In-memory measurement-device-independent quantum key distribution

Nicolas Gauriot,[1,*] Samuel Morrissey,[1,2] Paul M. Burdekin,[3] Ilse Maillette de Buy Wenniger,[3] Sarah E. Thomas,[4] Ian A. Walmsley,[3] Robert I. Woodward,[1] and R. Mark Stevenson[1]

[1]*Toshiba Europe Limited, 208 Cambridge Science Park, Cambridge CB4 0GZ, United Kingdom*
[2]*School of Physics, Engineering, and Technology, University of York, York YO10 5FT, United Kingdom*
[3]*Department of Physics, Imperial College London, London SW7 2AZ, United Kingdom*
[4]*Department of Engineering Science, University of Oxford, Oxford OX1 3PJ, United Kingdom*

(Received 9 June 2026; revised 28 September 2026; accepted 15 October 2026; published 16 November 2026)

Quantum memories offer to significantly extend the performance of quantum networks, but their complexity and limited performance at telecommunication wavelengths have so far hindered practical deployment. We propose a memory-assisted measurement-device-independent quantum key distribution (MDI-QKD) protocol based on in-memory interference. Our approach leverages progress in atomic ensemble memories that enable the interference of a stationary matter qubit and a traveling photonic qubit inside a single quantum memory. We quantitatively evaluate the performance of this scheme against memoryless MDI-QKD and existing two-memory architectures, showing that it not only achieves superior secret key rates and surpasses the memoryless bound at substantially shorter distances, but also reduces hardware requirements by halving the number of memories and detectors. Our results further indicate that off-resonant cascaded absorption memories are a particularly promising platform for implementing this protocol, and we outline a concrete design compatible with their capabilities. These findings demonstrate that in-memory interference provides a higher performance and a resource efficient route toward near-term memory-enhanced quantum secure communications.

DOI: 10.1103/wdz1-sbgg

I. INTRODUCTION

As quantum key distribution (QKD) matures and the expansion of deployed real-world quantum-secure networks continues, one of the key challenges that remain is extending secure communication over longer distances. In point-to-point implementations, the secret key rate (SKR) is fundamentally limited by the repeaterless secret key capacity bound (SKC_0) [1]. This bound scales linearly with the channel transmissivity, which decreases exponentially with distance, leading to a rate-distance trade-off that constrains the practical reach of QKD links.

Quantum repeaters based on quantum memories are an effective solution to this problem [2,3]. By distributing entanglement in smaller segments of the channel and using entanglement swapping, quantum communication can be performed over arbitrary distances. Memories are required in this architecture to store states until entanglement distribution in adjacent segments succeeds. However, classical signaling between nodes imposes long memory lifetimes, on the order of the total link's communication delay [4,5]. The development of a quantum repeater that can compete against direct transmission, therefore, requires memory lifetimes beyond the reach of currently available practical systems. As a result its deployment in real world networks remains a distant goal.

Quantum memories could still enhance near-term QKD technology. For example, memory-assisted measurement-device-independent QKD (MA-MDI-QKD) [6–9] can be beneficial with much less demanding requirements on the memory performance, while also providing a stepping stone toward a quantum repeater. It is based on the architecture of the synchronous Bell-state-measurement (BSM)-based MDI-QKD [10,11]. In this latter protocol, two users, Alice and Bob, send phase-randomized weak coherent pulses (WCPs), encoded in at least two mutually unbiased bases, to a middle node, Charlie. Charlie performs a BSM to project the joint state onto one of the four Bell states. This architecture is depicted in Fig. 1(a). Notably, this protocol is measurement-device-independent (MDI) as Charlie, who performs the measurement, is not trusted. This has significant implementation security consequences as it removes all side-channel attacks on the detectors. Although this protocol introduces a central node, the key rate scales as a standard point-to-point protocol and is limited by the SKC_0 bound. This is because a successful BSM requires both pulses to arrive simultaneously at Charlie, which is as likely as a pulse traveling all the way from Alice to Bob.

In the memory-assisted version [6,7] [see Fig. 1(b)], Charlie stores the pulses he receives from Alice and Bob in two separate memories. Once both memories are loaded, the pulses are retrieved and sent to Charlie's measurement station for the BSM. In this way Alice and Bob are synchronized at Charlie, and pulses sent in different clock cycles can interfere. It has been shown that this protocol can outperform the

*Contact author: nicolas.gauriot@toshiba.eu

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

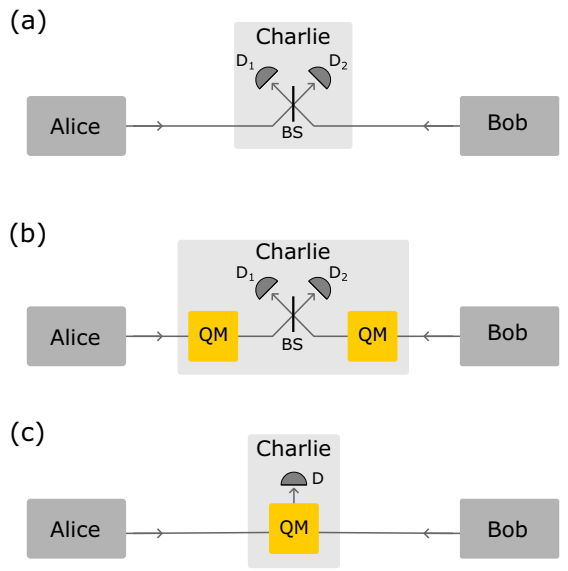


FIG. 1. Schematic architecture for (a) the MDI-QKD protocol, (b) the two-memory MA-MDI-QKD protocol, and (c) the in-memory protocol. QM, quantum memory; BS, beam splitter; D, detector.

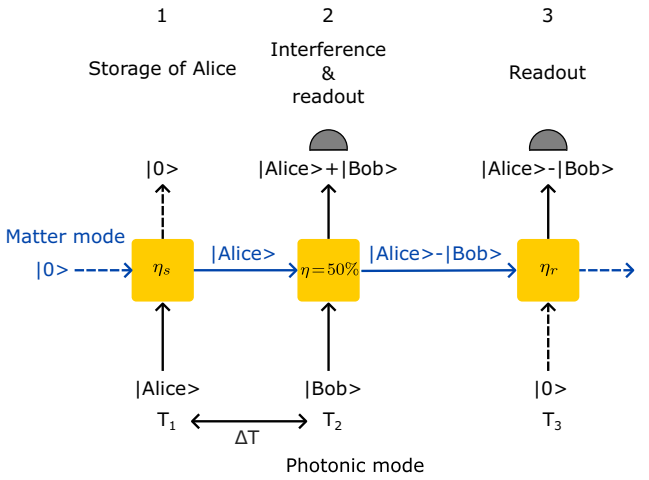


FIG. 2. Schematic of the in-memory Bell state measurement sequence.

memoryless version at certain distances depending on lifetime and efficiencies of the memories [6–8,12]. For infinite lifetime, the rate of successful BSM at Charlie is now proportional to the probability that a single photon travels through half of the link. The SKR then scales with $\sqrt{\eta}$, potentially allowing for a doubling of the maximum communication distance.

