Quantum Internet Architecture: unlocking Quantum-Native Routing via Quantum Addressing

Marcello Caleffi *and* Angela Sara Cacciapuoti
INVITED PAPER

***Abstract*—The Quantum Internet introduces a fundamental shift in the network design, since its key objective is the distribution and manipulation of quantum entanglement, rather than the transmission of classical information. This shift breaks key classical Internet design principles, such as the end-to-end argument, due to the inherently stateful and non-local nature of entangled states that require coordinated in-network operations. Consequently, in this paper we propose a novel hierarchical Quantum Internet architecture centered around the concept of *entanglement-defined controller*, enabling scalable and efficient management of the aforementioned in-network operations. However, architecture alone is insufficient for network scalability, which requires a *quantum-native control plane* that fundamentally rethinks addressing and routing. Consequently, we propose a *quantum addressing scheme* that embraces the principles and quantum phenomena within the node identifiers. Built upon this addressing scheme, we also design a *quantum-native routing protocol* that exhibits scalable and compact routing tables, by efficiently operating over entanglement-aware topologies. Finally, we design a *quantum address splitting* functionality based on Schrödinger's oracles that generalizes classical match-and-forward logic to the quantum domain. Together, these contributions demonstrate, for the first time, the key advantages of quantum-by-design network functioning.**

***Index Terms*—Quantum Internet, Quantum Network Architecture, Network Architecture, Quantum Networking, Entanglement, Addressing, Quantum Addressing, SDN, Quantum Routing, quantum-native functionalities.**

I. INTRODUCTION

The Quantum Internet [1]–[6] promises unprecedented capabilities, including unconditionally secure communication, distributed quantum computing, and enhanced sensing [7]. At the core of these capabilities lies quantum entanglement, the fundamental communication resource of the Quantum Internet [3], which demands a radical architectural departure from the classical Internet design principles [8], [9].

Specifically, the end-to-end principle [10] breaks down in the Quantum Internet, since the network paradigm shifts from transmitting information to distributing and managing entangled states [1], [9]. Unlike classical bits, entanglement generation, distribution, storage and exploitation inherently require in-network operations and the maintenance of related information inside the network. To elaborate more, classical bits are inherently stateless: intermediate network nodes can process and forward them, without the need to retain any additional information or detail about the state of the end-to-end communication. In contrast, the temporal constraints imposed by decoherence, along with the inherent complex mechanisms underlying the generation and maintenance of quantum entanglement, necessitate that the network nodes retain *state information*. As a pivotal example, nodes must be aware of the residual coherence time of the stored entangled qubits (e-bits), for properly operating on them. Hence, e-bits are fundamentally *stateful* [9], directly contradicting the classical end-to-end argument proposed by Saltzer in [11], outlining the classical stateless design.

Furthermore, the *non-local nature* of quantum entanglement requires additional state information for its effective exploitation, beyond merely identifying which nodes initially shared the initial entangled states. Indeed, in EPR-based networks, entanglement can be swapped, by changing at run-time the identities of the entangled nodes. In more complex scenarios where multipartite entanglement is exploited, the above dynamism is further enriched. In fact, multipartite entanglement can be manipulated and reconfigured across subsets of network nodes, namely, across entire sub-networks. Accordingly, since entanglement is not information per-se but rather a communication resource [3], its value and utility extend well beyond the original source-destination pair or the original initiating sub-network. If left uncoordinated, these non-local effects can trigger the so-called amplification principle [12], where uncontrolled entanglement resources lead to routing ambiguities, resource inefficiencies, and ultimately network instability, thereby undermining scalability. Therefore, effective tracking and management of entanglement resources are essential for scalable quantum network architectures.

Building on the above considerations, to efficiently manage the in-network entanglement operations while at the same time preserving the *simplicity principle* that shaped classical Internet design, we propose a novel architectural framework, which integrates *quantum-native* functionalities at its core. Specifically, by extending the architectural vision pioneered in [8], this paper formalizes a two-tier Quantum Internet architecture as depicted in Fig. 1, which centralizes control of entanglement operations through an *Entanglement-Defined Controller (EDC)*. Analogous to the SDN controller in classical networks, the EDC coordinates in-network operations and supports scalable entanglement management, as further detailed in Sec. II.

However, while such a proposed architecture provides a foundational framework, it is not sufficient on its own to ensure

M. Caleffi and A.S. Cacciapuoti are with the www.QuantumInternet.it research group, University of Naples Federico II, Naples, 80125 Italy. E-mail: angelasara.cacciapuoti@unina.it, marcello.caleffi@unina.it.

This work has been funded by the European Union under the ERC grant QNattyNet, n.101169850. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

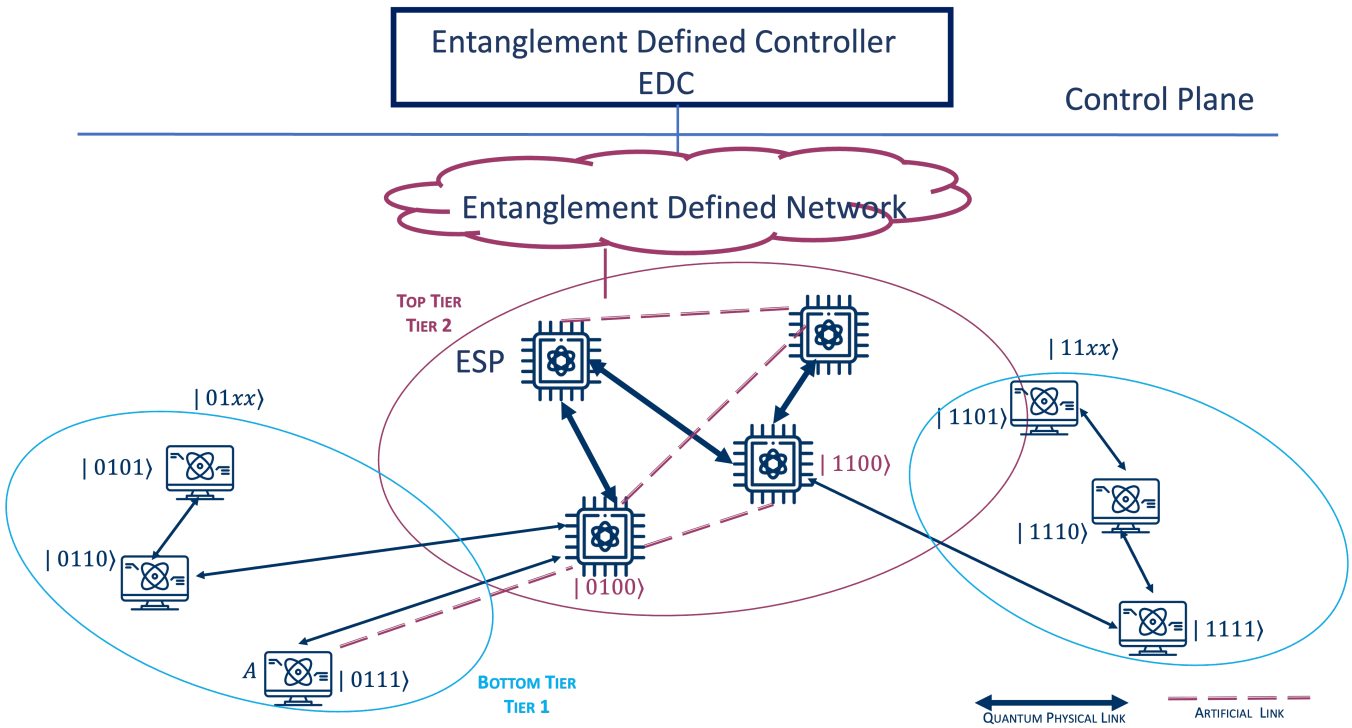


Fig. 1: Proposed architecture showing: (1) ESPs forming a virtual mesh via proactive entanglement sharing (dashed lines), (2) EDC in the control plane, and (3) end-user quantum nodes connected to the serving ESP.

scalability. Crucially, scalability depends on how *the control plane* is designed [3]. Specifically, the control plane must embrace quantum principles and phenomena to manage entanglement dynamics effectively. The rationale behind this claim stems again from the aforementioned non-locality of quantum entanglement: entanglement proximity cannot be confined to physical locality or restricted to fixed neighborhoods. This leads to a profound implication: *entanglement redefines the very same concept of topological neighborhood* [8], [13], [14]. As a result, a conventional control plane, built upon classical assumptions of locality and IP-like addressing, cannot timely track, respond to, or propagate entanglement state changes across the network.

In other words, without a fundamental rethinking of network addressing and control mechanisms, such a control plane will become a bottleneck to scalability. This issue is not entirely unfamiliar: even in classical networks where entanglement is absent, it has been shown that the number of control messages per topology change, namely, the updating communication overhead, cannot scale better than linearly on Internet-like topologies [15]. In the quantum setting, the challenge is even more severe due to the intrinsic statefulness and fragility of entanglement.

To address this, we complement the proposed architecture with a *quantum-native* routing protocol, built upon a novel addressing scheme that leverages quantum principles directly within the node identifiers. It is indeed the here-designed addressing scheme that makes possible to operate over entanglement-aware topologies with compact, dynamic, and scalable routing tables. More in detail, we design a quantum addressing scheme that leverages quantum superposition. This enables a paradigmatic departure from the underlying assumption beyond classical IP addressing: *a sequence of (32/128) bits encoding a single network address*, reflecting the node location within the physical network topology. In fact, a sequence of (say n) qubits can encode a single node identity, i.e., a single quantum network address, but it can also encode a superposition of node identities, i.e., a superposition of multiple quantum states with each state denoting a network address. Thus, a sequence of n qubits can represent a set of quantum nodes. Crucially and completely differently from classical IP subnet addresses, a quantum network address can represent a set of quantum nodes regardless of their physical position within the quantum network.

The rationale for such a *quantization* of the network addressing functionality is that superposed network addresses are key to *achieve scalable routing tables*. This is detailed in Sec. III.

In this light, it becomes crucial to allow a node to extract and act on individual node identities within a superposed address. Consequently, we design a *quantum addressing splitting functionality* in Sec. IV, which generalizes for the quantum domain the classical match-and-forward paradigm of classical networks. The proposed quantum addressing splitting has been conceived by modifying Grover's quantum search algorithm in order to tailor quantum-superposed data structures. In our proposal, the oracle is coherently controlled by quantum sub-network addresses, so that the presence or the absence of phase inversion becomes a controllable quantum degree of freedom.

Contributions. To sum up, the contributions of this paper are fourfold:

- the design of an architectural framework to manage in-network entanglement operations efficiently, centered around the Entanglement-Defined Controller (EDC);
- the design of a *quantum addressing scheme*, which encodes quantum properties and behaviors directly into node identifiers;
- the design of a *quantum-native* routing protocol that fully exploits the quantum nature of the addressing scheme to achieve scalable and compact routing tables;
- the proposal of a *quantum addressing splitting functionality* that extends classical forwarding operations to quantum-superposed identifiers, through the design of a Schrödinger's oracle for Grover's search algorithm, coherently controlled by quantum addresses.

To the best of our knowledge, this is the first work demonstrating with concrete architectural and protocol design the advantages of the quantum-native functioning of the network, that was previously only anticipated at a conceptual level in [8].

The remaining part of the paper is organized as follows. In Sec. II, we detail the two-tier architecture, by clarifying and substantiating some key deviations from the classical Internet design principles. In Sec. III, we design the quantum-native routing protocol, built upon the quantum addressing scheme. In Sec. IV, we propose Quantum Addressing Splitting functionality. Finally, in Sec. V we discuss some key aspects of the proposal along with future directions.

II. Quantum Internet Architecture

In this section, we propose some architectural principles for the Quantum Internet design, which should not be interpreted as dogmas but rather as pragmatic guidelines and criteria for harvesting the unique properties of quantum entanglement. Our design perspective departs from the classical Internet, yet it aligns with its most long-standing insight described in RFC 1958 [10]:

> "*the principle of constant change is perhaps the only principle of the Internet that should survive indefinitely*",

arguably the most valuable lesson from the classical Internet evolution.

A. Internet Architecture in a nutshell

By oversimplifying, in the classical Internet, Internet Service Providers (ISPs) form a multi-tiered hierarchy where lower-tier ISPs connect to higher-tier ones, creating a mesh of packet-switched networks, that provide end-to-end connectivity to end users. This hierarchy design has been the result of a long and largely unforeseen evolution, where a complex interplay of technical and economic factors played a crucial role.

In this context, the *end-to-end principle* [10] emerged as a key architectural tenet. Accordingly, end-to-end protocol design should not rely on state maintenance inside the network, i.e., on information about the state of the end-to-end communication. Rather, such a state should be maintained only at the end points, so that the state can be destroyed only if the end point itself breaks [12]. The overall effect[1] of the *end-to-end principle*, coupled with the *simplicity principle* [10], resulted in the classical Internet architecture, where intelligence is localized at the network edges rather than hidden inside the core network. In other words, Internet has smart edges where applications and operating systems reside and provide complex communication functionalities, and a simple core, consisting of stateless packet-forwarding engines and a control plane offering mainly best-effort datagram delivery [12].

B. The need for Entanglement-Packet Switching

The aforementioned design principles – centered around stateless routing, best-effort delivery, and end-to-end principle – are fundamentally inadequate for the Quantum Internet, which requires a radical departure from classical networking paradigms. This shift fundamentally transforms the core network operations, as detailed in the following.

First, a preliminary consideration must be introduced: informational qubits cannot be simply adapted to packet-switching paradigm, due to the fundamental constraints imposed by the no-cloning theorem and the quantum measurement postulate. In fact, unknown qubits cannot be perfectly copied or amplified, and any measurement irreversibly alters their state. This prohibits the adoption of conventional store-and-forward packet-switching paradigm, and it makes traditional best-effort delivery, where packets may be lost and retransmitted, inherently incompatible with quantum information transfer.

Entanglement distribution circumvents these limitations, by establishing quantum correlations as the fundamental network resource, allowing us to move beyond the direct transmission of informational qubits [1], [2]. Indeed, since entanglement is a communication resource rather than information itself, it is not constrained by the no-cloning theorem [9]. And, through the quantum teleportation protocol [2], pre-shared entanglement enables reliable quantum information transfer, without the need of physically transmitting an information carrier.

From the above, it follows that the Quantum Internet must embrace a fundamentally new paradigm:

> *ebits serve as the basic network "packets", carrying quantum correlations across network nodes. We term this paradigm as entanglement-packet switching.*

This transition is not merely an optimization or a modification of the current packet-switching paradigm, but a

[1] It is worthwhile to observe that this observation is not intended to imply a strict causal relationships between design principles and historical evolution. In fact, as authoritatively described in [16], the architectural foundations of the Internet were profoundly shaped by the combination of specific sets of priorities and a military context, characterizing the early stages of ARPANET development. For example, [16] explicitly identifies "survivability" as a design objective, directly emerging from the original military context in which the network was conceived. This requirement for survivability is widely recognized as the most essential rationale behind the end-to-end principle ("fate-sharing" concept in [16]).

