Supporting information of

Photon statistics of incoherent cathodoluminescence with continuous and pulsed electron beams

Magdalena Solà-Garcia,[1,*] Kelly W. Mauser, [1] Matthias Liebtrau, [1] Toon Coenen, [1,2] Silke Christiansen, [3] Sophie Meuret, [4] Albert Polman[1]

[1] Center for Nanophotonics, AMOLF, Science Park 104, 1098 XG Amsterdam, The Netherlands

[2] Delmic BV, Kanaalweg 4, 2628 EB, Delft, The Netherlands

[3] Fraunhofer Institute for Ceramic Technologies and Systems IKTS, Äußere Nürnberger Strasse 62, 91301 Forchheim, Germany

[4] CEMES-CNRS, 29 Rue Jeanne Marvig, 31055 Toulouse, France

[*] Corresponding author: m.sola@amolf.nl

Table of contents:

S1. Comparison of analytical model to Monte Carlo simulations

Previous $g^{(2)}(\tau)$ measurements in cathodoluminescence (CL) have been modelled using Monte Carlo (MC) simulations. Here we demonstrate the accuracy of our analytical model by comparing $g^{(2)}(0)$ results obtained with our analytical model to those produced by MC simulations. The comparison is performed for the three different electron beam configurations (continuous, pulsed with beam blanker and pulsed through photoemission). In all cases, the following steps were considered after the arrival of an electron to the sample:

1. Creation of b_i bulk plasmons, according to a Poisson distribution with expectation value b.
2. Decay of each plasmon into m_i electron-hole pairs, described with a Poisson distribution with expectation value m.
3. Excitation of a quantum well by an electron-hole pair with probability η.
4. Emission of a photon, following a given decay mechanism.

The MC simulations with a continuous electron beam have been performed using the same code as in refs. [1–3]. The code was adapted to represent the blanker experiment, in which only part of the initial continuous beam reaches the sample, thus generating (relatively long) electron pulses. In the MC simulations for the blanker case the current in continuous mode and the repetition rate were set to 20 pA and 1 MHz, respectively, and the pulse width was varied from 8 up to 500 ns, similar to the experiments. We also adapted the initial MC code to simulate electron pulses generated by photoemission. In this case, no continuous electron beam is initially generated, but instead we directly create pulses containing a certain number of electrons per pulse, given by a Poisson distribution with expectation value n_e. The pulse width is assumed to be Gaussian, with $\sigma = 1$ ps. The exact value of the pulse width is not critical, given that it is much shorter than the emitter lifetime τ_{emitter}. The MC simulations with photoemission were performed assuming a repetition rate of $F = 5.04$ MHz. In all cases, the results from the simulations have been analyzed using the same procedure as for the experimental data.

Figure S1 shows the values of $g^{(2)}(0) - 1$ obtained from the MC simulations using a continuous electron beam (a), and a pulsed electron beam generated by beam blanking (b) and photoemission (c). In the three cases we show $g^{(2)}(0) - 1$ as a function of electron beam current (a) and number of electrons per pulse (b, c). In all cases we consider an exponential decay for the emitter, with lifetime $\tau_{\text{emitter}} = 12$ ns and an average number of $b = 0.2$ bulk plasmons per electron that interact with the quantum wells, corresponding to an excitation efficiency of $\gamma = 0.18$. We also assume $m = 1$ and $\eta = 1$, even though it has already been shown that these parameters do not play a role in the final result of the MC simulation[2]. In this work we explain this fact by showing that m and η cancel out in the development of the analytical model. The time step in the simulations was set to 512 ps, the same as in our experiments.

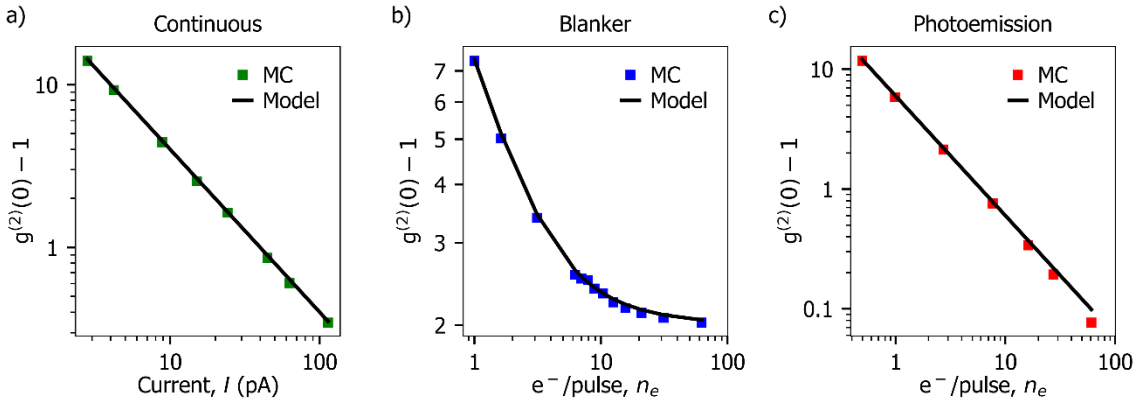

Figure S1: Monte Carlo simulations of $g^{(2)}(0)$ amplitude vs. electron beam current or number of electrons per pulse, together with the results from the analytical model. (a) Continuous electron beam, (b) electron beam pulsed using the blanking technique and (c) pulsed electron beam generated by photoemission.

Figure S1 also shows the results of our analytical model, in which we used the same parameters as in the MC simulations. We should note that here no fitting is needed, given that we just fix all the parameters (including b). The results show a very good agreement between the MC simulations and the analytical model.

S2. Analytical model – Continuous electron beam

A $g^{(2)}(\tau)$ experiment measures the photon statistics of a given emitter or source, and it is therefore based on random processes: the emission of a photon is stochastic, following a certain probability distribution (for example, an exponential decay). Moreover, in typical experimental setups (such as the HBT experiment), the emitted photons are split randomly towards the two detectors, with a 50% probability of being detected by each detector[4]. The measurement is based on collecting enough statistics such that it can accurately represent the chances of having a correlation at a given delay compared to any other. It is therefore not possible to predict exactly how many photons will correlate with photons from the same electron (thus leading to bunching), and how many with photons from other electrons (uncorrelated events). Instead, we can calculate how likely it is that one scenario happens with respect to the other one. Hence, our analytical model is based on calculating the average number of combinations of correlations that lead to bunching (A_b) compared to the average number of combinations that lead to uncorrelated events (photons coming from different electrons, or pulses, in the pulsed case, A_uncorr).

