Enhancing multiphoton rates with quantum memories

J. Nunn,[1,*] N. K. Langford,[2] W. S. Kolthammer,[1] T. F. M. Champion,[1]
M. R. Sprague,[1] P. S. Michelberger,[1] X.-M. Jin,[1,3] D. G. England,[1] and I. A. Walmsley[1]

[1] *Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, United Kingdom*
[2] *Department of Physics, Royal Holloway, University of London, Egham Hill, Egham TW20 0EX, United Kingdom*
[3] *Centre for Quantum Technologies, National University of Singapore, 117543, Singapore*

(Dated: August 9, 2012)

Single photons are a vital resource for optical quantum information processing. Efficient and deterministic single photon sources do not yet exist, however. To date, experimental demonstrations of quantum processing primitives have been implemented using non-deterministic sources combined with heralding and/or postselection. Unfortunately, even for eight photons, the data rates are already so low as to make most experiments impracticable. It is well known that quantum memories, capable of storing photons until they are needed, are a potential solution to this 'scaling catastrophe'. Here, we analyze in detail the benefits of quantum memories for producing multiphoton states, showing how the production rates can be enhanced by many orders of magnitude. We identify the quantity ηB as the most important figure of merit in this connection, where η and B are the efficiency and time-bandwidth product of the memories, respectively.

PACS numbers: 42.50.Ex, 42.50.Ct, 42.50.-p

After two decades of rapid advances, quantum optics experiments are becoming increasingly challenging. As the interests of the community shift to higher-dimensional entanglement [1, 2] and information processing tasks beyond mere proof-of-principle [3, 4], the demand for large numbers of simultaneous single photons is outstripping the capabilities of parametric sources [5, 6]. These sources, which so far have been the workhorse of the quantum optics lab, produce photons in pairs, but they also produce multiple unwanted photon-pairs with a probability that scales with the single-pair generation rate, which must therefore be kept low, so that most often no photons are emitted. The current record for photonic resources is an eight-photon experiment involving four parametric sources, in which statistics were accumulated over 40 hours [7]. Besides the exorbitant time required to run experiments of this kind, the quality of the multiphoton states produced is limited by multi-pair emission. One solution is to operate the sources in a heralded fashion, in which one of each photon pair is detected, indicating the presence of the other photon. This halves the number of usable photons but it removes the non-emission events, which makes noise events a much smaller fraction of the finally-detected photons. The use of heralded parametric sources is extremely convenient, since they operate at room temperature and at extremely high repetition rates. Nonetheless the low success probability means they cannot be efficiently combined to produce multiphoton states.

The scalability problem is easily understood. Suppose that a single photon is heralded with probability $q \ll 1$. The probability of producing N single photons simultaneously using N sources is then simply q^N, which becomes exponentially small as N increases, thus rendering complex experiments impossible.

One method to mitigate this problem is by multiplexing many sources via active switching [8–13], but large overheads — *i.e.* many identical sources — are required to achieve efficient operation [14]. It is well-known that temporal multiplexing using quantum memories offers an alternative solution [15–17].

To see how quantum memories can increase the rate of N-fold coincidences, consider the array of N sources coupled to $N-1$ memories shown in Fig. 1. We suppose that each source produces photons in pairs by means of a parametric scattering process such as downconversion [6] or spontaneous four-wave mixing [5], with one of each pair directed to a herald detector. With no memories, all N heralds must fire simultaneously to produce an N-fold coincidence. However with memories, heralded photons can be stored whenever they are produced. Once all $N-1$ memories are charged with a photon, one only has to wait for the final source to produce a photon, and then all the memories can be read out and one has, again, an N-fold coincidence. This protocol is probably not optimal, but it is amenable to a straightforward analysis that captures the scaling enhancement: by lifting the requirement for simultaneous emission, the memories greatly enhance the coincidence probability. Our purpose in this paper is to quantify the gain in coincidence rate afforded by using quantum memories to synchronize photon sources in this way. The time-bandwidth product $B = \delta\tau$ proves to be critical in this context, where δ is the acceptance bandwidth of the memories, and τ is their coherence lifetime [32]. If postselection on the final detection of N photons is used, even relatively inefficient memories can dramatically enhance the multiphoton rate.