Classical Internet	Quantum Internet
complexity located at the network edges	complexity concentrated in the core network
stateless core network	stateful core network
end-to-end protocol design	network-mediated protocol design

TABLE I: Classical vs Quantum Internet architecture: concise comparative overview.

fundamental departure, imposed by the unique constraints of quantum mechanics:

> *While the goal of the classical packet-switching paradigm is to determine the best next-hops toward a set of nodes (routing) and to forward packets through these next-hops from source to destination, entanglement-packet switching aims to distribute and manipulate entanglement among quantum nodes, ultimately entangling the source and destination regardless of their physical location.*

Accordingly, this paradigm shift enables the following two key features.

- *Decoupled design:* entanglement-packet switching fully *decouples* the quantum information transfer from the physical transmission of information carriers, overcoming the limitations imposed by the no-cloning theorem and the quantum measurement postulate on the direct informational-qubit transmission.
- *Scalability and compatibility:* entanglement-packet switching *retains* the flexibility and inherent scalability of packet-switching architectures, while ensuring backward compatibility with the classical Internet [17]–[19]. Hence, the approach is at the same time forward-looking and backward-compatible.

Given this fundamental transition towards entanglement-packet switching, we can now better digest the rationale for the incompatibility between the end-to-end principle (as said, substratum of the classical Internet design) and the Quantum Internet, highlighted in Sec. I. This breakdown stems directly from the unique properties of quantum entanglement.

Indeed, as described in Sec. I, the complex, challenging stateful nature of entanglement along with its non-local effects call for in-network operations and persistent state awareness across all phases of the entanglement lifecycle – from generation through distribution to storage and final utilization – by directly contradicting Saltzer's end-to-end argument.

Accordingly, the Quantum Internet demands a complete inversion of the classical design philosophy, as summarized in Table I.

Remark 1. Indeed, when the above considerations are combined with the sophisticated and resource-intensive setups required by the state-of-the-art hardware, it becomes both practical and efficient to advocate for concentrating the complexity inside the core network, while leaving the edges of the network simpler.

C. The Two-Tier Hierarchy

Building on the above perspective, we propose a two-tier Quantum Internet architecture, by distinguishing between *entanglement service providers (ESPs)* and *quantum edge nodes*, as represented in Fig. 1.

- *Top Tier*: ESPs act as the highest tier, referred to also as tier-2, analogous to ISPs in the classical Internet. They form the "entangled-backbone"[2], by providing end-to-end entanglement connectivity to the lowest tier, via proactively maintenance of entangled resources. The EPSs are interconnected via long-distance quantum links, such as optical fibers and they are equipped with the sophisticated and resource-intensive infrastructure required for entanglement generation and distribution.
- *Bottom Tier*: edge quantum nodes, which include quantum processors, sensors, cryptographic devices, acting as the edge tier (referred to also as tier-1) consuming entanglement resources to fulfill the quantum applications needs. They are mainly connected to the nearest ESP via short-range quantum links.

Remark 2. The two-level hierarchy introduced so far is not intended to represent a definitive architecture for the Quantum Internet, which is still in its early stage conceptualization[3]. Rather, it serves as a reference model for capturing the distinguishing features of quantum entanglement in contrast to classical information.

In the remaining part of this manuscript we focus on the design of the *top tier*, given its pivotal role in enabling large-scale deployment of the Quantum Internet.

To this aim, we must first recall a key point from a networking perspective. Specifically, it has been extensively shown in literature that entanglement, once shared, enables a new and richer form of connectivity, with no-counterpart in classical networks. The activated entanglement-proximity gives rise to an overlay topology, often referred to as *artificial topology*, established upon the physical topology [9], [13], [14], [20]. And differently from the physical topology, the entanglement-enabled overlay is inherently dynamic[4], since entanglement can be locally manipulated while still producing global coverage effects, due to its inherently non-local nature. Thus, the artificial topology can be reconfigured on-the-fly via local quantum operations (with entanglement swapping the archetypal example), by enabling global connectivity changes without modifying physical links. This adaptability is essential to accommodate evolving end-to-end entanglement requests, reduce latency – key in scenarios very sensitive to decoher-

[2] It is worth emphasizing that, in the classical Internet, the terms "core network" and "backbone network" are often used interchangeably, despite a subtle, yet important, distinction between them. In this paper, we adopt the same interchangeable usage in the context of the Quantum Internet, where these terms are borrowed primarily by analogy. Indeed, given that the Quantum Internet remains in its infancy and early conceptual stages, a formal distinction between the two has yet to emerge or become necessary.

[3] It is likely that the Quantum Internet will evolve into a multi-level hierarchy of increasing complexity, as happened to the classical Internet.

[4] We refer the reader to [8], [9] for an in-depth discussion about the artificial topology dynamics, which differ profoundly from the ones induced by node mobility in classical networks.

ence – and mitigate the limitations imposed by the physical topology [14].

In this perspective, we envision the ESPs to proactively establish and maintain entanglement resources among each others, by exploiting the underlying physical network topology. The activated artificial topology serves as the functional equivalent of the physical mesh network interconnecting ISPs. Crucially, this topology is continuously refreshed and reconfigured to reflect the current status and availability of entangled resources, thereby enhancing the overall reliability, adaptability, and robustness of the network. Moreover, the proactive nature of the strategy allows to face with the non-persistence of entangled resources, since the Quantum Internet must operate under the constraint of entanglement depletion, upon use. The overall result is a dynamic artificial mesh, where links are continually created and consumed, in response to quantum application demands.

D. Entanglement-Defined Controller

To efficiently manage in-network operations and state maintenance while addressing scalability, an Entanglement-Defined Controller (EDC) orchestrates the entanglement resources among the ESPs, by mirroring the role of a Software-Defined Networking (SDN) controller in classical architectures. The EDC oversees three main crucial functions:

- *Reconfiguration:* the dynamic management and reconfiguration of entangled resources among ESPs;
- *Monitoring:* the monitoring of the fidelity and availability of entanglement resources across ESPs;
- *Policy enforcement:* the enforcement of global policies for routing, resource allocation, and entanglement loss recovery.

By centralizing the control logic within the EDC, we lay the foundation for a scalable, adaptable, and programmable Quantum Internet architecture, necessary to manage quantum entanglement as a dynamic, non-persistent network resource. Indeed, the EDC[5] enables coordinated management of entangled resources across the network, thereby supporting the inherent network-mediated nature of quantum communication protocols. Moreover, the EDC enables a dynamic response to evolving application-demands and to the intrinsic volatility of entangled resources. It is important to highlight that, for maintaining the overlay topology, we do not force any EDC to have a persistent global network knowledge. In fact, in Sec. IV, we design a quantum-native routing protocol that operates effectively under limited or local network visibility.

Stemming from the above, we propose an architecture with a clear distinction between *quantum control* and *quantum data plan*, as summarized in Table II, where the EDC implements control plane functionalities, while the ESPs implement data plane functionalities.

As for the control plane, it is worthwhile to note that state-of-the-art quantum routing literature [21] has designed so far control functions, by exploiting only classical communications [22]. Conversely, we advocate for *quantum-native control functionalities*. More in detail, we design a quantum control plane where routing information is encoded into quantum states by leveraging the quantum addressing scheme, designed in Sec. II-F. These “quantum-encoded routes” are then stored within quantum routing tables and manipulated through quantum operations for the forwarding. This approach allows scalable and efficient path selection, without requiring persistent global network knowledge, as shown in Sec. III.

In this way, *we scale the network to a quantum-native functioning*.

As for the quantum data plane, it generalizes to the quantum domain the classical forwarding functionality, as highlighted in [22]. However, for accounting for the quantum-native functioning, key extensions beyond [22] must be introduced.

Specifically, in classical networks, the forwarding logic follows a match-and-forward paradigm, where the destination address is extracted from the packet header and matched against the routing table. In our architecture, this paradigm is extended as follows:

- *Entanglement-packet forwarding via quantum addressing:* ESP identifiers are encoded into quantum states via quantum addresses, as described in the next subsection. Thus, *forwarding decisions* require the presence of a quantum header in the packet as shown in Fig. 2, carrying the quantum equivalent of the source-destination address. To extract and manipulate this type of information, appropriate quantum operations are applied to the quantum addresses stored in the routing tables, by enabling path selection without classical header parsing. This mechanism is detailed in Secs. III and IV.
- *End-to-end entanglement establishment:* The data plane actively supports the establishment of end-to-end (E2E) entanglement across ESPs for the bottom-tier demands. This involves generating elementary hop-by-hop entangled links, which may not yet exist at the time of the request, and performing quantum operations, such as entanglement swapping and purification. Consequently, quantum packets, which carry ebits, are propagated through sequences of quantum operations, rather than merely through transmission over physical links.
- *Artificial topology maintenance:* The maintenance of the artificial topology among the ESPs requires the data plane to continuously regenerate virtual links depleted during the forwarding. This is achieved through elementary entanglement-link generation, quantum operations and refreshing of entangled resources.

This separation between data and control plane ensures scalability while facing with the ephemeral nature of quantum entanglement. For instance, a bottom-tier request is fulfilled, by manipulating and hence by reconfiguring the overlay among ESPs (e.g., stitching virtual links via swaps), while the control plane concurrently instructs the data plane to repair the overlay, by redistributing elementary entanglement. This maintains the topological robustness. Thus, while the data plane consumes pre-established entanglement for applications (e.g.,

[5]Please refer to Sec. V, in which we provide further details on the EDC, by including how the proposed architecture can support multiple, potentially federated EDCs rather than relying on a single controller.

network feature	classical Internet	Quantum Internet
resource persistence	communication links are permanent, i.e., their dynamics largely exceed the forwarding dynamics	entanglement is ephemeral and depleted upon use
control plane	populates routing tables with best hops toward destinations	encodes routing information via quantum superposition exploiting the quantum addressing scheme and orchestrates entanglement resources so that they are efficiently distributed
data plane	packet forwarding	quantum operations and entanglement-packet forwarding

TABLE II: Control and data plane: classical Internet vs the proposed Quantum Internet architecture.

teleportation), the control plane continuously refills depleted resources.

E. Quantum Addressing Scheme

As aforementioned, we scale the network to a quantum-native functioning, by designing control plane functionalities via quantum state manipulations. To this aim, we embrace quantumness within the node addresses [8]. It is worthwhile to note that the quantum addressing is not a substitute of the classical network addressing. Indeed, each network node must be equipped with two types of addresses: i) a classical address, such as an IP address, which is mandatory for the classical communications and signaling required by any quantum communication protocol; ii) and a quantum address, which facilitates efficient and scalable control functions.

This dual-addressing framework ensures backward compatibility with the classical infrastructure, while unlocking quantum advantages in control-plane scalability and entanglement orchestration, as proved in Sec. III.

In assigning the quantum addresses, we hybridize classical hierarchical principles with quantum features. Specifically, each node in the network, regardless of being either an EPS or a tier-1 node, is assigned a quantum address represented by a computational basis state of an N-qubit system. Thus, the addresses are orthogonal quantum states, ensuring so perfect distinguishability. We account for the two-tier hierarchy introduced so far by imposing tier-1 addresses to share a common prefix with the address of the serving ESP. We envision that this addressing scheme enables efficient routing and entanglement management across the quantum network.

In Fig. 1, we provide a pictorial representation of a toy model for the above scheme, by considering a network of $n = 16$ nodes with 4 ESPs. Each ESP serves up to 3 tier-1 nodes, and their corresponding address prefix determining the clustering are:

$$|00xx\rangle\,, |01xx\rangle\,, |10xx\rangle\,, |11xx\rangle\,. \tag{1}$$

Formally, the quantum network is modeled as an undirected graph:

$$G = (V, E), \tag{2}$$

where:

- the vertex set V represents the $|V| = n$ quantum nodes, partitioned in two disjoint subsets: V_1, representing edge quantum (tier-1) nodes with $|V_1| = n_u$, and V_2, representing ESPs (tier-2 nodes), with $|V_2| = n_e$. Accordingly:

$$V = V_1 \cup V_2, \quad V_1 \cap V_2 = \emptyset, \tag{3}$$

with $n = n_u + n_e$.

- the edge set $E \subseteq V \times V$ denotes the set of links interconnecting the quantum nodes.

Definition 1 (**Quantum Address).** Let $\mathcal{B} \triangleq \left\{ |x\rangle \;\middle|\; x \in \{0,1\}^N \right\}$ denote the computational basis of the Hilbert space associated with an N-qubit system, where $N = \lceil \log_2 n \rceil$.
Let $\mathcal{E} \triangleq \{|v_1\rangle, \ldots, |v_{n_e}\rangle\} \subset \mathcal{B}$ be a set of n_e distinct computational basis states. Let define $p = \lceil \log_2 n_e \rceil$ as the prefix length. For any $x \in \mathcal{B}$, we define $\text{prefix}_p(x)$ as the string consisting of the first p bits of x, i.e.:

$$\text{prefix}_p : \{0,1\}^N \to \{0,1\}^p. \tag{4}$$

The elements of $\mathcal{E}$ are selected to satisfy the condition:

$$\text{prefix}_p(v_i) \neq \text{prefix}_p(v_j), \quad \text{for } i \neq j, \tag{5}$$

ensuring that each element in $\mathcal{E}$ is assigned a unique prefix. Each ESP in V_2 is uniquely identified by a *quantum address* $|v_i\rangle$, selected in $\mathcal{E}$:

$$|v_i\rangle \in \{|v_1\rangle, \ldots, |v_{n_e}\rangle\}\,. \tag{6}$$

In the following, we refer to an ESP as either v_i or $|v_i\rangle$, depending on the context.
Each edge (tier-1) quantum node in V_1 served by a given ESP $|v_i\rangle$ is assigned with a quantum address that shares the same k-prefix with $|v_i\rangle$, thereby preserving hierarchical coherence in the quantum address space. Formally, for each ESP in V_2 with address $|v_i\rangle \in \mathcal{E}$, we define its *serving cluster* as:

$$\mathcal{C}_{|v_i\rangle} \triangleq \left\{ |x\rangle \in \mathcal{B} \;\middle|\; \text{prefix}_p(x) = \text{prefix}_p(v_i) \right\}. \tag{7}$$

That is, each cluster contains all and only those computational basis states that share the first k bits with $|v_i\rangle$. By construction[6]

[6]By construction, each cluster has the same size equal to $2^{N-p} - 1$, by excluding the address of the serving ESP. To support heterogeneous cluster sizes while keeping the prefix length p fixed, one would need to define a maximum cluster size M_{max} and set the quantum address length as $N = p + \lceil \log_2 M_{\max} \rceil$. This approach inevitably results in address spaces with unused quantum addresses in smaller clusters. However, exploring such generalizations is beyond the scope of this paper, which instead focuses on showing the powerful setup allowed by quantum addressing schemes.

