Minimum Cost Swapping and Purification Scheduling over Quantum Repeater Chains

Jiyao Liu*, Lei Fan†‡, Yuanxiong Guo§, Zhu Han‡, and Yu Wang*
*Department of Computer and Information Sciences, Temple University, Philadelphia, USA
†Dept. of Engineering Technology, ‡Dept. of Electrical and Computer Engineering, University of Houston, Houston, USA
§Department of Information Systems and Cyber Security, University of Texas at San Antonio, San Antonio, USA
{jiyao.liu, wangyu}@temple.edu, {lfan8, zhan2}@uh.edu, yuanxiong.guo@utsa.edu.

***Abstract*—Entanglement swapping and purification are the two fundamental primitives for long-distance entanglement distribution in quantum networks. While prior work has established that purification should precede swapping in Binary systems to minimize the expected number of consumed link-level entanglements for an E2E fidelity guarantee, the critical problem of finding the optimal swapping and purification strategy remains open. To fill this gap, we study the Optimal Swapping and Purification Tree (OSPT) problem for a quantum repeater chain in Binary systems and derive the exact optimal solution. By leveraging the associativity of swapping and purification fidelity, we decompose the problem into three subproblems: optimal swapping tree construction, optimal purification scheduling, and global budget allocation. We show that although the exact solution can be obtained via brute-force search over budget distributions, its complexity grows exponentially. To address this challenge, we further propose an efficient greedy budget allocation algorithm. Extensive evaluations demonstrate that the greedy algorithm closely approximates the optimal solution and outperforms existing solutions while significantly reducing computational complexity.**

***Index Terms*—Entanglement Distribution, Entanglement Swapping, Purification, Fidelity, Quantum Network.**

I. Introduction

Quantum networks [1]–[3] distribute entanglement between a source and a destination via a chain of quantum repeaters, so that quantum states can be transferred between them via teleportation. This entanglement, typically represented as Einstein–Podolsky–Rosen (EPR) pairs, is the fundamental resource for quantum communication [4]–[8] and distributed quantum computing [9], [10]. Establishing high-quality long-distance entanglement over a multi-hop quantum repeater chain, however, is difficult, since (1) quantum states cannot be copied due to the no-cloning theorem, and (2) direct photon transmission over long fibers is impractical because the photon survival probability decays exponentially with distance.

To overcome such challenges, two fundamental quantum operations, *entanglement swapping* and *entanglement purification*, are used to generate long-distance entanglement in quantum networks. *Entanglement swapping* extends entanglement across multiple links by performing a Bell-state measurement at an intermediate node: given two entangled pairs (e.g., n_3–n_4 and n_4–n_5 in Fig. 1 at time slot t_1), swapping entangles the end nodes (n_3–n_5) without them directly interacting. The swapping operation can be repeated hop-by-hop until an end-to-end (E2E) entanglement is obtained on the desired source and destination pair. Therefore, it forms the basis of quantum repeater chains and multi-hop quantum networks. However, swapping typically reduces fidelity due both to its inherent nature and to noisy, imperfect operations. To address this, *entanglement purification* (or distillation) probabilistically improves the quality of entangled states by consuming multiple low-fidelity pairs to produce fewer high-fidelity pairs through local operations and classical communication (e.g., two entanglements between n_2 and n_3 are purified into one higher-fidelity pair at t_1 by Scheme 1 in Fig. 1). Together, swapping enables scalable connectivity, while purification preserves reliability, making both essential for robust quantum networks.

Because swapping and purification can be interleaved in many ways, their interaction yields a large design space of *joint swapping and purification scheduling* (SPS) [11]–[15]. As shown in Figs. 1a and 1b, different scheduling can realize the same target E2E entanglement, yet achieve different fidelity and incur different expected costs once operation failures are considered. Prior work has studied joint swapping and purification [11]–[13], [16]–[18], but often within restricted settings or adopting a particular SPS pattern without justification. Several recent studies [19]–[21] investigated joint swapping and purification under over-simplified assumptions (e.g., neglecting operation failures), which can obscure the true expected resource cost.

Most recently, Liu *et al.* [14], [15] modeled the joint swapping and purification optimization while explicitly accounting for swapping and purification failures, and introduced a tree-based representation that can capture arbitrary hybrid SPS operation orders (as shown in Figs. 1c and 1d). Particularly, they studied the optimal swapping and purification problem to minimize the expected number of consumed entanglements while satisfying the E2E fidelity threshold. In the *Binary system* (similar to the *bit-flip noise* model), they established a key structural result: the optimal SPS always follows a certain pattern, i.e., *PS scheme* where purification should be performed before any swapping (e.g. Scheme 1 in Fig. 1). This result significantly narrows the space of candidate strategies,

The work is partially supported by the US National Science Foundation (Grant No. CNS-2106761, CNS-231868, and OAC-2417716).

