007), pp. 256–259.

[77] Ashlesha Patil et al. “Entanglement generation in a quantum network at distance-independent rate”. In: *npj Quantum Information* 8.1 (2022), p. 51.

[78] Álvaro G Iñesta and Stephanie Wehner. “Performance metrics for the continuous distribution of entanglement in multiuser quantum networks”. In: *Physical Review A* 108.5 (2023), p. 052615.

[79] Silvestre Abruzzo et al. “Quantum repeaters and quantum key distribution: Analysis of secret-key rates”. In: *Physical Review A—Atomic, Molecular, and Optical Physics* 87.5 (2013), p. 052315.

[80] Wolfgang Dür et al. “Quantum repeaters based on entanglement purification”. In: *Physical Review A* 59.1 (1999), p. 169.

[81] Pei-Shun Yan et al. “Advances in quantum entanglement purification”. In: *Science China Physics, Mechanics & Astronomy* 66.5 (2023), p. 250301.

[82] Norbert Kalb et al. “Entanglement distillation between solid-state quantum network nodes”. In: *Science* 356.6341 (2017), pp. 928–932.

[83] Haoxiong Yan et al. “Entanglement purification and protection in a superconducting quantum network”. In: *Physical Review Letters* 128.8 (2022), p. 080504.

[84] Jeroen Dehaene et al. “Local permutations of products of Bell states and entanglement distillation”. In: *Physical Review A* 67.2 (2003), p. 022310.

[85] Jian-Long Liu et al. “A multinode quantum network over a metropolitan area”. In: *arXiv preprint arXiv:2309.00221* (2023).

[86] Gayane Vardoyan et al. “On the capacity region of bipartite and tripartite entanglement switching”. In: *ACM Transactions on Modeling and Performance Evaluation of Computing Systems* 8.1-2 (2023), pp. 1–18.

[87] Philippe Nain et al. “On the analysis of a multipartite entanglement distribution switch”. In: *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 4.2 (2020), pp. 1–39.

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

[89] Michelle Victora et al. “Entanglement purification on quantum networks”. In: *Physical Review Research* 5.3 (2023), p. 033171.

[90] Sylvia Bratzik et al. “Quantum repeaters and quantum key distribution: The impact of entanglement distillation on the secret key rate”. In: *Physical Review A* 87.6 (2013), p. 062335.

[91] 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*. IEEE. 2023, pp. 1–10.

[92] Simon C Benjamin et al. “Brokered graph-state quantum computation”. In: *New Journal of Physics* 8.8 (2006), p. 141.

[93] Yuan Lee et al. “A quantum router architecture for high-fidelity entanglement flows in quantum networks”. In: *npj Quantum Information* 8.1 (2022), p. 75.

[94] Hannes Bernien et al. “Heralded entanglement between solid-state qubits separated by three metres”. In: *Nature* 497.7447 (2013), pp. 86–90.

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

[96] Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. “General teleportation channel, singlet fraction, and quasidistillation”. In: *Physical Review A* 60.3 (1999), p. 1888.

[97] Sarah Jansen et al. “Enumerating all bilocal Clifford distillation protocols through symmetry reduction”. In: *Quantum* 6 (2022), p. 715.

[98] Kenneth Goodenough et al. “Near-term n to k distillation protocols using graph codes”. In: *arXiv preprint arXiv:2303.11465* (2023).

[99] Kosto V Mitov et al. *Renewal processes*. Springer, 2014.

[100] Daniel Gottesman. “Theory of fault-tolerant quantum computation”. In: *Physical Review A* 57.1 (1998), p. 127.

[101] Daniel Gottesman. “The Heisenberg representation of quantum computers”. In: *arXiv preprint quant-ph/9807006* (1998).

[102] Asher Peres. “Separability criterion for density matrices”. In: *Physical Review Letters* 77.8 (1996), p. 1413.

[103] Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. "Separability of mixed states: necessary and sufficient conditions". In: *Physics Letters A* 223.1 (1996), pp. 1–8. ISSN: 0375-9601. DOI: https://doi.org/10.1016/S0375-9601(96)00706-2. URL: https://www.sciencedirect.com/science/article/pii/S0375960196007062.

