

# Latency of multipartite entanglement distribution in a quantum SDN architecture

A. Gore<sup>1\*</sup> and A. Beghelli<sup>1</sup>

<sup>1</sup> Optical Networks Group, Department of Electronic & Electrical Engineering, University College London, London, WC1E 7JE, UK

\*anuj.gore.22@ucl.ac.uk

**Abstract:** We quantify the latency to distribute multipartite entanglement over a quantum SDN architecture. Results show that classical communication can account for more than 80% of the delay, highlighting the need for efficient control plane design. © 2025 The Author(s)

## 1. Introduction

Quantum networks will enable the distribution of information via entanglement across long distances, supporting applications in enhanced security and quantum computing [1]. Multi-user quantum applications are enabled by multipartite entanglement distribution (MED) protocols. The latency of these MED protocols, i.e. the time taken to distribute a GHZ state, is key as it impacts both, network performance and feasibility.

Previous work on latency has mostly focused on bipartite entanglement distribution protocols [2–7]. Latency of MED protocols has also been investigated, but with some simplifying assumptions that affect the validity of results for wider networks, such as absence of repeaters [8]. Other works use the number of time-steps rather than actual time [8, 10] or assume that failure in a section restarts the whole system, which increases latency unnecessarily [9].

In this paper, we present a novel result for the MED protocol latency and show that classical communication strongly dominates in repeater-based quantum networks. We assumed a quantum Software-Defined Network (SDN) architecture where an SDN controller receives multipartite entanglement requests and orchestrates entanglement distribution among user nodes. Analytically validated Monte-Carlo simulation results show that, for an 8x8 grid network, classical communication accounts for more than 80% of the latency as soon as the link length exceeds 9 km, highlighting the need to design protocols with minimum control plane signalling requirements.

## 2. Quantum SDN Architecture Model

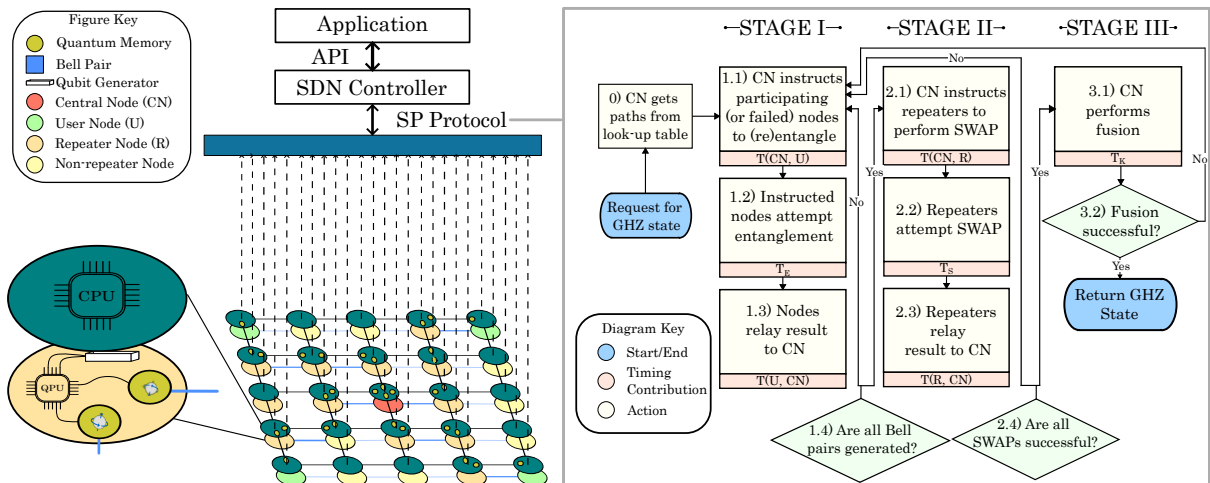


Fig. 1: Schematic of a quantum SDN architecture (left) and SP MED protocol (right).  $T(A, B)$ : signal propagation time from node  $A$  to  $B$ . CN: Central node. R: Repeater node. U: User node.  $T_x$ : time to execute operation  $x$ .

Fig. 1 (left) shows the SDN architecture used to evaluate the MED protocol latency. Network applications send multipartite entanglement requests to the SDN controller via an Application Programming Interface (API),

specifying the set of  $N$  users needed to be entangled by sharing a  $N$ -qubit Greenberger-Horne-Zeilinger (GHZ) state. The SDN controller communicates classically with the network nodes (vertically dashed lines) and triggers the execution of the MED protocol. This execution is delegated to the central node (CN), located at the centroid of the users. Once a GHZ state is generated, CN signals the SDN controller, which signals the network application.

In the network, each node is assumed to be made of two layers: a classical electronic layer and a quantum layer. The classical layer is equipped with a classical communication channel to every neighbour node and to its respective quantum layer. It also hosts a copy of the MED protocol. The quantum layer can generate Bell pairs with coherence time (lifetime)  $T_c$ , perform entanglement swapping (to enable long-distance entanglement) and fusion [11] with success probability  $p$ ,  $q$  and  $k$ , respectively. Heralded Bell pair generation is performed as described in [12].