We start from Eq. (2) of the main text, in which:

$$g^{(2)}_{cont}(0) = 1 + \frac{A^c_\mathrm{b}}{A^c_\mathrm{uncorr}} \frac{(2B+1)t_B}{2\alpha_\mathrm{b}}, \tag{S1}$$

where A^c_b contains the mean number of combinations of correlations between photons from the same electron (i.e., giving bunching), A^c_uncorr is the mean number of combinations of correlations between photons from different electrons (uncorrelated), B is the total number of bins during the acquisition time T, t_B is the bin size and α_b is the shape factor of the bunching peak, defined as the ratio between the area and height of the peak. Hence, we now need to calculate A^c_b and A^c_uncorr.

S2a. Correlations between photons from the same electron (A^c_b, bunching peak)

We consider that when one electron interacts with an emitter (semiconductor, quantum wells, atomic defect, etc), a certain number of photons k will be emitted, each of them with a certain arrival time t_k. We should note

that this value k is not fixed, but will be different for each electron, given the stochastic nature of all the processes (creation of bulk plasmon, decay into electron-hole pairs, radiative decay of emitter). We want to count the number of combinations of pair-correlation events between photons from the same electron. We define one correlation as the detection of a pair of photons, thus we need to take subsets of 2 from k photons, each photon with a fixed arrival time. Moreover, the order matters, given that this will determine whether the measured delay between photons is positive or negative, and there are no repetitions, i.e., a photon cannot correlate with itself. This is a common problem in combinatorics[5], sometimes referred to as variation without repetition, from which we extract that the number of possible combinations is

$$A_k = \binom{k}{2} = k(k-1). \tag{S2}$$

Next, we want to relate A_k to physical variables, i.e., expected value of number of bulk plasmons per electron (b), expected value of number electron-hole pairs created per plasmon (m) and radiative decay efficiency (η). We will follow steps 1-3 described in the main text, starting from step 3 and building up.

3. Given m_i electron-hole pairs, each of them with a probability η of exciting a QW that emits a photon, the expected value of the number of possible combinations of correlations of photons becomes

$$A_3 = \sum_{k=0}^{m_i} A_k \mathrm{Bin}(k; m_i, \eta) = \sum_{k=0}^{m_i} k(k-1) \frac{m_i!}{k!\,(m_i-k)!} \eta^k (1-\eta)^{m_i-k} = \eta^2 m_i (m_i - 1). \tag{S3}$$

2. Each bulk plasmon will create m_i electron-hole pairs, described with a Poisson distribution with expected value m $\left(\mathrm{Poiss}(m_i; m) = \frac{e^{-m} m^{m_i}}{m_i!}\right)$. Hence, we need to account for all the possible values of m_i, weighted by their probability. The expected value of the number of possible combinations correlations of photons produced by m_i electron-hole pairs is then

$$A_{2,b_i=1} = \sum_{m_i=0}^{\infty} A_3 \mathrm{Poiss}(m_i; m) = \sum_{m_i=0}^{\infty} \eta^2 m_i (m_i - 1)\, \frac{e^{-m} m^{m_i}}{m_i!} = m^2 \eta^2. \tag{S4}$$

If an electron creates more than one bulk plasmon, each of these plasmons will decay into a certain amount of electron-hole pairs, with likelihood given by a Poisson distribution with expectation value m, as already described. Thus, we need to account for all the possible combinations of correlations of photons produced by an arbitrary number b_i of bulk plasmons. We start with the case of two bulk plasmons, in which the expected value of the number of possible combinations of correlations of photons (the correlations can be from photons from the same or different plasmon) becomes

$$A_{2,b_i=2} = \sum_{m_1=0}^{\infty} \sum_{m_2=0}^{\infty} \eta^2 (m_1 + m_2)(m_1 + m_2 - 1)\ \mathrm{Poiss}(m_1; m) \mathrm{Poiss}(m_2; m) = 4m^2 \eta^2. \tag{S5}$$

It can be shown by induction (see S2d) that in the general case of b_i bulk plasmons, which produce photons that can correlate with other photons from the same plasmon or a different plasmon, the expectation value of the number of possible combinations of correlations is

$$A_{2,b_i} = b_i^2 m^2 \eta^2. \tag{S6}$$

1. Finally, the number of bulk plasmons produced by a single electron also follows a Poisson distribution with expected value b ($\mathrm{Poiss}(b_i; b)$). Therefore, averaging again over all possible values of b_i, we obtain that the average number of possible combinations of correlations produced by one electron is

$$A_1 = \sum_{b_i=0}^{\infty} b_i^2 m^2 \eta^2 \mathrm{Poiss}(b_i; b) = \; bm^2\eta^2(b+1). \tag{S7}$$

In the case of n electrons, the mean number of possible combinations of correlations between photons from the same electron becomes:

$$A_{\mathrm{b}}^{c} = nA_1 = nbm^2\eta^2(b+1). \tag{S8}$$

S2b. Correlations between photons from different electrons (A_1^c)

Next, we need to calculate the number of possible combinations of correlations between photons from different electrons. Taking into account the statistical distributions of each parameter involved in the emission of a photon (bulk plasmons, carriers, emission efficiency), the average number of photons emitted per electron is

$$N_{\mathrm{ph}} = \sum_{b_i=0}^{\infty} \sum_{m_i=0}^{\infty} \sum_{\mathrm{k}=0}^{\mathrm{m_i}} k b_i \mathrm{Poiss}(b_i; b)\ \mathrm{Poiss}(m_i; m)\ \mathrm{Bin}(k; m_i, \eta) \; = bm\eta. \tag{S9}$$

We should note that the result is the same as if we would just consider the expected values b, m and η given that the number of photons scales linearly with these parameters. We now need to create pairs between two photons from different electrons. In this case the order is still important. We calculate the average number of combinations of correlations of photons coming from different electrons as

$$A_{\mathrm{uncorr}}^{c} = [nbm\eta][(n-1)bm\eta] = n(n-1)b^2m^2\eta^2. \tag{S10}$$

S2c. $g^{(2)}(0)$ for a continuous beam

Finally, we can insert Eqs. (S8) and (S10) into Eq. (S1), rewrite n as a function of the electron current, $I = nq/(Bt_b)$. Given a typical acquisition time of a $g^{(2)}(\tau)$ experiment of at least 1min, and bin size of $t_b = 0.512$ ns, the total number of bins becomes $B \approx 10^{11}$. It is therefore reasonable to take the limit $B \to \infty$ to obtain

$$g_{\mathrm{cont}}^{(2)}(0) = \lim_{B\to\infty}\left(1 + \frac{(2B+1)t_b}{2\alpha_{\mathrm{b}}}\frac{b+1}{\left({}^{I}/_{q}\, t_b B - 1\right)b}\right) = 1 + \frac{q}{I\alpha_{\mathrm{b}}}\frac{b+1}{b}. \tag{S11}$$