We first fix the protocol by assuming that we always at-

*j.nunn1@physics.ox.ac.uk

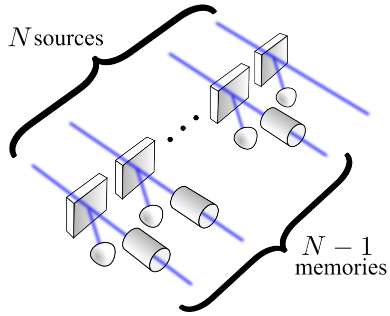

FIG. 1: An array of N heralded parametric sources synchronized by $N-1$ quantum memories. The sources are repeatedly pumped, and each photon emitted is stored until all memories are charged. Then emission of a photon by the final source triggers retrieval from the memories, in order to generate an N-fold coincidence.

tempt to store a photon, if a herald detector fires, regardless of whether or not the memory concerned is already charged. This ensures that we always use the most recently emitted photons, which mitigates photon loss due to decoherence in the memories. To avoid 'clashes' (in fact, interference [18, 19]) between incident and stored photons, we clean each memory before storage is attempted (*e.g.* by readout of the memory, or optical pumping) so that we are always attempting to charge an empty memory. This allows to use a classical model, in which individual photons are treated as particles that are probabilistically emitted, stored and retrieved. Finally, we adopt the policy that if we are ready to read out the memories — that is, if all memories have been charged and a photon is emitted from the N^{th} source — and at the same time one or more of the other sources emits a photon, we bypass the relevant memories and use these 'serendipitous' photons, rather than attempting to read out the memories.

The photon sources are pumped at a rate $\mathcal{R} \sim \delta$, limited by the minimum pulse duration that can be stored by the memories. The average waiting time $1/\mathcal{R}c$ between N-photon events can then be computed if we can find an expression for the N-fold coincidence probability $c = qp_{\text{sync}}^{N-1}$, which is the probability that one photon is obtained from each of the N sources. The leading factor of q describes the probability that the N^{th} source emits a photon, and we have defined $p_{\text{sync}} = q + \overline{q}\eta_{\text{r}}P$ as the probability that any one of the $N-1$ sources equipped with a memory provides a photon on demand, either directly, or through successful retrieval of a stored photon. Here P is the steady-state probability that any memory is charged with a stored photon, η_{r} is the retrieval efficiency, and the overbar notation denotes the probabilistic complement, $\overline{X} \equiv 1 - X$. The problem of computing the waiting time then reduces to that of finding P. To proceed we assume that the decoherence processes in the memories are Markovian (*i.e.* exponential), since then the stochastic evolution of the charge-state $\boldsymbol{x}^{(m)} = [\overline{P}^{(m)}, P^{(m)}]^{\mathsf{T}}$ of each memory can be tracked using a transfer matrix:

$$\boldsymbol{x}^{(m)} = T\boldsymbol{x}^{(m-1)}; \qquad T = \begin{pmatrix} \overline{r} & s \\ r & \overline{s} \end{pmatrix}, \tag{1}$$