[104] Charles H. Bennett, Gilles Brassard, and N. David Mermin. "Quantum Cryptography without Bell's Theorem". In: *Physical Review Letters* 68.5 (Feb. 1992), pp. 557–559. DOI: 10.1103/PhysRevLett.68.557. URL: https://link.aps.org/doi/10.1103/PhysRevLett.68.557 (visited on 10/23/2020).

[105] Seth Lloyd. "Enhanced sensitivity of photodetection via quantum illumination". In: *Science* 321.5895 (2008), pp. 1463–1465.

[106] Kevin Qian et al. "Heisenberg-scaling measurement protocol for analytic functions with quantum sensor networks". In: *Phys. Rev. A* 100.4 (2019), p. 042304.

[107] Duncan G England, Bhashyam Balaji, and Benjamin J Sussman. "Quantum-enhanced standoff detection using correlated photon pairs". In: *Phys. Rev. A* 99.2 (2019), p. 023828.

[108] Bo-Han Wu, Saikat Guha, and Quntao Zhuang. "Entanglement-assisted multi-aperture pulse-compression radar for angle resolving detection". In: *Quantum Sci. Tech.* 8.3 (2023), p. 035016.

[109] Gilles Brassard, Anne Broadbent, and Alain Tapp. "Quantum pseudo-telepathy". In: *Found. Phys.* 35.11 (2005), pp. 1877–1907.

[110] Anne Broadbent and Alain Tapp. "Can quantum mechanics help distributed computing?" In: *ACM SIGACT News* 39.3 (2008), pp. 67–76.

[111] Kaushik Chakraborty et al. "Distributed Routing in a Quantum internet". In: *arXiv preprint arXiv:1907.11630* (2019).

[112] Mohammad Ghaderibaneh et al. "Pre-distribution of entanglements in quantum networks". In: *2022 IEEE International Conference on Quantum Computing and Engineering (QCE)*. IEEE. 2022, pp. 426–436.

[113] Mohsen Falamarzi Askarani, Kaushik Chakraborty, and Gustavo Castro Do Amaral. "Entanglement distribution in multi-platform buffered-router-assisted frequency-multiplexed automated repeater chains". In: *New J. Phys.* 23.6 (2021), p. 063078.

[114] Stav Haldar et al. "Policies for multiplexed quantum repeaters: theory and practical performance analysis". In: *arXiv preprint arXiv:2401.13168* (2024).

[115] Conor E Bradley et al. "A ten-qubit solid-state spin register with quantum memory up to one minute". In: *Phys. Rev. X* 9.3 (2019), p. 031045.

[116] Mohamed H Abobeih et al. "One-second coherence for a single electron spin coupled to a multi-qubit nuclear-spin environment". In: *Nat. Commun.* 9.1 (2018), p. 2552.

[117] Sören Wengerowsky et al. "An entanglement-based wavelength-multiplexed quantum communication network". In: *Nature* 564.7735 (2018), pp. 225–228.

[118] Kevin C Chen et al. “Zero-added-loss entangled-photon multiplexing for ground- and space-based quantum networks”. In: *Phys. Rev. Applied* 19.5 (2023), p. 054029.

[119] Jacob Mower and Dirk Englund. “Efficient generation of single and entangled photons on a silicon photonic integrated chip”. In: *Phys. Rev. A* 84.5 (2011), p. 052326.

[120] V Krutyanskiy et al. “Multimode ion-photon entanglement over 101 kilometers”. In: *PRX Quantum* 5.2 (2024), p. 020308.

[121] OA Collins et al. “Multiplexed memory-insensitive quantum repeaters”. In: *Phys. Rev. Lett.* 98.6 (2007), p. 060502.

[122] WJ Munro et al. “From quantum multiplexing to high-performance quantum networking”. In: *Nature Photonics* 4.11 (2010), pp. 792–796.

[123] Suzanne B van Dam et al. “Multiplexed entanglement generation over quantum networks using multi-qubit nodes”. In: *Quantum Sci. Tech.* 2.3 (2017), p. 034002.