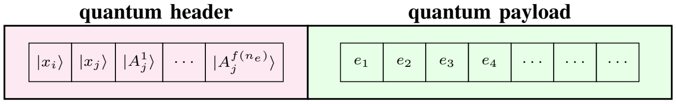


Fig. 2: Quantum packet structure. The header carries quantum information for the quantum routing logic, such as source-destination quantum network addresses $|x_i\rangle, |x_j\rangle$ and quantum superpositions $|A_j\rangle$ representing set of quantum nodes introduced in Sec. IV-B, while the payload carries entangled qubits (ebits) to be shared among network nodes.

the serving clusters are mutually disjoint and they constitute a partition of the computational basis set $\mathcal{B}$, i.e.:

$$\mathcal{B} = \cup_{i=1}^{n_e} \mathcal{C}_{|v_i\rangle} \quad \wedge \quad \mathcal{C}_{|v_i\rangle} \cap \mathcal{C}_{|v_j\rangle} = \emptyset, \; \forall \, i \neq j \in V_2. \tag{8}$$

F. Quantum Packet Structure

The proposed quantum addressing scheme requires the definition of a *quantum packet structure*, represented in Fig. 2. Differently from existing proposals [17], [23], our proposal consists of both a quantum header and a quantum payload, thus enabling a fully quantum-native processing. Specifically, the quantum packet is organized into:

- a *quantum header*, carrying quantum addresses[7]. This header enables the network nodes to interpret and forward entangled packets according to the quantum routing logic, based on the hierarchical address structure;
- and a *quantum payload*, carrying entanglement as multiple ebits, intended for distribution to the destination node(s).

As aforementioned, this design allows the quantum packet to be processed in a fully quantum-native manner, without fallback to classical parsing mechanisms. This facilitates scalable quantum routing, as analyzed in the following section. Furthermore, since in this paper we focus on tier-2 nodes, we restrict our attention to packets carrying ESP-related information. Accordingly, the source and destination addresses $\{|x_\ell\rangle\}$ in Fig. 2 should be interpreted as $\{|v_\ell\rangle\}$ and $\{|A_j^l\rangle\}$ denote superposed quantum addresses representing set of nodes.

As a final remark, although a comprehensive treatment is beyond the scope of this work and left for future investigation, the proposed quantum packet structure can be generalized to the multipartite entanglement case. In such a scenario, we envision the header carrying source address and a variable-length list of destination addresses, while the payload carrying a multipartite entangled state, which is intended for simultaneous distribution across the specified destination nodes.

[7]As detailed in the next two sections, a quantum address can be the univocal identifier of a network node, which by design is an orthogonal basis states. But it can also be a superposition of quantum states constituting the identifiers of a set of nodes – as instance, the superposed address stored within the e-neighborhood entry of the routing table as shown in Fig. 4. As aforementioned, there may be also the need of sharing classical information, as instance, for classical signaling. Although quantum/classical coexistence is a key open problem, we envision that this can be achieved by multiplexing the quantum header/payload with classical header through quantum/classical multiplexing techniques [17], [23], [24].

III. Quantum-Native Routing

Here we design two different versions of the quantum-routing protocol tasked with proactively maintaining and utilizing the entanglement-activated overlay topology among the ESPs. Although these protocols differs in terms of overall control-plane complexity, they both leverage the quantum addressing scheme introduced in the previous section.

Before delving into the technical details of the proposal, we refer the reader to Box 1, which provides an outline of the overall routing logic for tier-1 nodes, which, as already described, delegate routing complexity to their serving ESPs.

A. Preliminaries

As pioneered in [8] and reflected in the previous sections, the *goal* of a quantum routing protocol for ESPs fundamentally diverges from classical routing paradigm. In contrast to the classical objective of determining physical routes toward destinations for packet forwarding, a native-quantum routing shifts the goal to:

> *proactively maintaining the overlay topology activated by the entanglement and ensuring end-to-end entanglement distribution, by managing and tracking the entangled resources* shared within the tier-2 network.

Achieving this goal requires a paradigmatic departure in the structure and semantics of routing tables. Specifically, routing entries no longer store next-hop interfaces toward destination addresses. Rather, they list the locally-available entangled qubits[8] at each node, together with the identities[9] of the ESPs sharing this entanglement.

This shift does not imply that physical topology knowledge is irrelevant. On the contrary, such knowledge remains essential for the generation and distribution of link-level entanglement, needed to replenish or restore depleted entangled resources. In this perspective, by leveraging the dual-addressing framework described above in which classical addresses are augmented with quantum addresses, each ESP maintains two distinct routing tables:

- a classical routing table, used to manage physical topological information related to the classical communication infrastructure and populated using classical routing protocols;
- a quantum routing table, used to manage the entangled overlay topology information and populated by the EDC.

The focus of the remaining part of the manuscript is on the design of such quantum routing tables and on their central role in enabling scalable, quantum-native forwarding decisions within the tier-2 network. To this aim, we first collect some definitions used in the following.

Definition 2 (**Link Entanglement Metric**)**.** The "cost" associated with generating and distributing *link entanglement* over

[8]As instance, as pointers to the "addresses" of the communication qubits [3] storing the entangled state.

[9]Indeed, any quantum communication protocol requires a tight cooperation between the network nodes storing the entangled qubits for being able to exploit the quantum correlation provided by entanglement, and thus nodes must be aware of each other identities.

Box 1: Outlook for Bottom Tier Nodes

As aforementioned, by hybridizing classical hierarchical design principles with quantum features in the addressing scheme, the proposed architecture enables a *tier-aware differentiation of routing information* stored locally at each node. Specifically, the structure of the quantum addresses, based on a shared prefix between tier-1 nodes and their serving ESP, allows tier-1 nodes to operate with minimal routing intelligence, effectively delegating forwarding decisions and complexity to tier-2 nodes. As a result, tier-1 nodes require only basic quantum-packet forwarding logic based on the generalization to the quantum domain of the classical prefix-matching mechanism. In contrast, ESPs maintain richer and more dynamic routing tables to manage inter-ESP overlay, activated by entanglement, as analyzed in Sec. III. These tables are used to execute entanglement operations/manipulations, such as entanglement swapping, enabling end-to-end distribution of ebits across the network. This tiered routing abstraction significantly enhances scalability, by localizing complexity at tier-2 nodes. Indeed, it reduces both memory and processing overhead on tier-1 nodes, without compromising global routing capabilities.

More in detail, by leveraging the proposed quantum addressing scheme, end-to-end entanglement requests initiated by tier-1 nodes are first routed to their respective serving ESPs. In the ideal case, where the ESP overlay topology activated by the entanglement is fully connected, the serving ESP can directly perform entanglement swapping on the ebit(s) within the quantum payload, to reach the destination's serving ESP, thereby enabling efficient end-to-end entanglement distribution.

In more general and realistic scenarios, where the ESP overlay is not fully connected, the serving ESP must inspect its quantum routing table to assess whether pre-established entanglement exists with the destination's serving ESP. If such a "link" is available, the ESP can immediately perform entanglement swapping to enable end-to-end entanglement between the tier-1 nodes. Otherwise, the ESP must identify an appropriate next-hop ESP, selected according to the adopted routing metric, that brings the quantum packet closer to the destination. Thus, the ebit(s) in the quantum packet payload is eventually forwarded through an entangled path, i.e., through a sequence of intermediate ESPs, until an ESP already sharing entanglement with the destination ESP is found. This interplay between packet forwarding and swapping mechanism, ruled by the quantum addressing scheme and the selected routing metric, is detailed in Sec. III.

While several alternative design choices for tier-1 nodes are possible, including the option to handle their requests only with classical communications[a], this paper focuses on the core routing functionality of ESPs, which ensures end-to-end entanglement distribution across the network. Accordingly, the exploration of alternative tier-1 design choices lies beyond the scope of this work.

[a] In such a case the proposed quantum packet structure would apply only to ESPs.

a quantum link $(i,j) \in V_2 \times V_2$ is modeled by a general *entangling-cost metric* $w(i,j)$. This metric is equipped with two operations [25]:

i) an order relation $<$, allowing pairwise comparison of link entanglement costs;
ii) a binary operation $\oplus$, enabling the composition of entanglement costs over multiple links.

In particular, the order operation $<$ allows us to express relative entangling difficulty across links. For instance:

$$w(i,k) < w(j,k), \tag{9}$$

implies that establishing link entanglement between ESP v_i and v_k is less costly than between v_j and v_k. The binary operation $\oplus$ extends the cost function from individual links to multi-hop entanglement paths. For instance:

$$w\big((i,k) \oplus (k,j)\big) \tag{10}$$

denotes the cost of distributing end-to-end entanglement between remote ESPs v_i and v_j, via an intermediate ESP v_k.

This abstraction allows us to support arbitrary entanglement cost models, which in turn reflect the chosen quantum routing metric [26]. In other words, with this abstraction, *we decouple the routing logic from any specific cost formulation*. Hence, the framework accommodates a broad class of quantum metrics, ranging from fidelity degradation to quantum memory usage or decoherence effects. Consequently, the entangling cost $w(\cdot,\cdot)$ encapsulates the link/route "quality" under the selected model, allowing the routing protocol to adapt to the specific characteristics and constraints of the underlying quantum network.

Remark 3. For notational and conceptual simplicity, we refer to routes with the lowest entangling cost as *shortest paths*. This is in line with the conventional terminology of classical routing, originally developed under the assumption that the "quality" of a communication path is determined by its hop count, with shorter paths being preferable.

We assume that $w(\cdot,\cdot)$ satisfies the axiomatic properties of a metric, i.e.:

$$\text{definiteness:} \quad w(i,i) = 0 \iff v_i = v_j \tag{11}$$

$$\text{non-negativity:} \quad w(i,j) \leq 0 \tag{12}$$

$$\text{symmetry:} \quad w(i,j) = w(j,i) \tag{13}$$

$$\text{triangle inequality:} \quad w(i,j) \leq w\big((i,k) \oplus (k,j)\big), \tag{14}$$

plus an additional property to ensure the metric being isotone, namely, being both left-isotone:

$$w(j,l) < w(j,k) \Longrightarrow w\big((i,j) \oplus (j,l)\big) < w\big((i,j) \oplus (j,k)\big), \tag{15}$$

and right-isotone:

$$w(l,j) < w(k,j) \Longrightarrow w\big((l,j) \oplus (j,i)\big) < w\big((k,j) \oplus (j,i)\big). \tag{16}$$

Remark 4. Isotonicity ensures that the relative ordering of entangling costs between two quantum links (or paths), sharing a common origin or destination, is preserved when both are extended by the same quantum link. Meanwhile, triangle inequality, also referred to as monotonicity in [25], implies that the entangling cost cannot decrease, when it is extended by a new quantum link. These properties are not only rationale for capturing the constraints[10] of entanglement distribution, but they are required for guaranteeing the convergence [26] of quantum routing protocols to optimal entanglement paths, without resorting to exhaustive enumeration of all possible routes [25].

Definition 3 (End-to-end Entangling Metric). Let us consider two pair of non-adjacent ESPs, say v_i and v_j and let R denote a set of ESPs forming a valid swapping path from v_i to v_j. $w_R(i,j)$ denotes the accumulated entangling cost along that specific path, using the $\oplus$ operator. Clearly, end-to-end entanglement between v_i and v_j can be achieved through multiple possible routes, and an effective routing protocol should select the route with the lowest entangling cost. Specifically, swapping should occur at the most favorable, accordingly to the selected metric, sequence of intermediate ESPs acting as quantum repeaters, i.e.:

$$w(i,j) = \min_R \{w_R(i,j)\}. \tag{17}$$

Stemming from Def. 3, we are now ready to formally define the *entangling stretch* introduced by a quantum routing protocol. Broadly speaking, the purpose of a quantum routing protocol is to discover and select, among the available ESPs, a suitable sequence of intermediate nodes that enables two remote ESPs to establish shared entanglement through entanglement swapping. This selection is driven by a predefined cost metric, which models the quality or difficulty of establishing entanglement along a link or path.

Let $\mathcal{QR}$ denote an arbitrary quantum routing protocol. And let us define with $R = \mathcal{QR}(i,j)$ the set of ESPs selected as repeaters by $\mathcal{QR}$ to establish end-to-end entanglement between the remote (in the overlay topology) ESPs v_i and v_j.

Definition 4 (Entangling Stretch). The entangling stretch induced by the quantum routing protocol $\mathcal{QR}$ between ESPs v_i and v_j is defined[11] as:

$$\mathcal{ES}(i,j) = \frac{w_R(i,j)}{w(i,j)}, \quad \text{with } R = \mathcal{QR}(i,j), \tag{18}$$

where $w_R(i,j)$ denotes the entangling cost incurred by the path selected by the routing protocol $\mathcal{QR}$, and $w(i,j)$ represents the optimal entangling cost in (17). Accordingly, the *overall entangling stretch* $\mathcal{ES}$ of the quantum routing protocol $\mathcal{QR}$ is defined as the worst-case stretch over all ESP pairs in the network:

$$\mathcal{ES} = \max_{v_i, v_j \in V_2} \{\mathcal{ES}(i,j)\}. \tag{19}$$

From Eq. (19), it follows that a quantum routing protocol is optimal if and only if its entangling stretch satisfies $\mathcal{ES} = 1$, since it always selects the lowest-cost entangled path for any ESP pair. Conversely, an entangling stretch strictly greater than 1 indicates that the protocol may route entanglement through suboptimal entangled paths, for at least some pairs of ESPs.

Remark 5. We emphasize that the term *stretch*, used in Def. 4, stems from classical routing terminology. In classical networks, the *path stretch* of a routing scheme is typically defined as the ratio between the actual path length followed by a packet (usually measured in number of hops) and the shortest possible path length[12] [15]. Here we adopt this classical terminology, by reinterpreting the concept: the stretch is defined as the ratio between the entanglement cost incurred by the routing protocol, based on the selected sequence of repeaters, and the minimum possible entanglement cost under optimal swapping sequence. This generalization preserves the intuition behind the term "stretch", while capturing the fundamentally different nature of cost in quantum networks, reflecting entanglement quality/complexity.

B. Design Principles

Design Principle 1 (Compact). Our goal is to design a quantum routing protocol with *sublinear quantum memory requirements*. This is achieved by defining: i) routing tables that, for each ESP, have at most[13] $\tilde{\mathcal{O}}(\sqrt{n_e})$ entries, and ii) *logarithmic* quantum network addresses, i.e., addresses that scales in size as $\mathcal{O}(\log n)$, with n denoting the number of ESPs within the quantum network.

The design principle 1 is very reasonable, since maintaining entanglement between every pair of ESPs would determine prohibitive memory requirements at each ESP, as well as significant network overhead in terms of both quantum and classical resources for entanglement generation and distribution. Consequently, we adopt the strategy whereby each ESP shares entanglement with only a subset of the other ESPs. And the size of this subset scales sublinear with the total number of ESPs, i.e., $\sqrt{n_e} \log n_e$ or equivalently[13] $\tilde{\mathcal{O}}(\sqrt{n_e})$.