The last expression can also be expressed in terms of the excitation efficiency γ (Eq. (6) in the main text, further explained in section S2e) such that

$$g_{\mathrm{cont}}^{(2)}(0) = 1 + \frac{q}{I\alpha_{\mathrm{b}}}\frac{\log(\gamma-1)-1}{\log(\gamma-1)}. \tag{S12}$$

S2d. Bunching peak: mean number of possible combinations of photon correlations from b_i bulk plasmons

We want to find the expected value for number of combinations of correlations for an arbitrary number of bulk plasmons. Similar to Eq. (S5), in the case of $j+1$ bulk plasmons, we have

$$A_{2,j+1} = \eta^2 \sum_{m_1=0}^{\infty} \dots \sum_{m_j=0}^{\infty} \sum_{m_{j+1}=0}^{\infty} \left(\bar{m} + m_{j+1}\right)\left(\bar{m} + m_{j+1} - 1\right)\ \mathrm{P}_{\bar{m}}\mathrm{P}_{j+1} \tag{S13}$$

$$= \eta^2 \sum_{m_1=0}^{\infty} \dots \sum_{m_{b_j}=0}^{\infty} \sum_{m_{j+1}=0}^{\infty} (m_{j+1}^2 + m_{j+1}(2\bar{m} - 1) - \bar{m} + \bar{m}^2)\ \mathrm{P}_{\bar{m}} \mathrm{P}_{j+1},$$

where we have defined $\bar{m} = m_1 + .. + m_j$ and $\mathrm{P}_{\bar{m}}$ is the product of Poisson distributions, i.e., $\mathrm{P}_{\bar{m}} = \mathrm{Poiss}(m_1; m) \dots \mathrm{Poiss}(m_j; m) = \prod_{i=1}^{j} \frac{e^{-m} m^i}{i!}$. Eq. (S13) can be further developed into

$$A_{2,\mathrm{b=j+1}} = \eta^2 \left[m(m+1) + 2jm^2 - m - jm + \sum_{m_1=0}^{\infty} \dots \sum_{m_j=0}^{\infty} \bar{m}^2 \mathrm{P}_{\bar{m}} \right]. \tag{S14}$$

Therefore, we need to find an analytical expression for the last term in Eq. (S14). We assume that

$$\sum_{m_1=0}^{\infty} \dots \sum_{m_j=0}^{\infty} \bar{m}^2 \mathrm{P}_{\bar{m}} = jm(jm+1), \tag{S15}$$

which we will prove by induction. In the case of $j = 1$,

$$\sum_{m_1=0}^{\infty} m_1^2\, \mathrm{P}(\mathrm{m}_1; \mathrm{m}) = m(m+1). \tag{S16}$$

Assuming that Eq. (S15) is true, we need to prove that in the case of $j+1$ bulk plasmons, it becomes $(j+1)m[(j+1)m]$. Hence

$$\sum_{m_1=0}^{\infty} \dots \sum_{m_j=0}^{\infty} \sum_{m_{j+1}=0}^{\infty} \left(\bar{m} + m_{j+1}\right)^2 P_{\bar{m}} P_{j+1} = \sum_{m_1=0}^{\infty} \dots \sum_{m_j=0}^{\infty} \sum_{m_{j+1}=0}^{\infty} (\bar{m}^2 + 2\bar{m}m_{j+1} + m_{j+1}^2)\, P_{\bar{m}} P_{j+1} \tag{S17}$$

$$= jm(jm+1) + 2jm^2 + m(m+1) = (j+1)m[(j+1)m].$$

Finally, inserting Eq. (S15) into Eq. (S14), we obtain

$$A_{2,\mathrm{b=j+1}} = (j+1)^2 m^2 \eta^2. \tag{S18}$$

S2e. Obtaining the excitation efficiency (γ)

For each electron, the probability of interacting, i.e., creating at least one plasmon that can excite the quantum wells, or any other emitter, is

$$P_{int} = 1 - \mathrm{Poiss}(0; b) = 1 - \mathrm{e}^{-b}, \tag{S19}$$

where b is the average number of bulk plasmons generated per electron (around the emitter). We define γ as the fraction of electrons that create at least one bulk plasmon near the emitter. Given a certain number of electrons n_{total}, from which $n_{interacting}$ interact with the emitter, γ becomes

$$\gamma = \frac{n_{interacting}}{n_{total}} = \frac{n_{total} P_{int}}{n_{total}} = 1 - \mathrm{e}^{-b}. \tag{S20}$$

S2f. Number of correlations at long delays

We consider that electrons interact with the sample during a certain (square) time window $T = Bt_b$, where B is the total number of bins and t_B is the bin size. The distribution of the electrons in time can be represented as a uniform random distribution. The number of possible correlations between photons coming from different

electrons as a function of delay τ exhibits a triangular shape, with base corresponding to $2T$. This shape results from the convolution of two squared signals with width T. Thus, in the model, the total number of correlations are spread within an area corresponding to a triangle, with base $(2B+1)t_b$ and height h_1. Figure S2 shows an example of this effect. In the experiments, the typical acquisition time (at least seconds) is much larger than the time window within which we acquire correlations (30 μs in our case for $t_b = 0.512$ ps), and thus this effect becomes negligible in the narrow time window in which we analyze the data, i.e., we only see the very top of the triangle. However, in the case of a pulsed electron beam, the time window corresponds to the pulse width Δp. The correlations between photons from the same or different pulses then exhibit a triangular shape, with base $2\Delta p$, as a function of τ. This is the shape that we observe in our $g^{(2)}(\tau)$ measurements with the blanker (inset of Fig. 3, main text).

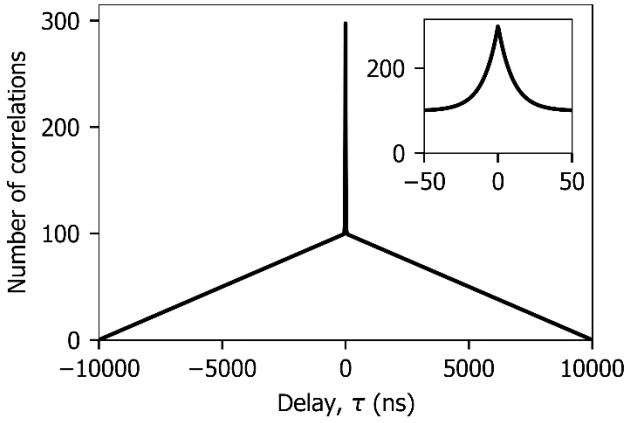

Figure S2: Theoretical shape of the number of possible correlations as a function of delay τ between photons emitted after the exposure of the sample to an electron beam during a time T. Here, $T = 10\ \mu s$. The observed triangular shape results from the convolution of two square pulses with base T. The inset shows the curve at small delay, in which the triangular background is not observed.