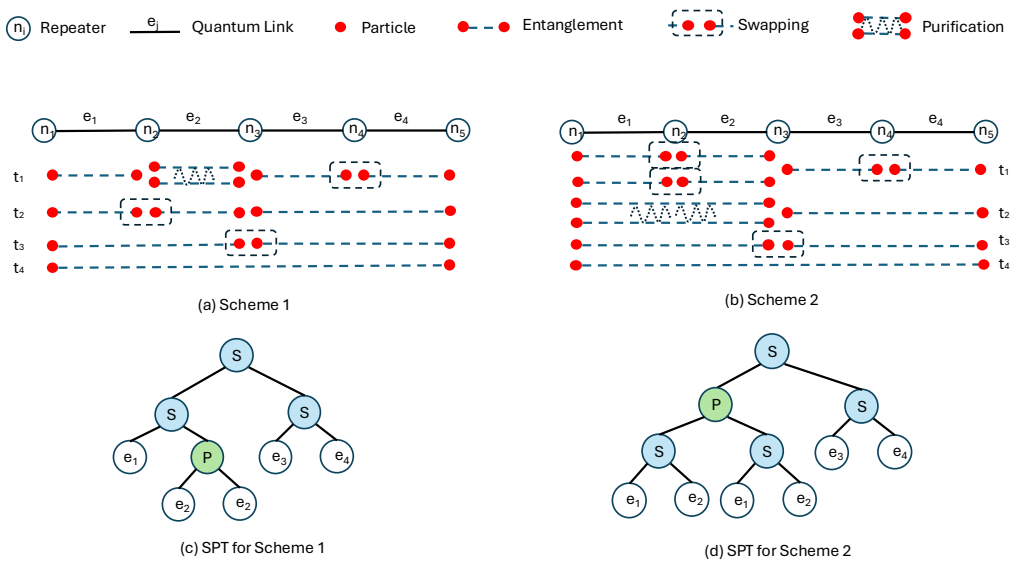


Fig. 1. Modeling of joint swapping and purification scheduling over a quantum repeater chain via the swapping and purification tree (SPT): (a) and (b) Scheme 1 and Scheme 2, (c) and (d) SPTs of Scheme 1 and Scheme 2.

but it does *not* by itself identify the *exact* optimal SPS strategy. The authors of [14], [15] proposed a tree-based swapping and purification scheme following the PS pattern, however it cannot guarantee the optimum for the studied problem.

This paper addresses this gap by targeting the *exact* optimal SPS strategy, rather than only the optimal structural pattern. Note that even when restricted to PS strategies, the expected cost depends on (i) how many entanglements (budget) to allocate to purification on each link/segment on the repeater chain, (ii) the order in which purifications are performed, and (iii) how swapping is organized across the chain. We formulate the *Optimal Swapping and Purification Tree* problem, whose goal is to minimize the expected number of consumed link-level entanglements (i.e. the cost) while guaranteeing a required E2E fidelity threshold. Leveraging associativity properties together with the tree-based model, we decompose the problem into three coupled subproblems: (i) optimal swapping tree construction, (ii) optimal purification scheduling under a given per-link budget, and (iii) global budget allocation across links/segments. The global optimal solution can be obtained by exhaustive search over budget distributions, but the resulting complexity grows exponentially with both the budget and the path length. To make the approach practical, we further develop an efficient greedy budget-allocation algorithm that closely tracks the optimal solution in practice.

In summary, this paper makes the following contributions:

1) We develop an exact solution framework by decomposing the problem into optimal swapping tree construction, optimal purification scheduling, and exhaustive budget allocation, and characterize its computational complexity.
2) We propose an efficient greedy budget allocation algorithm that achieves near-optimal performance with substantially reduced runtime.
3) Through extensive simulations, we demonstrate that the proposed method significantly reduces entanglement consumption compared with representative baselines under realistic fidelity targets.

The remainder of the paper is organized as follows. Section II reviews related work. Section III presents the system model and necessary background. Section IV first provides the problem definition, then develops the exact solution framework and the greedy method. Section V reports evaluation results, and finally Section VI concludes the paper.

II. Related Works

A. Entanglement Distribution and Routing

Studies on entanglement distribution or entanglement routing have focused on throughput maximization, latency minimization, or fidelity optimization.

Dai *et al.* [16] design remote entanglement distribution protocols that maximize the entanglement distribution rate (EDR), i.e., the average amount of entanglement distributed between two specified nodes per time slot, by solving formulated linear programming problems. Zeng *et al.* [22] develop an entanglement routing protocol to maximize the number of served users and their expected throughout. Their method uses a two-step approach which solves two sequential integer programming steps. Chang *et al.* [23] show that swapping order affects the expected throughput of a quantum path, and prove that computing a path with maximum expected throughput under any entanglement swapping order does not have the sub-path optimality property.

Liu and Wang [24] investigate the stochastic behavior of entanglement swappings and design a swapping scheduling method for path-level latency minimization. Yang *et al.* [25] propose an online routing algorithm to enhance efficiency and scalability by calculating the entanglement paths for each request when it arrives. Farahbakhsh and Feng [26] propose an opportunistic routing approach to reduce the E2E delay by forwarding the request along the path as soon as it is possible.

Besides throughput and delay, fidelity is a key metric for quantifying entanglement quality, accordingly, fidelity-aware routing has also been investigated. For example, Gu *et al.*

[27] explore the trade-off between the throughput and achieved fidelity and study how to maximize the worst-case E2E fidelity subject a lower bound on the expected EDR. Ghaderibaneh *et al.* [28] and Fan *et al.* [29] use a swapping tree to model swapping orders along a path and develop a dynamic programming algorithm to select an optimal swapping-tree to minimize entanglement generation latency under the given capacity and fidelity constraints.

B. Purification and Joint Optimization with Swapping

One possible way to improve E2E fidelity is to use purification to enhance the fidelity at either link-level or path-level entanglements. Dür *et al.* [17] consider a few fixed purification schemes, purifying after a fixed number of swapping steps, and provide network-level study of them. Zhao *et al.* [12] and Li *et al.* [13] both consider use purification to improve the E2E fidelity in entanglement routing, and design greedy algorithms that iteratively identify a "critical" link (e.g., based on fidelity increase or gradient) and purify it until a fidelity threshold is met. Panigrahy *et al.* [18] study purification in the context of quantum switches and capacity considerations. These existing purification approaches typically either assume a fixed purification pattern or optimize purification via heuristics, and often do not consider the exact expected cost under both swapping and purification failures.