[10] As an example, the triangle inequality correctly models the preference for direct link entanglement over entanglement swapping at an intermediate node.

[11] For simplicity, we omit the explicit dependence of $\mathcal{ES}(i,j)$ on $\mathcal{QR}$, i.e., $\mathcal{ES}(i,j) \stackrel{\triangle}{=} \mathcal{ES}_{\mathcal{QR}}(i,j)$.

[12] In this context, the research area in classical networks named as *compact routing* aims at designing scalable routing schemes able to guarantee a path stretch upper-bounded by a *constant factor* c independent from the network size n, such as $c = 5$ or even $c = 3$ as in [27]. In this research area, both the table-scaling and the path-stretch bounds are derived with a worst-case analysis, i.e., the routing table of each node scales sub-linearly in n and the path stretch is at most c among all the source-destination paths

[13] In this manuscript, we adopt the computer science notation for classifying routing protocols time/space complexity. Accordingly, $\mathcal{O}(\cdot)$ (namely, big O) denotes the asymptotical growth rate of the number of communication qubits stored at each node as the network size grows, while $\tilde{\mathcal{O}}(\cdot)$ (namely, soft O) denotes the asymptotical growth rate when logarithmic factors are ignored.

ENTANGLING STRETCH UPPER BOUND

routing scheme	entangling cost metric			section
	arbitrary	additive	min	
Partial-Anchor	$\frac{w\big(\oplus_5\,(i,j)\big)}{w(i,j)}$	5	1	Sec. III-C
Full-Anchor	$\frac{w\big(\oplus_3\,(i,j)\big)}{w(i,j)}$	3	1	Sec. III-D

TABLE III: Entangling stretch vs. quantum routing scheme. With a "Partial-Anchor" scheme, only a subset of ESPs is responsible for proactively creating long-range (i.e., high-cost) artificial links as shown in Fig. 3, yet at the price of a slight increased entangling stretch.

Clearly, the choice of which ESP belongs to such a fraction crucially dictates the overall quantum routing performance. This aspect is thoroughly analyzed in Secs. III-C-Sec. III-D, where we introduce and compare two distinct design strategies, offering different trade-offs in terms of complexity and routing efficiency.

Design Principle 2 (**Constant Stretch**)**.** Our objective is to design a quantum routing protocol that guarantees an *overall entangling stretch* $\mathcal{ES}$ upper-bounded by a constant factor for an arbitrary isotonic entangling cost metric:

$$\mathcal{ES} \leq c \triangleq \frac{w_{\tilde{R}}(i,j)}{w(i,j)}, \tag{20}$$

where v_i and v_j denote the ESPs exhibiting the worst-case entangling stretch, i.e., the pair of nodes satisfying (19), and $w_{\tilde{R}}(i,j)$ denotes an upper bound on the cost of entangling v_i and v_j through the entangled path discovered by the protocol. The exact expression of the constant factor depends on the specific design choices.

Remark 6. In Secs. III-C and III-D, we show that our protocol is able to assure a constant stretch of 5 and 3, respectively, i.e.:

$$\mathcal{ES} \leq c = 3 \vee 5. \tag{21}$$

These results are obtained under the assumption that the entangling cost metric is *additive*, namely, when the binary operation $\oplus$ in (4) denotes the standard addition: $w\big((i,k) \oplus (k,j)\big) \triangleq w(i,k) + w(k,j)$. It is worthwhile to note that routing protocol metrics (classical or quantum) are mainly additive, with delay and the quantum version of the hop count as representative examples. Other popular metrics could be either multiplicative, e.g., packet loss, or concave metrics, e.g, bandwidth. As for the former class, multiplicative metrics can be straightforwardly converted into additive metrics via logarithmic scaling. Differently, for the latter class – namely, whenever the binary operation $\oplus$ in (4) denotes the minimum operator [28] and $w\big((i,k) \oplus w(k,j)\big) \triangleq \min\{w(i,k), w(k,j)\}$ – the constant factor is unitary:

$$\mathcal{ES} = 1. \tag{22}$$

From the above, it is evident that restricting the attention on additive entangling-cost metric $w(\cdot,\cdot)$ is not a limitation. On the contrary, it represents a conservative design choice: the concave case trivially achieves optimality, while additive metrics pose a more realistic and challenging setting for protocol design.

Although the next design principle has been already introduced and justified in the previous section, here we report it once again for the sake of completeness.

Design Principle 3 (**Quantum Addressing**)**.** Our objective is to design a quantum routing protocol that takes advantage of quantum principles and phenomena via quantum addressing.

Summarizing, regardless of the particulars of the adopted entangling cost metric $w(\cdot,\cdot)$, which determines the final form of (19), our design principles enforce a quantum routing protocol characterized by the following key features and summarized in Table IV:

i) *Sublinear size of the routing tables*: each ESP maintains entanglement with an ESP set of size $\tilde{\mathcal{O}}\left(\sqrt{n_e}\right)$, where n_e is the total number of ESPs. This implies a sublinear quantum memory overhead.
ii) *Logarithmic quantum address length*: the quantum address length scales as $\mathcal{O}\left(\log n\right)$, with n denoting the total number of nodes in the network.
iii) *Constant entangling stretch*: the quantum communication overhead, measured in terms of entangling stretch, and quantifying the entanglement "wasted" for not having a fully-connected overlay mesh among the ESPs, is upper bounded by a constant, independent of the network size.

C. Partial-Anchor Scheme

Here, as aforementioned, we design the first version of the quantum-routing protocol, in which a small subset of ESPs, referred to as *anchor ESPs*, is responsible for proactively establishing and maintaining long-range (i.e., high-cost) entangled links. The remaining ESPs, constituting the majority, are instead in charge of proactively creating and maintaining short-range (i.e., low-cost) artificial links.

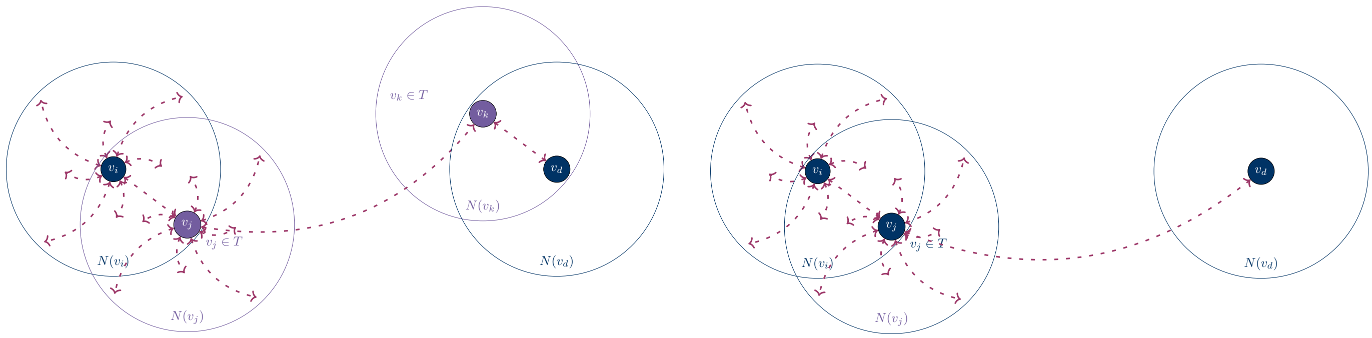
As proved in the following, this design yields an overlay topology, where any ESP is at most 3-hop-entanglement away from any other ESP (see Remark 8). This worst-case guarantee stems from the strategic connectivity established through the anchor ESPs, which serve as entanglement hubs within the overlay.

Conversely, when we relax the constraint of (very) limited number of anchor ESPs, as explored in Sec. III-D, the overlay topology is restructured to have every ESP 2-hop-entanglement distant from any other ESP (see Remark 12). Thus, we reduce the communication cost in terms of entangling stretch, at the price of increasing the maintenance complexity of the overlay. Remarkably, this enhancement in routing performance is achieved without affecting the asymptotic scaling of the routing table size.

The aforementioned trade-off between routing performance and control complexity, summarized in Table III, gives rise to two distinct protocol variants: the first, governed by the following last-but-not-least design principle, is referred to

metric	scaling	description
quantum memory overhead	$\tilde{\mathcal{O}}(\sqrt{n_e})$	size of the ESP Routing Table defining the number of entangled links maintained by each ESP node, with n_e denoting the total number of ESPs
quantum address length	$\mathcal{O}(\log n)$	number of qubits required to encode a quantum address, with n denoting the total number of network nodes
entangling stretch	$\mathcal{ES} \leq c$	maximum ratio between the actual and optimal entangling cost bounded by a constant c, independent of network size

TABLE IV: Key performance indicators of the proposed quantum-native routing protocol.



(a) Partial-Anchor Scheme: only the fraction of ESPs in T (depicted in purple) is responsible for proactively creating high-cost artificial links each others, whereas all the ESPs maintain low-cost artificial links with their e-neighborhood $N(\cdot)$.

(b) Full-Anchor Scheme: each ESP is responsible for proactively creating both low- and high-cost artificial links, the formers with their e-neighborhood $N(\cdot)$ and the latters with a specific subset of nodes.

Fig. 3: Schematic view of the two schemes underlying our compact quantum routing protocol: *Partial-Anchor* (presented in Sec. III-C) vs *Full-Anchor* (presented in Sec. III-D). Artificial links are denoted with red arrows, whereas circles around ESPs denote their e-neighborhood as defined in Def. 5.

as *partial-anchor scheme*; the second, discussed in the next section and formalized by Design Principle 5.B, is referred to as the *full-anchor scheme*.

Design Principle 5.A (Partial-Anchor Scheme). Our objective is to design a quantum routing protocol where most of the nodes proactively creates and maintains short-range (i.e., low-cost) artificial links, and only a fraction of nodes is responsible for proactively creating and maintaining long-range (i.e., high-cost) artificial links.

Accordingly, we begin by defining two sets of nodes: the *e-neighborhood* and the *anchor-nodes*.

Definition 5 (e-neihborhood). The e-neighborhood[14] $N(v_i)$ of the arbitrary ESP $v_i \in V_2$ denotes the set of the $k = |N(v_i)|$ nodes closest to v_i, according to the entangling-cost metric $w(\cdot,\cdot)$. The nodes in $N(v_i)$ are referred to as the *e-neighbors* of v_i.

Intuitively, the e-neighborhood of an ESP captures also localized information about the underlying physical topology, as inferred through the entangling-cost $w(\cdot,\cdot)$. In this sense, the concept of e-neighborhood inherently reflects the low-cost regime associated with the proactive creation and maintenance of entangled links. Conversely, only the nodes in the following set are requested to proactively create and maintain high-cost entangled links.

[14]We omit the explicit dependence on k for notational simplicity.

Definition 6 (Anchor Set). A subset $T \subseteq V_2$ of ESPs is defined as *anchor set*, and its elements are equivalently referred to as *anchor ESPs* or simply *anchors*.

From the above, it is evident that the two key parameters influencing the overall routing behavior are: i) the size of each ESP e-neighborhood, denoted by k, and ii) the cardinality of the anchor-node set $|T|$, as summarized in Tab. V. Building on this observation, we now propose a constructive method to set these parameters to induce the compact routing Property 1, which is subsequently exploited by the routing protocol.

Specifically, we leverage a result from classical combinatorial set theory concerning the covering set problem, originally introduced in [29], but in the formulation given in [30] as follows. Let P be a finite set of ℓ elements, and let $\mathcal{A} = \{A_1, \dots, A_h\}$ be a collection of subsets of P, each with cardinality $|A_i| = s$. A set $D \subseteq P$ is said to cover $\mathcal{A}$ if:

$$\forall A_i \in \mathcal{A}, \quad A_i \cap D \neq \emptyset. \tag{23}$$

By constructing the set D using a randomized cover algorithm, then with high probability $1 - \mathcal{O}(h^{1-m})$, for any constant $m > 1$, D covers $\mathcal{A}$ and its size is upper bounded as:

$$|D| = \mathcal{O}\left(\frac{\ell \log h}{s}\right). \tag{24}$$

In our context, we map this result to the quantum routing problem, by interpreting P as the set of ESPs, and the collection $\mathcal{A}$ as the set of e-neighborhoods, each with size

feature	e-neighborhood	anchor
connectivity scope	local	global (long-range)
entanglement	toward the k-closest ESPs according to entangling cost $w(\cdot,\cdot)$	toward the other $\|T\|-1$ anchors
entanglement cost	low	high
routing role	capturing local physical topology structure and ensuring short-entanglement hop redundancy	enabling global connectivity across distant ESPs
overlay topology impact	supporting dense local clustering	guaranteeing that any pair of ESP is within at most 3-entanglement-hops
maintenance complexity	low	high
design parameter	number of e-neighbors $k=\|N(v_i)\|$	number of anchors $\|T\|$
scalability/performance impact	sublinear routing tables	constant entangling stretch

TABLE V: Comparison between e-neighborhood $N(\cdot)$ and anchor set T, highlighting their respective contributions to local and global entanglement reachability in the overlay topology.

$k=|N(v_j)|$. The goal is to construct the anchor set T, such that each e-neighborhood $N(v_j)$ contains at least one anchor ESP. Thus, in this analogy, T plays the role of the covering set D. By applying the result from [30] and by neglecting lower-order terms, we ensure that if the e-neighborhood size is set to:

$$k=(1+m)\sqrt{n_e}\log n_e, \tag{25}$$

then there exists a cover set T of size:

$$|T|=\sqrt{n_e}, \tag{26}$$

that covers all the e-neighborhoods with high probability. In other words, with this parametrization, we are able to ensure that the anchor set T serves as a dominating set over the graph induced by e-neighborhoods, thereby enabling the following crucial Property 1, while maintaining both sublinear memory complexity and constant stretch.

Property 1. *Any ESP has in its e-neighborhood at least one anchor with high probability (w.h.p.), i.e.:*

$$P\big(\nexists v_j\in T: v_j\in N(v_i)\big)=\mathcal{O}(n_e^{1-m}),\quad \forall v_i\in V_e. \tag{27}$$

Remark 7. As aforementioned, the anchor set T can be constructed by using the randomized algorithm proposed in [30], which does not require any a-priori knowledge nor any particular property. And the probability of violating Property 1 can be made arbitrarily small, by appropriately tuning the constant m. In practical scenarios, when the anchors are equipped with enhanced quantum hardware capabilities[15], it may be preferable to deterministically construct the anchor set. This can be accomplished either by de-randomizing the aforementioned algorithm using techniques such as those proposed in [31], or by adopting a classical greedy approach as originally introduced in [29]. In the latter case, a tighter bound on the anchor set size can be achieved, such as $|D|\leq\frac{\ell(1+\log h)}{s}$, yet at the price of some global network knowledge. Indeed, more efficient covering algorithms can be devised, by exploiting particular properties of the overlay topology. However, the investigation of optimized techniques for the anchor-set construction lies beyond the scope of this paper. Our goal here is to demonstrate that Property 1 can be enforced with high probability through careful parameterization, and that both randomized and deterministic construction methodologies exist to support this design objective.

