Supplementary Information
Purification of single photons by temporal heralding of quantum dot sources

Hamza Abudayyeh, Boaz Lubotzky, and Ronen Rapaport
Racah Institute for Physics and Center for Nanoscience and Nanotechnology,
The Hebrew University of Jerusalem, Jerusalem 9190401, Israel

Somak Majumder and Jennifer A. Hollingsworth
Materials Physics & Applications Division: Center for Integrated Nanotechnologies,
Los Alamos National Laboratory, Los Alamos, New Mexico 87545, United States

* ronenr@phys.huji.ac.il
I. THEORETICAL DERIVATION OF HERALDING PARAMETERS

The quantum yield of an emission process \( i \) is given by the ratio of the radiative decay rate \( \Gamma_i^r \) to total decay rate \( \Gamma_i \):

\[
QY_i = \frac{\Gamma_i^r}{\Gamma_i}
\]

For a biexciton-exciton emission cascade simple statistical scaling implies that the radiative decay rate of the biexciton should be 4 times that of the exciton. In reality other factors play a role such as the degree of confinement of the electrons and holes to the core of a core/shell quantum dot for example. Therefore in general this scaling factor will be called \( \beta \), i.e. \( \Gamma_{BX}^r = \beta \Gamma_X^r \) and thus:

\[
\frac{QY_{BX}}{QY_X} = \frac{\Gamma_{BX}^r \Gamma_X}{\Gamma_X^r \Gamma_X} = \beta \frac{\Gamma_X}{\Gamma_{BX}}
\]

Therefore if the \( \beta \)-scaling factor is known a simple lifetime measurement can be used to find the ratio of the biexciton to exciton quantum yield.

We will assume that our nanocrystal quantum dot can emit at most two photons based on the biexciton-exciton cascade and that it is being pumped well above saturation. The probability that a biexciton photon is emitted \( T_{BX} \) after the laser pulse is given by:

\[
p(T_{BX}) = \alpha QY_{BX} \Gamma_{BX} e^{-\Gamma_{BX} T_{BX}}
\]

whereas the exciton emission time \( (T_X) \) is conditioned on the emission time of the biexciton as follows:

\[
p(T_x|T_{BX}) = \alpha QY_X \Gamma_X e^{-\Gamma_X (T_x - T_{BX})}
\]

where:
\( \alpha \): overall detection efficiency of the system
\( QY_i \): quantum yield of the \( i \)th emission process
\( p(T_{BX}) \): probability of biexciton event at local time \( T_{BX} \)
\( p(T_x|T_{BX}) \): conditional probability of exciton event at local time \( T_X \) given that a biexciton event was detected at \( T_{BX} \)

Therefore for a standalone NQD the probability to emit one and two photons \( (P_1 \) and \( P_2 \) respectively) are:

\[
P_1 = \alpha QY_X + \alpha QY_{BX} - 2\alpha^2 QY_X QY_{BX}
\]

\[
P_2 = \alpha^2 QY_X QY_{BX}
\]

I.A. Fixed gate techniques

If we assume that a fixed time gate is applied as in the case of TGF and TIMED than we can ask what is the probability that a photon will take a certain route. For example what is the probability that one photon will arrive within \( T \) from the laser pulse and the other photon will arrive after \( T \). This probability is given by the following integral:

\[
\mathcal{P}_1(T) = \alpha^2 QY_X QY_{BX} \int_0^T dT_{BX} \int_T^\infty dT_X p(T_{BX}) p(T_x|T_{BX})
\]

\[
= \alpha^2 QY_X QY_{BX} \frac{\Gamma_{BX}}{\Gamma_X - \Gamma_{BX}} \exp (-\Gamma_X T) \{1 - \exp (- (\Gamma_{BX} - \Gamma_X) T)\}
\]
TABLE S1. Possible events resulting from an emitter that has a maximum of a two photon cascade by applying a fixed time gate