For this protocol to be advantageous the memories need to have a large time-bandwidth product so that their lifetime is long compared to the repetition period of the transmitters. A spectral acceptance bandwidth large enough to store the short pulses (≈100 ps) of state-of-the-art QKD transmitters [13–15] is also required, as slowing down the transmitter's repetition rate risks losing any advantage provided by the memories. With GHz bandwidth and low noise, memories based on atomic ensembles have appeared as promising candidates for this application [16–23]. However, their complexity and limited performance at standard telecommunication wavelengths have so far hindered practical deployment.

Typically, the BSM is performed with linear optics, in which case a maximum of two Bell states can be distinguished [24]. It is convenient to use the two antisymmetric states $\psi^{\pm} = \frac{1}{\sqrt{(2)}}(|0,1\rangle \pm |1,0\rangle)$, then the BSM simply amounts to an interference on a beam splitter followed by a measurement of the two beam-splitter outputs. Remarkably, some of the atomic ensemble memories have been shown to behave as a beam splitter with tunable reflectivity, allowing interference between a stored matter mode and a propagating optical mode [25–29].

The two input modes of this effective beam splitter are the matter mode already in the memory (which could be empty or populated by a state previously loaded) and the incoming photonic mode. The output modes are the photonic mode coming out of the memory and the matter mode stored in the memory. The reflectivity of the effective beam splitter corresponds to the storage and retrieval efficiencies of the memory and can be tuned by a control pulse.

Here we propose a MA-MDI-QKD protocol based on this beam-splitter analogy. Our protocol uses only one memory and half the number of detectors compared to the two-memory version, while providing higher performance. By hosting this memory in a central, untrusted, location, incoming photons from Alice and Bob can be interfered "in-memory," even if arriving within different clock cycles [Fig. 1(c)]. We simulate the achievable SKR of our protocol and critically evaluate it against the memoryless MDI and two-memory MA-MDI-QKD protocols. We also consider a specific practical implementation with an off-resonant cascaded absorption (ORCA) memory.

II. PROTOCOL DESCRIPTION

In the following we compare three protocols: the original memoryless MDI protocol [10,11], the two-memory assisted MDI protocol [6,7], and our protocol which we refer to as "in-memory MDI-QKD." In all three protocols, Alice and Bob send phase-randomized WCPs to Charlie, who performs a two-particle interference. In all cases Charlie is not trusted. Security is then provided by the same postselected reversed entanglement principle for all the protocols [30]. In the memoryless protocol, interference occurs only when pulses from Alice and Bob arrive simultaneously at Charlie. These are interfered on a standard beam splitter, whose two output modes are measured simultaneously using detectors determined by the encoding. In the two-memory protocol, pulses are stored in separate memories and later released for interference and measured using the same Bell-state analyzer as in the memoryless case. In our protocol, only one memory is used and the two qubits are interfered in-memory. Figure 2 presents the protocol steps in more detail. The BSM part of the protocol proceeds in three main steps.

(i) A pulse is received from Alice (Bob) and stored in the memory with the memory efficiency set to its maximum.

(ii) A pulse is received from Bob (Alice) and stored in the memory. Here the memory efficiency is nominally set to

50%. The memory then behaves as a beam splitter, and it is at this step that the interference occurs. The photonic state at the output of the memory is measured on a set of detectors. The spin wave that is left in the memory corresponds to the other output of the beam splitter. If the maximum storage and retrieval efficiencies are less than 50%, the protocol can still proceed; however, the beam splitter is no longer symmetric, which results in a reduction of the SKR. We investigate in more detail this effect in the Supplemental Material [31].

(iii) The memory is emptied, and the output state is measured on the same set of detectors as earlier. Here the readout efficiency can be set to its maximum or left unchanged. If the efficiency is less than 100%, the memory can be read multiple times to increase the effective readout efficiency [27]. In the following, we assume that the readout efficiency at this step is effectively 100%.

Because the two output modes of the beam splitter are not measured simultaneously but sequentially on the same set of detectors, our protocol requires half the number of detectors compared to the other two protocols. For example, for time-bin qubits, one detector is required instead of two, while for polarization-encoded qubits two detectors are required instead of four. Once the quantum communication stage is finished, Alice and Bob proceed with sifting, error correction, and privacy amplification in the same way for all three protocols.

Careful timing and synchronization of the signal and control pulses are required to ensure that the pulses are stored, interfered, and read out as intended. In particular, during steps (i) and (ii) the memory should receive only the pulse from the intended user, and no input signal should be present during the final readout [step (iii)]. This could be achieved in various ways. For example, optical switches may be used to selectively accept or reject pulses from Alice and Bob. Alternatively, signals at the input of the memory can be rejected by not applying a control pulse. In this case, the memory is transparent to the signal and the pulse simply passes through it. Coordination of the arrival time of the different pulses can ensure that Alice's and Bob's pulses arrive at Charlie in different time slots and that the final readout [step (iii)] is performed when no signal pulses are expected to arrive.

III. PROTOCOL PERFORMANCE

We now simulate and compare the SKR achievable with the three protocols. In Sec. III A we first consider the case where a heralding mechanism informs Charlie of successful storage or arrival of a pulse. We consider the case where such a mechanism is not built in the memory in Sec. III B. For simplicity we use polarization-encoded qubits but the main results remain valid for any other encoding. We also use the efficient BB84 encoding scheme where Alice and Bob use the Z basis (H and V) as the coding basis, while the X basis ($|H+V\rangle, |H-V\rangle$) is used much less frequently and exclusively for quantum bit error rate (QBER) estimation and security checks [32]. The numerical results presented below assume a state-of-the-art system clocked at 1 GHz, a single-photon avalanche photodiode with 15% efficiency, and a dark counts rate of 1 kHz. We also assume a 0.5% misalignment probability and 0.2 dB/km loss in the optical fibers. Details of the SKR calculations are provided in the Supplemental Material [31] (see also Refs. [33,34] therein).

A. Directly heralding memories

With the decoy state method [35] and in the asymptotic limit, the key rate for all three protocols is given by [36]

$$\mathrm{SKR} = Q_{11}[1 - H(e_{11,x})] - fQ_{\mu\nu}H(E_{\mu\nu,z}), \tag{1}$$

where Q_{11} is the single-photon gain, i.e., the probability that both Alice and Bob send single photons and that a successful BSM is performed at Charlie. $Q_{\mu\nu}$ is the overall gain, i.e., the probability of successful a BSM when Alice and Bob use WCPs with μ and ν mean photon numbers. $e_{11,x}$ is the single-photon QBER measured in the X basis, $E_{\mu\nu,z}$ is the overall QBER in the Z basis, and H is the binary Shannon entropy function. $f = 1.16$ is the the error correction efficiency factor.