With the above in mind, the key design choice for the routing protocol is the following:

> *each anchor node $v_j\in T$ proactively establishes and maintains entanglement with any other anchor node in T,*

via the optimal end-to-end entangling metric in Def. 3. Accordingly, $v_i,v_j\in T$ are connected by an artificial link within the artificial topology, with the lowest entangling stretch, i.e., $\mathcal{ES}(i,j)=1$. Additionally:

> *each ESP v_i proactively establishes and maintains entanglement with any ESP within its e-neighborhood $N(v_i)$,*

again, by exploiting the optimal entangling metric in Def. 3. Such artificial links between v_i and any $v_j\in N(v_i)$ likewise exhibit the minimum entangling stretch, i.e., $\mathcal{ES}(i,k)=1$. This design is depicted in Fig. 3a, where non-anchor ESPs, i.e., nodes in $V_2\setminus T$, and ESPs in T are represented in blue and purple, respectively, and artificial links are denoted with red arrows.

As a consequence, an arbitrary ESP $|v_i\rangle$ maintains an *entangling routing table* with $\tilde{\mathcal{O}}(\sqrt{n_e})$ entries, as enforced by our *compact* design principle. This table is organized as shown in Fig. 4 and detailed below:

- $\mathcal{O}(\sqrt{n_e}\log n_e)$ "entries" are dedicated to e-neighbors in $N(v_i)$, with each entry storing a certain number[16] of e-bits shared with v_j in $N(v_i)$, along with:
  - i) the quantum address $|v_j\rangle$,
  - ii) and the superposition of the quantum addresses $\{|v_m\rangle\}$ of the e-neighbors in $N(v_j)$.

[15] An assumption arguably reasonable given their role in proactively maintaining long-range entangled links.

[16] The number of e-bits per entry is a design parameter, that depends on the number of communication qubits available at each ESP. Thus this choice is dictated by the hardware-specific complexity, characterizing the ESPs.

QUANTUM ROUTING TABLE AT ESP v_i

ebits	e-hop	e-neghborhood	Anchor	
e_1 | e_2 | $\cdots$ | $\cdots$	$\cdots$	$\cdots$	$\cdots$	for any $v_i \in V_e$: $\mathcal{O}(\sqrt{n_e}\log n_e)$ entries one for each $v_j : v_i \in N(v_j)$
e_1 | e_2 | $\cdots$ | $\cdots$	$\vert v_j\rangle$	$\frac{1}{\sqrt{\vert N(v_j)\vert}} \sum_{v_m \in N(v_j)} \vert v_m\rangle$	yes/no	
e_1 | e_2 | $\cdots$ | $\cdots$	$\cdots$	$\cdots$	$\cdots$	
e_1 | e_2 | $\cdots$ | $\cdots$	$\cdots$	$\cdots$	$\cdots$	PARTIAL-ANCHOR SCHEME: if $v_i \in T$: $\mathcal{O}(\sqrt{n_e})$ entries one for each $v_k \in T$; if $v_i \notin T$: no entries. FULL-ANCHOR SCHEME: $\mathcal{O}(\sqrt{n_e})$ entries one for each $v_k \in t(v_i)$
e_1 | e_2 | $\cdots$ | $\cdots$	$\vert v_k\rangle$	$\frac{1}{\sqrt{\vert N(v_k)\vert}} \sum_{v_m \in N(v_k)} \vert v_m\rangle$	yes	
e_1 | e_2 | $\cdots$ | $\cdots$	$\cdots$	$\cdots$	$\cdots$	

Fig. 4: Schematic view of the entangling table of ESP v_i. The former portion is devoted to track entanglement within e-neighborhood (low-cost artificial links) and it is common in both *Partial-Anchor* and *Full-Anchor* schemes. The latter portion is devoted to maintain high-cost artificial links, and it differs depending on the adopted scheme.

- If $v_i \in T$ then there exist $\mathcal{O}(\sqrt{n_e})$ additional "entries", storing a certain number of e-bits shared with each anchor ESP $v_k \in T$, together with:
  i) the quantum address $|v_k\rangle$,
  ii) and the superposition of the quantum addresses of the e-neighbors in $N(v_k)$.

It is important to note that an artificial link between v_i and v_j, representing a shared entangled state, requires that (at least) one ebit is stored at each node. Thus, whenever an ESP v_i establishes a link with its e-neighbor v_j, there is an entry in v_i quantum routing table as well as a corresponding entry in v_j quantum routing table. However, the neighbor sets of v_i and v_j may differ: the entangling cost function $w(\cdot,\cdot)$ may induce the selection of v_i among the top $\sqrt{n_e}$ entanglement-neighbors of v_j, i.e., $v_i \in N(v_j)$, but the reverse is not true, i.e., $v_j \notin N(v_i)$. This asymmetry may arise, since the neighbors for v_i and v_j could be constructed independently and possibly from partial views of the network. To preserve routing consistency and full usability of the links established by v_j, node v_i must be aware of such asymmetry. This motivates the introduction of the *reverse neighborhood* of node v_i, defined as $R(v_i) \stackrel{\triangle}{=} \bigcup \{v_j \in V : v_i \in N(v_j)\}$. In practice this means that, while the primary entries in v_i entangling table are determined by its own neighbor-set $N(v_i)$ and utilized for the packet forwarding at v_i, the table must also accommodate auxiliary entries for $R(v_i)$, to enable correct handling of routing requests terminating at or passing through v_i due to its inclusion in another node's neighbor-set. Thus, the number of entries at v_i scales as $|R(v_i) \cup N(v_i)|$. There exist some degenerate cases, where the cardinality of $|R(v_i)|$ is not upper-bounded by $\tilde{\mathcal{O}}(\sqrt{n_e})$, such as when the entangling cost $w(\cdot,\cdot)$ does not satisfy the axiomatic properties of a metric given in (11)-(14) or when the cost induces strong centralization, with a few privileged nodes being entangled-efficient for a disproportionately large number of others. In the following, we reasonably assume that such degenerate cases can be properly handled by fairness policies enforced by the EDC. These policies can induce a maximum number of entangling table entries per node, prioritizing fairness and scalability.

Quantum Routing Logic:
We are ready now to present the logic underlying our compact quantum routing protocol for the partial-anchor scheme, achieving entanglement-stretch $\mathcal{ES}$ upper bounded by 5, as proved in Lemma 1. More in detail, end-to-end entanglement between an "initiating" ESP, say node v_i, and a "target" ESP v_d proceeds through one of the following three cases, examined in order.

- *Case I.* v_d belongs to $N(v_i)$, or equivalently if $v_i \in T$, v_d may belong to T as well. This case can be checked by simply searching for quantum address $|d\rangle$ within the *e-hop* field of the entangling table entries, by exploiting the orthogonality of the addresses. If this condition holds, then v_i has already established entanglement with v_d, through the optimal entangling metric by design. Thus, an artificial link exists with unitary entangling stretch, i.e., $\mathcal{ES}(i,d) = 1$.
- *Case II.* This case is illustrated in Fig. 5a. If *Case I* does not hold, the targeted ESP v_d may still belong to the e-neighborhood of some $v_j \in N(v_i)$. This can be checked by searching for the quantum address $|v_d\rangle$ within the superposed quantum addresses stored in the *e-neighborhood* field of the entangling table entries, by exploiting the *quantum addressing splitting* functionality designed in Sec. IV. If $|v_d\rangle$ is found, by denoting with $v_j \in N(v_i)$ the *e-hop* of $|v_d\rangle$ as in Fig. 5a, then both v_i, v_j and v_j, v_d are entangled through the optimal entangling metric by design. Thus, two artificial links with unitary

stretch exist – $\mathcal{ES}(i,j) = \mathcal{ES}(j,d) = 1$ – and they enable end-to-end entanglement between v_i and v_d via entanglement swapping at v_j.
- *Case III*. This case is represented in Fig. 5b. Whenever the previous two cases do not apply then, with a probability that can be made as close to 1 as wished accordingly to Property 1, there exists an anchor ESP, say $v_l \in T$, within the e-neighborhood of v_i[17]. By design, v_l has already established entanglement with any other anchor in T, including some anchor $v_k \in T$ that belongs to the e-neighborhood of v_d. This can be verified by searching for the quantum address $|v_d\rangle$ in the *e-neighborhood* fields associated with the e-hops field of $v_l \in T$. As a consequence, there exist three artificial links characterized by unitary entangling stretch – i.e. $\mathcal{ES}(i,l) = \mathcal{ES}(l,k) = \mathcal{ES}(k,d) = 1$ – and thus end-to-end entanglement between v_i and v_d can be straightforwardly obtained by entanglement swapping at both v_l and v_k[18].

Remark 8. The configuration described in *Case III* substantiates the key insight anticipated at the beginning of Sec. III-C. Specifically, with high probability, *every ESP is at most 3-hop-entanglement distant from any other ESP* within the overlay topology, and this confirms the strategic role of the anchors as entanglement hubs. Importantly, this 3-hop value represents a worst-case scenario, derived under the assumption that both v_i and v_d are neither anchor nor k-closest neighbor nodes, thus without direct entanglement. In many practical configurations, the hop-entanglement can be strictly smaller. Nevertheless, this worst-case guarantee offers strong evidence of the routing performance, achieved with only sublinear memory overhead and local decision-making. Thus, it reinforces the effectiveness of the anchor-based overlay design.

Remark 9. Although we fairly impose an *optimal entangling metric* $w(\cdot,\cdot)$ for the selection of the e-neighbors, by ensuring that each ESP maintains entanglement with its most favorable peers according to a selected performance criterion, the entangling stretch $\mathcal{ES}$ of the proposed routing scheme is not necessarily equal to one. This is the consequence of two structural relaxations deliberately introduced in our design to preserve the *compactness and scalability* of our protocol.
First, we *relax the requirement of a fully connected entangled mesh among the ESPs*. While such a mesh would guarantee unit stretch for all the aforementioned cases, even in *Case III*, maintaining all-to-all entanglement among ESPs would be practically unsustainable, due to the complexity related to the hardware and control overhead. Instead, our protocol relies on a sparse backbone, built through localized decisions and bounded connectivity.
Second, but most importantly, we do *not optimize the construction of the anchor set* T, as mentioned above. The anchor ESPs are selected through the *randomized, distributed algorithm*, described in Remark 7. While this approach requires no global knowledge, it may yield a suboptimal anchor placement from the perspective of minimizing end-to-end entangling stretch. In this light and in agreement with Remark 7, it is reasonable to assume the nodes in T equipped with *enhanced quantum hardware capabilities*. Under this assumption, it may be more appropriate to construct the anchor set *deterministically*. Such a deterministic approach, potentially combined with a joint optimization of the entangling cost metric and the anchor topology, could further improve the routing efficiency, by achieving even lower stretch, even in scenarios corresponding to *Case III*. However, a detailed investigation of this joint optimization strategy falls beyond the scope of this paper. We leave it as future work.

The entangling stretch of the proposed partial-anchor scheme is upper-bounded formally in the following Lemma.

Lemma 1. *By denoting with v_i and v_d the ESPs exhibiting the worst-case entangling stretch defined in* (19)*, the stretch of the corresponding entangling path is upper bounded as follow:*

$$\begin{aligned}\mathcal{ES} \leq c &= \frac{w\big((i,d)\oplus(i,d)\oplus(i,d)\oplus(i,d)\oplus(i,d)\big)}{w(i,d)}\\ &\triangleq \frac{w\left(\bigoplus_5 (i,d)\right)}{w(i,d)}.\end{aligned} \tag{28}$$

For additive metrics, i.e., when the composition operator is such that $w(a\oplus b) = w(a)+w(b)$, (28) *simplifies to:*

$$\mathcal{ES} \leq c = 5. \tag{29}$$

Conversely, for concave metrics where the composition operator is defined as $\oplus \triangleq \min$, the stretch in (28) *becomes unitary, i.e., $\mathcal{ES} = 1$.*

Proof: See Appendix A. ■

D. Full-Anchor Scheme

Here we design a second quantum routing scheme, which relies on enhanced ESP capabilities to establish long-range artificial links. To this aim, while the definition of e-neighborhood in Def. 5 remains unchanged, we introduce the following design principle, as an alternative to Design Principle 5.A.

Design Principle 5.B (Full-Anchor Scheme). The design of the quantum routing protocol requires each ESP to proactively generate and sustain both the two "species" of artificial links: short-range (i.e., low-cost) and long-range (i.e., high-cost) links.

Accordingly, in this second routing scheme, each ESP serves as anchor. To emphasize this aspect, we termed it *full-anchor scheme*. However, to avoid unsustainable routing-table growth, we induce a sparse tracking paradigm, where

[17] The underlying and non-restrictive hypothesis is that each anchor ESP reveals its anchor-status to its e-neighbors. In other words, if $v_l \in T$ belongs to the e-neighborhood $N(v_i)$, then v_i is aware that v_l is an anchor. This information is required to allow forwarding decisions that leverage the entanglement already shared among the anchors. Importantly, this does not entail any global coordination, only local exchanges during the entanglement establishment phase, achieved by embedding such metadata into the exchanged quantum packets, carrying also the superposed state encoding the identities of the neighbors of each e-hop.

[18] There exists some special configurations of *Case III* – such as when $v_i \in T$ as shown in Fig. 5c or when $v_d \in T$ – that yield a tighter upper bound for the entangling stretch, i.e., $\mathcal{ES} < 3$. In particular, when both $v_i, v_d \in T$, the configuration actually falls under Case I, since anchor ESPs are directly entangled by design, resulting in unitary stretch $\mathcal{ES} = 1$.

each ESP monitors only a subset, randomly selected, of other ESPs. A naive design of this scheme may result in pathological scenarios, such as for instance, all ESPs tracking the same subset, leaving parts of the overlay unreachable. Consequently, to ensure that the sparse tracking preserves full-network coverage, we leverage the result for extended dominating sets, introduced in the previous subsection. To formalize this approach, the following definitions are needed.

Definition 7 (**Tracked-Sets**)**.** A collection of disjoint subsets $\mathcal{T} = \{T_1, \ldots, T_{\sqrt{n_e}}\}$ is said to form a partition of the ESP set V_2 if:

$$\bigcup_i T_i = V_2 \ \wedge \ T_i \cap T_j = \emptyset. \tag{30}$$

In the following, we refer to the elements in $\mathcal{T}$ as *tracked-sets*.

We assume that the mapping function $V \longrightarrow \mathcal{T}$, determining the partition of the ESP set V_2, is globally known by the nodes.

In the following, for the sake of simplicity we assume the tracked-sets are constructed via an arbitrary flat partitioning of the ESP set V_2. One natural choice is to partition nodes based on their quantum addresses, i.e., a node with quantum address $|i\rangle$ is assigned to the j-th tracking-set T_j if and only if $(j-1)\sqrt{n_e} < i \leq j\sqrt{n_e}$.