Modeling joint swapping and purification is challenging because operation outcomes are probabilistic and purification success depends on input fidelity. Jia and Chen [19]–[21] study purification scheduling and fidelity-constrained routing. Their analysis provides valuable insights, but it typically assumes no operation failures and/or relies on simplified metrics, making it difficult to directly obtain the expected entanglement cost. Recently, Liu *et al.* [14], [15] explicitly models swapping and purification with failures using a unified tree framework and establishes structural optimality results under the Binary system. Particularly, they prove that the optimal SPS solution under Binary systems follows the PS scheme (i.e., performing purification before swapping). They also propose a branching tree method to generate a swapping and purification scheduling which follows PS scheme. However, such tree-based method does not guarantee the optimal expected cost for the formulated optimization. Building on the foundation from [14], [15], this paper focuses on the remaining open challenge: computing the *exact* optimal strategy (rather than only the optimal pattern) by optimizing tree construction and, critically, the budget allocation across links/segments.

III. System Model

We now introduce the network model, noise assumption, quantum operations, and tree-based model used by this paper.

A. Network Model

We consider a quantum repeater chain with N repeaters modeled as a simple path $\pi : n_1 \leftrightarrow n_2 \leftrightarrow \cdots \leftrightarrow n_N$, where each adjacent node pair (n_i, n_{i+1}) is connected by a quantum link e_i (typically an optical fiber). Each link can probabilistically generate entangled photon pairs between its two endpoints. Following common practice [12]–[14], we model entanglement generation in an expected-value sense: the number of successfully generated link-level entanglements over a scheduling period is represented by its expectation, which serves as the basic resource unit.

An E2E entanglement between n_1 and n_N is established by applying entanglement swapping and purification operations on these link-level entanglements. The goal is to generate E2E entanglements over the quantum repeater chain that satisfy a target fidelity requirement F^* while minimizing the expected consumption of link-level entanglements.

B. Quantum State and Noise Model

We focus exclusively on the *Binary* [14], [15] noise model, where the behaviors of swapping and purification are similar to the *bit-flip* model. This model has been widely adopted in the literature due to its analytical tractability and relevance to practical purification protocols [12], [14], [17]. Under this model, an entangled pair is represented by a density matrix in which only one undesired Bell state has non-zero probability mass. The quality of an entanglement is fully characterized by its *fidelity* $f \in (0.5, 1)$, defined as the probability that the pair is in the desired Bell state. We assume that all link-level entanglements are initially in the Binary state.

C. Entanglement Swapping and Purification

The two fundamental quantum operations are entanglement swapping and entanglement purification.

a) Entanglement Swapping: Entanglement swapping combines two adjacent entanglements into a longer one via a Bell-state measurement (BSM) at the intermediate node. Under the Binary model, if two input entanglements have fidelities f_1 and f_2, respectively, the output fidelity after swapping is given by

$$f_{\text{swap}} = f_1 f_2 + (1 - f_1)(1 - f_2). \tag{1}$$

The swapping operation succeeds with probability $q_s \in [0.5, 1]$, which depends on the physical implementation of the BSM (e.g., (non-)linear-optical or memory-assisted realizations) [14], [17]. Upon failure, the input entanglements are discarded.

b) Entanglement Purification: Entanglement purification consumes two entanglements between the same pair of nodes and probabilistically produces one entanglement with higher fidelity. We consider standard two-to-one purification protocols designed for Binary states [17]. Given two input entanglements with fidelity f_1 and f_2, the output fidelity after successful purification is

$$f_{\text{pur}} = \frac{f_1 f_2}{f_1 f_2 + (1 - f_1)(1 - f_2)}, \tag{2}$$

and the success probability of the purification operation is

$$q_p = f_1 f_2 + (1 - f_1)(1 - f_2). \tag{3}$$

If purification fails, both input entanglements are discarded.

D. Swapping and Purification Tree

Any joint swapping and purification strategy along the repeater chain can be represented by a *Swapping and Purification Tree (SPT)* [14], [15], as shown in Figs. 1c and 1d). In an SPT, leaf nodes correspond to link-level entanglements, and internal nodes correspond to swapping or purification operations, each consuming two child entanglements and producing one output entanglement upon success.

The fidelity of the entanglement produced at each internal node is determined by applying the corresponding operation recursively from the leaves to the root, using Eq. (1) or Eq. (2). The *expected cost* of an SPT is defined as the expected number of link-level entanglements consumed to produce one successful output at the root, accounting for operation failures. As shown in [14], [15], this cost can be computed recursively in a bottom-up manner by dividing the sum of the expected costs of the two children by the success probability of the operation at the parent node.

IV. Optimal Swapping and Purification

In this section, we first formally define the optimal swapping and purification problem which aims to minimize the cost. Then leveraging the structural properties under the Binary (bit-flip) model, we derive an *exact* solution of the problem via exhaustive budget allocation over a *finite* budget range. Finally, we will give a greedy budget allocation algorithm, which closely approximates the optimal solution in practice.

A. Problem Definition

We now formally define the minimum cost optimization problem studied in this paper. Our model follows the same basic setting in [14], [15], but with a narrower focus on the Binary (bit-flip) noise model and exact optimality.

Definition 1 (Optimal Swapping and Purification Tree)**.** *Given a quantum repeater chain $\pi : n_1 \leftrightarrow \cdots \leftrightarrow n_N$ with initial link fidelities $\{f_i\}$, swapping success probability q_s, and a target E2E fidelity threshold $F^* \in (0.5, 1)$, the* **Optimal Swapping and Purification Tree (OSPT)** *problem under Binary model aims to find an SPT such that:*

1) *the fidelity of the root entanglement is at least F^*, and*
2) *the expected number of consumed link-level entanglements (i.e. the expected cost of SPT) is minimized.*

Unlike prior work [14], [15] only identifying optimal structural patterns (e.g., purify-then-swap), this paper aims to generate the *exact* optimal solution of OSPT by explicitly optimizing the organization of swapping and the allocation of purification budget across links and segments.