S2g. Calculation α_0. Discussion discrete/continuous

A key parameter in our analytical model is the relation between the height and the area of the bunching peak, α_b (shape factor). Given a known decay function of the bunching peak, the shape factor can typically be easily calculated. In the case of a simple exponential we obtain $\alpha_\mathrm{b} = 2\tau_\mathrm{b}$, while for stretched exponential the shape factor becomes $\alpha_\mathrm{b} = 2\frac{\tau}{\beta}\Gamma\left(\frac{1}{\beta}\right)$. Nevertheless, $g^{(2)}(\tau)$ measurements are discrete, and thus these expressions for α_b are only valid if the bin size t_bin is much smaller than the typical decay time, such that we can assume an almost continuous function. Otherwise, the discretized nature of the measurement should be taken into account. For example, the generalized expression for α_bunching in the case of an exponential decay with arbitrary bin size is $\alpha_\mathrm{b} = \frac{2t_b}{1-e^{-t_b/\tau_\mathrm{b}}}$, which becomes $2\tau_\mathrm{b}$ when $t_b \ll \tau_\mathrm{b}$.

S3. Analytical model – Pulsed electron beam

In the case of a pulsed electron beam we need to adapt the definition of $g^{(2)}(0)$ given in Eq. (1) of the main text. Here, we need to normalize the height at $\tau = 0$, with respect to the height of any other peak, H_i, which represents the uncorrelated events. Hence

$$g^2_{\text{pulsed}}(0) = \frac{H(0)}{H_i(i \neq 0)}. \tag{S21}$$

$H(0)$ is the height of the peak at $\tau = 0$. This peak will contain two contributions: correlations between photons from the same electron (with mean number of possible combinations A^p_{b}) and between photons from different electron but same pulse ($A^p_{\text{uncorr},0}$). As discussed in main text, the first contribution will be distributed over a temporal shape (with height h^p_{b} and area A^p_{b}) determined by the emitter decay, through Eq. (S36). Hence, the ratio between the area and the height is: $h^p_{\text{b}} = \alpha_{\text{b}} A^p_{\text{b}}$, similar to the bunching contribution in the continuous case. In contrast, the temporal distribution of the correlations between photons from different electrons but same pulse (second contribution of $\tau = 0$ peak) depends not only on the emitter decay but also the shape of the electron pulse. Hence, the shape of this contribution is given by the convolution of two electron pulses, convoluted also with the emitter decay (see S4b). We define the ratio between the area ($A^p_{\text{uncorr},0}$) and the height ($h^p_{\text{uncorr},0}$) of this part as: $A^p_{\text{uncorr},0} = \alpha_{conv} h^p_{\text{uncorr},0}$.

$h_{\text{uncorr},i}(i \neq 0)$ is the height of any peak at $\tau \neq 0$, i.e., containing correlations between photons from consecutive pulses ($i = \pm 1$), from every second pulse ($i = \pm 1$) and so on. The shape of any of these peaks is also determined by the electron pulse shape and emitter decay, hence we can define: $A^p_{\text{uncorr,i}} = \alpha_{conv} h^p_{\text{uncorr,i}}$. $A^p_{\text{uncorr,i}}$ contains the possible correlations between photons from different pulses.

Taking the previous definitions into account, we can rewrite (S21) as

$$g^{(2)}_{\text{pulsed}}(0) = \frac{h^p_{\text{b}} + h^p_{\text{uncorr},0}}{h^p_{\text{uncorr,i}}} = \frac{\alpha_{conv} A^p_{\text{b}} + \alpha_{\text{b}} A^p_{\text{uncorr},0}}{\alpha_{\text{b}} A^p_{\text{uncorr,i}}}. \tag{S22}$$

S3a. Correlations between photons from the same pulse

Correlations between photons from the same electron (A^p_{b})

The mean number of possible combinations of correlations between photons from the same electron, i.e., leading to bunching, is given in Eq. (S7), which we have to multiply by the number of electrons per pulse and the total number of pulses r

$$A^p_{\text{b}} = rb(b+1)m^2\eta^2 \sum_{n_i=0}^{\infty} n_i \text{Poiss}(n_i; n_e) = r n_e b(b+1)m^2\eta^2. \tag{S23}$$

Here we have assumed that the number of electrons per pulse n_i follows a Poisson distribution with expected value n_e. This will be the case in most experiments, such as in the beam blanker and photoemission of electron pulses described in the main text. However, we would obtain the same result if we consider the number of electrons per pulse fixed, given that A^p_{b} scales linearly with n_i.

Correlations between photons from different electron within the same pulse ($A^p_{\text{uncorr},0}$)

Given the average number of emitted photons per electron N_{ph} (Eq. (S9)), the number of combinations of correlations between photons from the same pulse, but different electron, becomes

$$A^p_{\text{uncorr},0} = b^2 m^2 \eta^2 \sum_{n_i=0}^{\infty} n_i(n_i - 1)\text{Poiss}(n_i; n_e) = r n_e^2 b^2 m^2 \eta^2. \tag{S24}$$

S3b. Correlations between photons from different pulses ($A^p_{\mathrm{uncorr},i}$)

Following from Eq. (S9), which gives the average of photons emitted per electron, and assuming n_i electrons per pulse (Poisson-distributed), the average number of photons emitted per pulse is

$$N_p = bm\eta \sum_{n_i=0}^{\infty} n_i \mathrm{Poiss}(n_i; n_e) = n_e bm\eta. \tag{S25}$$

The number of possible correlations between photons from different pulses is therefore

$$\mathrm{A}^p_{\mathrm{uncorr,r}} = r(r-1)n_e^2 b^2 m^2 \eta^2, \tag{S26}$$

which is distributed over $2(r-1)$ peaks, given that we do not count the peak at $\tau = 0$, which would contain correlations between photons from the same pulse. We also need to take into account that the peaks at τ_i are contained within a triangular envelope, given that the number of possible correlations decreases as the delay between pulses increases, as explained in section S2f.

Hence, the area below each peak at τ_i becomes

$$A^p_{\mathrm{uncorr},i} = \frac{2\,\mathrm{A}^p_{\mathrm{uncorr,r}}}{2(r-1)} = r n_e^2 b^2 m^2 \eta^2. \tag{S27}$$

S3c. $g^{(2)}(0)$ for a pulsed electron beam

Finally, inserting Eqs. (S23), (S24) and (S27) into Eq. (S22) we obtain

$$g^{(2)}_{\mathrm{pulsed}}(0) = 1 + \frac{\alpha_{\mathrm{conv}}}{\alpha_{\mathrm{b}}} \frac{b+1}{n_e b} = 1 + \frac{\alpha_{\mathrm{conv}}}{n_e \alpha_{\mathrm{b}}} \frac{\log(\gamma - 1) - 1}{\log(\gamma - 1)}. \tag{S28}$$

In which again we have used the relation between b and γ given in Eq. (6) of the main text.