[124] Emre Togan et al. “Quantum entanglement between an optical photon and a solid-state spin qubit”. In: *Nature* 466.7307 (2010), pp. 730–734.

[125] Yiru Zhou et al. “Long-lived quantum memory enabling atom-photon entanglement over 101 km of telecom fiber”. In: *PRX Quantum* 5.2 (2024), p. 020307.

[126] Filip Rozpedek et al. “Near-term quantum-repeater experiments with nitrogen-vacancy centers: Overcoming the limitations of direct transmission”. In: *Phys. Rev. A* 99.5 (2019), p. 052330.

[127] Earl T Campbell and Simon C Benjamin. “Measurement-based entanglement under conditions of extreme photon loss”. In: *Physical review letters* 101.13 (2008), p. 130502.

[128] Cody Jones et al. “Design and analysis of communication protocols for quantum repeater networks”. In: *New J. Phys.* 18.8 (2016), p. 083015.

[129] Wolfgang Dür and Hans J Briegel. “Entanglement purification and quantum error correction”. In: *Rep. Prog. Phys.* 70.8 (2007), p. 1381.

[130] Liangzhong Ruan et al. “Efficient entanglement distillation for quantum channels with polarization mode dispersion”. In: *Phys. Rev. A* 103.3 (2021), p. 032425.

[131] Rodney Van Meter et al. “System design for a long-line quantum repeater”. In: *IEEE/ACM Transactions On Networking* 17.3 (2008), pp. 1002–1013.

[132] R Timothy Marler and Jasbir S Arora. “Survey of multi-objective optimization methods for engineering”. In: *Struct. Multidiscip. Optim.* 26 (2004), pp. 369–395.

[133] Lan Zhou, Wei Zhong, and Yu-Bo Sheng. “Purification of the residual entanglement”. In: *Optics Express* 28.2 (2020), pp. 2291–2301.

[134] Yu-Bo Sheng and Fu-Guo Deng. “Deterministic entanglement purification and complete nonlocal Bell-state analysis with hyperentanglement”. In: *Phys. Rev. A* 81.3 (2010), p. 032307.

[135] Reinhard F Werner. “Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model”. In: *Physical Review A* 40.8 (1989), p. 4277.

[136] Karl Sigman and Ronald W Wolff. "A review of regenerative processes". In: *SIAM review* 35.2 (1993), pp. 269–288.

[137] Armand Makowski, Benjamin Melamed, and Ward Whitt. "On averages seen by arrivals in discrete time". In: *Proceedings of the 28th IEEE Conference on Decision and Control,* IEEE. 1989, pp. 1084–1086.

[138] Maria Vlasiou. "Regenerative Processes". In: *Wiley Encyclopedia of Operations Research and Management Science.* John Wiley & Sons, Ltd, 2011. ISBN: 9780470400531. DOI: https://doi.org/10.1002/9780470400531.eorms0713. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470400531.eorms0713.

[139] Raymond Laflamme et al. "Perfect quantum error correcting code". In: *Phys. Rev. Letters* 77.1 (1996), p. 198.

[140] Ashley M Stephens et al. "Hybrid-system approach to fault-tolerant quantum communication". In: *Phys. Rev. A* 87.5 (2013), p. 052333.

[141] SLN Hermans et al. "Entangling remote qubits using the single-photon protocol: an in-depth theoretical and experimental study". In: *New Journal of Physics* 25.1 (2023), p. 013011.

[142] Sergey Bravyi and Alexei Kitaev. "Universal quantum computation with ideal Clifford gates and noisy ancillas". In: *Physical Review A* 71.2 (2005), p. 022316.

[143] Simon Baier et al. "Realization of a Multi-Node Quantum Network of Remote Solid-State Qubits". In: *Quantum Information and Measurement.* Optica Publishing Group. 2021, M2A–2.

[144] V. Krutyanskiy et al. "Entanglement of Trapped-Ion Qubits Separated by 230 Meters". In: *Phys. Rev. Lett.* 130 (5 Feb. 2023), p. 050803. DOI: 10.1103/PhysRevLett.130.050803. URL: https://link.aps.org/doi/10.1103/PhysRevLett.130.050803.