B. Preliminaries: Associativity and Tree Decomposition

Under the Binary model, both swapping and purification admit associativity properties that greatly simplify fidelity analysis. As established in [14], [15], for any fixed multiset of input fidelities, the final fidelity produced by a sequence of swapping operations is invariant to the swapping order; similarly, for repeated purification over the same pair of nodes, the final fidelity is determined by the purification budget, rather than the specific purification tree shape. These properties imply a key separation:

- **Fidelity** is primarily governed by (i) the per-link purified fidelities and (ii) the set of links on the chain, rather than the detailed tree structure.
- **Expected cost** depends on how probabilistic operations are organized (tree shapes), because swapping succeeds with probability q_s and purification succeeds with a fidelity-dependent probability [14], [15], [17].

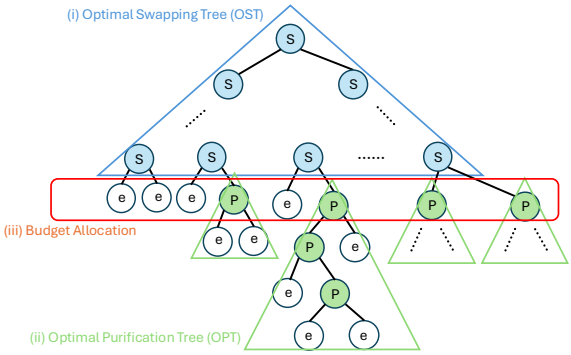


Fig. 2. Three subproblems of OSPT: OST construction, OPT construction, and budget allocation.

Therefore, we decompose the OSPT problem into three coupled subproblems: (i) *optimal swapping-tree construction*, (ii) *optimal purification-tree construction on each link for a given budget*, and (iii) *global budget allocation across links/segments*. See Fig. 2 for illustration. Next we will introduce each of them in details.

C. Optimal Swapping Tree and Optimal Purification Tree

We now introduce how to solve the first two subproblems.

Optimal swapping tree (OST): Consider a chain subpath from node n_i to n_j $(i < j)$ whose adjacent segments have fixed fidelities (after purification is decided). Since swapping order does not affect the resulting E2E fidelity under the Binary model [14], [15], the best swapping tree is the one minimizing expected cost. Following the tree-cost recursion in [14], [15], the optimal swapping tree can be obtained via dynamic programming (DP):

$$C(i,j) = \min_{i<v<j} \frac{C(i,v) + C(v,j)}{q_s}, \tag{4}$$

where $C(i,j)$ denotes the minimum expected cost to produce one entanglement spanning (n_i, n_j) from already-available entanglements on shorter subpaths. When swapping success probabilities vary by node, q_s above is replaced by the corresponding q_v at split point v.

Optimal purification tree (OPT): For each link (or short segment) e, given an integer budget $b \geq 1$ (number of link-level entanglements participated in successful purifications), we seek the purification strategy that yields the lowest expected cost for that resulting fidelity. Under the Binary model and standard two-to-one purification [17], an optimal purification tree for budget b can be computed via dynamic programming (details follow the same recursion style as [14], [15]):

$$C_e(b) = \min_{1 \le k < b} \frac{C_e(k) + C_e(b-k)}{q_p(\, f_e(k), f_e(b-k)\,)}, \tag{5}$$

where $f_e(b)$ and $C_e(b)$ denote the best achievable minimum expected cost on link e using budget b, and $q_p(\cdot,\cdot)$ is the purification success probability under Binary model (Equ. (3)).

We will use a table T (called purification table or OPT table) to save pre-link solutions $(f_e(b_e), C_e(b_e))$ of each edge e with a budget b_e from OPT. This table can support quick look-up and used by our budget allocation methods.

Note that the final entanglement is obtained only via succeeded operations (swapping and/or purification). Those whose operation fails only contributes to the extra cost, i.e., *cost* minus *budget*. Therefore, *budget* is an integer that tells the number of entanglements that actually participated in generating the final E2E entanglement, while *cost* is a real number that tells the expected number of total entanglements needed (including failed ones).

Key implication: Once we restrict to (i) per-link OPTs and (ii) a global OST built on the resulting purified segments, the only remaining degree of freedom for exact optimality of OSPT is the *budget distribution* across links/segments.

D. Budget Distribution and Finite Search Range

Let π denote a repeater chain with L links (edges). A *budget distribution* is a vector $\mathbf{b} = \{b_e \mid e \in \pi\}$, where $b_e \ge 1$ is an integer to denote the budget assigned to link e. The total budget is

$$B(\mathbf{b}) = \sum_{e \in \pi} b_e. \tag{6}$$

Given $\mathbf{b}$, we construct a candidate strategy by: (i) applying OPT on each link e with budget b_e to obtain purified fidelity $f_e(b_e)$ and cost $C_e(b_e)$, and (ii) applying OST on the chain using those purified segments to obtain total expected cost $C(\mathbf{b})$ and E2E fidelity $F(\mathbf{b})$.

A crucial step for an *exact* algorithm is to show that the optimal total budget lies within a finite range by proving the following theorem.

Theorem 1 (Finite Budget Range)**.** *Consider a repeater chain π with L links. Let $f_{\min} = \min_{e\in\pi} f_e(1)$ be the minimum initial link fidelity (before purification), and let $q_{\min} = \min q_s$ be the minimum swapping success probability on the chain. For any target E2E fidelity threshold $F^* \in (0.5, 1)$, there exists a finite upper bound $B_{\max}$ such that an exact optimal solution will use a total budget of B^**

$$B^* \in [B_{\min}, B_{\max}], \tag{7}$$

where $B_{\min} = L$ and $B_{\max} = \left\lceil \frac{L \cdot b \cdot 2^{b-1}}{q_{\min}^{L-1}} \right\rceil$. Here $b = \left\lceil \log_{(\frac{1}{f_{\min}}-1)} (\frac{1}{f} - 1) \right\rceil$, and f is a per-link purified fidelity that suffices to make the E2E fidelity no smaller than F^* under swapping (e.g., a conservative choice based on the chain length). Refer to Appendix for how to calculate f.