S3d. Alternative calculation of γ for a pulsed electron beam

In the case of a pulsed electron beam, we don't need to calculate $\mathrm{g}^{(2)}(0)$ to retrieve the excitation efficiency γ, but we can simply divide the sum of Eqs. (S23) and by Eq. (S27), which results in

$$\frac{A^p_{\mathrm{b}} + A^p_{\mathrm{uncorr},0}}{A^p_{\mathrm{uncorr},i}} = 1 + \frac{b+1}{n_e b} = 1 + \frac{1 - \log(1-\gamma)}{n_e \log(1-\gamma)}, \tag{S29}$$

which corresponds to Eq. (10) in the main text. In experiments, this ratio would be equivalent to dividing the sums of all the counts below the peak at 0 delay with the sum of the counts below any other peak.

S3e. Alternative derivation of $g^{(2)}_{\mathrm{pulsed}}(0)$ in photoemission

The previous derivation of $g^{(2)}(0)$ assumes that bunching comes only from correlations between photons from the same electron. However, in the case of electron pulses obtained by photoemission, several electrons might excite the sample instantaneously (i.e., within a ps timescale, much smaller than the emitter decay). In this case bunching comes from correlations between photons from the same pulse, and it doesn't matter whether they come from the same or different electrons. Here we show a derivation of $g^{(2)}(0)$ starting from the point that all electrons within a pulse will create bunching, and show that it results in the same expression as Eq. (12) (main text).

We assume that the duration of the electron pulses is much smaller than the emitter lifetime. Hence, all peaks will have the same shape. In particular, the area below the bunching peak (peak at $\tau = 0$) is related to its height as: $A'_{\mathrm{b}} = \alpha_{\mathrm{b}} h'_{\mathrm{b}}$. Similarly, any other peak $\tau_i (i \neq 0)$ follows the same relation: $A'_{\mathrm{i}} = \alpha_{\mathrm{i}} h'_{\mathrm{i}}$. Eq. (S22) can now be written as

$$g^{(2)}_{\mathrm{ultrashort,v2}}(0) = \frac{A'_{\mathrm{b}}}{A'_i}. \tag{S30}$$

We first calculate the area below the bunching peak, i.e., A'_{b}. The first steps are the same as in the continuous case. From Eq. (S6) we know that given b_i bulk plasmons, the mean number of combinations of correlations is $b_i^2 m^2 \eta^2$. Assuming that we have n_i electrons per pulse, each of them can create a different number of plasmons b_i. The case of $n_i = 1$ is derived in Eq. (S7). In the case of $n_i = 2$,

$$A_{n_i=2} = \sum_{b_1=0}^{\infty} \sum_{b_2=0}^{\infty} (b_1 + b_2)^2 m^2 \eta^2 \mathrm{Poiss}(b_1; b)\, \mathrm{Poiss}(b_2; b) = 2bm^2\eta^2(2b+1). \tag{S31}$$

In the general case of n_i electrons per pulse, it can be shown (through a similar demonstration as in section S2d) that

$$A_{n_i} = n_i b m^2 \eta^2 (b n_i + 1). \tag{S32}$$

Finally, the number of electrons per pulse is not fixed but follows a Poisson distribution with expected value n_e. Moreover, we need to multiply this by the total number of pulses exciting the sample during a measurement (r). Hence, the average number of combinations of pair-correlations leading to bunching becomes

$$A'_{\mathrm{b}} = rbm^2\eta^2 \sum_{n_i=0}^{\infty} n_i(bn_i + 1)\mathrm{Poiss}(n_i; n_e) = rn_e bm^2\eta^2(n_e b + b + 1). \tag{S33}$$

The area below each peak i, containing the number of combinations of pair-correlations between photons from different pulses (consecutive pulses, every second pulse, etc) was already calculated in Eq. (S27). Hence,

$$A'_i = A^p_{\mathrm{uncorr},i} = \frac{2\, \mathrm{A}^p_{\mathrm{uncorr,r}}}{2(r-1)} = rn_e^2 b^2 m^2 \eta^2. \tag{S34}$$

Inserting Eqs. (S34) and (S34) into Eq. (S30)(S30) yields

$$g^{(2)}_{\mathrm{ultrashort,v2}}(0) = 1 + \frac{b+1}{n_e b}, \tag{S35}$$

which is the same as $g^{(2)}_{\mathrm{ultrashort,}}(0)$ given in Eq. (12) of the main text, which was obtained by setting $\alpha_{conv} = \alpha_{\mathrm{b}}$ in Eq. (S28).

S4. Full description of $g^{(2)}(\tau)$

In the previous sections we have derived the value of $g^{(2)}(0)$, but we have not discussed yet the full shape of the autocorrelation function as a function of delay (i.e., $g^{(2)}(\tau)$). In the continuous case, the shape of $g^{(2)}(\tau)$ only depends on the bunching peak, while in pulsed experiments $g^{(2)}(\tau)$ depends only on the temporal shape of the electron pulses, as will be seen below.

S4a. Shape of the bunching peak

Given a certain function $y(t)$, the result of its autocorrelation is[4]

$$h(\tau) = \int_{-\infty}^{\infty} y(t)y(t+\tau)\,dt = y(-\tau) * y(\tau). \tag{S36}$$

In the case of the bunching peak in a $g^{(2)}(\tau)$ measurement, $y(t) = y_{\text{emitter}}(t)$ and $h(\tau) = y_{\text{bunching}}(\tau)$, as given in Eq. (7) in the main text.

S4b. Shape of uncorrelated peaks in a pulsed electron beam

As we have already discussed, a $g^{(2)}(\tau)$ measurement in pulse shows peaks centered at 0-delay and delays τ_i $(i = \pm 1, \pm 2 \dots)$ corresponding to the time between pulses. The peak at $\tau = 0$ has contributions from bunching, which result in a shape determined by the emitter (as shown in Eq. (S36)(S36) and Eq. (7) in the main text), and from uncorrelated photons, i.e., coming from different electrons. The peaks at τ_i $(i \neq 0)$ contain uncorrelated photons, i.e., coming from different pulses. In all cases in which there are correlations between photons from different electrons, the shape of the electron pulse also plays a role, together with the emitter decay. The probability of emitting a photon coming from a pulsed electron beam is given by the convolution between the electron pulse shape ($p(t)$) and emitter decay (y_{emitter}), i.e.,

$$y(t) = p(t) * y_{\text{emitter}}(t). \tag{S37}$$

And from Eq. (S36), we know that the correlation between two photons with temporal spread $y(t)$ is

$$h^{p}_{\text{uncorr}}(\tau) = [p(-\tau) * y_{\text{emitter}}(-\tau)] * [p(\tau) * y_{\text{emitter}}(\tau)] = [p(-\tau) * p(\tau)] * h_{\text{b}}(\tau), \tag{S38}$$

where in the last step we have used the definition of $h_{\text{b}}(\tau)$ from Eq. (7) in the main text.