[145] Tim van Leent et al. "Long-distance distribution of atom-photon entanglement at telecom wavelength". In: *Physical Review Letters* 124.1 (2020), p. 010510.

[146] Yu Ma et al. "One-hour coherent optical storage in an atomic frequency comb memory". In: *Nature communications* 12.1 (2021), p. 2381.

[147] Dominik Leichtle et al. "Verifying BQP computations on noisy devices with minimal overhead". In: *PRX Quantum* 2.4 (2021), p. 040302.

[148] Joseph Glaz, Joseph I Naus, and Sylvan Wallenstein. *Scan statistics.* Springer, 2001.

[149] Joseph Glaz, Vladimir Pozdnyakov, and Sylvan Wallenstein. *Scan statistics: Methods and Applications.* Birkhäuser, 2009.

[150] Shuo-Yen Robert Li. "A martingale approach to the study of occurrence of sequence patterns in repeated experiments". In: *the Annals of Probability* 8.6 (1980), pp. 1171–1176.

[151] James C Fu and WY Wendy Lou. *Distribution theory of runs and patterns and its applications: a finite Markov chain imbedding approach.* World Scientific, 2003.

[152] Morteza Ebneshahrashoob, Tangan Gao, and Milton Sobel. "Sequential window problems". In: *Sequential Analysis* 24.2 (2005), pp. 159–175.

[153] Vladimir Pozdnyakov and J. Michael Steele. “Martingale Methods for Patterns and Scan Statistics”. en. In: *Scan Statistics: Methods and Applications*. Boston, MA: Birkhäuser, 2009, pp. 289–317. ISBN: 978-0-8176-4749-0.

[154] In: (). Our code can be found at https://github.com/bethanydavies234/tools-paper.

[155] Vedran Dunjko et al. “Composable security of delegated quantum computation”. In: *Advances in Cryptology–ASIACRYPT 2014: 20th International Conference on the Theory and Application of Cryptology and Information Security, Kaoshiung, Taiwan, ROC, December 7-11, 2014, Proceedings, Part II 20*. Springer. 2014, pp. 406–425.

[156] Guus Avis et al. “Requirements for a processing-node quantum repeater on a real-world fiber grid”. In: *NPJ Quantum Information* 9.1 (2023), p. 100.

[157] David Williams. *Probability with Martingales*. en. Cambridge University Press, Feb. 1991. ISBN: 978-0-521-40605-5.

[158] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985. DOI: 10.1017/CBO9780511810817.

[159] Charles H Bennett et al. “Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels”. In: *Physical review letters* 70.13 (1993), p. 1895.

[160] Claude Crépeau, Daniel Gottesman, and Adam Smith. “Secure multi-party quantum computation”. In: *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*. 2002, pp. 643–652.

[161] Kevin C Chen, Eric Bersin, and Dirk Englund. “A polarization encoded photon-to-spin interface”. In: *npj Quantum Information* 7.1 (2021), p. 2.

[162] CM Knaut et al. “Entanglement of nanophotonic quantum memory nodes in a telecom network”. In: *Nature* 629.8012 (2024), pp. 573–578.

[163] Richard S Sutton and Andrew G Barto. “Reinforcement learning: An introduction”. In: vol. 17. 2. 1999, pp. 229–235.

[164] Csaba Szepesvári. *Algorithms for reinforcement learning*. Springer nature, 2022.

[165] Sumeet Khatri. “On the design and analysis of near-term quantum network protocols using Markov decision processes”. In: *AVS Quantum Science* 4.3 (2022).

[166] Linh Le et al. “Entanglement routing for quantum networks: a deep reinforcement learning approach”. In: *ICC 2022-IEEE International Conference on Communications*. 2022.

[167] Simon D Reiß and Peter van Loock. “Deep reinforcement learning for key distribution based on quantum repeaters”. In: *Physical Review A* 108.1 (2023), p. 012406.