| # | Condition on BX | Condition on X | Prob. of Process |
|---|-----------------|----------------|------------------|
| 1. | $T_{BX} \leq T$ | $T_{X} \geq T$ | $\mathcal{P}_1(T) = \alpha^2 QY_X QY_{BX} \frac{\Gamma_{BX}}{\Gamma_{BX} - \Gamma_X} \{\exp(-\Gamma_X T) - \exp(-\Gamma_{BX} T)\}$ |
| 2. | $T_{BX} \leq T$ | $T_{X} \leq T$ | $\mathcal{P}_2(T) = \alpha^2 QY_{BX} QY_{X} \{1 - \frac{\Gamma_{BX}}{\Gamma_{BX} - \Gamma_X} \exp(-\Gamma_X T) + \frac{\Gamma_X}{\Gamma_{BX} - \Gamma_X} \exp(-\Gamma_{BX} T)\}$ |
| 3. | $T_{BX} \geq T$ | $T_{X} \geq T$ | $\mathcal{P}_3(T) = \alpha^2 QY_{BX} QY_{X} \exp(-\Gamma_{BX} T)$ |
| 4. | $T_{BX} \leq T$ | $T_{X} \leq T$ | $\mathcal{P}_4(T) = \alpha QY_{BX} (1 - \alpha QY_{X}) \{1 - \exp(-\Gamma_{BX} T)\}$ |
| 5. | $T_{BX} \geq T$ | $T_{X} \leq T$ | $\mathcal{P}_5(T) = \alpha QY_{BX} (1 - \alpha QY_{X}) \exp(-\Gamma_{BX} T)$ |
| 6. | $T_{BX} \leq T$ | $T_{X} \geq T$ | $\mathcal{P}_6(T) = \alpha QY_{X} (1 - \alpha QY_{BX}) \{1 - \exp(-\Gamma_{BX} T)\}$ |
| 7. | $T_{BX} \geq T$ | $T_{X} \leq T$ | $\mathcal{P}_7(T) = \alpha Q Y_{X} (1 - \alpha Q Y_{BX}) \exp(-\Gamma_{BX} T)$ |
| 8. | $T_{BX} \geq T$ | $T_{X} \leq T$ | $\mathcal{P}_8 = (1 - \alpha Q Y_{BX}) (1 - \alpha Q Y_{X})$ |

On the other hand the probability that both will arrive prior to $T$ ($\mathcal{P}_2(T)$) or after $T$ ($\mathcal{P}_3(T)$) are given by the following integrals:

$$\mathcal{P}_2(T) = \alpha^2 QY_X QY_{BX} \int_0^T dT_{BX} \int_{T_{BX}}^T dT_X p(T_{BX}) p(T_x | T_{BX})$$

$$\mathcal{P}_3(T) = \alpha^2 QY_X QY_{BX} \int_T^\infty dT_{BX} \int_{T_{BX}}^\infty dT_X p(T_{BX}) p(T_x | T_{BX})$$

These events in addition to all other possibilities are summarized in table S1.

In Time Gated Filtering (TGF) shown in figure S1, the switch is open to an optical dump from the beginning of each pulse up to a filtering time $T_F$ after which the photons are routed to the signal port. In such a manner the short lifetime components of an optical signal can be filtered out. This can be used to filter out the biexciton emission where the the efficiency of the source would be the probability of only one photon events after a time gate of $T_F$ namely:

$$\eta_{TGF} = \mathcal{P}_1(T_F) + \mathcal{P}_5(T_F) + \mathcal{P}_7(T_F)$$

where as the purity is the same quantity normalized by the probability of obtaining one or two photons after $T_F$ which is given by:

$$S_{TGF} = \frac{\mathcal{P}_1(T_F) + \mathcal{P}_5(T_F) + \mathcal{P}_7(T_F)}{\mathcal{P}_1(T_F) + \mathcal{P}_3(T_F) + \mathcal{P}_5(T_F) + \mathcal{P}_7(T_F)}$$