with $P^{(m)}$ the probability that the memory is charged at the m^{th} time step, r the probability that an empty memory becomes charged over the course of one time step, and s the probability that a charged memory becomes empty. The steady state probabilities are given by the eigenvector $\boldsymbol{x}_{\text{s}}$ of T with eigenvalue 1, $\boldsymbol{x}_{\text{s}} = [s, r]^{\mathsf{T}}/(r+s)$, so that we have $P = r/(r+s)$. The probability that an empty memory becomes charged is the probability that a heralded photon is emitted and that it is stored, provided that the rest of the set-up is not primed for readout, so we have $r = q\eta_{\text{s}}\overline{R}$, where R is the probability that the system is ready to be read out (the evaluation of R is described in the Appendix), and η_{s} is the storage efficiency. The loss probability that a charged memory is emptied is more complicated. There are four processes involved. First, decoherence in the memory during standby, second, readout of the memory when we attempt to generate a coincidence, third, the loss of a stored photon during standby when a new photon comes along and we attempt to replace the stored photon but fail, and finally decoherence in the memory during the readout stage, when a photon is heralded and the memory is bypassed, leaving the memory charged and vulnerable to decay. Denoting the decoherence probability by b, the total loss probability works out to be $s = \overline{q}[b\overline{R} + R] + q[\overline{\eta}_{\text{s}}\overline{R} + Rb]$. Generally the time-bandwidth product will be much larger than one, so that $b = 1 - e^{-1/B} \approx 1/B$, and we finally obtain

$$c = q^N \left\{ 1 + \frac{\overline{R}\overline{q}\eta B}{1 + (B-1)\left[R(\overline{q} - q) + q\right]} \right\}^{N-1}. \tag{2}$$

This is the main analytic result of this paper. In the limit of small photon generation rates such that $\{RB, qB\} \ll 1$, we have $c \approx q^N(\eta B)^{N-1}$, which supports the intuition that each memory effectively boosts the photon generation probability by B, moderated by its efficiency $\eta = \eta_{\text{s}}\eta_{\text{r}}$. In this regime the gain in the multiphoton rate is therefore exponential in the quantity ηB, which highlights the importance of the time-bandwidth product for synchronization applications. As B is increased so that $qB \gg 1$, the rate eventually saturates and becomes independent of B, limited finally by η.

To make a fair comparison with the unsynchronized case, we now consider the effect of higher photon number components on the quality of the states produced. Typically, parametric sources generate photon pairs according to a thermal distribution, where $p_{\text{source}}(n) = \overline{p}p^n$ is the probability of emitting n photon pairs, and p is a small real number. We also assume non-photon-number-resolving heralding detectors, such as APDs, so that the conditional probability that n photons are sent towards a memory, given a herald click, is $p_{\text{h}}(n) =$

$[1-\overline{h}^n]p_{\text{source}}(n)/q$ where h is the efficiency of the heralding detector, and as before $q = hp/[1-p\overline{h}]$ is the probability of a herald click. For simplicity we assume that detector dark counts are negligible. The charge state $\boldsymbol{x}^{(m)}$ of the memory is now a vector of probabilities that the memory contains n photons, with $n = 0, 1, 2...$, which we truncate for numerical convenience. The transition probability that the number of excitations stored in a memory changes from k to j over the course of any time step is given by the transfer matrix element

$$T_{jk} = \theta_{jk} b^{k-j} \overline{b}^j \binom{k}{j} (\overline{R}\overline{q} + Rq) + \overline{q}R\delta_{j0} + q\overline{R}p_{\text{s}}(j), \quad (3)$$

where the three terms represent decoherence, readout, and storage, respectively. Here $\theta_{jk} = 1$ for $k \geq j$ and zero otherwise, and δ_{j0} is a Kronecker delta. We have also defined $p_{\text{s}}(n)$ as the probability that n photons are stored in the memory when read-in is attempted after a herald, $p_{\text{s}}(n) = \sum_{k=n}^{\infty} p_{\text{h}}(k)\eta_{\text{s}}^n \overline{\eta}_{\text{s}}^{k-n}\binom{k}{n}$. Repeated application of T to an arbitrary initial charge state converges to the steady state $\boldsymbol{x}_{\text{s}}$, and the probability that n photons are retrieved from the memory is then given by $p_{\text{r}}(n) = \sum_{k=n}^{\infty} x_{\text{s}}(k)\eta_{\text{r}}^n \overline{\eta}_{\text{r}}^{k-n}\binom{k}{n}$. We can then write $c = qp_{\text{h}}(1)p_{\text{sync}}(1)^{N-1}$, where $p_{\text{sync}}(n) = qp_{\text{h}}(n) + \overline{q}p_{\text{r}}(n)$. This result for c represents only a minor correction to Eq. (2), but the treatment of multi-pair emissions is important for the fidelity calculation below.