The quantum device is connected to each neighbour node via a physical channel and is equipped with one quantum memory per physical edge to store entanglement. Each physical edge (for quantum and classical communication) is assumed to be a single mode optical fibre, with attenuation being the only source of loss.

### 3. Multipartite Entanglement Distribution (MED) Protocol

The execution of the single-path (SP) MED protocol is performed in 3 stages, as shown in Fig. 1 (right).

- \* STAGE I: Bell pairs are established between adjacent nodes on pre-calculated routes between CN and users.
- \* STAGE II: Entanglement swapping is performed in the repeater nodes to produce long-distance Bell pairs between the CN and the users. Failed swapping requires affected nodes to return to STAGE I. Stages I-II are repeated until all users share a long-distance Bell pair with the CN.
- \* STAGE III: Fusion operation is performed in the CN. If fusion fails, the protocol returns to STAGE I

### 4. Measuring quantum and classical delays: Methodology

To account for the probabilistic nature of the quantum operations, we run the MED protocol described in Fig 1 as a Monte-Carlo simulation 10000 times for different values of  $p$ ,  $q$  and  $k$  and count the total number of times entanglement ( $n_e$ ), swapping ( $n_s$ ) and fusion ( $n_f$ ) operations are attempted. We verify our Monte-Carlo simulation results with a Markov Chain model, extended from previous work developed for the bipartite case [2, 3]. Fig 2 shows the histogram of the total number of quantum operations required to distribute a GHZ state obtained via the Monte Carlo simulation in an  $8 \times 8$  grid topology with the users in the corners and  $T_c = \infty$ . The vertical lines show the excellent match between the Monte Carlo simulation and Markov Chain model mean values.

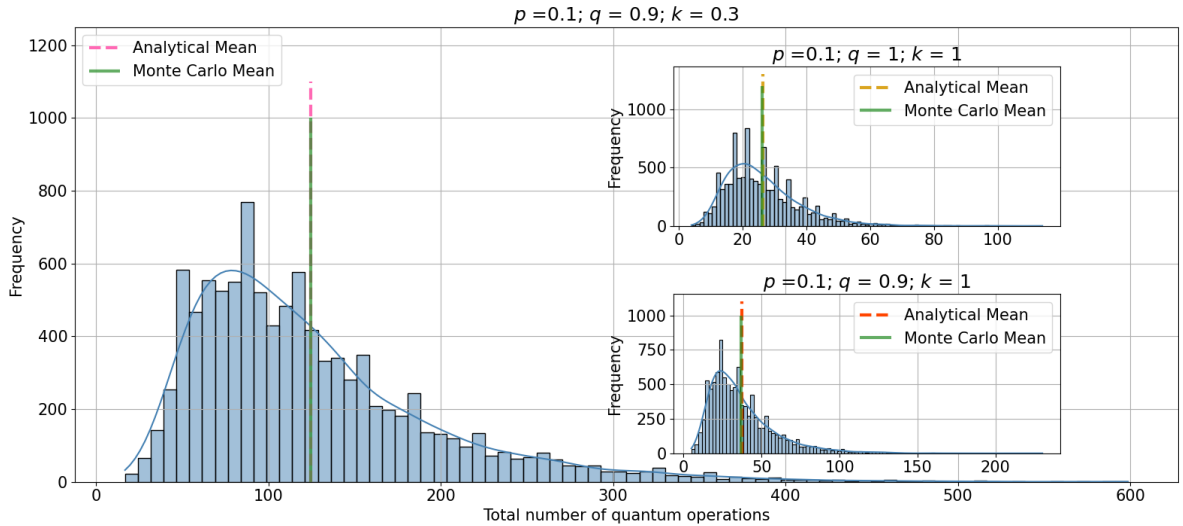


Fig. 2: Simulation histogram and simulation vs. analytical mean values of the number of quantum operations required by the MED protocol to generate a 4-user GHZ state in an  $8 \times 8$  grid.

### 5. Numerical Results

The verified Monte-Carlo simulation is then used to obtain latency results for the same grid topology, with distance between adjacent nodes  $L \in [1, 29]$  km. We assume a worst-case latency scenario: the users are located at the corner nodes; classical signal latency (signalling for entanglement and swapping) is calculated for the farthest

away node/user. We split the total latency between classical ( $\tau_{\text{classical}}$ ) and quantum ( $\tau_{\text{quantum}}$ ) latency.  $\tau_{\text{classical}}$  depends on  $n_e, n_s$  and the propagation time of classical signals. Considering the worst case, classical signals travel  $2(V-1)\frac{L}{c'}$  for the farthest user, and  $2(V-2)\frac{L}{c'}$  for the farthest repeater, with  $V$  is the number of vertices traversed from CN to the user,  $L$  the length (in km) of one edge and  $c'$  the speed of light in fibre. Similarly,  $\tau_{\text{quantum}}$  depends on  $n_e, n_s, n_f$  and photon propagation time from Bell pair generation. Entangling, swapping and fusion times are denoted by  $\tau_e, \tau_s$  and  $\tau_f$ .