This case of TGF with $QY_X = 0.6$ and $QY_{BX} = 0.71$ is plotted in figure S1 where the trade-off between purity and efficiency in terms of $T_F$ is clear. For the special case of unity collection efficiency and unity quantum yields the expressions reduce to:

$$\eta_{TGF}^{ideal} = \frac{\Gamma_{BX}}{\Gamma_{BX} - \Gamma_X} \{\exp(-\Gamma_X T_F) - \exp(-\Gamma_{BX} T_F)\}$$

$$S_{TGF}^{ideal} = \frac{\Gamma_{BX} \{\exp(-\Gamma_X T_F) - \exp(-\Gamma_{BX} T_F)\}}{\Gamma_{BX} \exp(-\Gamma_X T_F) - \Gamma_X \exp(-\Gamma_{BX} T_F)}$$

On the other hand for the TIMe resolved heraldED (TIMED) scheme only the first process constitutes a successful heralding event namely:

$$\eta_{TIMED} = \mathcal{P}_1(T_C) = \alpha^2 QY_X QY_{BX} \frac{\Gamma_{BX}}{\Gamma_{BX} - \Gamma_X} \{\exp(-\Gamma_X T_C) - \exp(-\Gamma_{BX} T_C)\}$$
By using $\tau_i = \Gamma_i^{-1}$ for $i = X, BX$ we arrive at the equation stated in the main text. We can choose the optimum cutoff time $T_{c}^{opt}$ that will maximize $\eta$. This is given in units of the exciton lifetime as:

$$\Gamma_x T_{c}^{opt} = \left( \frac{\Gamma_{BX}}{\Gamma_X} - 1 \right)^{-1} \ln \left( \frac{\Gamma_{BX}}{\Gamma_X} \right)$$

Using this the optimum efficiency $\eta^{opt}$ is:

$$\eta_{TIMED}^{opt} = \alpha^2 QY_X QY_{BX} \Gamma_{BX}/\Gamma_X \int_{0}^{\infty} dT_X p(T_X \mid T_{BX}) p(T_{BX} \mid T_{BX}) \right) = \alpha^2 QY_X QY_{BX} \exp(-\Gamma_X T_R)$$

An important parameter for the case of multiplexed sources is the determinicity defined as the ratio of true heralded events to overall trigger events. This parameter defines the reliability of the trigger signal in heralding a signal photon. In statistics this often called the Positive Predictive Value (PPV) defined as the ratio of true positives to overall positives, but we will stick to the term determinicity for comparison with non-heralded sources. For the TIMED scheme the determinicity $D$ is given by:

$$D_{TIMED} = \frac{\mathcal{P}_1(T_C)}{\mathcal{P}_1(T_C) + \mathcal{P}_2(T_C) + \mathcal{P}_4(T_C) + \mathcal{P}_6(T_C)}$$

Again for the idealistic case of unity collection efficiency and quantum yields this reduces to:

$$D_{ideal}^{ TIMED} = \frac{\Gamma_{BX}}{\Gamma_{BX} - \Gamma_X} \exp(-\Gamma_X T_R) - \exp(-\Gamma_{BX} T_R)}{1 - \exp(-\Gamma_{BX} T_R)}$$

I.B. Active Switching Heralded (ASH) scheme

Due to the active switching the only parameter of importance for this scheme is the resolution time $T_R$ of the system. It is the difference in the arrival time between the two photons that needs to be taken into consideration. There are five possible outcomes:

- Two photons with temporal separation more than $T_R$ (successful heralding event):

  $$\mathcal{P}_{2S} = \alpha^2 QY_X QY_{BX} \int_{0}^{\infty} dT_B X \int_{T_B X + T_R}^{\infty} dT_X \ p(T_B X) p(T_X \mid T_{BX})$$

- Two photons with temporal separation arriving within $T_R$ from each other:

  $$\mathcal{P}_{2F} = \alpha^2 QY_X QY_{BX} \int_{0}^{\infty} dT_B X \int_{T_B X + T_R}^{\infty} dT_X \ p(T_B X) p(T_X \mid T_{BX})$$

- Exciton photon emission only:

  $$\mathcal{P}_{1X} = \alpha QY_X (1 - \alpha QY_{BX})$$

- Biexciton photon emission only:

  $$\mathcal{P}_{1BX} = \alpha QY_{BX} (1 - \alpha QY_X)$$
FIG. S1. (a) Schematic of Time-Gated Filtering technique (TGF). (b) The purity (solid line) and efficiency (dashed lines) of the TGF technique for the quantum yields corresponding to a gQD coupled to a nanocone [1]. This is compared to the efficiency of the ASH and TIMED technique