Proof Sketch: $B_{\min} = L$ is necessary since each link requires at least one entanglement to make the chain connectable. To obtain the upper bound, we first determine a sufficient per-link purification budget that raises each link to a conservative fidelity f, ensuring that the end-to-end fidelity meets F^* after $L-1$ swapping operations. We then upper-bound the resulting increase in expected cost due to (i) purification failures (each two-to-one purification step succeeds with probability at least $1/2$ in the Binary model [17]) and (ii) swapping failures along the chain, captured by $q_{\min}^{-(L-1)}$. A full proof is provided in Appendix. □

Remark: Theorem 1 is intentionally conservative, and guarantees finiteness of the search space and ensures exactness. Tighter bounds can be obtained in practice by using the purified-fidelity tables $f_e(b)$ computed by OPT and stopping early once increasing B cannot reduce cost.

E. Exact Solution: Optimal Budget Allocation (OBA)

With Theorem 1, we can define an exact algorithm that is guaranteed to terminate. For each total budget $B \in [B_{\min}, B_{\max}]$, we enumerate all integer compositions of B into L positive parts, i.e., all $\mathbf{b}$ such that $\sum b_e = B$ and $b_e \ge 1$.

For a fixed B, the number of feasible distributions (denoted by $\mathcal{A}_{B,L}$) is:

$$|\mathcal{A}_{B,L}| = \binom{B-1}{L-1}. \tag{8}$$

For each $\mathbf{b} \in \mathcal{A}_{B,L}$, we compute $F(\mathbf{b})$ and $C(\mathbf{b})$ by: (i) looking up per-link $(f_e(b_e), C_e(b_e))$ from OPT tables and (ii) running OST DP on the chain to obtain total expected cost. Among all feasible distributions that satisfy $F(\mathbf{b}) \ge F^*$, we keep the one with minimum $C(\mathbf{b})$.

Exactness: Because (i) OST/OPT are optimal for any fixed purified segments and budgets, and (ii) we enumerate *all* budget distributions within a range guaranteed to contain an optimum (Theorem 1), the best feasible strategy found by OBA is globally optimal.

Complexity: The above exact algorithm is computationally expensive. Even for a fixed B, the enumeration size $\binom{B-1}{L-1}$ grows rapidly with L, and $B_{\max}$ itself increases with the fidelity target and chain length. Therefore, while OBA provides the *ground truth* optimum for evaluation and small-to-moderate chains, it is impractical for larger chains. This motivates efficient approximation algorithms.

F. Greedy Budget Allocation (GBA)

We now present a greedy budget allocation algorithm that follows the structure implied by previous section: it incrementally increases the total budget and assigns the next unit of budget to the link that yields the largest gain in final E2E fidelity (or equivalently, the largest marginal improvement toward meeting F^*), while using OPT/OST to evaluate each step. Empirically, this approach closely tracks the exact optimum while reducing complexity. Algorithm 1 shows the detailed design.

We need to first use OPT DP algorithm to calculate the purification tables T, where $T[e][b]$ is the fidelity of the entanglement over edge e purified with b entanglements. Then, in one iteration of the outer while loop, we allocate one more budget the the edge that increases the total fidelity the most (when purified with one more entanglement) by trying every edge. Here, BUILD-OST(A,T) is the dynamic programming algorithm to calculate the optimal swapping tree.

Algorithm 1 GREEDY BUDGET ALLOCATION

Input: Path π with edges' fidelity f_e for $\forall e \in \pi$. E2E fidelity threshold F^*.

Output: Budget allocation A.

Calculate purification tables T by OPT
Initialize $A[e] = 1$ for $\forall e \in \pi$ ▷ *initial allocation*
$f = \text{BUILD-OST}(A, T)$
while $f < F^*$ **do**
  for $e \in \pi$ **do** ▷ *find edge with most increased fidelity*
    $A[e] = A[e] + 1$
    $f' = \text{BUILD-OST}(A, T)$
    $A[e] = A[e] - 1$
    if $f' > f$ **then**
      $e_m = e, f = f'$
  $A[e_m] = A[e_m] + 1$ ▷ *increase budget of that edge*
return A

V. EVALUATIONS

We now evaluate the performance of the proposed method.

A. Experiment Setup

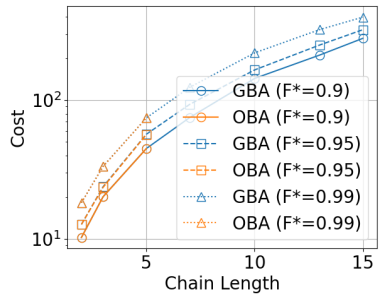
We generate random quantum repeater chains with varying numbers of links. Unless otherwise specified, link-level fidelity are independently drawn from $[0.7, 0.95]$, consistent with practical settings in prior work [12]. Swapping success probabilities q_s are in $[0.5, 1]$. For each configuration, results are averaged over 100 independent runs. Reported runtime includes all required subroutines, including purification table construction (OPT) and swapping-tree computation (OST). We evaluate performance under multiple E2E fidelity thresholds $F^* \in \{0.9, 0.95, 0.99\}$, representing low, medium, and stringent reliability requirements.