Ref. [12] experimentally demonstrated a single repeater node by entangling a trapped ion and the photon emitted from the excited ion, where swapping is performed via the Mølmer-Sørensen (MS) gate. The Monte Carlo simulation uses the following values reported in [12]:  $p = 0.018 \times 10^{-\frac{0.2L}{10}}$ ;  $\tau_e = 123\mu s$  and  $\tau_s = 2157\mu s$  (without error correction). Ref. [13] generates a GHZ state from Bell pairs on trapped ion qubits by applying three 2-qubit gate rotations, with each gate operation taking  $100\mu s$ . Thus,  $\tau_f = 300\mu s$ . To account for experimental imperfections, we assume  $q = 0.7$  and  $k = 0.5$ . Our code is available at [14].

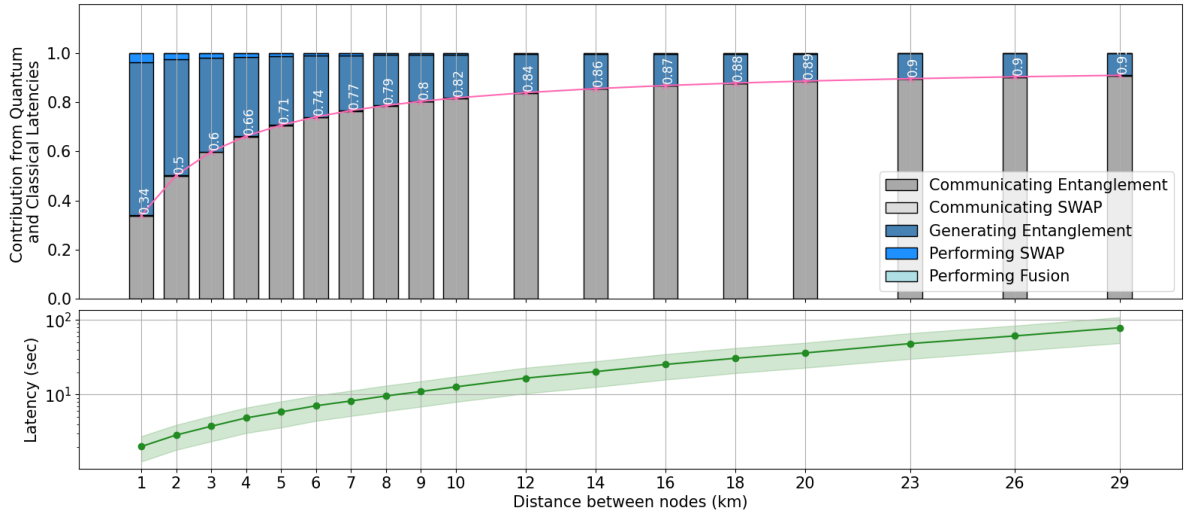


Fig. 3: Ratio between quantum (blue) and classical (grey) latency contribution (top plot) and total latency (bottom plot) as a function of link length for GHZ state generation between 4 corner users in an 8x8 grid topology.

The top plot of Fig. 3 shows the contribution of classical and quantum latency. It can be seen that for  $L > 2$  km the majority of the delay is due to classical communication (grey bars) used for entanglement signalling, accounting for  $\geq 80\%$  of latency for  $L \geq 9$  km. While quantum operation related variables depend on experimental details, our analysis (not shown here due to space constraints) shows the same overall trend: classical communication dominates latency. The bottom plot of Fig. 3 shows the total latency as a function of link length  $L$ . It can be seen that latency quickly reaches the second timescales with  $L \leq 6$  km. Given current experimental values of  $T_c$  for ion-based repeaters ( $^{40}\text{Ca}^+$ ) below 4 seconds [15], for link lengths  $> 4$  km, longer qubit coherence is needed in the studied network. These results highlight the need for protocol design with minimum control plane signalling.

**Acknowledgements.** Financial support from the NNTI Joint Lab, EPSRC Grant EP/W032643/1 and Innovate UK Quantum Data Centre of the Future (10004793) is gratefully acknowledged.

## References

1. Wehner et al *Science* 362.6412, 2018.
2. Shchukin et al, *Physical Review A* 100, 2019 .
3. Brand et al, *IEEE J. Sel. Areas Commun.* 38, 2020.
4. Khatri et al, *Phys. Rev. Res.* 1.2, 2019.
5. Kao et al, *npj Quantum Information* 10.73, 2024.
6. Pirandola et al, *Nature Communications* 8, 2017.
7. Pirandola et al, *Nature Communications* 2, 2019.
8. Avis et al, *Physical Review A* 107, 2023.
9. Bugalho et al, *Quantum* 7, 2023.
10. Bäuml et al, *Communications Physics* 3, 2020.
11. Nielsen et al, *Cambridge University Press*, 2010.
12. Krutyanskiy et al, *Phys. Rev. Lett.* 130, 21, 2023.
13. Kaufmann et al, *Phys. Rev. Lett.* 119, 15, 2017.
14. <https://github.com/AnujGore/Timing-Multiuser-Protocols>
15. Chen et al, *Chin. Phys. Lett.*, Accepted, 2024.