- No photon emission:

\[ \mathcal{P}_0 = (1 - \alpha QY_X)(1 - \alpha QY_{BX}) \]

Only the first of these events is a successful heralding event therefore the efficiency is given by:

\[ \eta_{ASH} = \alpha^2 QY_X QY_{BX} \exp(-\Gamma_X T_R) \]

On the other hand all but the last of these outcomes still will consitute a trigger event therefore the determinicity can be written as:

\[ D_{ASH} = \frac{\mathcal{P}_{2S}}{\mathcal{P}_{2S} + \mathcal{P}_{2F} + \mathcal{P}_{1X} + \mathcal{P}_{1BX}} = \frac{\alpha^2 QY_X QY_{BX} \exp(-\Gamma_X T_R)}{\alpha QY_X + \alpha QY_{BX} + \alpha^2 QY_X QY_{BX}} \]

To establish the relevance of these techniques we compare the efficiency of ASH and TIMED to the simpler TGF technique for the case of \( QY_X = 0.61 \) and \( QY_{BX} = 0.7 \) (see main text) [1] in figure S1b). For TGF \( T_F > 1.8 \) (indicated by the dotted line) is needed to produce the same purity (0.995) measured in our heralded experiments. For this value of \( T_F \) it is clear that the efficiency of both ASH and TIMED are superior to TGF.

Figure S2 displays the determinicity as a function of the quantum yield for equal quantum yields (solid line) and for the case where \( QY_{BX} = 0.5QY_X \) (dashed line). As can be clearly seen the ASH scheme has a determinicity of unity for unity quantum yields but TIMED reaches to a maximum of around 75% when the optimum cutoff time is
Determinicity as a function of the quantum yield for ASH ($T_R << \tau_X$) and TIMED ($T_C^{opt}$) as compared to an ideal SPS used. It should be noted however that this cutoff time was optimized for efficiency not determinicity and in principle one can operate at higher determinicities at the cost of lower efficiency. In the non-ideal case ($QY_{BX} = 0.5QY_X$) the TIMED scheme overcomes the ASH scheme since in this case false positives start to appear. ASH is more susceptible to false triggers since it is always open to the idler port until a trigger photon has arrived, whereas the TIMED scheme switches after $T_C$ regardless of the presence of a trigger photon or not.

Determinicity is an important parameter for a multiplexed heralded source. This parameter is important because it defines the upper limit for the overall efficiency of the multiplexed source since the optical router will rely on the trigger signal to switch between different components. The schemes proposed in this paper have maximum theoretical determinicities that approach unity for optimized parameters.

II. EXPERIMENTAL DETAILS

II.A. Setup and data format

The measurement is conducted by spin coating a sample of core/thick shell CdSe/CdS nanocrystal quantum dots dispersed in PMMA onto a silicon wafer. The input laser light is a femtosecond 2 MHz 405 nm pulsed laser generated by second harmonic generation of a femtosecond Ti:Sapphire laser operating at 810 nm. The sample is scanned using a periscopic system of scanning stages and the laser light is focused onto the sample using a 0.9 NA objective (Olympus MPLFLN100xBD). The emission from a single NQD is collected using the same objective and directed to the collection arm using a 600nm shortpass dichroic mirror. The sample is then imaged using a CMOS camera to verify the excitation of a single NQD. After that the emission is redirected to a set of beam-splitters and single photon avalanche photodiodes (Excelitas SPCM-AQRH-14-FC) as shown in figure 3a in the main text.