S4c. Comparison to experiments: stretched exponential decay

In order to test the validity of Eq. (7) (main text) (same as Eq. (S36)), describing the shape of the bunching peak, we performed time-correlated single-photon counting measurements (TCSPC) on the sample. We subsequently acquired a $g^{(2)}(\tau)$ measurement with exactly the same conditions. The TCSPC measurements were performed using a pulsed electron beam obtained by photoemission, with the same conditions as in the $g^{(2)}(\tau)$ photoemission experiments (here, $n_e = 347$ electrons/pulse), and the data is collected in the same way as explained in refs.[6,7]. In these measurements, a histogram of the arrival time of photons following the electron pulse is built, and thus they directly show the emission decay. Figure S3a shows the resulting decay trace. We observe that the trace can be best fitted using a stretched exponential

$$y_{emitter}(t) = y_0 e^{-\left(\frac{t}{\tau_{\text{emitter}}}\right)^{\beta_{\text{emitter}}}}. \tag{S39}$$

In this case we find that $\tau_{\text{emitter}} = 11$ ns and $\beta_{\text{emitter}} = 0.73$. The histogram obtained in the corresponding $g^{(2)}(\tau)$ measurement is shown in Figure S3b. We observe that the bunching peak cannot be properly described with a stretched exponential using τ_{emitter} and β_{emitter} as parameters (green curve). Instead, the result of solving numerically Eq. (S36) with the emitter parameters exhibits a very good agreement with the data (red curve).

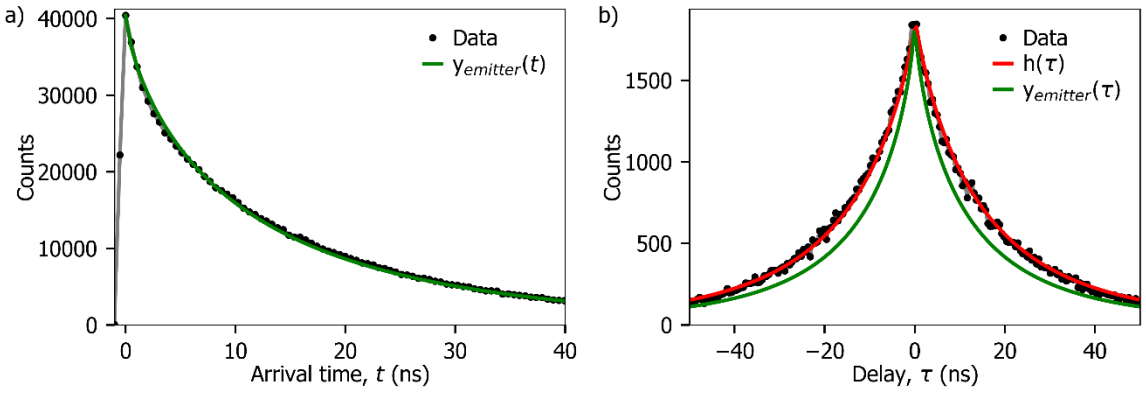

Figure S3: (a) TCSPC measurement on quantum wells, performed using a pulsed electron beam generated by photoemission (black) together with the corresponding fit with a stretched exponential (green). The best fit is obtained when $\tau_{\text{emitter}} = 11$ ns and $\beta_{\text{emitter}} = 0.73$. (b) $g^{(2)}(\tau)$ measurement performed on the same area on the sample and identical conditions as in (a) (black), together with the result from Eq. (S36) using the emitter parameters from (a) (red) and a stretched exponential using the emitter parameters. The model (Eq. (S36)) shows a very good agreement with the data.

The discrepancy between the shape of the $g^{(2)}(\tau)$ curve and the actual emitter decay when the latter follows a stretched exponential could explain the different lifetimes obtained in ref. [7] when comparing $g^{(2)}(\tau)$ and decay trace measurements.

Single and double exponential decays

In most systems, the decay mechanism can be approximated with a single or double exponential decay. Solving Eq. (S36) in those cases yields: $y_{\text{bunching}}(\tau) \propto e^{-\tau/\tau_1}$ and $y_{\text{bunching}}(\tau) \propto e^{-\frac{\tau}{\tau_1}} + e^{-\tau/\tau_2}$, for single and double exponential decays, respectively. Therefore, in both cases the decay of the $g^{(2)}(\tau)$ function directly gives the decay of the emitter.

S4d. Comparison to experiments: square electron pulse

The experiments using the beam blanker are performed using square electron pulses, with pulse width Δp determined by the blanking conditions (repetition rate and duty cycle) (see Section S6a). Hence, the pulse shape is given by

$$p(t) = \begin{cases} 1, & 0 \le x \le \Delta p \\ 0, & \text{otherwise,} \end{cases} \tag{S40}$$

and the emitter decay $y_{\text{emitter}}(t)$ follows the expression in Eq. (S39). The shape of the peaks at $\tau_i (i \neq 0)$ then become (Eq. (S38))

$$h^{p}_{\text{uncorr}}(\tau) = T(\tau) * [y_{\text{emitter}}(-\tau) * y_{\text{emitter}}(\tau)] = T(\tau) * h_{\text{b}}(\tau), \tag{S41}$$

where $T(\tau)$ is a triangular function with base Δp, resulting from the convolution of $p(t)$ with $p(-t)$.

S5. Correction at long delays

In a $g^{(2)}(\tau)$ measurement, when the delay is longer than the typical correlation time (in our case, the emitter lifetime), we expect all events to be uncorrelated, thus exhibiting a constant amplitude. In the case of a continuous electron beam, this means that the $g^{(2)}(\tau)$ curve is constant for $\tau \gg 0$, while in the pulsed case, we

still observe peaks at the delays corresponding to the time between pulses, all of them with the same amplitude. Nevertheless, this is not typically what we observe in experiments. Figure S4 shows the raw data of two $g^{(2)}(\tau)$ measurements, in continuous (a) and pulsed (b) mode. In both cases we observe that the number of counts decreases with increasing τ, contrary to what we would expect from the theory. This is due to an experimental artifact in the Hanbury-Brown and Twiss experiment. In the experiment, the emitted light is split into two beams with a 50:50 beam splitter. Each beam is directed towards one detector, connected to the time correlator. When one of the detectors receives a photon, the time correlator starts counting until a photon is received on the second detector. Therefore, having a count at a certain delay τ means that the second detector does not receive any photon during the time τ. This becomes very unlikely with increasing τ, thus producing the effect observed in the figure.