[168] Julius Wallnöfer et al. “Machine learning for long-distance quantum communication”. In: *PRX quantum* 1.1 (2020), p. 010301.

[169] Guus Avis and Stefan Krastanov. “Optimization of Quantum-Repeater Networks using Stochastic Automatic Differentiation”. In: *arXiv preprint arXiv:2501.06291* (2025).

[170] SLN Hermans et al. “Qubit teleportation between non-neighbouring nodes in a quantum network”. In: *Nature* 605.7911 (2022), pp. 663–668.

[171] [Online] Available: https://gitlab.com/adaptivealgo.

[172] Charles H Jones. “Generalized hockey stick identities and N-dimensional block-walking”. In: *The Fibonacci Quarterly* 34.3 (1996), pp. 280–288.

[173] William J Munro et al. “Inside quantum repeaters”. In: *IEEE Journal of Selected Topics in Quantum Electronics* 21.3 (2015), pp. 78–90.

[174] Emily A Van Milligen et al. “Utilizing probabilistic entanglement between sensors in quantum networks”. In: *arXiv preprint arXiv:2407.15652* (2024).

[175] Xiangyi Meng et al. “Quantum Networks Enhanced by Distributed Quantum Memories”. In: *arXiv preprint arXiv:2403.16367* (2024).

[176] Matheus Guedes de Andrade et al. “On the Analysis of Quantum Repeater Chains with Sequential Swaps”. In: *arXiv preprint arXiv:2405.18252* (2024).

[177] Vaishnavi L Addala, Shu Ge, and Stefan Krastanov. “Faster-than-Clifford simulations of entanglement purification circuits and their full-stack optimization”. In: *npj Quantum Information* 11.1 (2025), p. 12.

[178] Michael Zwerger et al. “Quantum repeaters based on trapped ions with decoherence-free subspace encoding”. In: *Quantum Science and Technology* 2.4 (2017), p. 044001.

[179] S. Bose, V. Vedral, and P. L. Knight. “Purification via entanglement swapping and conserved entanglement”. In: *Phys. Rev. A* 60 (1 July 1999), pp. 194–197. DOI: 10.1103/PhysRevA.60.194. URL: https://link.aps.org/doi/10.1103/PhysRevA.60.194.

[180] Luis Roa Oppliger et al. “Threshold effect for probabilistic entanglement swapping”. In: *Annals of Physics* 451 (2023), p. 169257.

[181] Wei Song, Ming Yang, and Zhuo-Liang Cao. “Purifying entanglement of noisy two-qubit states via entanglement swapping”. In: *Physical Review A* 89.1 (2014), p. 014303.

[182] Chuanmei Xie et al. “Quantum Correlation Swapping between Two Werner States Undergoing Local and Nonlocal Unitary Operations”. In: *Entropy* 24.9 (2022), p. 1244.

[183] Luis Roa, Ariana Muñoz, and Gesa Grüning. “Entanglement swapping for X states demands threshold values”. In: *Physical Review A* 89.6 (2014), p. 064301.

[184] Joanna Modławska and Andrzej Grudka. “Increasing singlet fraction with entanglement swapping”. In: *Physical Review A—Atomic, Molecular, and Optical Physics* 78.3 (2008), p. 032321.

[185] Shu-Cheng Li et al. “Violation of Clauser–Horne–Shimony–Holt inequality and teleportation enhanced for a class of two-qubit X-states resulting from entanglement swapping”. In: *Quantum Information Processing* 14 (2015), pp. 3845–3855.

[186] Brian T Kirby et al. “Entanglement swapping of two arbitrarily degraded entangled states”. In: *Physical Review A* 94.1 (2016), p. 012336.

[187] Aditi Sen et al. “Entanglement swapping of noisy states: A kind of superadditivity in nonclassicality”. In: *Physical Review A—Atomic, Molecular, and Optical Physics* 72.4 (2005), p. 042310.

[188] J Dajka and J Łuczka. “Swapping of correlations via teleportation with decoherence”. In: *Physical Review A—Atomic, Molecular, and Optical Physics* 87.2 (2013), p. 022301.