B. Baselines

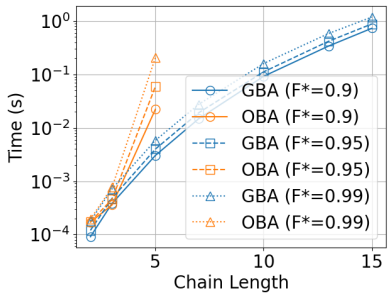
We consider the following baselines.

1) **OBA**: our exhaustive optimal budget allocation algorithm, employing the dynamic programming algorithms for the optimal swapping tree and the optimal purification tree. For efficient comparison, in experiments, OBA does not iterate over $[B_{\min}, B_{\max}]$; instead, it starts with the budget given by GBA and decreases by one each time until reaching the minimum possible budget L.
2) **GBA**: our greedy budget allocation Algorithm 1, which differs from OBA only in budget search and allocation.
3) **TREE**: the TREE algorithm working on the swapping and purification tree in [14], [15].
4) **FER**: Fidelity-Guaranteed Entanglement Routing, the greedy purification scheduling used in Q-PATH and Q-LEAP [13], which greedily selects the link with the highest final fidelity increase.
5) **EPP**: an entanglement path preparation method [12] that greedily finds the link with the largest fidelity gradient.

Note that we do not include the algorithm in [21] for comparison because they do not consider the additional cost incurred by failed swappings and purifications (so their 'cost' is actually the budget in our design), and it is non-trivial to find the real cost for their algorithm.



(a) Cost



(b) Time

Fig. 3. Cost and time for OBA and GBA to achieve different F^*.

C. OBA and GBA

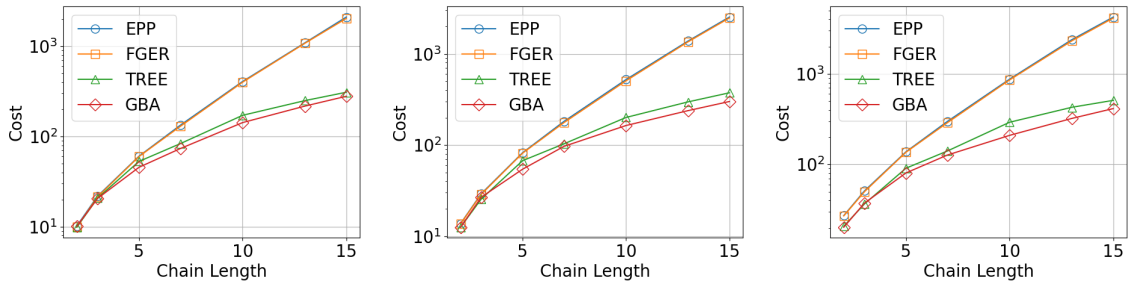
We first compare OBA and GBA to understand the optimality gap and computational overhead. Swapping probability q_s is set to 0.75 for this set of simulations . We repeat the same set of simulations for $q_s = 0.5$ and the results are similar.

From Fig. 3a, we observe that under all tested thresholds, GBA matches OBA in terms of the achieved minimum expected entanglement cost. Empirically, this indicates that the greedy allocation rule in Algorithm 1 aligns well with the structure of the Binary-system objective: allocating one additional unit of budget to the link that maximizes the marginal increase of E2E fidelity is sufficient to reach the globally optimal allocation in these instances.

From Fig. 3b, OBA consistently requires more runtime than GBA for the same target F^*, and the gap widens quickly as the path length increases. We stop OBA before $L = 7$ because for $F^* = 0.99$ it exceeds 100 seconds per run. Note that in these experiments, OBA only searches budget allocations starting from the budget returned by GBA (B_{GBA}) and then checks whether a feasible solution exists for $B_{\text{GBA}} - 1$, $B_{\text{GBA}} - 2$, etc. In all tested settings, OBA terminates immediately at B_{GBA} because it cannot find a feasible solution for $B_{\text{GBA}} - 1$. This suggests that GBA not only finds a near-optimal allocation, but also identifies the minimal feasible budget in practice.

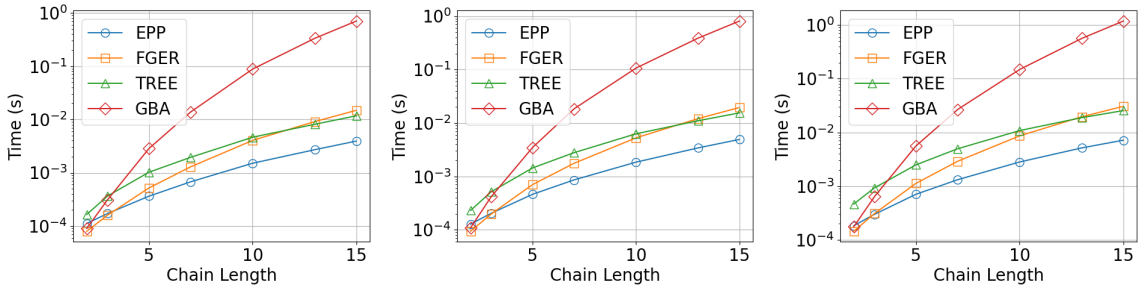
D. Comparison with Baselines

Since OBA is inefficient on longer chains, we compare GBA with the baselines. Fig. 4 reports the expected entanglement cost required to achieve the same F^*. We observe that GBA always consumes the least number of entanglements across all thresholds. Again, q_s is set to 0.75 for the results, but we also tested for 0.5 and the results are similar.



(a) $F^* = 0.9$ (b) $F^* = 0.95$ (c) $F^* = 0.99$

Fig. 4. Cost incurred by all baselines to achieve different E2E fidelity.



(a) $F^* = 0.9$ (b) $F^* = 0.95$ (c) $F^* = 0.99$

Fig. 5. Time consumed for all baselines to achieve different E2E fidelity.