Remark 10. The aforementioned flat partitioning has the desirable property of being independent of the network topology and yet it guarantees that each tracked-set T_j contains at most $\sqrt{n_e}$ nodes, i.e. $T_j \leq \sqrt{n_e}$ for any j. Nevertheless, any alternative flat partitioning that respects this cardinality constraint is equally valid.

We now define the *tracking-nodes*.

Definition 8 (**Tracking-nodes**)**.** Each ESP $v_i \in V_2$ tracks only one tracked-set $T_j \in \mathcal{T}$, chosen uniformly at random. This set is denoted as $t(v_i)$:

$$\forall\, v_i \in V_2 : \exists!\, T_j \in \mathcal{T} \text{ s.t. } t(v_i) = T_j \tag{31}$$

Any node $v_d \in t(v_i)$ is referred to as *one of the nodes tracked by* v_i.

The key observation is that the union of tracked-sets originating from the e-neighborhood of any given ESP must collectively cover the entire ESP set. This ensures reachability and global coverage under sparse tracking regime. To enforce this coverage, we leverage again the result on extended dominating sets [29], introduced in the previous section. Indeed, by enforcing the same parametrization of the e-neighborhood cardinality $k = (1+m)\sqrt{n_e}\log n_e$, we ensure the following property with high-probability:

Property 2. *Any ESP v_i has in its neighborhood at least one node tracking any other ESP w.h.p., i.e.:*

$$P\big(\ \nexists\, v_j \in N(v_i) : v_d \in t(v_j)\big) = \mathcal{O}(n_e^{1-m}),\ \forall\, v_i, v_d \in V_2. \tag{32}$$

This property guarantees that, although each ESP tracks only a limited number of other ESPs, the collective coverage provided by the e-neighborhood is sufficient to ensure full overlay reachability.

Remark 11. As noted in Remark 7, a similar observation applies here as well: the specific construction of the tracked sets lies outside the scope of this work. Our primary objective is to demonstrate that Property 2 can be ensured with high probability, through appropriate parametrization. Although more optimized constructions are possible, our analysis deliberately avoids assuming any particular structure in the tracked set formation. As such, the presented results should be interpreted as worst-case guarantees.

With the above in mind, the full-anchor scheme relies on the following:

> *each ESP v_i proactively establishes and maintains entanglement with any node within its tracked-set $t(v_i)$,*

where entanglement is established using the optimal end-to-end entangling metric in Def. 3. Accordingly, v_i and any $v_j \in t(v_i)$ are connected by an artificial link within the artificial topology characterized by unitary entangling stretch, i.e., $\mathcal{ES}(i,j) = 1$. Furthermore:

> *each ESP v_i proactively establishes and maintains entanglement with any ESP within its e-neighborhood $N(v_i)$,*

again via the optimal end-to-end entangling metric in Def. 3. Consequently, each artificial link between v_i and any $v_k \in N(v_i)$ also exhibits a unitary entangling stretch, i.e., $\mathcal{ES}(i,k) = 1$.

As a consequence, an arbitrary ESP $|v_i\rangle$ maintains an *entangling routing table* with $\tilde{\mathcal{O}}(\sqrt{n_e})$ entries, as enforced by our *compact* design principle. This table is organized as shown in Fig. 4. The distinguishing factor between the *partial-anchor scheme* and the *full-anchor scheme* is the latter part of the quantum routing table devoted to high-cost artificial links. Specifically, the *full-anchor scheme* allows each ESP, rather than only the nodes in T, to proactively establish and maintain such links. This maps into an "enhanced" connectivity within the artificial topology, which enables our compact quantum routing protocol to guarantee entangling stretch $\mathcal{ES}$ upper bounded by 3, as in (21), with at most $\tilde{\mathcal{O}}(\sqrt{n_e})$ qubits to be stored at each node.

Quantum Routing Logic:

We are ready now to describe the logic underlying *full-anchor scheme*. More in detail, end-to-end entanglement between "initiating" ESP v_i and "target" ESP v_d proceeds through one of the following two cases, examined in order.

- *Case I.* v_d may belong to $N(v_i)$ or $t(v_i)$; if so, similarly to *Case I* of the *Partial-Anchor Scheme*, then v_i already established some entanglement with v_d through the optimal entangling metric by design. Thus, an artificial link exists with unitary entangling stretch: $\mathcal{ES}(i,d) = 1$.
- *Case II.* It is similar to the case represented in Fig. 5a. Whenever the first case does not hold, according to Property 2 which holds w.h.p., v_d belongs to the e-neighborhood of at least one node in $N(v_i)$. This can be checked by searching for the quantum address $|v_d\rangle$

within the superposed quantum addresses stored in the *e-neighborhood* field of the entangling table entries, by exploiting the *addressing splitting* functionality designed in Sec. IV. If $|v_d\rangle$ is found, by denoting with $v_j \in N(v_i)$ the *e-hop* of v_d as in Fig. 5a, then both v_i, v_j and v_j, v_d are entangled with unitary stretch by design. Thus, two artificial links with unitary stretch exist. As a consequence, end-to-end entanglement between v_i and v_d can be straightforwardly obtained by entanglement swapping at v_j.

Remark 12. Similarly to the observation in Remark 8, the configuration described in *Case II* substantiates the key insight anticipated at the beginning of Sec. III-C. Specifically, with high probability, *every ESP is at most at 2 entanglement-hops from any other ESP* in the overlay topology.

Lemma 2. *By denoting with v_i and v_d the ESPs exhibiting the worst-case entangling stretch defined in* (19)*, the stretch of the corresponding entangling path is upper bounded as follow:*

$$\mathcal{ES} \leq c = \frac{w\big((i,d) \oplus (i,d) \oplus (i,d)\big)}{w(i,d)} \triangleq \frac{w\left(\bigoplus_3 (i,d)\right)}{w(i,d)} \quad (33)$$

which, for additive metrics (28)*, simplifies to:*

$$\mathcal{ES} \leq c = 3. \quad (34)$$

Conversely, for concave metrics ($\oplus \triangleq \min$) the stretch in (25) *becomes unitary, i.e., $\mathcal{ES} = 1$.*

Proof: The proof follows by adopting the same reasoning as in App. A ■

IV. Quantum Address Splitting via Schrödinger's Oracle

As detailed in the previous section, the proposed routing protocol, by exploiting superposed quantum addresses, requires the availability of a functionality we termed *quantum address splitting*.

This functionality generalizes classical forwarding operations to quantum-superposed identifiers. Specifically, the quantum address splitting mechanism determines whether a given quantum address, say $|v_d\rangle$, is embedded within the superposed address stored in the *e-neighborhood* field of a the quantum routing table entry, as illustrated in Fig. 4. And this check must be performed by satisfying two key constraints:

i) *non-reliance on prior knowledge*: the addressing splitting functionality must not require any preliminary knowledge about the specific quantum addresses that have been superposed, since they depend on the topology via the e-neighborhood $N(\cdot)$;
ii) *non-destructive search*: the addressing splitting functionality should not alter any superposed address that does not contain the target address, $|v_d\rangle$, ensuring so that the superposed address remains available for subsequent queries involving different target addresses.

With the above two key constraints in mind, we propose a modified version of Grover's quantum search algorithm [32], *tailored for quantum-superposed data structures*. In our design, the oracle operates coherently on quantum-superposed entries and is thus referred to as the *Schrödinger's oracle*. Specifically, the oracle is implemented via controlled quantum gates, where the control is provided by the quantum registers encoding the superposed addresses stored in the *e-neighborhood* field of the routing table. The control condition is determined by the target quantum address $|v_d\rangle$ we aim to identify. As illustrated in Fig. 6, and detailed in the following, the oracle evaluates whether $|v_d\rangle$ is included in any superposed e-neighborhood associated with each routing entry. Specifically, it performs a phase inversion on the state representing the label of a given entry only when the corresponding superposed e-neighborhood contains the address $|v_d\rangle$.
This modified oracle causes the algorithm to evolve into a superposition of parallel search paths: one where the marked entry undergoes a phase flip, and another where it does not. Thus, the presence of the phase inversion becomes a coherent, quantum-controllable degree of freedom. This mechanism enables the selective identification of $|d\rangle$ without collapsing the superposition of unrelated addresses.

A. Quantum Search Register

To fulfill the second key constraint, namely, the ability to perform non-destructive queries, we encode the labels of the routing table entries into a dedicated quantum register, which then undergoes the Grover search iterations. Formally, our search space consists of the[19] n_{T} routing table entries, each uniquely identified by a label encoded as a computational basis state $|y\rangle$, with $y \in \{0,1\}^{N_{\mathrm{T}}}$, and $N_{\mathrm{T}} = \lceil \log_2(n_{\mathrm{T}}) \rceil$. This quantum register, referred to as the *entry label register*, holds the label of each entry. The register is initialized following the standard Grover preparation step, as shown in Fig. 6, resulting in an equal superposition of all the n_{T} basis states:

$$|\psi_0\rangle = \frac{1}{\sqrt{n_{\mathrm{T}}}} \sum_{y=0}^{n_{\mathrm{T}}-1} |y\rangle \,. \quad (35)$$

B. Quantum Superposed Address Registers

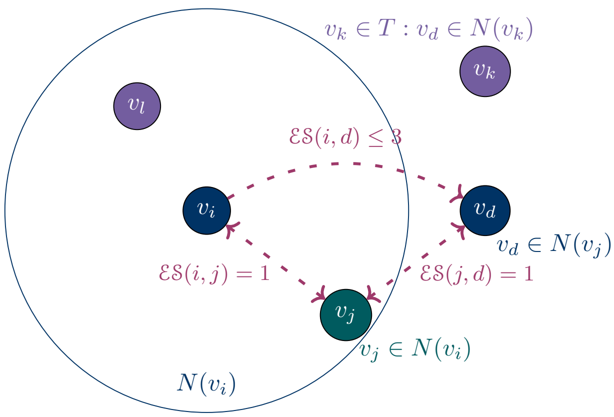
Let us consider the arbitrary entry labeled as $|y\rangle$ and denote the quantum address of its associated e-hop as $|v_j\rangle$.

The e-neighborhood $N(v_j)$ of node v_j is partitioned by the node itself into $f(n_e)$ disjoint sub-e-neighborhoods, such that any given quantum node $v_k \in N(v_j)$ belongs to exactly one of these partitions. Each of these partitions is encoded as a quantum-superposed address, denoted as $|A_j^l\rangle$, and announced by $|v_j\rangle$. The superposition representing l-th partition of $N(v_j)$ is defined as follows:

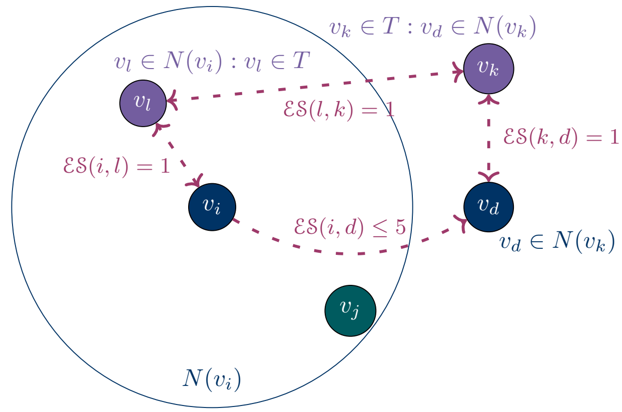
$$|A_j^l\rangle = \sqrt{\frac{f(n_e)}{\sqrt{n_e}\log n_e}} \sum_{v \in S_j^l} |v\rangle \,, \quad (36)$$

where S_j^l is the set of nodes in the l-th partition of $N(v_j)$. Accordingly, each node $|v\rangle$ in S_j^l appears in the quantum superposition $|A_j^l\rangle$ with amplitude $\sqrt{\frac{f(n_e)}{\sqrt{n_e}\log n_e)}}$.

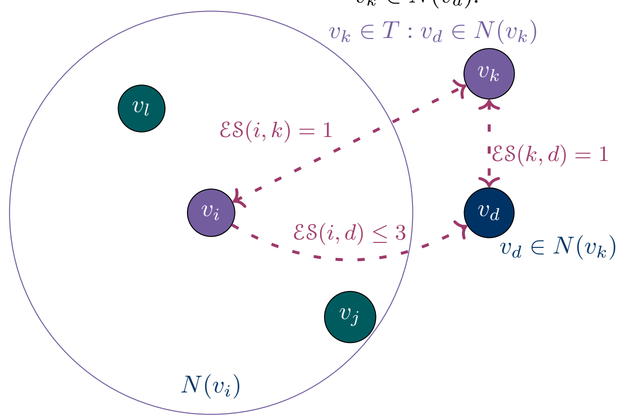
[19]We introduce a more compact notation for the number of routing entries, with respect to the one adopted in Sec. III, for the sake of brevity.



(a) Case II: there exists $v_j \in N(v_i)$ so that $v_d \in N(v_j)$.



(b) Case III: there exists $v_l, v_k \in T$ so that $v_l \in N(v_i)$ and $v_k \in N(v_d)$.



(c) Case IV as particular configuration of case III where either v_i belongs to T.

Fig. 5: Partial-Anchor Scheme: schematic view of the logic underlying our compact quantum routing protocol for establishing end-to-end entanglement between "initiating" ESP v_i and "target" ESP v_d. Tracking nodes in T are denoted in purple, artificial links are denoted with red arrows, and circles around quantum nodes denote their e-neighborhood.

Remark 13. We highlight that $f(n_e)$ serves as a key tunable design parameter of our quantum address splitting functionality. Indeed, $f(n_e)$ allows us to drive the success probability of the Grover search via our Schrödinger's oracle toward one: the larger is $f(n_e)$, the higher is the success probability.

The $f(n_e)$ quantum addresses $\{|A_j^l\rangle\}$, jointly describing the e-neighborhood $N(v_j)$ of e-hop $|v_j\rangle$, control our Schrödinger's oracle for the j-th entry label, as shown in Fig. 6.

C. Schrödinger's Oracle

The Schrödinger's oracle O_S for the e-hop $|v_j\rangle$ acts on the composite quantum system $|\psi_0\rangle \otimes |A_j\rangle$, with $|A_j\rangle \triangleq |A_j^1\rangle \otimes \ldots \otimes |A_j^{f(n_e)}\rangle$, as shown in equation (37) shown[20] within next page. The amplitude associated with the target address in the first sub-e-neighborhood is denoted as $\sqrt{\alpha} \triangleq \langle v_d | A_j^1 \rangle = \sqrt{\frac{f(n_e)}{\sqrt{n_e}\log n_e)}}$.

[20]Where we omitted the superposed addressed for the e-neighborhoods of the other entries for the sake of notation simplicity and brevity. For the same reasons, we assume, without loss of generality, that v_d is in A_j^1.