We assume that each memory stores a pulse with an efficiency η_s and releases it with an efficiency $\eta_r(\Delta T) = \eta_{r_0} e^{-\frac{\Delta T}{\tau}}$ where ΔT is the storage time and τ is the memory lifetime. We refer to η_s as the storage efficiency and η_{r_0} as the retrieval efficiency. Both memory-assisted protocols allow to interfere qubits received in different clock cycles, so the rate will follow the same scaling with distance [$\mathcal{O}\sqrt{(\eta)}$]. However, for a given lifetime and memory efficiency the in-memory protocol will lead to a higher key rate. This can be understood as follows. Consider the two-memory case when Charlie receives a pulse from Alice first. Alice's pulse gets stored for a time ΔT before being released, while Bob's pulse is stored and released right away. So, Alice's pulse suffers a loss of $\eta_{a_2} = 1 - \eta_s \eta_{r_0} e^{-\frac{\Delta T}{\tau}}$, while Bob's pulse suffers a loss pf $\eta_{b_2} = 1 - \eta_s \eta_{r_0}$. In contrast, in the in-memory case Alice's pulse incurs a loss of $\eta_{a_1} = 1 - \eta_s e^{-\frac{\Delta T}{\tau}}$ and Bob's pulse does not suffer any losses. Therefore, when the storage and retrieval efficiencies are ideal ($\eta_s = \eta_{r_0} = 100\%$), both protocols yield the same key rate. However, under realistic conditions, the in-memory protocol outperforms the two-memory version. This is demonstrated in Fig. 3, which compares the achievable key rates for both protocols assuming infinite memory lifetimes. Figure 4 illustrates the key rate achievable with our protocol, across varying memory lifetimes and efficiencies. The corresponding comparison for the two-memory protocol is shown in Supplemental Fig. S2 [31]. At a given distance our protocol provides a higher key rate, and for a given memory efficiency it surpasses the memoryless protocol at a shorter distance. For example, with 50% storage and readout efficiency and a 10-ns lifetime, our protocol surpasses the memoryless SKR after 50 km (which is more than 150 km earlier than the two-memory protocol), enabling significantly improved key rates and ultimately longer distances.

B. Indirectly heralding memories

A key requirement of the memory-assisted protocols described above is that either the successful storage of a pulse in the memory or its arrival at the memory input be heralded. While some memories do not offer intrinsic direct heralding mechanisms, it is still possible to consider them for MA protocols using an indirect scheme. Illustrated in Fig. 5(a), a

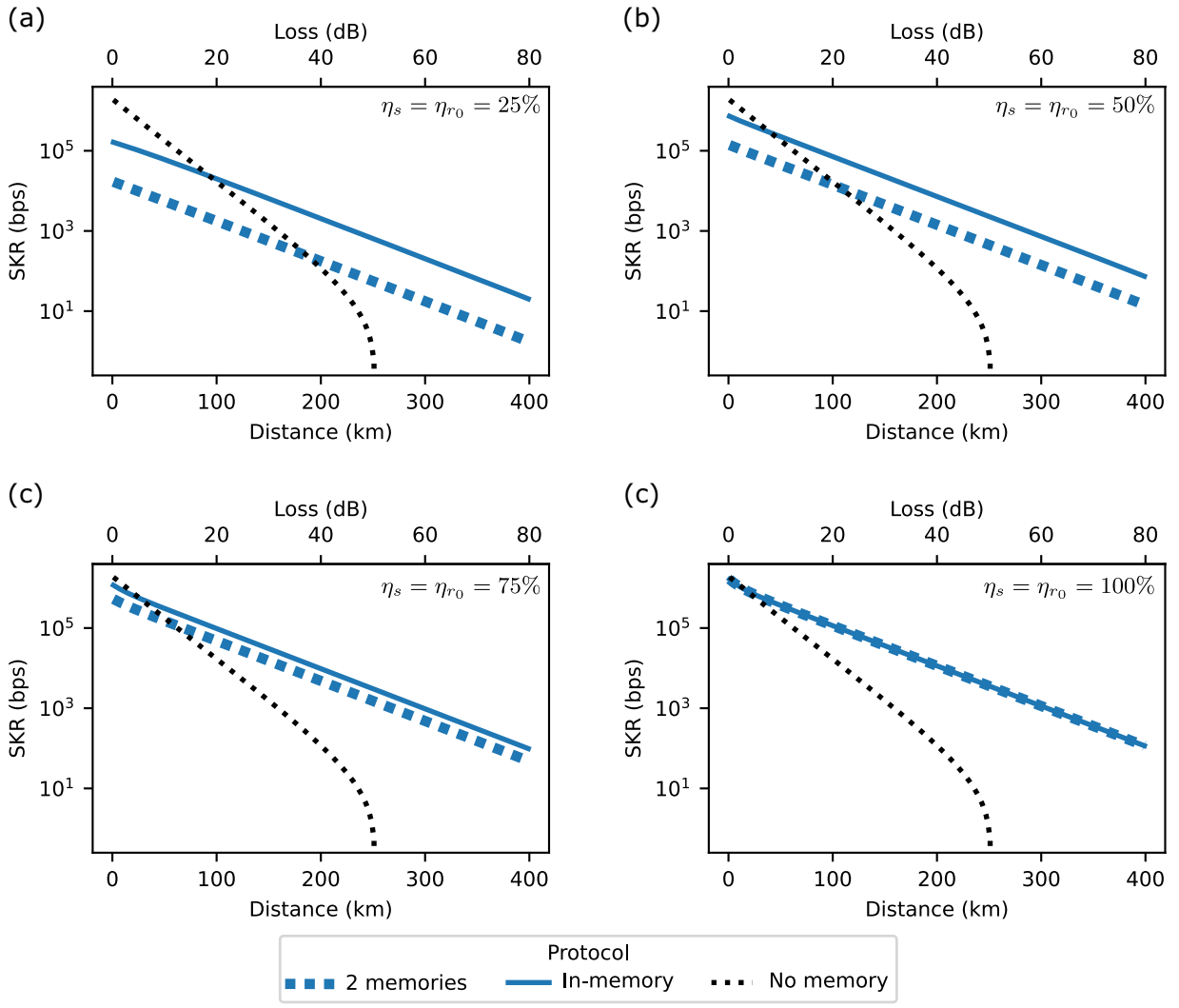


FIG. 3. Comparison of the SKR between the memoryless, two-memory and in-memory protocols for (a) 25%, (b) 50%, (c) 75%, and (d) 100% storage/retrieval efficiencies ($\eta_s = \eta_{r_0}$).