The signal from each detector is routed to a different channel on the timetagger device (Swabian TimeTagger 2.0). The output from the timetagger to the computer is a vector containing the arrival times of all the photons within the set exposure time with respect to the beginning of the measurement in addition to a tag labeling which channel this count came from. These arrival times are commonly referred to as global times. Another channel on the time-tagger records the excitation pulse times. By comparing the global time of each count with the nearest preceding laser pulse one can find the local time for the count. A histogram of these local times is what constitutes a lifetime measurement. Therefore at the end of this step we have information about which pulse and what channel the count came from, and its global and local times. This constitutes all the information needed for subsequent analysis steps.
II.B. Photon number correction

Due to the deadtime of each detector, counts from the same detector within the same pulse must be neglected. If this was not conducted then a bias would be introduced since two photons arriving with a time separation less than the dead time would not be detected whereas photons separated by more than the deadtime would be. However this does introduce an error in estimating the two and three photon counts due to the finite probability that two or more photons would arrive at the same detector. To correct for this error we must calculate the probability for the photons to arrive at different detectors and correct the photon counts by dividing by this probability. We have a two beamsplitter scheme the first with a reflectivity $R_1$ and the second with a reflectivity $R_2$. Assuming two photons were emitted the probability that they would arrive at different detectors is a simple exercise in probabilities and is given by:

\[ P_{\text{diff}}^{(2)} = 2R_1 (1 - R_1) + (1 - R_1)^2 [2R_2 (1 - R_2)] \]

If the pulse contains three photons the probability of them each arriving at a different detector is given by:

\[ P_{\text{diff}}^{(3)} = 3R_1 (1 - R_1)^2 [2R_2 (1 - R_2)] = 6R_1 R_2 (1 - R_1)^2 (1 - R_2) \]

Therefore the “true” photon numbers ($N_2$ and $N_3$) are related to the measured ones ($N_{2m}$ and $N_{3m}$) by:

\[ N_2 = \frac{N_{2m}}{P_{\text{diff}}^{(2)}} \]
\[ N_3 = \frac{N_{3m}}{P_{\text{diff}}^{(3)}} \]

In our case $R_1 = 0.4$, $R_2 = 0.5$ and therefore:

\[ N_2 = 1.52 N_{2m} \]
\[ N_3 = 4.63 N_{3m} \]

From now on it will be assumed that this correction is conducted whenever there is a two photon or three photon event.

II.C. Analysis Method

Prior to purification the purity of the QD was calculated by comparing the number of multiphoton ($N_{\geq 2}$) events to single photon events ($N_1$) using this formula:

\[ S = \frac{N_1}{N_1 + \frac{N_2}{\alpha} + \frac{N_3}{\alpha^2}} \]

Higher order terms were neglected due to the much lower probability of four or more emission events. This also is consistent with the ability to resolve up to three photons in our setup. The factor of $\alpha$ in the denominator is to account for the extra optical loss that a two photon and three photon states encounter.

In the heralded schemes we first filter out all counts having local times less than $T_F=300$ ps to reduce the effect of correlated noise. Then all pulses containing two or more photons were chosen for further analysis.

For the TIMED scheme the condition implemented afterwards is that one (or more) photons had a local time less than $T_C$ and one (or more) photons had a local time more than $T_C$. The number of these events constituted the number of successful heralding events and divided by the overall number of excitation pulses yields the heralding efficiency. This is what is plotted in figure 3d in the main text. From these figures we chose the $T_C$ that yields the maximum efficiency and for the successful heralding events we count the number of photons in each case that arrive after $T_C$ which will be called the signal photons. The number of cases in which we have two signal photons per heralding event $N_2$ compared to the number of cases where there is only one $N_1$ gives the purity: $S = 1 - \frac{N_2}{\alpha N_1}$
TABLE S2. System collection efficiency estimation

| Component                   | Method | Efficiency |
|-----------------------------|--------|------------|
| Collection into objective   | sim    | 39.0 %     |
| Objective transmission      | meas   | 90.0 %     |
| 600 nm SP dichroic          | meas   | 96.6 %     |
| 700 nm SP filter            | meas   | 88.5 %     |
| 550 nm LP filter            | meas   | 95.2 %     |
| 600 nm LP filter            | meas   | 84.3 %     |
| Beam splitters (2)          | meas   | 86.3 %     |
| Mirrors (6)                 | meas   | 75.9 %     |
| Fiber coupling              | meas   | 80.0 %     |
| Detector efficiency         | fact   | 70.0 %     |
| **Total**                   |        | **8.83 %** |

which can be plotted as a function of the original filtering time $T_F$ as in figure 3e in the main text. The factor of $\alpha$ in the denominator is to account for the extra optical loss that a two photon state encounters.