One way to avoid this artifact is by having a very low count rate on each detector, such that the probability of having two (uncorrelated) photons emitted within a time smaller than τ becomes very low. Nevertheless, this can result in very long acquisition times (in the order of hours) or low signal-to-noise ratios. In our case, we decided to keep the number of counts relatively high (typically 10^4 counts/s) and correct for this artifact during the data analysis. We observe that the evolution of the signal over τ due to this artifact follows an exponential decay, with average decay τ_{long}. The fits obtained when applying this decay are shown in Figure S4, for which we obtained $\tau_{long} = 31$ and 769 µs, respectively. This procedure is valid as long as τ_{long} is much larger than the bunching decay and pulse width, in the case of a pulsed electron beam. Otherwise, artifacts due to this effect would also affect the value of $g^{(2)}(0)$.

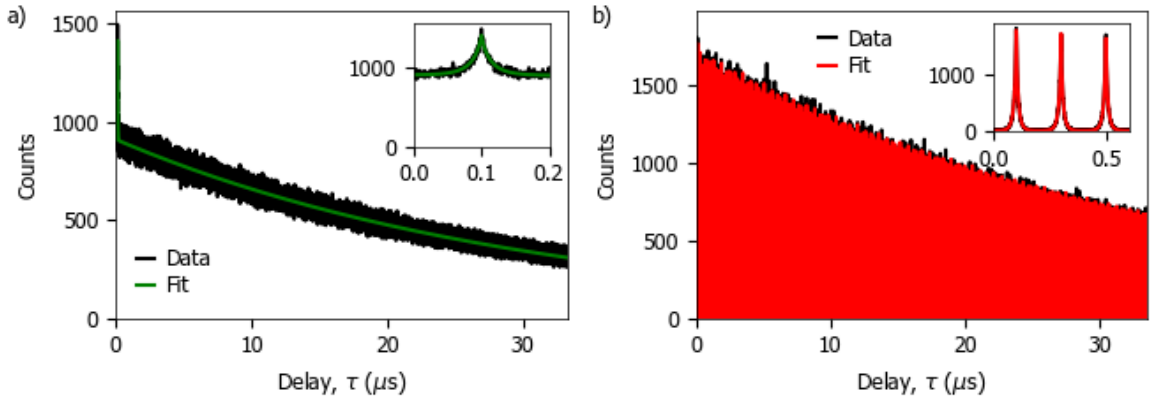

Figure S4: Raw data from a $g^{(2)}(\tau)$ experiment with a (a) continuous and (b) pulsed (by photoemission) electron beam. The data show that the number of counts decrease for long delays due to an experimental artifact in the HBT experiment. The green and red curves show the fits using an exponential decay to account for this artifact. The obtained average decays are $\tau_{long} = 31$ and 769 μs for (a) and (b), respectively. The insets show a zoom in for small delays, in which this artifact is not visible.

S6. Experimental details

All measurements are performed while focusing the electron beam on a single spot on the sample. The electron current is measured by collecting the beam current through a Faraday cup and reading it with a picoammeter.

S6a. Beam blanker

The experiments using a beam blanker are performed using the same microscope as in ref. [7]. In our case, a 400 µm aperture is placed right below the pole piece. The distance between the blanking plates is kept to 2 mm for all experiments. In contrast to previous work, here we apply a square signal on one of the blanking plates, with peak-to-peak amplitude of 5V and offset 2.5V. The other plate is grounded. This results in a square

electron pulse, with pulse width determined by the duty cycle D and repetition rate F, i.e., $\Delta p = (1 - D)/F$. In order to confirm the shape of the electron pulse, we performed decay trace measurements on the QWs while blanking the beam. Figure S5 shows two examples of traces, both obtained using $D = 0.6$ and repetition rate $F = 0.5$ and 6 MHz, respectively. We fitted the data using the following equation

$$f(x) = \begin{cases} B, & t < t_0 \\ A\left(1 - e^{-\left(\frac{(t-t_0)}{\tau}\right)^{\beta}}\right) + B, & t_0 \le t \le t_1 \\ Ae^{-\left(\frac{(t-t_1)}{\tau}\right)^{\beta}} + B, & t > t_1 \end{cases}, \tag{S42}$$

where $\tau = 8.6$ ns and $\beta = 0.63$ are the parameters describing the QW radiative decay, A is the amplitude of the signal and B is the background signal. The pulse width can be obtained from $\Delta p = t_1 - t_0$. In the experiments we obtain pulse widths of 796 and 62 ns, for Figure S5(a) and (b) respectively, which are very close to the theoretical values at these conditions (800 and 66 ns, respectively). These experiments were performed using an electron energy of 10 keV, but we do not expect significant deviations when changing the electron energy to 8 keV.

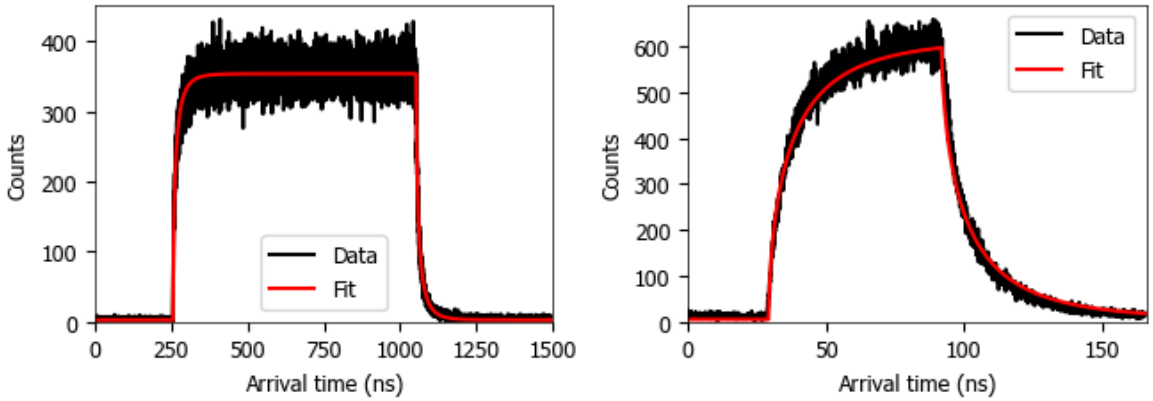

Figure S5: Decay traces on the QWs obtained using an electron beam blanker, with repetition rate (a) 0.5 and (b) 6 MHz. The fits are obtained using Eq. (S42), showing that the electron pulses can be described as a square pulse with pulse width of 796 and 62 ns, for (a) and (b) respectively.