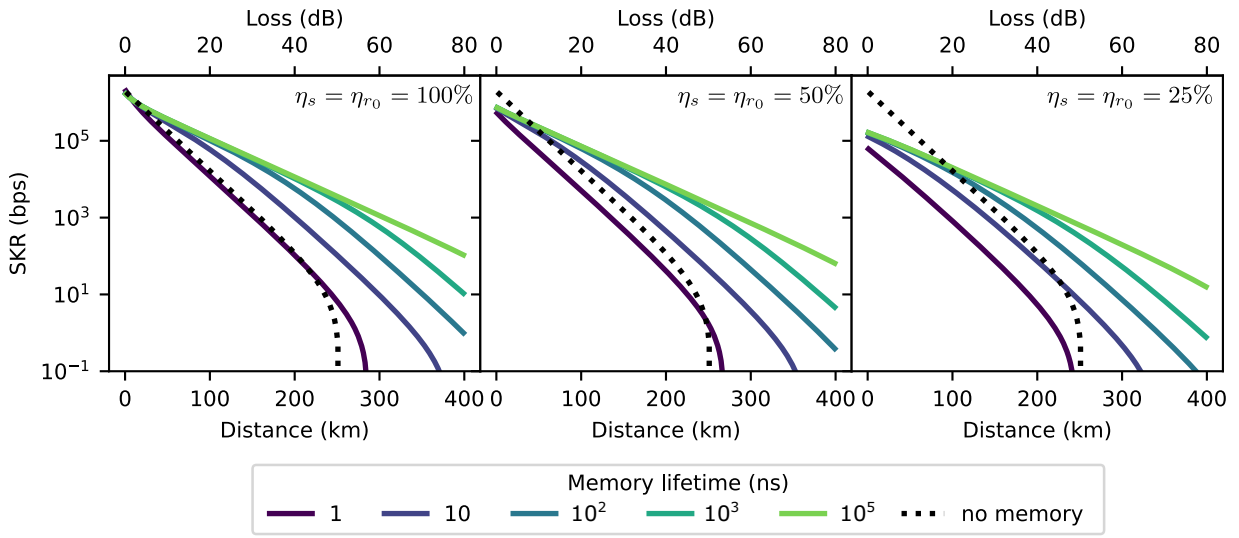


FIG. 4. Comparison of the secret key rate of the in-memory protocol for different memory lifetimes and memory efficiencies and the memoryless protocol.

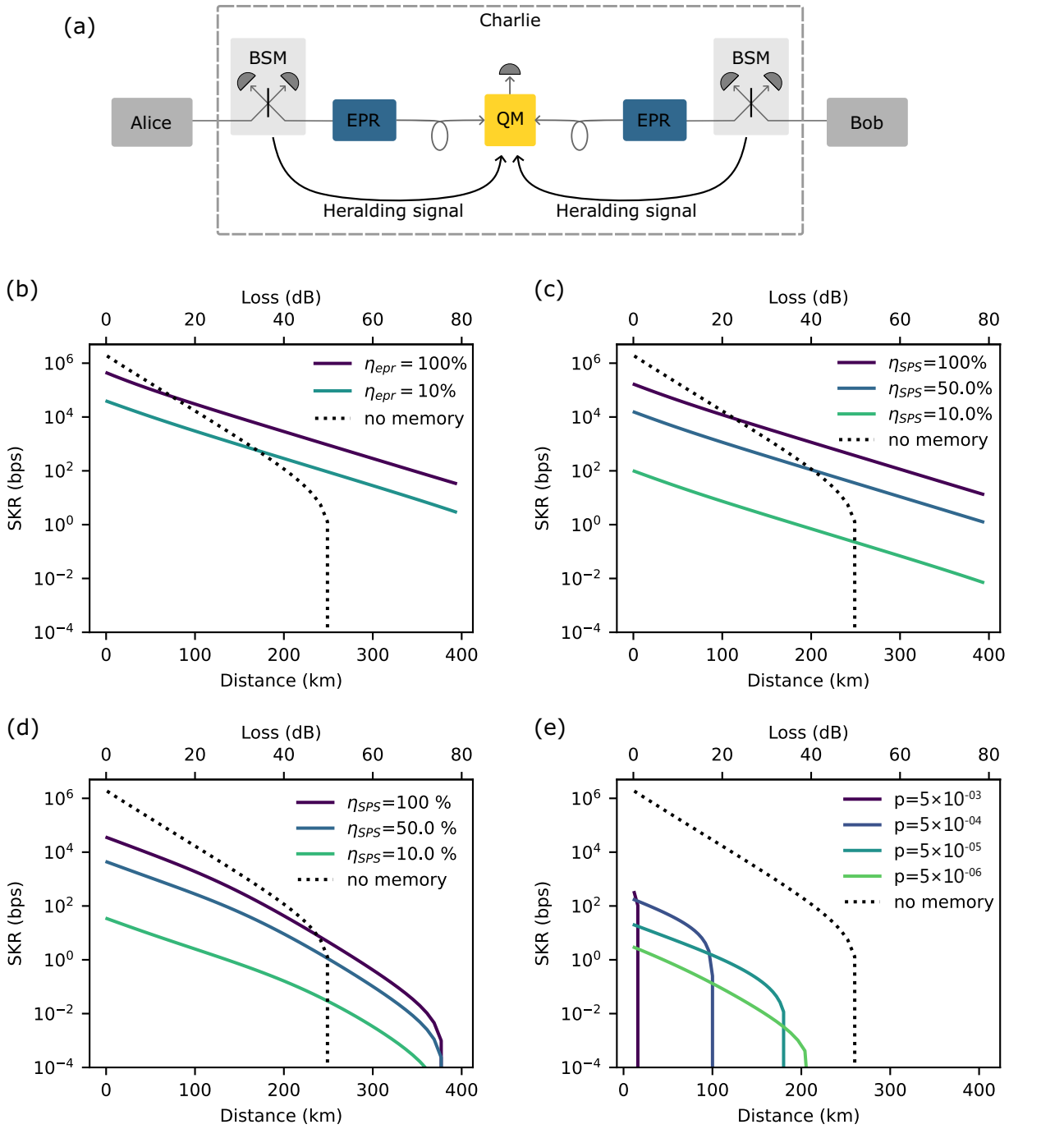


FIG. 5. (a) Schematic network architecture with indirect heralding provided by side Bell-state measurements. The resulting SKR strongly depends on the type of entanglement source used in the side BSM (EPR). (b)–(e) Comparison of the SKR for the memoryless protocol and the in-memory protocol with ideal memories for (b) a "true EPR" source with an efficiency η_{epr}, (c) a "quasi-EPR" source built with two single-photon sources with η_{SPS} efficiency, (d) A noiseless linear amplifier built with single-photon sources of efficiency η_{SPS}, and (e) an SPDC source.

potential solution is to herald the arrival of a photon via a BSM performed before the memory to teleport the traveling qubit to the memory [7,8,12]. The performance of this teleportation step and its effect on the whole protocol depend on the characteristics of the entanglement source [8,12]. To assess the suitability of this approach with our protocol, we have investigated different types of entanglement sources and we compare the SKR obtained in each case with ideal memories in Fig. 5(b)–5(e). We begin with an idealized entanglement source and then consider increasingly realistic implementations that reflect current or near-term technological capabilities.