In the ASH technique the time differences between the two (or more) photons is calculated. If this time difference is more than $T_R$ then this is counted as a successful heralding event and the efficiency is calculated as in the previous case. Again in these heralding events the number of photons following the trigger photon is counted ($N_2$ if there is 2 and $N_1$ if there is 1) and the purity can be calculated using the same formula as above as a function of the filtering time.

II.D. Collection efficiency estimation

To estimate the collection efficiency in our setup we take into account the efficiencies of the various optical components using three techniques:

- Simulation (sim) of collection efficiency into objective by using a commercial FDTD software (Lumerical)
- Transmission/Reflection measurement using a 655 nm diode laser (meas).
- Factory data (fact)

The efficiency of our system is estimated in table S2.

II.E. Effect of noise on single photon purity

Experimentally we expect that the sole source reducing $S$ in the signal port is from noise. To check this point we show the purity, $S$, of both schemes as a function of the filtering time $T_F$ after the excitation pulse. We can see from figure 3e in the main text that the ASH purity improves dramatically with even 1 ns of filtering which fits well with the extracted correlated noise lifetime. The TIMED scheme, on the other hand, is hardly affected by the filtering. This difference in sensitivity can be understood since in the TIMED scheme we apply a passive gate, so all short-lifetime counts will by default be directed to the idler port without affecting the fraction of photons arriving to the signal port. On the other hand, in the ASH scheme a short lifetime count will lead to premature "switching" to the signal port causing both bi-exciton and exciton photons to be directed to the signal port. To confirm that that the short-lifetime counts are indeed from correlated noise counts and not from higher order multiexcitons we estimate their rate based on the measured ASH purity at $T_F = 0$. This is equivalent to finding the ratio of three photon to two photon events in the raw data. If indeed correlated noise is the source for the reduced $T_F = 0$ purity, this ratio should give the probability per pulse to get a correlated noise count which turns out to be $1.4 \times 10^{-3}$ or correspondingly the rate is $\sim 2800$ cps. This agrees with the correlated noise level measured at the same excitation power on the same substrate in a region with no quantum dots ($\sim 2500 - 3500$ cps). For $T_F > 1$ ns, $S$ reaches a maximum of around 0.995, after filtering out nearly all correlated noise. Uncorrelated noise is what limits this value from reaching unity. Using the procedure as above we find that this noise rate is around 880 cps which is in agreement with our uncorrelated noise rate $\sim 600-800$ cps.
To confirm these statements we will attempt to reconstruct the measured purities based on the error rates stated above. In general the single photon purity of a signal can be defined as:

\[ S = \frac{P_1}{P_{\geq 1}} = 1 - \frac{P_{\geq 2}}{P_{\geq 1}} \]

where \( P_1, P_{\geq 1} \) and \( P_{\geq 2} \) are the probabilities per relevant event of obtaining a single photon count, at least one photon, and at least two photons respectively.

Due to the low quantum yields in our case we can assume that \( P_{\geq 2} \approx P_2 \) and therefore:

\[ S = 1 - \frac{P_2}{P_{\geq 1}} \]

Note the absence of the \( \alpha \) factor here since we are theoretically considering the events obtained directly from the emitter. Now assume we have a certain probability per pulse \( \eta_{cn} = 1.4 \times 10^{-3}/\alpha \) and \( \eta_{un} = 4.4 \times 10^{-4}/\alpha \) of obtaining correlated and uncorrelated noise event respectively. We will now study the effect of these noise terms on our measured purity:

\[ \text{Heralded schemes} \]

For the heralded schemes in principle the purity should be unity unless there is noise. \( P_2 \) and \( P_{\geq 1} \) are the probabilities of getting two photons and at least one photon conditioned on the presence of another trigger photon.