As evident from equation (37) shown within next page, whenever the target $|v_d\rangle$ does not belong to the e-neighborhood $N(v_j)$ of the e-hop $|v_j\rangle$, then our Schrödinger's oracle O_S satisfies both the aforementioned constraints, by acting trivially on the entry label $|x\rangle$ (namely $|x\rangle$ is unchanged) and, crucially, leaving any non-matching superposed address unaltered. This ensures the non-destructive property of the search.

Conversely, whenever $|v_d\rangle \in N(v_j)$, then O_S generates an inversion of the phase of the associated entry label $|x\rangle$. More precisely, O_S entangles the entry label $|x\rangle$ with the "hitting " sub-e-neighborhood, inducing a conditional phase flip on $|x\rangle$. Thus, two computational evolutions are possible, one *with* and the other *without* the oracle phase inversion, as follows (by

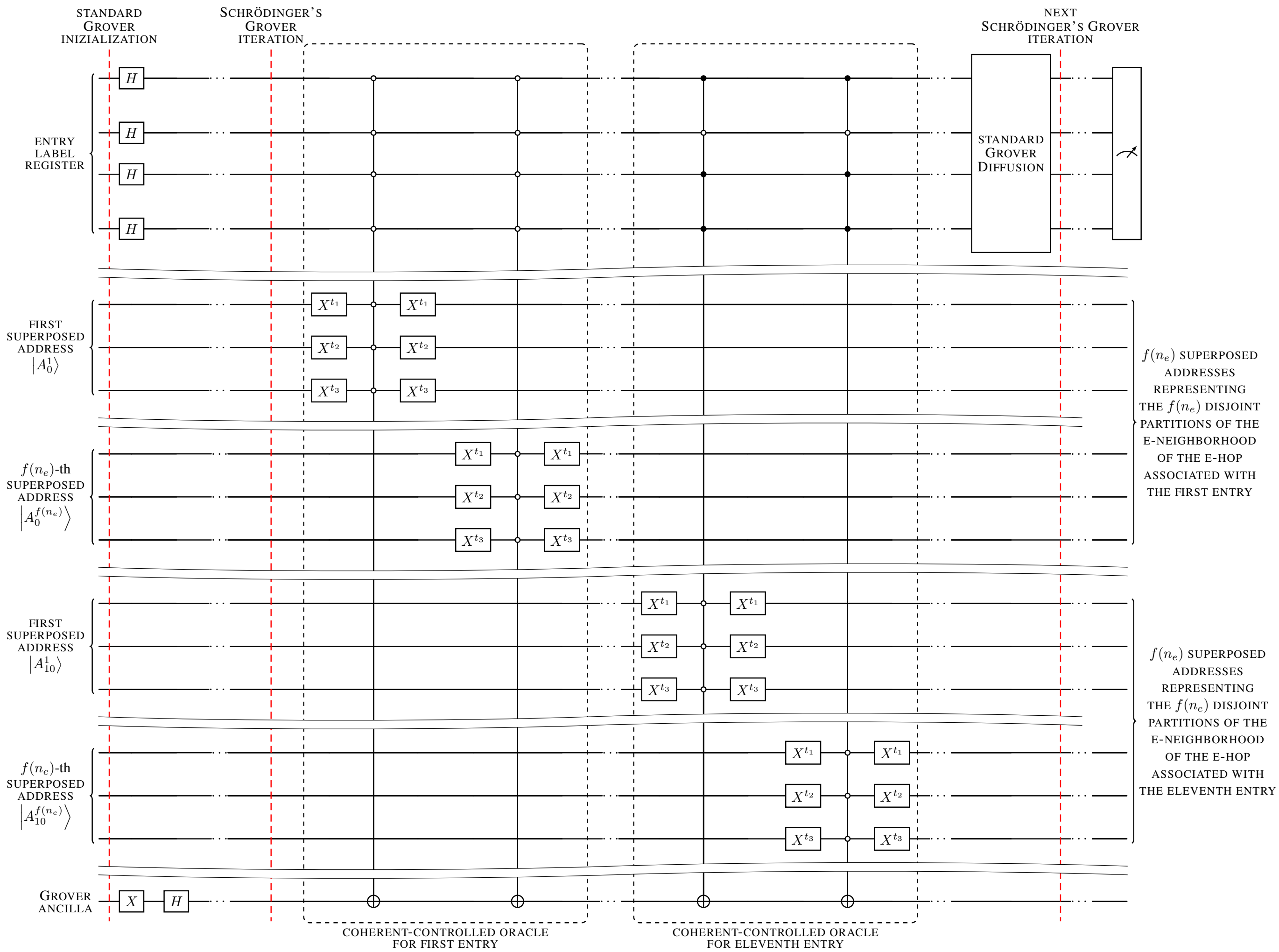


Fig. 6: Quantum circuit representing the oracle of the Grover's search algorithm for quantum address $|v_d\rangle$, with binary coefficient $\{t_i\}$ given in (39). The labels of the $n_T = \sqrt{n_e}\log n_e$ entries of the quantum routing table are encoded within the entry register – represented in the figure as four qubits rather than $\lceil \log_2 \left(\sqrt{n_e}\log n_e\right)\rceil$ qubits for the sake of simplicity. The oracle performs a phase flip on the quantum state representing the label of entry associated to e-hop $|v_j\rangle$, whenever one of the superposed addresses, representing the disjoint e-neighborhoods of $|v_j\rangle$, contains the target quantum address $|v_d\rangle$. In the figure, we assumed for the sake of simplicity a number of entries equal to 16. Thus, the entry register is 4 qubits and each superposed address is 3-qubits long. Furthermore, the top wire is the leftmost significant qubit, thus register state $|1011\rangle$ labels the eleventh entry.

$$
\begin{aligned}
&O_S\left(\frac{1}{\sqrt{n_\mathrm{T}}}\sum_{y=0}^{n_\mathrm{T}-1}|y\rangle\otimes|A_j\rangle\right) = O_S\left(\frac{1}{\sqrt{n_\mathrm{T}}}\sum_{y=0}^{n_\mathrm{T}-1}|y\rangle\otimes|A_j^1\rangle\otimes\ldots\otimes|A_j^l\rangle\otimes\ldots\otimes|A_j^{f(n_e)}\rangle\right) = \\
&= \begin{cases}
= \dfrac{1}{\sqrt{n_\mathrm{T}}}\displaystyle\sum_{y=0}^{n_\mathrm{T}-1}|y\rangle\otimes|A_j\rangle & \text{if } v_d \notin N(v_i) \\
= \begin{cases}
= O_S\left(\dfrac{1}{\sqrt{n_\mathrm{T}}}\displaystyle\sum_{y=0}^{n_\mathrm{T}-1}|y\rangle\otimes\left(\sqrt{1-\alpha}\displaystyle\sum_{v\in S_j^1\setminus\{v_d\}}|v\rangle+\sqrt{\alpha}\,|v_d\rangle\right)\otimes|A_j^2\rangle\otimes\ldots\otimes|A_j^{f(n_e)}\rangle\right) = \\
= \dfrac{1}{\sqrt{n_\mathrm{T}}}\displaystyle\sum_{y=0,y\neq x}^{n_\mathrm{T}-1}|y\rangle\otimes|A_j\rangle+\dfrac{1}{\sqrt{n_\mathrm{T}}}\left(\sqrt{1-\alpha}\,|x\rangle\otimes\displaystyle\sum_{v\in S_j^1,v\neq v_d}|v\rangle-\sqrt{\alpha}\,|x\rangle\otimes|v_d\rangle\right)\otimes \\
\qquad\otimes|A_j^2\rangle\otimes\ldots\otimes\otimes|A_j^{f(n_e)}\rangle
\end{cases} & \text{if } v_d \in A_j^1 \subset N(v_i)
\end{cases}
\end{aligned}
\tag{37}
$$

omitting all the irrelevant quantum subsystems):

$$O_S\left(\frac{1}{\sqrt{n_\mathrm{T}}}\left|x\right\rangle\otimes\left(\sqrt{1-\alpha}\left|v_d^\perp\right\rangle+\sqrt{\alpha}\left|v_d\right\rangle\right)\right)= \\ =\sqrt{\frac{1-\alpha}{n_\mathrm{T}}}\left|x\right\rangle\otimes\left|v_d^\perp\right\rangle-\sqrt{\frac{\alpha}{N_\mathrm{T}}}\left|x\right\rangle\otimes\left|v_d\right\rangle. \tag{38}$$

Remark 14. The Schrödinger's oracle O_S allows Grover's algorithm to evolve in a coherent superposition of oracle-inverting and oracle-trivial computational dynamics. Further discussion in reported in Sec. V.

In Fig. 6 we provide a straightforward, un-optimized circuit-level realization of the Schrödinger's oracle for 3-qubits quantum addresses. The control conditions are determined by the binary coefficients $\{t_i\}$ driving the X-gates, derived from the classical representation of the target state $|v_d\rangle$:

$$d=\sum_{i=0}^{N-1} b_i 2^i,\ b\in\{0,1\}, \tag{39}$$

with N defined in Def. 1.

D. Diffusion Operator

As shown in Fig. 6, each application of the Schrödinger's oracles is followed by the standard Grover diffusion operator U_D, acting on the entry label register. U_D performs the conventional inversion about the mean, a key mechanism in amplitude amplification.

As shown in (38), what distinguishes our setting, however, is that the overall system evolves into a coherent superposition of two computational branches:

i) a non-inverting branch, corresponding to the component $|x\rangle\otimes|v_d^\perp\rangle$, in which the label has not undergone a phase flip, and
ii) an inverting branch, corresponding to $|x\rangle\otimes|v_d\rangle$, where the label has experienced a phase inversion, namely has been marked.

Here $|x\rangle$ denotes the label of the entry whose e-neighborhood contains the target address $|v_d\rangle$. In the non-inverting branch, since no phase-inversion occurs, the subsequent application of U_D leaves the amplitude distribution unchanged. In contrast in the inverting branch, the diffusion operator U_D acts as in standard Grover: it amplifies the probability amplitude of hitting entry $|x\rangle$, while it suppresses those of the non-marked entries.

E. Measurement

As shown in Fig. 6, the search is concluded upon the measurement of the entry label register, after a prescribed[21] number of iterations. This measurement collapses the superposition over the entry labels, yielding a specific outcome $|x^*\rangle$.

The probability of successfully identifying a hitting entry associated with the target address $|v_d\rangle$ benefits from two contributing factors: one coming from the non-inverting branch and the other from the inverting branch. Specifically, in the non-inverting branch, the amplitudes remain uniform, i.e., $\frac{1}{\sqrt{n_\mathrm{T}}}$ and the measurement behaves like a uniform random choice over all the entries. Differently, in the phase-inverting branch, the amplitude of the hitting entry label associated to $|v_d\rangle$ is amplified toward unity, following standard Grover dynamics.

[21]This prescribed number can be computed by averaging over the optimal number of iterations for the two different branches.

Overall, the probability of successfully measuring the marked entry $|x\rangle$ is enhanced by the Grover-amplified branch weighted by α, while the no-inverting branch contributes with a uniform background distribution. Since α is proportional to $f(n_e)$, i.e., to the number of superposed addressed announced by each e-hop, $f(n_e)$ becomes pivotal in steering the algorithm's success probability toward one.

Furthermore, while we assumed a single hitting entry $|x\rangle$ for simplicity, the expected number of such entries is much larger, on the order of $\log n_e$. This follows from our parametrization of the routing tables in terms of e-neighborhood size k and number of anchors (or tracking nodes in the second routing scheme), which is $\sqrt{n_e}$. Under this design, each quantum address appears redundantly in approximately $\log n_e$ distinct entries of a routing table. As a result, the probability of successfully detecting a hitting entry is further boosted by this built-in redundancy.

Although a comprehensive analytical characterization remains open for future work, preliminary simulations suggest that a poly-logarithmic scaling of $f(n_e)$, when combined with the inherent redundancy of approximately $\log n_e$ hitting entries, is sufficient to ensure w.h.p the discovery of a matching entry, even in very large-scale quantum networks with a number of nodes in the order of millions or more.

V. Discussion

We conclude the paper by discussing several key aspects and implications of our proposed approach. For clarity and readability, each point is discussed individually, by highlighting specific aspects of the proposal and by providing context for rational and future directions.

Compact Routing Design: In 1977, Kleinrock and Kamoun published their pioneering paper [33] on hierarchical routing. Since then, hierarchical routing has been the foundation of both inter-domain and intra-domain routing techniques adopted in Internet, such as CIDR and OSPF/ISIS [15]. And Kleinrock and Kamoun's hierarchical approach was essentially the first *name-dependent* routing scheme[22]. In the same paper, they were the first to analyze the stretch/routing-table size trade-off, by showing that the routing stretch produced by the hierarchical approach is satisfactory only for specific topologies. In other words, hierarchical routing is optimal by achieving shortest paths (namely, paths minimizing the route cost according to the adopted metric) with scalable routing table (routing tables that scale sublinearly with the number of nodes) for topologies with specific characteristics, such as trees or grids. But Internet does not satisfy this

[22]In a nutshell, a *name-dependent* routing scheme embeds some topological information within node addresses, which, thus, cannot be arbitrary. This topological information is, then, exploited to reduce the amount of information to be stored in each routing table. Conversely, *name-independent routing* works with topologically-agnostic node addresses. The differences between the two routing approaches are summarized in Table VI

property	name-dependent	name-independent
address structure	topology-aware	topology-agnostic
routing table size	topology-dependent	sublinear (universal)
robustness to topology changes	limited	high

TABLE VI: Name-dependent vs name-independent routing: concise comparative overview

type of topologies. And indeed, analytical estimates show that applying hierarchical routing to the Internet topology incurs a $\sim$ 15-times path length increase [15].

With this lesson learned from the classical Internet history, and by accounting for the unsuitability of topological-addresses for tracking entanglement as pioneered in [8], we focus our attention on the design of *universal quantum routing* schemes, namely, accordingly to the notation introduced in [15], schemes that work correctly and satisfy promised scaling bounds on all graphs. The rationale for our design choice does not limit to the attractive feature of generality exhibited by universal routing protocols. But it accounts also for the conflicting preliminary results in terms of topology properties of quantum networks reported by recent literature [34], [35], as well as for the absence of experimental studies on quantum network topologies (excepting extremely-small quantum networks such as [24]).

Last but not least, it is worthwhile to note that our proposal is the first compact routing protocol proposed in literature that achieves fully name-independent routing, i.e., that does not exploit an underlying name-dependent scheme [15]. This achievement is made possible by the design of the quantum addressing scheme, which encodes quantum properties directly into node identifiers. We believe that many other network functionalities could benefit similarly from quantum-native design principles.

Worst-Case Baseline for Clustering: The clustering of ESPs into the anchor set (or tracked-sets for the second scheme) has been performed using a flat, non-topology-aware partitioning (see Remarks 9 and 11). This results in a worst-case baseline, as no effort has been made to optimize the placement or connectivity of anchors/tracking ESPs. Indeed, our goal in this first-case analysis is to demonstrate that, even under such an unaware clustering, the proposed schemes achieve coverage w.h.p. and constant-bounded entangling stretch. It follows that future clustering optimizations will definitively lead to even better performance.