1. True EPR source

This source emits a maximally entangled photon pair with η_{epr} efficiency. The density matrix of the emitted state is

$$\rho_{\text{epr}} = (1-\eta_{\text{epr}})|0\rangle\langle 0| + \eta_{\text{epr}}|\psi\rangle\langle\psi|, \tag{2}$$

where ψ is one of the four Bell states. With a 100% efficient source, a perfect Bell state is emitted at every clock cycle, and the side BSM succeeds 50% of the time. Therefore, with an infinite memory lifetime, the SKR is simply divided by two compared to the directly heralded case. As shown in Fig. 5(b), the repeater scaling is preserved and the memoryless bound is surpassed at a distance that depends on the efficiency of the EPR sources used. Quantum dots can be used to generate these entangled photon pairs at telecommunication wavelengths [37–39]. However, although they are rapidly advancing, their development remains a technological challenge, which motivates the exploration of alternatives.

2. Quasi-EPR source

Single-photon sources at telecommunication wavelenghts are more mature [40–42] and can be used to generate approximations to the ideal EPR source. Here we propose a quasi-EPR source based on the interference of two single photons on a 50:50 beam splitter. The two photons should be indistinguishable in all their degrees of freedom except their polarization. With one photon in the $|D\rangle = |H + V\rangle$ polarization state and the other in the $|A\rangle = |H - V\rangle$ state, the resulting state after interference is

$$\psi_{\text{quasi}} = [(c_h^\dagger)^2 - (d_h^\dagger)^2 + (c_v^\dagger)^2 - (d_v^\dagger)^2 + 2(d_v^\dagger c_h^\dagger - c_v^\dagger d_h^\dagger)]|0\rangle, \tag{3}$$

where $c_h^\dagger$ and $d_h^\dagger$ ($c_v^\dagger$ and $d_v^\dagger$) are the creation operators in the two output modes of the beam splitter for H-polarized (V-polarized) photons. We then obtain a photon statistic similar to that of the quasi-EPR source proposed in Ref. [12]. This is not a perfect Bell state, and as a result it leads to lower performance than the ideal EPR source as shown in Fig. 5(c). Nevertheless, it can still overcome the memoryless bound.

3. Noiseless linear amplifier

An alternative way to herald the arrival of pulse with two single-photon sources is based on the heralded qubit amplifier [43]. In this case an H-polarized and a V-polarized photon are reflected off an asymmetric beam splitter. The reflected mode is used for the BSM while the transmitted mode is sent to the memory. Figure 5(c) shows the resulting SKR when such a source is used with a beam splitter with 0.2 transmission, for different efficiency of the single-photon sources (η_{SPS}). While this source cannot outperform the memoryless protocol at short distance, it could extend the range of QKD beyond 250 km, with efficient enough single-photon sources.

4. Spontaneous parametric down-conversion source

Entanglement sources based on spontaneous parametric down-conversion (SPDC) can be convenient to use as they are widely available and operate at room temperature. They emit a state characterized by the density matrix:

$$\rho_{\text{spdc}} = (1 - p)\sum_{n=0}^{\infty} p^n |\Psi_n\rangle\langle\Psi_n|, \tag{4}$$

with

$$|\Psi_n\rangle = \frac{1}{\sqrt{n!}}(\hat{a}_H^\dagger \hat{b}_V^\dagger + \hat{a}_V^\dagger \hat{b}_H^\dagger)^n |0\rangle. \tag{5}$$

This source emits a single pair of entangled photons with probability $\approx p$ but also n pairs with probability $\approx p^n$. This parameter can be controlled in practice by varying the power of the pump laser. A high emission rate is desirable to obtain a high gain but multipairs introduce noise which increases the QBER and reduces the SKR [8]. As shown in Fig. 5(e), no value of p allows this source to surpass the memoryless bound. We note that this is not specific to the in-memory protocol and a similar conclusion is reached in the two-memory case [8].

This survey of entanglement sources demonstrates that heralding the arrival of qubits at Charlie with a teleportation protocol is an effective solution and retains the advantages of our protocol. However it will require advanced sources either providing high-fidelity single-photon pairs or based on multiple single-photon sources.

IV. IMPLEMENTATION WITH THE ORCA MEMORY

Ensemble-based quantum memories operated in a Raman configuration can be described as time-dependent beam splitters with control-field-tunable reflectivity [26,29,44]. In the short cell limit, the ORCA memory is equivalent to the Raman memory and thus can likewise be understood within this effective beam-splitter picture. The telecom-ORCA memory provides a particularly compelling platform for our in-memory MDI-QKD protocol. Operating with a GHz acceptance bandwidth in the telecom C-band, it has demonstrated storage and retrieval efficiencies of 69.13(1)% and 30.23(1)% respectively, while adding negligible excess noise ($<10^{-6}$ photons per memory operation) [45]. This low noise operation is a vital prerequisite to obtaining high-fidelity BSM. In addition, its room-temperature operation eliminates the need for cryogenic cooling or high-vacuum infrastructure, and its compatibility with simple laser systems makes it uniquely suited for deployment in existing fiber communication networks. Importantly, the memory has also shown to be compatible with solid-state quantum dots, a leading platform for deterministic quantum light generation [22]. Motivated by the close match between these characteristics and the requirements of our in-memory protocol, we detail a practical implementation based on the telecom-ORCA memory, which is illustrated in Fig. 6.