In many photonic networks, successful operations can be postselected on the final detection of at least N photons. In this case the fidelity of the postselected states is the fraction of these which contain 1 photon per mode. We normalise this to the number of modes by taking the N^{th} root,

$$\widetilde{\mathcal{F}} = \left[\frac{c}{p_{\geq N}}\right]^{1/N} = \left[\frac{c}{q - p_{<N}}\right]^{1/N}. \quad (4)$$

Here $p_{\geq N}$ is the probability that the state obtained from the memories/sources comprises N photons or more, and we have re-written this in terms of $p_{<N}$, the probability that fewer than N photons in total are emitted, given by

$$p_{<N} = \sum_{j=0}^{N-1} \sum_{k=0}^{j} qp_{\text{h}}(j-k) \sum_{\boldsymbol{s}_k} \prod_{l=1}^{N-1} p_{\text{sync}}(s_k(l)). \quad (5)$$

Here $s_k(l)$ is the l^{th} element of a vector $\boldsymbol{s}_k$ containing $N-1$ real, non-negative integers whose sum is equal to k. The summation $\sum_{\boldsymbol{s}_k}$ runs over all such vectors.

If postselection is not used, we consider the fraction of readouts that our desired N photon state, with one photon per mode, is produced, to obtain $\mathcal{F} = [c/RY]^{1/N}$, where RY is the probability that we believe we have produced an N photon state (see Appendix).

Without memories, the N-fold coincidence rate is $c_{\text{no mem}} = [qp_{\text{h}}(1)]^N$ and the unpostselected fidelity is simply given by $\mathcal{F}_{\text{no mem}} = p_{\text{h}}(1)$. The postselected fidelity is identical, $\widetilde{\mathcal{F}}_{\text{no mem}} = \mathcal{F}_{\text{no mem}}$, since the heralding completely removes the vacuum component.

In general, neither measure of fidelity for either synchronized or unsynchronized systems will reach 1, except in the limit $p \to 0$. Therefore one must choose a threshold fidelity Θ that is acceptable, and then one should choose the largest value p_Θ of p such that $\mathcal{F} = \Theta$, for each system. Having done this, one can then compare the N-fold coincidence rates. For N unsynchronized sources we have that $p_\Theta = \{2 - h - [(2-h)^2 - 4\overline{h\Theta}]^{1/2}\}/2\overline{h}$, independent of N. For the same number of synchronized sources, p_Θ depends on N and needs to be determined by a numerical optimisation. Figure 2 shows the resulting comparison of synchronized and unsynchronized systems. The waiting times scale exponentially with the number N of photons required, and without synchronization a 12 photon experiment would require more than 30 years in between coincidence events, so that quantum computing with photons using such a system is totally unfeasible. However the use of memories reduces the waiting time quite dramatically. For a postselected experiment one can use inefficient memories with $\eta = 56\%$ and reduce the 12-fold waiting time to $\sim$10 ms. Quantum memories based on Raman scattering have already been demonstrated with $\delta >$ GHz [20], $B > 1000$ [21, 22] and $\eta > 50\%$ [23], while highly efficient and multiplexed storage in rare-earth memories is maturing [24–28], and so these dramatic enhancements lie well within the reach of current technology. Without postselection more efficient memories are required to achieve the fidelity threshold, which puts the implementation beyond current techno-