Quantum Address Splitting: The Address Splitting functionality proposed in this paper serves as a foundational proof of concept for extending classical forwarding operations to quantum-superposed identifiers. While potential optimizations, such as tuning the redundancy factor $\log_2 n_e$ and the parameter $f(n_e)$ have been briefly discussed in Sec. IV-E, one important optimization dimension remains unexplored. Specifically, the current design does not account for the impact of anchor set (tracked sets in the second scheme) structure on the effectiveness of address splitting. This opens the door to more efficient variants that could emerge from a joint optimization of the address splitting mechanism and the clustering strategies. Exploring this synergy is a promising future direction.

Schrödinger's Oracle: The Schrödinger's oracle O_S designed to implement the quantum address splitting allows Grover's algorithm to evolve in a coherent superposition of oracle-inverting and oracle-trivial computational dynamics. The presence or absence of the oracle's action becomes itself a quantum degree of freedom, which introduced a new layer of coherence within the search process. We strongly believe this approach generalizes standard Grover search and opens new directions for exploring quantum algorithms operating in superpositions of operational regimes, enabled by quantum-superposed data structures. Indeed, the potential advantages offered by this additional dimension of quantumness are reminiscent of the benefits observed in scenarios involving superpositions of quantum operations or tasks [36]–[39]. However, this remains an open question and warrants further investigation.

Multipartite Entanglement: The proposed quantum-native routing protocol considered only bipartite-entanglement, as key communication resource. Yet, the full potential of the proposed quantum addressing scheme becomes even more evident in multipartite entanglement scenarios. Indeed, in such cases, the ESPs must address not just individual nodes, but entire subsets of nodes involved in the entanglement relation. Thus, the quantum addressing and the tracking mechanisms proposed here are particularly well-suited for the multipartite setting, as they can significantly reduce routing table sizes that, without superposed addresses, would scale more than linearly. Exploring this extension is part of our future work.

Architecture Scalability and Generality: The proposed architecture integrates hierarchical principles with SDN principles, resulting in a two-tier structure that clearly separates control from data plane. This separation simplifies network management, by localizing decisions and reducing control overhead. The two-tier hierarchy enables scalable orchestration of entanglement generation and routing, while preserving flexibility and extensibility. Although our work focuses primarily on routing, the architectural foundation here presented is general and it can support other quantum-native functions, such as scheduling, fault tolerance, and resource provisioning. Thus, it is a robust and adaptable foundation for future Quantum Internet design and deployment.

Furthermore, the proposed architecture is not constrained by the assumption of a single, centralized EDC with complete network knowledge. Instead, multiple EDCs can coexist, either hierarchically organized or operating in a distributed fashion, with each EDC orchestrating a portion of the network. Thus, the EDCs can coordinate to share partial topological knowledge and enforce consistent entanglement resource policies, while still enabling local autonomy and scalability.

In this perspective, the aforementioned approach mirrors the design philosophy of distributed SDN controllers in classical networks, where control logic is logically centralized but physically distributed.

Overall, our proposal is flexible and supports scalability and resilience. It allows even for a decentralized control infrastructure that can better adapt to the operational demands of large-scale quantum networks, as exemplified by the full-anchor scheme.

Appendix A
Proof of Lemma 1

By denoting with v_i and v_d the ESPs exhibiting the worst-case entangling stretch in (19), we have two cases[18] with a entangling stretch larger than one: either *Case II* or *Case III*.

Let us focus first on *Case III*. Accordingly, it results:

$$w_{\mathcal{R}}(i,d) = w\big((i,l) \oplus (l,k) \oplus (k,d)\big) \tag{40}$$

By applying the axiomatic monotonic property of a metric given in (14) to quantum path (l,k), we can upper bound (40) as:

$$\begin{aligned} w_{\mathcal{R}}(i,d) &= w\big((i,l) \oplus (l,k) \oplus (k,d)\big) \leq \\ &\leq w\big((i,l) \oplus (l,i) \oplus (i,k) \oplus (k,d)\big) \end{aligned} \tag{41}$$

Clearly, it results $v_d \notin N(v_i)$ – otherwise *Case I* would hold – whereas $v_l \in N(v_i)$ by design. Thus, being $w(i,l) \leq w(i,d)$ from Def. 6 and by accounting for the axiomatic symmetry of a metric given in (13), we have:

$$w\big((i,l) \oplus (l,i)\big) \leq w\big((i,d) \oplus (d,i)\big) \tag{42}$$

By extending both the entangling paths given in (42) with the common path $(i,k) \oplus (k,d)$ and by accounting for the right isotonicity property given in (16), we can upper bound (41) as:

$$\begin{aligned} w_{\mathcal{R}}(i,d) &\leq w\big((i,l) \oplus (l,i) \oplus (i,k) \oplus (k,d)\big) \\ &\leq w\big((i,d) \oplus (d,i) \oplus (i,k) \oplus (k,d)\big) \end{aligned} \tag{43}$$

By using the axiomatic monotonic property of a metric given in (14) to the entangling path (i,k), we can upper bound (43) as follows:

$$\begin{aligned} w_{\mathcal{R}}(i,d) &\leq w\big((i,d) \oplus (d,i) \oplus (i,k) \oplus (k,d)\big) \\ &\leq w\big((i,d) \oplus (d,i) \oplus (i,d) \oplus (d,k) \oplus (k,d)\big) \end{aligned} \tag{44}$$

We have that $v_i \notin N(v_d)$ – otherwise there would have been a corresponding entry for v_d in v_i quantum routing table, as pointed out with the remark after Lemma 1, and *Case I* would hold – while $v_k \in N(v_d)$ by design. Thus, being $w(d,k) \leq w(d,i)$ from Def. 6 and by accounting for the axiomatic symmetry of a metric given in (13), we have:

$$w\big((d,k) \oplus (k,d)\big) \leq w\big((d,i) \oplus (i,d)\big) \tag{45}$$

By extending both the entangling paths given in (45) with the common path $(i,d) \oplus (d,i) \oplus (i,d)$ and by accounting for the left isotonicity property given in (15), we have the thesis, i.e.:

$$\begin{aligned} w_{\mathcal{R}}(i,d) &\leq w\big((i,d) \oplus (d,i) \oplus (i,d) \oplus (d,i) \oplus (i,d)\big) \\ &\triangleq w\left(\bigoplus_{5}(i,d)\right) \triangleq w_{\tilde{R}}(i,d). \end{aligned} \tag{46}$$

As for *Case II*, it follows straightforward from *Case III* by denoting $v_k \triangleq v_k$ and by setting $w(i,l) = 0$. Indeed, in this particular case the entangling stretch exhibits an even tighter upper bound, i.e.:

$$\begin{aligned} w_{\mathcal{R}}(i,d) &= w\big((i,j) \oplus (j,d)\big) \leq w\big((i,j) \oplus (j,i) \oplus (i,d)\big) \\ &\leq w\big((i,d) \oplus (d,i) \oplus (i,d)\big) \triangleq w\left(\bigoplus_{3}(i,d)\right). \end{aligned} \tag{47}$$

References

[1] A. S. Cacciapuoti, M. Caleffi, F. Tafuri, F. S. Cataliotti, S. Gherardini, and G. Bianchi, "Quantum internet: Networking challenges in distributed quantum computing," *IEEE Network*, vol. 34, no. 1, pp. 137–143, 2020.

[2] A. S. Cacciapuoti *et al.*, "When entanglement meets classical communications: Quantum teleportation for the quantum internet," *IEEE Trans. on Communications*, vol. 68, no. 6, pp. 3808–3833, 2020. invited paper.

[3] W. Kozlowski, S. Wehner, R. V. Meter, B. Rijsman, A. S. Cacciapuoti, M. Caleffi, and S. Nagayama, "Architectural Principles for a Quantum Internet." RFC 9340, Mar. 2023.

[4] H. J. Kimble, "The quantum internet," *Nature*, vol. 453, no. 7198, pp. 1023–1030, 2008.

[5] R. Van Meter, *Quantum Networking*. Wiley-IEEE Press, 1st ed., 2014.

[6] W. Dür, R. Lamprecht, and S. Heusler, "Towards a quantum internet," *European Journal of Physics*, vol. 38, no. 4, p. 043001, 2017.

[7] C. Wang, A. Rahman, R. Li, M. Aelmans, and K. Chakraborty, "Application Scenarios for the Quantum Internet." RFC 9583, June 2024.

[8] A. S. Cacciapuoti, J. Illiano, and M. Caleffi, "Quantum internet addressing," *IEEE Network*, 2023.

[9] J. Illiano, M. Caleffi, A. Manzalini, and A. S. Cacciapuoti, "Quantum internet protocol stack: A comprehensive survey," *Computer Networks*, vol. 213, p. 109092, 2022.

[10] B. E. Carpenter, "Architectural Principles of the Internet." RFC 1958, June 1996.

[11] J. H. Saltzer, D. P. Reed, and D. D. Clark, "End-to-end arguments in system design," *ACM Trans. Comput. Syst.*, vol. 2, no. 4, 1984.

[12] D. Meyer and R. Bush, "Some Internet Architectural Guidelines and Philosophy." RFC 3439, 2002.

[13] S.-Y. Chen, J. Illiano, A. S. Cacciapuoti, and M. Caleffi, "Entanglement-based artificial topology: Neighboring remote network nodes," *IEEE Open Journal of the Communications Society*, vol. 6, pp. 2220–2238, 2025.

[14] F. Mazza, M. Caleffi, and A. S. Cacciapuoti, "Intra-QLAN Connectivity via Graph States: Beyond the Physical Topology," *IEEE Transactions on Network Science and Engineering*, pp. 1–18, 2025.

[15] D. Krioukov, k. c. claffy, K. Fall, and A. Brady, "On compact routing for the internet," *SIGCOMM Comput. Commun. Rev.*, vol. 37, p. 41–52, jul 2007.

[16] D. Clark, "The Design Philosophy of the DARPA Internet Protocols," *ACM SIGCOMM '88 Comput. Commun. Rev.*, vol. 18, p. 106–114, aug 1988.

[17] S. DiAdamo, B. Qi, G. Miller, R. Kompella, and A. Shabani, "Packet switching in quantum networks: A path to the quantum internet," *Phys. Rev. Res.*, vol. 4, Oct 2022.

[18] S. J. Ben Yoo, S. K. Singh, M. B. On, G. Gül, G. S. Kanter, R. Proietti, and P. Kumar, "Quantum wrapper networking," *IEEE Communications Magazine*, vol. 62, 2024.

[19] F. Vista, D. Holme, and S. DiAdamo, "Quantum backbone networks for hybrid quantum dataframe transmission," *IEEE Communications Magazine*, vol. 62, 2024.

[20] J. Miguel-Ramiro and W. Dür, "Delocalized information in quantum networks," *New Journal of Physics*, vol. 22, no. 4, p. 043011, 2020.

[21] L. Chen, K. Xue, J. Li, Z. Li, R. Li, N. Yu, Q. Sun, and J. Lu, "Redp: Reliable entanglement distribution protocol design for large-scale quantum networks," *IEEE Journal on Selected Areas in Communications*, vol. 42, no. 7, 2024.

[22] A. Abane, M. Cubeddu, V. S. Mai, and A. Battou, "Entanglement routing in quantum networks: A comprehensive survey," *IEEE Transactions on Quantum Engineering*, 2025.

[23] J. M. Thomas, G. S. Kanter, and P. Kumar, "Designing noise-robust quantum networks coexisting in the classical fiber infrastructure," *Opt. Express*, vol. 31, pp. 43035–43047, Dec 2023.

[24] J. M. Thomas, F. I. Yeh, J. H. Chen, J. J. Mambretti, S. J. Kohlert, G. S. Kanter, and P. Kumar, "Quantum teleportation coexisting with classical communications in optical fiber," *Optica*, vol. 11, pp. 1700–1707, Dec 2024.

[25] J. Sobrinho, "An algebraic theory of dynamic network routing," *IEEE/ACM Transactions on Networking*, vol. 13, no. 5, pp. 1160–1173, 2005.

[26] M. Caleffi, "Optimal routing for quantum networks," *IEEE Access*, vol. 5, pp. 22299–22312, 2017.

[27] M. Thorup and U. Zwick, "Compact routing schemes," in *Proceedings of the Thirteenth Annual ACM Symposium on Parallel Algorithms and Architectures*, SPAA '01, p. 1–10, Association for Computing Machinery, 2001.

[28] A. Riedl and D. A. Schupke, "Routing optimization in ip networks utilizing additive and concave link metrics," *IEEE/ACM Trans. Netw.*, vol. 15, p. 1136–1148, Oct. 2007.

[29] L. Lovász, "On the ratio of optimal integral and fractional covers," *Discrete Mathematics*, vol. 13, no. 4, pp. 383–390, 1975.

[30] B. Awerbuch, A. Bar-Noy, N. Linial, and D. Peleg, "Compact distributed data structures for adaptive routing," in *Proceedings of the Twenty-First Annual ACM Symposium on Theory of Computing*, STOC '89, (New York, NY, USA), p. 479–489, Association for Computing Machinery, 1989.

[31] I. Abraham, C. Gavoille, D. Malkhi, N. Nisan, and M. Thorup, "Compact name-independent routing with minimum stretch," *ACM Trans. Algorithms*, vol. 4, July 2008.

[32] L. Grover, "A fast quantum mechanical algorithm for database search," in *ACM Symposium on Theory of Computing*, pp. 212–219, 1996.

[33] L. Kleinrock and F. Kamoun, "Hierarchical routing for large networks performance evaluation and optimization," *Computer Networks*, vol. 1, no. 3, pp. 155–174, 1977.

[34] S. Brito, A. Canabarro, R. Chaves, and D. Cavalcanti, "Statistical properties of the quantum internet," *Phys. Rev. Lett.*, vol. 124, p. 210501, May 2020.

[35] S. Brito, A. Canabarro, D. Cavalcanti, and R. Chaves, "Satellite-based photonic quantum networks are small-world," *PRX Quantum*, vol. 2, p. 010304, Jan 2021.

[36] G. Chiribella, G. M. D'Ariano, P. Perinotti, and B. Valiron, "Quantum computations without definite causal structure," *Phys. Rev. A*, vol. 88, Aug 2013.

[37] M. Caleffi and A. S. Cacciapuoti, "Quantum switch for the quantum internet: Noiseless communications through noisy channels," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 3, pp. 575–588, 2020.

[38] K. Simonov, M. Caleffi, J. Illiano, J. Romero, and A. S. Cacciapuoti, "Universal quantum computation via superposed orders of single-qubit gates," 2024.

[39] M. Caleffi, K. Simonov, and A. S. Cacciapuoti, "Beyond shannon limits: Quantum communications through quantum paths," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 8, pp. 2707–2724, 2023.