Alice and Bob asynchronously send polarization-encoded weak coherent pulses to a central node, Charlie. At Charlie, the pulses are combined on a polarization beam splitter followed by a dual-rail ensemble-based quantum memory. Control pulses are generated when a heralding signal is received from the Bell-state measurement performed before the memory stage [see Fig. 5(a)]. As shown in Fig. 6, orthogonal input polarizations are mapped onto the same spatial mode of the memory: a horizontally polarized photon from Bob copropagates in the same rail as a vertically polarized photon from Alice, and vice versa. As a result, two photons with orthogonal polarizations would not, in general, interfere within a single rail. To enable interference in this case, the polarization of the control field used during the second memory interaction is rotated. In the absence of a second input photon, this oper-

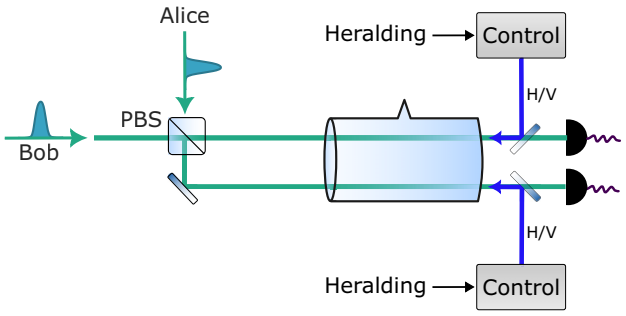


FIG. 6. Proposed memory-assisted MDI-QKD with an ensemble-based memory. A PBS maps incoming polarization-encoded photons onto the two spatial modes of a dual-rail ensemble-based quantum memory. Heralded signals from BSM performed before the memory generate control pulses for each rail. The first photon (Alice) is stored, while the second photon (Bob) interferes with the stored excitation via a memory beam-splitter interaction. Correlated detection events in the time domain herald successful Bell-state measurements.

ation causes the initially stored excitation to be retrieved with a polarization orthogonal to that in which it was stored, effectively implementing a rotation of the memory qubit. When a second photon is present with polarization orthogonal to the first, the rotated control field acts to transform the stored excitation such that its polarization becomes aligned with that of the incoming photon. Consequently, the two optical modes are rendered indistinguishable within the memory, allowing them to interfere during the beam-splitter-like interaction. Importantly, this feedforward operation renders the protocol insensitive to the order in which the photons arrive.

The first pulse from Alice (Bob) is stored via the two-photon absorption process ORCA, mediated by a strong control field, and mapped onto a collective matter excitation in the ensemble, forming a spatially distributed coherence (orbital wave). Separately, Bob (Alice) prepares their photonic state and subsequently sends it to Charlie. During the second memory interaction, the control field intensity is set such that the mapping efficiency is 50% and the results of two processes interfere: the partial retrieval (nonretrieval) of the initially stored coherence and the partial transmission (storage) of the incoming second photonic field. This interaction is analogous to interference on a balanced beam splitter, where the two output modes are given by the photonic field and the collective orbital-wave excitation [29,44]. Afterwards, the joint system state is projected onto a coherent superposition between a bunched orbital-wave excitation and a bunched retrieved optical field. The orbital-wave excitation is subsequently mapped back onto an optical field at a later time. Successful Bell-state projections are then heralded by correlated detection events across two well-defined temporal modes corresponding to the directly retrieved optical field (during the first memory interaction) and the delayed retrieval from the memory. In this way, the effective coincidence detection required for Bell-state measurements is realized in the time domain and across the two detectors, without requiring temporal overlap of the input photons.

As demonstrated above, with 50% storage and retrieval efficiencies and a 1-GHz clock rate, a memory lifetime of several nanoseconds is required to surpass the memoryless bound, and the need for indirect heralding will increase this threshold by severalfold. However, in its original implementation, the storage time of the telecom-ORCA memory is limited by Doppler dephasing, resulting in a lifetime of approximately 1 ns [45].

To address this limitation, we recently proposed a rephasing scheme that exploits Doppler dephasing to extend the storage time of the telecom-ORCA memory [46]. This approach extends the memory lifetime by more than a factor of 50. With Doppler dephasing suppressed, the dominant limitations to the memory lifetime arise from hyperfine beating and spontaneous emission from the electronically excited state. Nevertheless, even in this first implementation, the improved memory lifetime would already enable enhanced secret key rates compared to memoryless protocols at longer transmission distances. Suppressing hyperfine beating, either through optical pumping or by using magnetic fields to spectrally resolve individual hyperfine states, could extend the memory lifetime to approximately 140 ns. Beyond this, ground-state mapping of the stored optical excitation offers a route to orders-of-magnitude further improvement, as coherence times ranging from microseconds to seconds have been demonstrated in other atomic systems.

V. CONCLUSION

In this work we have introduced a new MDI-QKD protocol based on in-memory interference, achieving a $O(\sqrt{\eta})$ rate-distance scaling. By requiring only one memory and a single detector set, the protocol reduces hardware demands relative to the standard dual-memory architecture. It delivers significantly higher key rates and surpasses the memoryless protocol at much shorter distances. A central conceptual advance of this work is that the Bell-state measurement is performed between a matter qubit and a photonic qubit rather than relying on interference between two photons. Performing the interference directly inside the memory, between two nonidentical particles, avoids the losses associated with matter-photon conversion and provides performance advantages even with modest memory efficiencies. The simulations presented in this work correspond to symmetric links in the asymptotic regime. However, the advantages of our protocol are expected to persist in both finite-size scenarios and asymmetric network configurations [47–49] since the underlying mechanism of loss reduction will remain. We also note that other sources, for example, single-photon sources, could be used instead of weak coherent states. The architecture presented here aligns well with ongoing experimental progress in atomic ensemble memories. In particular, the combination of GHz acceptance bandwidth, low noise, and near room-temperature operation makes ORCA memories a promising platform for realizing the protocol, and we have outlined a concrete implementation compatible with their demonstrated capabilities. These results show that in-memory interference offers a practical and resource efficient pathway toward near-term QKD enhancements, lowering the barrier for deploying quantum memories within real-world quantum secure networks.

ACKNOWLEDGMENTS

S.M. gratefully acknowledges funding from the Engineering and Physical Sciences Research Council, as well as useful discussions with academic supervisor M. Lucamarini.

DATA AVAILABILITY

There are no publicly available research data or software supporting this manuscript. Requests for further information or data should be sent to the authors.

[1] S. Pirandola, R. Laurenza, C. Ottaviani, and L. Banchi, Fundamental limits of repeaterless quantum communications, Nat. Commun. **8**, 15043 (2017).

[2] L.-M. Duan, M. D. Lukin, J. I. Cirac, and P. Zoller, Long-distance quantum communication with atomic ensembles and linear optics, Nature (London) **414**, 413 (2001).

[3] H.-J. Briegel, W. Dür, J. I. Cirac, and P. Zoller, Quantum repeaters: The role of imperfect local operations in quantum communication, Phys. Rev. Lett. **81**, 5932 (1998).

[4] N. Sangouard, C. Simon, H. de Riedmatten, and N. Gisin, Quantum repeaters based on atomic ensembles and linear optics, Rev. Mod. Phys. **83**, 33 (2011).

[5] K. Azuma, S. E. Economou, D. Elkouss, P. Hilaire, L. Jiang, H.-K. Lo, and I. Tzitrin, Quantum repeaters: From quantum networks to the quantum internet, Rev. Mod. Phys. **95**, 045006 (2023).