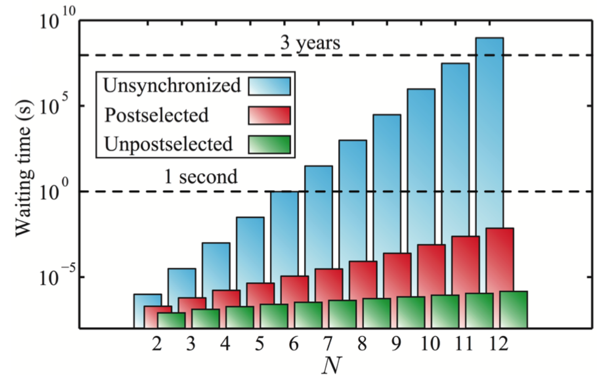

FIG. 2: Multiphoton waiting times. The blue bars show the average waiting time between N-photon events for a system of N unsynchronized downconversion sources, assuming a pulse repetition rate of $\mathcal{R} = 1$ GHz, a heralding efficiency of $h = 50\%$ and a threshold fidelity $\Theta = 90\%$. The red bars show the corresponding waiting times when the system is synchronized with $N-1$ memories, with memory efficiencies $\eta_{\text{s}} = \eta_{\text{r}} = 75\%$ and a time-bandwidth product $B = 1000$, where postselection on at least N photons is used (we set p_Θ so that $\widetilde{\mathcal{F}} = \Theta$). The green bars show the waiting times without postselection (we set p_Θ so that $\mathcal{F} = \Theta$), where to achieve the required unpostselected fidelity threshold we now assume memory efficiencies of $\eta_{\text{s}} = \eta_{\text{r}} = 99\%$.

logical capabilities, but this improved performance further reduces the waiting time to $\lesssim$1 μs.

In summary, we have analyzed the use of quantum memories for the synchronization of multiple single photon sources as a canonical application of quantum storage for the enhancement of photonic information processing. We derived an analytic formula for the multiphoton rate achievable and showed that the most important figure of merit for quantum memories is the product ηB of the memory efficiency with its time-bandwidth product. Finally we extended our model to include higher-order photon number contributions, so that the quality of the states produced with and without memories could be compared. We showed that even inefficient memories can produce enormous improvements in the multiphoton rate when combined with postselection. Without postselection, highly efficient memories are required to match the quality of unsynchronized sources, but if these are available the gain in multiphoton rate becomes larger still. It would be interesting to consider the effects of noise in the memories, or extensions to more complicated synchronization protocols. It is expected that similar advantages could pertain to the scaling of other heralded quantum operations, such as entanglement generation or two-photon gates. While much attention in the quantum memory community has focussed on the need for long storage times and high efficiencies in the context of quantum repeaters [29, 30], our analysis underlines the value of developing quantum memories for local synchronization, for which lower efficiencies still provide considerable advantages, and for which the time-bandwidth product B is much more important than the absolute storage time.

Acknowledgments

This work was supported by the EPSRC (EP/C51933/01), the EC project Q-ESSENCE (248095), the Royal Society, and the AFOSR EOARD. XMJ acknowledges support from NSFC (11004183) and CPSF (201003327). NKL was supported by an EC Marie-Curie Fellowship. MRS was supported by a Clarendon Scholarship. PSM and TFM were supported by the EC ITN FASTQUAST. JN is grateful to Marco Barbieri for wearing him down.

Appendix

The readout probability R can be computed by tracking the belief state $\boldsymbol{y}^{(m)} = [\overline{V}^{(m)}, V^{(m)}]^{\mathsf{T}}$, where $V^{(m)}$ is the probability that we believe the memory to be charged at the m^{th} time step. We have

$$\boldsymbol{y}^{(m)} = S\boldsymbol{y}^{(m-1)}; \qquad S = \begin{pmatrix} \overline{w} & z \\ w & \overline{z} \end{pmatrix},$$

where $w = \overline{R}q$ ($z = \overline{q}R$) is the probability that we believe an empty (charged) memory becomes charged (empty) over the course of one time step. In the steady state $V^{(m)} \to V = w/(w+z)$. On the other hand, readout occurs when we believe $N-1$ other photons to be available, so we can write $R = qY^{N-2}$, where $Y = q + \overline{q}V$ is the probability that we believe a source has provided a photon, either directly or through its memory. Combining these relations we obtain the consistency condition $(1-2q)Y^{N-1} + q^2Y^{N-2} + Y - 1 = 0$, the positive real root of which can be found numerically, which then fixes R. Note that when $N = 2$, we have $R = q$.