[189] János A Bergou et al. “Average concurrence and entanglement swapping”. In: *Physical Review A* 104.2 (2021), p. 022425.

[190] Run-Ze Cong and Shuang Xu. “Formal bounds of entanglement purification via entanglement swapping”. In: *Physical Review A* 111.1 (2025), p. 012409.

[191] Víctor HT Brauer and Andrea Valdés-Hernández. “Enhancing teleportation via noisy channels: Effects of the induced multipartite entanglement”. In: *Physical Review A* 109.5 (2024), p. 052606.

[192] Sangchul Oh, Soonchil Lee, and Hai-woong Lee. “Fidelity of quantum teleportation through noisy channels”. In: *Physical Review A* 66.2 (2002), p. 022316.

[193] Jeongho Bang, Junghee Ryu, and Dagomir Kaszlikowski. “Fidelity deviation in quantum teleportation”. In: *Journal of Physics A: Mathematical and Theoretical* 51.13 (2018), p. 135302.

[194] Piotr Badziag et al. “Local environment can enhance fidelity of quantum teleportation”. In: *Physical Review A* 62.1 (2000), p. 012311.

[195] Tal Mor and Pawel Horodecki. “Teleportation via generalized measurements, and conclusive teleportation”. In: *arXiv preprint quant-ph/9906039* (1999).

[196] Wan-Li Li, Chuan-Feng Li, and Guang-Can Guo. “Probabilistic teleportation and entanglement matching”. In: *Physical Review A* 61.3 (2000), p. 034301.

[197] Pankaj Agrawal and Arun K Pati. “Probabilistic quantum teleportation”. In: *Physics Letters A* 305.1-2 (2002), pp. 12–17.

[198] Raphael Fortes and Gustavo Rigolin. “Probabilistic quantum teleportation in the presence of noise”. In: *Physical Review A* 93.6 (2016), p. 062330.

[199] Michael R Geller and Zhongyuan Zhou. “Efficient error models for fault-tolerant architectures and the Pauli twirling approximation”. In: *Physical Review A—Atomic, Molecular, and Optical Physics* 88.1 (2013), p. 012314.

[200] Mauricio Gutiérrez and Kenneth R Brown. “Comparison of a quantum error-correction threshold for exact and approximate errors”. In: *Physical Review A* 91.2 (2015), p. 022335.

[201] Joseph Emerson, Robert Alicki, and Karol Życzkowski. “Scalable noise estimation with random unitary operators”. In: *Journal of Optics B: Quantum and Semiclassical Optics* 7.10 (2005), S347.

[202] Jonas Helsen et al. “General framework for randomized benchmarking”. In: *PRX Quantum* 3.2 (2022), p. 020357.

[203] Xiao-Min Hu et al. “Progress in quantum teleportation”. In: *Nature Reviews Physics* 5.6 (2023), pp. 339–353.

[204] Charles H Bennett and Gilles Brassard. "Quantum cryptography: Public key distribution and coin tossing". In: *Theoretical computer science* 560 (2014), pp. 7–11.

[205] Dagmar Bruß. "Optimal eavesdropping in quantum cryptography with six states". In: *Physical Review Letters* 81.14 (1998), p. 3018.

[206] Stefano Pironio et al. "Device-independent quantum key distribution secure against collective attacks". In: *New Journal of Physics* 11.4 (2009), p. 045021.

[207] Victoria Lipinska et al. "Verifiable hybrid secret sharing with few qubits". In: *Physical Review A* 101.3 (2020), p. 032332.

[208] https://gitlab.com/GuusAvis/swap_strategies.

[209] Armin Tavakoli et al. "Semidefinite programming relaxations for quantum correlations". In: *Reviews of Modern Physics* 96.4 (2024), p. 045006.

[210] Ryo Namiki et al. "Role of syndrome information on a one-way quantum repeater using teleportation-based error correction". In: *Physical Review A* 94.5 (2016), p. 052304.

[211] Yumang Jing, Daniel Alsina, and Mohsen Razavi. "Quantum key distribution over quantum repeaters with encoding: Using error detection as an effective postselection tool". In: *Physical Review Applied* 14.6 (2020), p. 064037.