[6] S. Abruzzo, H. Kampermann, and D. Bruß, Measurement-device-independent quantum key distribution with quantum memories, Phys. Rev. A **89**, 012301 (2014).

[7] C. Panayi, M. Razavi, X. Ma, and N. Lütkenhaus, Memory-assisted measurement-device-independent quantum key distribution, New J. Phys. **16**, 043005 (2014).

[8] N. L. Piparo, M. Razavi, and C. Panayi, Measurement-device-independent quantum key distribution with ensemble-based memories, IEEE J. Sel. Top. Quantum Electron. **21**, 138 (2015).

[9] M. K. Bhaskar, R. Riedinger, B. Machielse, D. S. Levonian, C. T. Nguyen, E. N. Knall, H. Park, D. Englund, M. Lončar, D. D. Sukachev, and M. D. Lukin, Experimental demonstration of memory-enhanced quantum communication, Nature (London) **580**, 60 (2020).

[10] H.-K. Lo, M. Curty, and B. Qi, Measurement-device-independent quantum key distribution, Phys. Rev. Lett. **108**, 130503 (2012).

[11] S. L. Braunstein and S. Pirandola, Side-channel-free quantum key distribution, Phys. Rev. Lett. **108**, 130502 (2012).

[12] N. Lo Piparo, N. Sinclair, and M. Razavi, Memory-assisted quantum key distribution resilient against multiple-excitation effects, Quantum Sci. Technol. **3**, 014009 (2018).

[13] G.-W. Zhang, S.-T. Zheng, Y. Xiao, F.-X. Wang, W.-J. Ding, D. Wang, P. Hao, L. Zhang, J.-L. Chen, Y.-Y. Ding, S. Wang, D.-Y. He, Z.-Q. Yin, Z. Zhou, H. Li, L. You, G.-C. Guo, W. Chen, and Z.-F. Han, 5-GHz chip-based quantum key distribution with 1Mbps secure key rate over 150 km, Laser Photonics Rev. **20**, e03250 (2026).

[14] G.-J. Fan-Yuan, W.-X. Xie, D.-Y. He, X.-J. Huang, S. Wang, Z.-Q. Yin, W. Chen, Q. Li, H.-K. Mao, G.-W. Zhang, Y.-L. Wang, G.-C. Guo, and Z.-F. Han, High-rate quantum key distribution with compact state preparation and detection, Proc. Natl. Acad. Sci. USA **123**, e2521590123 (2026).

[15] Z. Yuan, A. Murakami, M. Kujiraoka, M. Lucamarini, Y. Tanizawa, H. Sato, A. J. Shields, A. Plews, R. Takahashi, K. Doi, W. Tam, A. W. Sharpe, A. R. Dixon, E. Lavelle, and J. F. Dynes, 10-Mb/s quantum key distribution, J. Lightwave Technol. **36**, 3427 (2018).

[16] P. S. Michelberger, T. F. M. Champion, M. R. Sprague, K. T. Kaczmarek, M. Barbieri, X. M. Jin, D. G. England, W. S. Kolthammer, D. J. Saunders, J. Nunn, and I. A. Walmsley, Interfacing GHz-bandwidth heralded single photons with a warm vapour Raman memory, New J. Phys. **17**, 043006 (2015).

[17] J. Wolters, G. Buser, A. Horsley, L. Béguin, A. Jöckel, J.-P. Jahn, R. J. Warburton, and P. Treutlein, Simple atomic quantum memory suitable for semiconductor quantum dot single photons, Phys. Rev. Lett. **119**, 060502 (2017).

[18] G. Buser, R. Mottola, B. Cotting, J. Wolters, and P. Treutlein, Single-photon storage in a ground-state vapor cell quantum memory, PRX Quantum **3**, 020349 (2022).

[19] R. Finkelstein, E. Poem, O. Michel, O. Lahad, and O. Firstenberg, Fast, noise-free memory for photon synchronization at room temperature, Sci. Adv. **4**, eaap8598 (2018).

[20] K. T. Kaczmarek, P. M. Ledingham, B. Brecht, S. E. Thomas, G. S. Thekkadath, O. Lazo-Arjona, J. H. D. Munns, E. Poem, A. Feizpour, D. J. Saunders, J. Nunn, and I. A. Walmsley, High-speed noise-free optical quantum memory, Phys. Rev. A **97**, 042316 (2018).

[21] S. Gao, O. Lazo-Arjona, B. Brecht, K. T. Kaczmarek, S. E. Thomas, J. Nunn, P. M. Ledingham, D. J. Saunders, and I. A. Walmsley, Optimal coherent filtering for single noisy photons, Phys. Rev. Lett. **123**, 213604 (2019).

[22] S. E. Thomas, L. Wagner, R. Joos, R. Sittig, C. Nawrath, P. Burdekin, I. M. de Buy Wenniger, M. J. Rasiah, T. Huber-Loyola, S. Sagona-Stophel, S. Höfling, M. Jetter, P. Michler, I. A. Walmsley, S. L. Portalupi, and P. M. Ledingham, Deterministic storage and retrieval of telecom light from a quantum dot single-photon source interfaced with an atomic quantum memory, Sci. Adv. **10**, eadi7346 (2024).

[23] K. F. Reim, P. Michelberger, K. C. Lee, J. Nunn, N. K. Langford, and I. A. Walmsley, Single-photon-level quantum memory at room temperature, Phys. Rev. Lett. **107**, 053603 (2011).

[24] N. Lütkenhaus, J. Calsamiglia, and K.-A. Suominen, Bell measurements for teleportation, Phys. Rev. A **59**, 3295 (1999).

[25] G. Campbell, M. Hosseini, B. M. Sparkes, P. K. Lam, and B. C. Buchler, Time- and frequency-domain polariton interference, New J. Phys. **14**, 033022 (2012).

[26] K. Su, Y. Zhong, S. Zhang, J. Li, C.-L. Zou, Y. Wang, H. Yan, and S.-L. Zhu, Quantum interference between nonidentical single particles, Phys. Rev. Lett. **129**, 093604 (2022).

[27] K. F. Reim, J. Nunn, X.-M. Jin, P. S. Michelberger, T. F. M. Champion, D. G. England, K. C. Lee, W. S. Kolthammer, N. K. Langford, and I. A. Walmsley, Multipulse addressing of a Raman quantum memory: Configurable beam splitting and efficient readout, Phys. Rev. Lett. **108**, 263602 (2012).

[28] E. Saglamyurek, T. Hrushevskyi, A. Rastogi, K. Heshami, and L. J. LeBlanc, Coherent storage and manipulation of broadband photons via dynamically controlled Autler–Townes splitting, Nat. Photonics **12**, 774 (2018).