[1] Lanyon B and Langford N, *New Jour. Phys.*, 11:013008, 2009.
[2] Wieczorek W, Krischek R, Kiesel N *et al.*, *Phys. Rev. Lett.*, 103(2):20504, 2009.
[3] Thompson M, Politi A, Matthews J *et al.*, *Circuits, Devices & Systems, IET*, 5(2):94, 2011.
[4] Thomas-Peter N, Langford N, Datta A *et al.*, *New Jour. Phys.*, 13:055024, 2011.
[5] Cohen O, Lundeen JS, Smith BJ *et al.*, *Phys. Rev. Lett.*, 102(12):123603, 2009.
[6] Mosley PJ, Lundeen JS, Smith BJ *et al.*, *Phys. Rev. Lett.*, 100(13):133601, 2008.
[7] Yao X, Wang T, Xu P *et al.*, *arXiv:1105.6318*, 2011.
[8] Migdall A, Branning D and Castelletto S, *Phys. Rev. A*, 66(5):053805, 2002.
[9] O'Brien J and Akira Furusawa J, *Nat. Phot.*, 3(12):687, 2009.
[10] Jennewein T, Barbieri M and White A, *Jour. Mod. Opt.*, 58(3-4):276, 2011.
[11] Ma X, Zotter S, Kofler J *et al.*, *Phys. Rev. A*, 83(4):043814, 2011.
[12] McCusker KT and Kwiat PG, *Phys. Rev. Lett.*, 103:163602, 2009.
[13] Hall M, Altepeter J and Kumar P, *Phys. Rev. Lett.*, 106(5):53901, 2011.
[14] Christ A and Silberhorn C, *Phys. Rev. A*, 85(2):023829, 2012.
[15] Yuan Z, Chen Y, Chen S *et al.*, *Phys. Rev. Lett*, 98(18):180503, 2007.
[16] Pittman T, Jacobs B and Franson J, *Phys. Rev. A*, 66(4):042303, 2002.
[17] Mower J and Englund D, *Phys. Rev. A*, 84(5):052326, 2011.
[18] Campbell G, Hosseini M, Sparkes B *et al.*, *arXiv:1107.0507*, 2011.
[19] Datta A, Zhang L, Nunn J *et al.*, *Phys. Rev. Lett.*, 108(6):060502, 2012.
[20] Reim K, Nunn J, Lorenz V *et al.*, *Nat. Phot.*, 2010.
[21] Reim K, Michelberger P, Lee K *et al.*, *Phys. Rev. Lett.*, 107(5):53603, 2011.
[22] Bao X, Reingruber A, Dietrich P *et al.*, *Nat. Phys.*, 8(7):517, 2012.
[23] Hosseini M, Campbell G, Sparkes B *et al.*, *Nat. Phys.*, 7(10):795, 2011.
[24] Hedges M, Longdell J, Li Y *et al.*, *Nature*, 465(7301):1052, 2010.
[25] Usmani I, Afzelius M, De Riedmatten H *et al.*, *Nat. Comm.*, 1:12, 2010.

[26] Afzelius M, Usmani I, Amari A *et al.*, *Phys. Rev. Lett.*, 104(4):40503, 2010, ISSN 1079-7114.
[27] Saglamyurek E, Sinclair N, Jin J *et al.*, *Nature*, 469(7331):512, 2011.
[28] Simon C, de Riedmatten H, Afzelius M *et al.*, *Phys. Rev. Lett.*, 98:190503, 2007.
[29] Duan L, Lukin M, Cirac J *et al.*, *Nature*, 414:413, 2001.
[30] Sangouard N, Simon C, Zhao B *et al.*, *Phys. Rev. A*, 77(6):062301, 2008.
[31] Nunn J, Reim K, Lee KC *et al.*, *Phys. Rev. Lett.*, 101(26):260502, 2008.
[32] This is not the same as the multimode capacity [28, 31].