For simplicity let's consider the ASH technique with \( T_R = 0 \) in two regimes:

- In the presence of correlated noise \( (T_F = 0) \) in this case \( P_{\geq 1} \approx QY_X QY_{BX} + (\eta_{cn} + \eta_{un}) (QY_X + QY_{BX}) \) notice that due to the heralding requirement this is effectively the probability of two photons in a non heralding scheme. On the other hand the probability of obtaining two photons in the signal port of a heralding technique is just the correlation between the noise counts and the heralded counts given by \( P_2 = QY_X QY_{BX} (\eta_{cn} + \eta_{un}) \). Therefore by using the values obtained from our experiment we find that the purity is given by:

\[ S_{\text{with noise}} = 1 - \frac{QY_X QY_{BX} (\eta_{cn} + \eta_{un})}{QY_X QY_{BX} + (\eta_{cn} + \eta_{un}) (QY_X + QY_{BX})} = 0.9867 \]

- In the absence of correlated noise \( (T_F \gg \tau_{cn}) \) and assuming that \( T_F \) is still much shorter than the biexciton lifetime then we can effectively set \( \eta_{cn} = 0 \) in the previous equation to obtain:

\[ S_{\text{with noise}} = 1 - \frac{QY_X QY_{BX} \eta_{un}}{QY_X QY_{BX} + \eta_{un} (QY_X + QY_{BX})} = 0.996 \]

This is in good agreement with the results shown in figure 3e of the main text.

\[ \text{III. PRACTICAL DETAILS} \]

In this section we will discuss in detail the effect of response time on the performance of ASH. The diagram of the proposed optical and electronic circuit needed to implement this technique is displayed in S3a and b. We conduct a detailed analysis of the different components that will contribute to \( T_R \) in table S3 where the propagation delay of the S/R latch is the composite delay of two NOR gates [6]. We consider different situations with either single photon avalanche detectors (SPAD) or semiconductor nanowire single photon detectors (SNSPD) combined with the other components either on-chip or free space. We compare the efficiency and single photon rate of the resulting 4 scenarios in Fig. S3c and d with a comparison with TIMED. Depending on the specific implementation the ASH technique will
FIG. S3. Schemes for the (a) ASH scheme and (b) S/R latch. (c) The heralding efficiency and (d) single photon rate of ASH under different response times as compared to the TIMED technique.

| Circuit Component     | Type       | Description            | Value       |
|-----------------------|------------|------------------------|-------------|
| Detector              | SNSPD      | Latency                | ∼50 ps [2]  |
|                       |            | Jitter                 | 15 ps [3]   |
| Detector              | SPAD       | Latency                | ∼2 ns [4]   |
|                       |            | Jitter                 | 50 ps [5]   |
| Logic Circuit         | S/R Latch  | Propagation delay      | 185 ps [6]  |
| Optical Switch        | 25 GHz Modulator | Rise time          | 15 ps [7]   |
| Optical/Electronic Propagation | On Chip | Negligible          |             |
|                       |            | Free Space             | 500 ps      |

TABLE S3. (a) Contributions of various components to $T_R$. (b) The total response time under different configurations.

be better than TIMED for $\tau_X$ more than a certain value.

[1] K. Matsuzaki, S. Vassant, H.-W. Liu, A. Dutschke, B. Hoffmann, X. Chen, S. Christiansen, M. R. Buck, J. A. Hollingsworth, S. Götzinger, and V. Sandoghdar, “Strong plasmonic enhancement of biexciton emission: controlled coupling of a single quantum dot to a gold nanocone antenna,” *Scientific Reports*, vol. 7, p. 42307, 2 2017.
[2] J. P. Allmaras, A. G. Kozorezov, B. A. Korzh, K. K. Berggren, and M. D. Shaw, “Intrinsic timing jitter and latency in superconducting single photon nanowire detectors,” 4 2018.
[3] Single Quantum, “Single Quantum Eos.”
[4] I. Prochazka, J. Blazej, and J. Kodet, “Measurement of the optical to electrical detection delay in the detector for ground-to-space laser time transfer,” *Metrologia*, vol. 48, pp. L13–L16, 6 2011.
[5] PicoQuant GmbH, “PDM Series.”
[6] ON Semiconductor, “2.5V/3.3V 12 Gb/s Differential Clock/Data SmartGate.”
[7] iXblue, “NIR-MX800-LN-10.”