[212] Kah Jen Wo et al. "Resource-Efficient Fault-Tolerant One-Way Quantum Repeater with Code Concatenation". In: *npj Quantum Information* 9.1 (Dec. 2023), pp. 1–11. ISSN: 2056-6387. DOI: 10.1038/s41534-023-00792-8. URL: https://www.nature.com/articles/s41534-023-00792-8 (visited on 12/20/2023).

[213] Peter W. Shor and John Preskill. "Simple Proof of Security of the BB84 Quantum Key Distribution Protocol". In: *Physical Review Letters* 85.2 (July 2000), pp. 441–444. DOI: 10.1103/PhysRevLett.85.441. URL: https://link.aps.org/doi/10.1103/PhysRevLett.85.441 (visited on 03/15/2021).

[214] Bruce E Sagan. *The symmetric group: representations, combinatorial algorithms, and symmetric functions*. Vol. 203. Springer Science & Business Media, 2013.

ACKNOWLEDGEMENTS

The four years of my PhD flew by because of the brilliant people around me.

Firstly, I would like to thank Stephanie, for the opportunity to carry out this research in the first place, as well as your support throughout. Every meeting helped to shape the research direction due to your unique and broad perspective on the field, while leaving enough freedom to allow me to develop as a researcher. I would also like to thank my co-promotor, Ronald, as well as my doctoral committee, Rihan, Wolfgang, Mathijs, Michael and Lieven for investing the time in reviewing the dissertation and sharing their valuable perspective during the defence.

Thank you to every one of my close collaborators and co-authors, for the discussions and inspiration. Working with you was one of my favourite experiences of the PhD. More broadly, our research group and the other groups along the corridor have always formed a stimulating and entertaining research environment. I will look back fondly on quiet days, when there were only a few people in the office but the coffee time extended past 1pm because of the conversation, as well as moments on trips abroad, from exploring Las Vegas to admiring the view of Lake Geneva. It was also a pleasure to guide bachelor and master students: thank you for your motivation, enthusiasm, and valuable research contributions.

I would like to thank everyone who I shared a laugh with in Delft and made me feel comfortable and at home. From badminton tournaments to board game nights to calm walks through Delft, I really appreciate everyone who I share these memories with. Nancy and Álvaro, I would not want to have anyone else there with me during my defence. Siri, somehow our friendship strengthened even more after I moved abroad (possibly because we have more creative license in calls). Thank you to my family, especially my parents, for your unconditional support. Y el resto ya lo sabes.

CURRICULUM VITÆ

Bethany Jane DAVIES

2021-2025	PhD in Quantum Networks Delft University of Technology, the Netherlands *Thesis:* Performance analysis of near-term quantum networks *Promotor:* Prof. dr. S.D.C. Wehner *Co-promotor:* Prof. dr. ir. R. Hanson
2020–2021	Master of Mathematics Newnham College, University of Cambridge, UK
2017–2020	BA in Mathematics Newnham College, University of Cambridge, UK
28/07/1999	Born in Brighton, UK

List of Publications

5. **A. Tacettin*, T. Qu*, B. Davies*, B. Goranov, I. Draganescu, and G. Vardoyan**, *Adaptive policies for resource generation in a quantum network*, arXiv preprint arXiv:2509.17576.

4. **B. Davies, G. Avis, S. Wehner**, *On the accuracy of twirled approximations in repeater chains*, arXiv preprint arXiv:2509.16689.

3. **Á.G. Iñesta*, B. Davies*, S. Kar and S. Wehner**, *Entanglement buffering with multiple quantum memories*, npj Quantum Inf. (2025)

2. **B. Davies*, Á. G. Iñesta*, S. Kar and S. Wehner**, *Entanglement buffering with two quantum memories*, Quantum 8 (2024): 1458

1. **B. Davies, T. R. Beauchamp, G. Vardoyan and S. Wehner**, *Tools for the analysis of quantum protocols requiring state generation within a time window*, IEEE Transactions on Quantum Engineering, vol. 5, pp. 1-20, 2024

*These authors contributed equally.