The reason is twofold. First, as shown in [14], [15], TREE follows the optimal structural pattern (restricted tree), which already avoids inefficient hybrid placements of purification relative to swapping. This explains why TREE generally outperforms FER and EPP, which make purely local decisions based on per-link improvement (FER) or gradients (EPP). Second, even within the restricted-tree space, TREE does not explicitly optimize the optimal purification tree and the optimal swapping tree in the decomposed manner. GBA, by contrast, explicitly leverages the optimal substructures, which is exactly where the additional cost savings come from.

Fig. 5 reports runtime. GBA has the highest runtime among compared methods because each budget increment evaluates the marginal effect of allocating one extra unit to each link, and the corresponding E2E fidelity needs to be recomputed under the optimal substructures. Nevertheless, the optimization is performed offline; empirically, pre-solving a repeater chain within $\sim$1 second is practical for scheduling, especially given the significant reduction in entanglement consumption.

VI. Conclusion

This paper studies the Optimal Swapping and Purification Tree (OSPT) problem for quantum repeater chains under the Binary noise model with operation failures. Building on the structural insights in [14], [15], we develop an exact solution framework that decomposes OSPT into three components: optimal swapping tree construction, optimal purification tree construction, and global budget allocation. We further propose a Greedy Budget Allocation (GBA) algorithm that approximates the optimal budget allocation without exhaustive enumeration. Our evaluations show that GBA closely matches the globally optimal solution in practice while substantially reducing computational complexity. Compared with existing approaches, GBA achieves lower entanglement cost, particularly for long chains and stringent fidelity requirements.

References

[1] S. Wehner, D. Elkouss, and R. Hanson, "Quantum internet: A vision for the road ahead," *Science*, vol. 362, no. 6412, p. eaam9288, Oct. 2018.

[2] R. Van Meter, *Quantum networking*. John Wiley & Sons, 2014.

[3] A. Dahlberg, M. Skrzypczyk, T. Coopmans, L. Wubben, F. Rozpędek, M. Pompili, A. Stolk, P. Pawełczak, R. Knegjens, J. de Oliveira Filho *et al.*, "A link layer protocol for quantum networks," in *Proc. of ACM Special Interest Group on Data Communication (SIGCOMM)*, 2019.

[4] X. Wei, J. Liu, L. Fan, Y. Guo, Z. Han, and Y. Wang, "Optimal entanglement distribution problem in satellite-based quantum networks," *IEEE Network*, vol. 39, no. 1, pp. 97–103, 2025.

[5] Y. Zhang, Y. Gong, L. Fan, Y. Wang, Z. Han, and Y. Guo, "Efficient entanglement routing for satellite-aerial-terrestrial quantum networks," in *34th IEEE Int'l Conf. on Compu. Commu. and Net. (ICCCN)*, 2025.

[6] X. Wei, L. Fan, Y. Guo, Z. Han, and Y. Wang, "Entanglement from sky: Optimizing satellite-based entanglement distribution for quantum networks," *IEEE/ACM Transactions on Networking*, vol. 32, no. 6, pp. 5295–5309, 2024.

[7] J. Liu, X. Liu, X. Wei, and Y. Wang, "Topology design with resource allocation and entanglement distribution for quantum networks," in *21st Annual IEEE Int'l Conf. on Sensing, Commu., and Net. (SECON)*, 2024.

[8] X. Wei, L. Fan, Y. Guo, Z. Han, and Y. Wang, "Optimizing satellite-based entanglement distribution in quantum networks via quantum-assisted approaches," in *IEEE International Conf. on Quantum Communications, Networking, and Computing (QCNC)*, 2024.

[9] J. Liu, L. Fan, Y. Guo, Z. Han, and Y. Wang, "Co-design of network topology and qubit allocation for distributed quantum computing," in *IEEE Int'l Conf. on Quantum Commu., Net., & Compu. (QCNC)*, 2025.

[10] M. Caleffi, M. Amoretti, D. Ferrari, J. Illiano, A. Manzalini, and A. S. Cacciapuoti, "Distributed quantum computing: A survey," *Computer Networks*, vol. 254, p. 110672, 2024.