[29] P. M. Burdekin, I. Maillette de Buy Wenniger, S. Sagona-Stophel, J. Szuniewicz, A. Zhang, S. E. Thomas, and I. A. Walmsley, Enhancing quantum memories with light–matter interference, Optica **12**, 1514 (2025).

[30] E. Biham, B. Huttner, and T. Mor, Quantum cryptographic network based on quantum memories, Phys. Rev. A **54**, 2651 (1996).

[31] See Supplemental Material at http://link.aps.org/supplemental/10.1103/wdz1-sbgg for details on the SKR calculation and a comparison between the achievable SKR for the two-memory protocol and the in-memory protocol, which includes Refs. [7,29,33,34,44,50,51].

[32] H.-K. Lo, H. Chau, and M. Ardehali, Efficient quantum key distribution scheme and a proof of its unconditional security, J. Cryptology **18**, 133 (2005).

[33] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University, Cambridge, England, 2010).

[34] F. Xu, M. Curty, B. Qi, and H.-K. Lo, Practical aspects of measurement-device-independent quantum key distribution, New J. Phys. **15**, 113007 (2013).

[35] H.-K. Lo, X. Ma, and K. Chen, Decoy state quantum key distribution, Phys. Rev. Lett. **94**, 230504 (2005).

[36] S. Pirandola, U. L. Andersen, L. Banchi, M. Berta, D. Bunandar, R. Colbeck, D. Englund, T. Gehring, C. Lupo, C. Ottaviani, J. L. Pereira, M. Razavi, J. Shamsul Shaari, M. Tomamichel, V. C. Usenko, G. Vallone, P. Villoresi, and P. Wallden, Advances in quantum cryptography, Adv. Opt. Photonics **12**, 1012 (2020).

[37] T. Müller, J. Skiba-Szymanska, A. B. Krysa, J. Huwer, M. Felle, M. Anderson, R. M. Stevenson, J. Heffernan, D. A. Ritchie, and A. J. Shields, A quantum light-emitting diode for the standard telecom window around 1,550 nm, Nat. Commun. **9**, 862 (2018).

[38] G. Shooter, Z.-H. Xiang, J. R. A. Müller, J. Skiba-Szymanska, J. Huwer, J. Griffiths, T. Mitchell, M. Anderson, T. Müller, A. B. Krysa, R. Mark Stevenson, J. Heffernan, D. A. Ritchie, and A. J. Shields, 1 GHz clocked distribution of electrically generated entangled photon pairs, Opt. Express **28**, 36838 (2020).

[39] T. Lettner, S. Gyger, K. D. Zeuner, L. Schweickert, S. Steinhauer, C. R. Hedlund, S. Stroj, A. Rastelli, M. Hammar, R. Trotta, K. D. Jöns, and V. Zwiller, Strain-controlled quantum dot fine structure for entangled photon generation at 1550 nm, Nano Lett. **21**, 10501 (2021).

[40] C. Nawrath, R. Joos, S. Kolatschek, S. Bauer, P. Pruy, F. Hornung, J. Fischer, J. Huang, P. Vijayan, R. Sittig, M. Jetter, S. L. Portalupi, and P. Michler, Bright source of Purcell-enhanced, triggered, single photons in the telecom C-band, Adv. Quantum Technol. **6**, 2300111 (2023).

[41] P. Holewa, D. A. Vajner, E. Zięba-Ostój, M. Wasiluk, B. Gaál, A. Sakanas, M. G. Mikulicz, P. Mrowiński, B. Krajnik, M. Xiong, K. Yvind, N. Gregersen, A. Musiał, A. Huck, T. Heindel, M. Syperek, and E. Semenova, High-throughput quantum photonic devices emitting indistinguishable photons in the telecom C-band, Nat. Commun. **15**, 3358 (2024).

[42] P. Holewa, A. Sakanas, U. M. Gür, P. Mrowiński, A. Huck, B.-Y. Wang, A. Musiał, K. Yvind, N. Gregersen, M. Syperek, and E. Semenova, Bright quantum dot single-photon emitters at telecom bands heterogeneously integrated on Si, ACS Photonics **9**, 2273 (2022).

[43] N. Gisin, S. Pironio, and N. Sangouard, Proposal for implementing device-independent quantum key distribution based on a heralded qubit amplifier, Phys. Rev. Lett. **105**, 070501 (2010).

[44] W. Wasilewski and M. G. Raymer, Pairwise entanglement and readout of atomic-ensemble and optical wave-packet modes in traveling-wave Raman interactions, Phys. Rev. A **73**, 063816 (2006).

[45] S. E. Thomas, S. Sagona-Stophel, Z. Schofield, I. A. Walmsley, and P. M. Ledingham, Single-photon-compatible telecommunications-band quantum memory in a hot atomic gas, Phys. Rev. Appl. **19**, L031005 (2023).

[46] I. M. de Buy Wenniger, P. Burdekin, S. Zhang, M. J. Rasiah, A. Rastogi, O. T. P. Schmidt, P. M. Ledingham, I. A. Walmsley, and S. E. Thomas, Dynamic rephasing in a telecom warm vapor quantum memory, arXiv:2604.13900.

[47] W.-X. Xie, G.-J. Fan-Yuan, Z.-H. Wang, F.-Y. Lu, J.-X. Li, S. Wang, Z.-Q. Yin, W. Chen, D.-Y. He, G.-C. Guo, and Z.-F. Han, Higher key rate in asymmetric quantum-classical integrated measurement-device-independent quantum-key-distribution systems, Phys. Rev. Appl. **20**, 054042 (2023).

[48] H. Liu, W. Wang, K. Wei, X.-T. Fang, L. Li, N.-L. Liu, H. Liang, S.-J. Zhang, W. Zhang, H. Li, L. You, Z. Wang, H.-K. Lo, T.-Y. Chen, F. Xu, and J.-W. Pan, Experimental demonstration of high-rate measurement-device-independent quantum key distribution over asymmetric channels, Phys. Rev. Lett. **122**, 160501 (2019).

[49] W. Wang, F. Xu, and H.-K. Lo, Asymmetric protocols for scalable high-rate measurement-device-independent quantum key distribution networks, Phys. Rev. X **9**, 041012 (2019).

[50] J. Nunn, I. A. Walmsley, M. G. Raymer, K. Surmacz, F. C. Waldermann, Z. Wang, and D. Jaksch, Mapping broadband single-photon wave packets into an atomic memory, Phys. Rev. A **75**, 011401(R) (2007).

[51] A. V. Gorshkov, T. Calarco, M. D. Lukin, and A. S. Sørensen, Photon storage in Λ-type optically dense atomic media. IV. Optimal control using gradient ascent, Phys. Rev. A **77**, 043806 (2008).