[11] S. Pouryousef, N. K. Panigrahy, and D. Towsley, "A quantum overlay network for efficient entanglement distribution," *arXiv preprint arXiv:2212.01694*, 2022.
[12] Y. Zhao, G. Zhao, and C. Qiao, "E2E fidelity aware routing and purification for throughput maximization in quantum networks," in *IEEE Conference on Computer Communications (INFOCOM)*, 2022.
[13] J. Li, M. Wang, K. Xue, R. Li, N. Yu, Q. Sun, and J. Lu, "Fidelity-guaranteed entanglement routing in quantum networks," *IEEE Transactions on Communications*, vol. 70, no. 10, pp. 6748–6763, 2022.
[14] J. Liu, X. Zhang, X. Wei, X. Liu, Y. Chen, H. Gao, and Y. Wang, "Joint swapping and purification with failures for entanglement distribution in quantum networks," in *IEEE/ACM 33rd Int'l Symp. on Quality of Service (IWQoS)*, 2025.
[15] J. Liu, X. Zhang, X. Wei, X. Liu, Y. Chen, H. Gao, and Y. Wang, "Swapping and purification scheme optimization for entanglement distribution in quantum networks," *IEEE Transactions on Networking*, 2026.
[16] W. Dai, T. Peng, and M. Z. Win, "Optimal remote entanglement distribution," *IEEE J. on Selected Areas in Communi.*, vol. 38, no. 3, pp. 540–556, 2020.
[17] W. Dür, H.-J. Briegel, J. I. Cirac, and P. Zoller, "Quantum repeaters based on entanglement purification," *Physical Review A*, vol. 59, no. 1, p. 169, 1999.
[18] N. K. Panigrahy, T. Vasantam, D. Towsley, and L. Tassiulas, "On the capacity region of a quantum switch with entanglement purification," in *IEEE Conference on Computer Communications (INFOCOM)*, 2023.
[19] L. Chen and Z. Jia, "On optimum entanglement purification scheduling in quantum networks," *IEEE Journal on Selected Areas in Communications*, vol. 42, no. 7, pp. 1779–1792, 2024.
[20] Z. Jia and L. Chen, "From entanglement purification scheduling to fidelity-constrained entanglement routing," in *2024 IEEE 32nd International Conference on Network Protocols (ICNP)*, 2024.
[21] Z. Jia and L. Chen, "From entanglement purification scheduling to fidelity-constrained multi-flow routing," *arXiv preprint arXiv:2408.08243*, 2024.
[22] Y. Zeng, J. Zhang, J. Liu, Z. Liu, and Y. Yang, "Multi-entanglement routing design over quantum networks," in *IEEE Conference on Computer Communications (INFOCOM)*, 2022.
[23] A. Chang and G. Xue, "Order matters: On the impact of swapping order on an entanglement path in a quantum network," in *IEEE Conf. on Computer Communications Workshops (INFOCOM WKSHPS)*, 2022.
[24] J. Liu and Y. Wang, "Statistical modeling and latency optimization for entanglement routing in quantum networks," in *IEEE Int'l Conf. on Quantum Computing and Engineering (QCE)*, 2025.
[25] L. Yang, Y. Zhao, H. Xu, and C. Qiao, "Online entanglement routing in quantum networks," in *IEEE/ACM 30th International Symposium on Quality of Service (IWQoS)*, 2022.
[26] A. Farahbakhsh and C. Feng, "Opportunistic routing in quantum networks," in *IEEE Conf. on Computer Communi. (INFOCOM)*, 2022.
[27] H. Gu, Z. Li, R. Yu, X. Wang, F. Zhou, J. Liu, and G. Xue, "Fendi: Toward high-fidelity entanglement distribution in the quantum internet," *IEEE/ACM Trans. on Networking*, vol. 32, no. 6, pp. 5033–5048, 2024.
[28] M. Ghaderibaneh, C. Zhan, H. Gupta, and C. Ramakrishnan, "Efficient quantum network communication using optimized entanglement swapping trees," *IEEE Trans. on Quantum Eng.*, vol. 3, pp. 1–20, 2022.
[29] X. Fan, C. Zhan, H. Gupta and C. Ramakrishnan,"Optimized Distribution of Entanglement Graph States in Quantum Networks," in IEEE Trans. on Quantum Eng., vol. 6, pp. 1-17, 2025

APPENDIX

In this appendix, we derive a finite upper bound on the total entanglement budget required by an exact optimal solution in Theorem 1, thereby justifying the finite search range used in Section IV.

Lower Bound on the Budget: Consider a repeater chain π consisting of L links. To construct a connected swapping and purification tree that spans the entire chain, at least one entanglement must be available on each link. Therefore, any feasible strategy must satisfy

$$B \geq B_{\min} = L. \tag{9}$$

Per-Link Purification Requirement: Let $f_{\min} = \min_{e\in\pi} f_e(1)$ denote the minimum initial link-level fidelity before purification. Under the Binary model, repeated two-to-one purification increases fidelity monotonically. As shown in [20], [21], purifying a link from initial fidelity $f_{\min}$ to a target fidelity $f \in (0.5, 1)$ requires a purification budget b that satisfies

$$b \;\geq\; \left\lceil \log_{(\frac{1}{f_{\min}}-1)} \left(\frac{1}{f} - 1\right) \right\rceil. \tag{10}$$

This bound is conservative but sufficient for our purpose.

Target Per-Link Fidelity for End-to-End Requirement: We now derive a sufficient per-link fidelity f such that, after swapping over a chain of L links, the resulting E2E fidelity is no smaller than the target threshold F^*.

Under the Binary model, the associativity of swapping implies that the E2E fidelity after $L-1$ swapping operations over links with identical fidelity f is given by [21]

$$F_{\text{e2e}} = \frac{1}{2} + 2^{L-1}\left(f - \frac{1}{2}\right)^L. \tag{11}$$

To ensure $F_{\text{e2e}} \geq F^*$, it suffices to require

$$f \;\geq\; \frac{1}{2} + \frac{1}{2}\sqrt[L]{2F^* - 1}. \tag{12}$$

Combining (10) and (12) yields a conservative upper bound on the purification budget required for each link.

Bounding the Expected Cost Blow-Up: We now account for the impact of two operation failures on the expected cost.

a) Purification failures: Under the Binary model, the success probability of any two-to-one purification step is at least $1/2$ [17]. Therefore, purifying b entanglements into one output entanglement incurs an expected cost of at most $b\cdot 2^{b-1}$ link-level entanglements.

b) Swapping failures: Let $q_{\min}$ denote the minimum swapping success probability along the chain. Since $L-1$ swapping operations are required to connect L links, the expected cost is amplified by at most a factor of $q_{\min}^{-(L-1)}$.

Combining the above two factors, the expected total cost incurred by a strategy that allocates b entanglements to each of the L links is upper bounded by

$$C_{\max} \;\leq\; \frac{L\cdot b\cdot 2^{b-1}}{q_{\min}^{L-1}}. \tag{13}$$

Since an optimal solution cannot have a total budget larger than its expected cost, we obtain a valid upper bound on the total budget:

$$B_{\max} = \left\lceil \frac{L\cdot b\cdot 2^{b-1}}{q_{\min}^{L-1}} \right\rceil. \tag{14}$$

Together, the above arguments establish that any exact optimal solution must use a total budget

$$B^* \in [B_{\min}, B_{\max}],$$

where $B_{\min} = L$ and $B_{\max}$ is given above. This completes the proof of Theorem 1.