Even though the experimental data shows an almost perfect square electron pulses, small deviations from this can arise when changing parameters, especially when increasing the duty cycle and repetition rate. In order to account for this, we measured the electron current in continuous mode I_c (i.e., in blanking conditions but without any signal driving the blanking plates) and in pulsed I_p (square signal driving one of the plates). The relation between both magnitudes is given by $I_p = I_c(1 - D)$. Figure S6a shows the value of electron current in pulsed I_p measured at different repetition rates. These measurements were performed at 8 keV and $D = 0.95$, with the same blanking conditions as for the $g^{(2)}(\tau)$ measurements using the blanker in the main text. The figure also shows the expected value of I_p (red curve), given a continuous current of $I_c = 213.9$ pA. We observe that the measured values are slightly lower than the expected ones, and the discrepancy increases with increasing repetition rate. These measured values of I_p were used to calculate the number of electrons per pulse in Fig. 3 of the main text. The pulse duration of the electron beam can also be extracted from these measurements, given that $\Delta p = {I_p}/{I_c F}$. Figure S6b shows the value of pulse width obtained using the experimental values of I_p (black dots) compared to the theoretical values, given by $\Delta p = (1 - D)F$ (red curve).

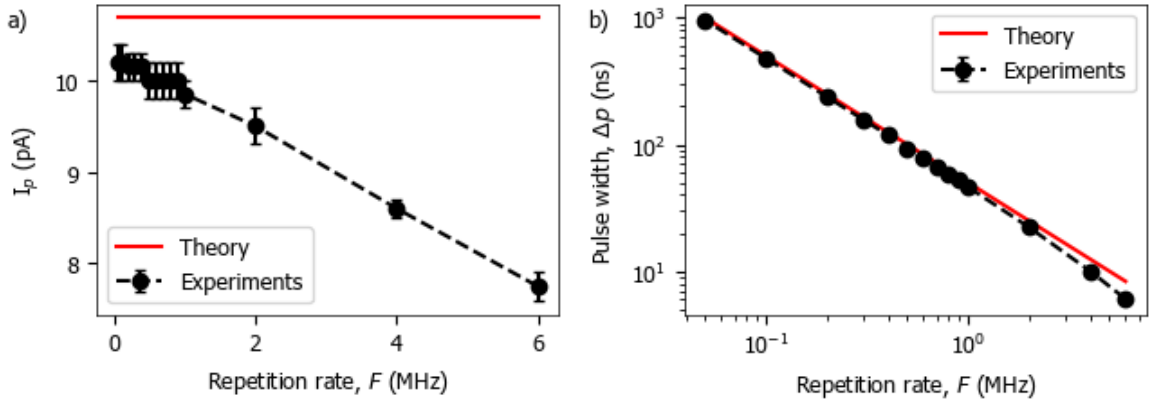

Figure S6: (a) Electron beam current measured in pulsed conditions using a beam blanker as a function of repetition rate. The red curve represents the theoretical current that we should obtain given a continuous current of 213.9 pA, and duty cycle of 0.95. (b) Pulse duration extracted from the experimental values of electron current in pulsed in (a), together with the theoretical value of the pulse width (red curve).

S6b. Laser-driven electron source (photoemission)

Ultrashort (ps) pulses are obtained by focusing the 4th harmonic (257 nm) of an Yb-doped femtosecond laser ($\lambda = 1035$ nm, 250 fs pulses) onto the electron cathode. The experiments are performed using a Quanta 250 FEG SEM. In order to suppress continuous emission, the filament current is reduced from 2.35 down to 1.7 A. The extractor voltage is also lowered from the typical 4550V value down to 650V. These settings allow us to achieve a high number of electrons per pulse, at the expense of lower spatial resolution, as explained in ref [7].

S7. Cathodoluminescence with 8 keV electrons

Figure S7 shows the CL spectrum obtained when exciting the sample with a continuous 8 keV electron beam, corresponding to the energy used in the experiments using the beam blanker. Most of the emission comes from the QW emission (410-490 nm). The inset shows a schematic of the structure of the sample together with Monte Carlo simulations of the trajectory of an 8 keV electron inside the sample, performed with the Casino software[8]. Each dot in the plot corresponds to an inelastic collision of the primary electron beam with the sample, while the color indicates the energy of the primary electron beam. We observe that barely any electron reaches the QWs, thus explaining the low excitation efficiency obtained at 8 keV ($\gamma = 0.05$).

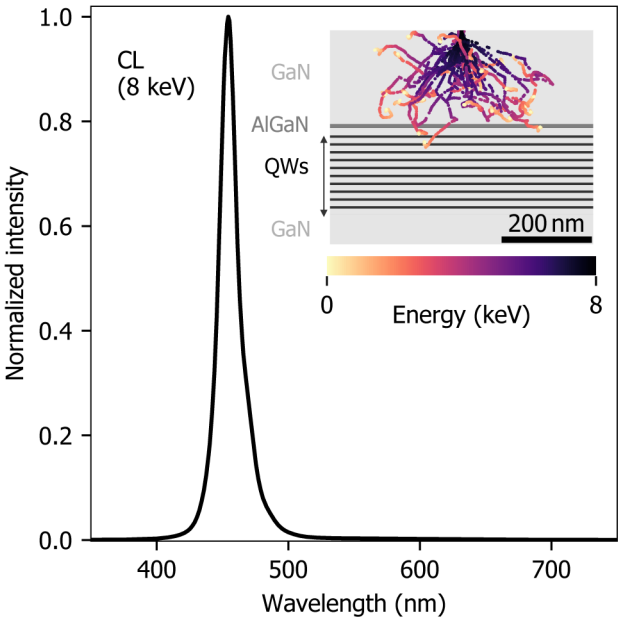

Figure S7: CL spectrum obtained after excitation with an 8 keV continuous electron beam (213.5 pA). Inset: schematic of the InGaN/GaN quantum well stack overlaid with the simulations of the trajectory of an 8 keV electron inside the sample.

S8. Dependence of QW emission decay on area

In the main text we show $g^{(2)}(\tau)$ measurements performed using the different electron beam configurations (continuous, pulsed with blanker and pulsed with photoemission), in each case exhibiting different decay lifetimes (τ_{emitter} and β_{emitter}). Here we prove that the main reason for this discrepancy is the inhomogeneity in the sample. Figure S8 shows $g^{(2)}(\tau)$ measurements performed using a continuous electron beam on different spots on the sample. The curves were obtained at 10 keV with beam currents of 10.6, 14.1 and 34 pA (green, blue and yellow curves, respectively). Each experimental curve (data points) is accompanied by the corresponding fit (solid lines), obtained by solving numerically Eq. (S36) when $y(t)$ is a stretched exponential. We observe that τ_{emitter} strongly depends on the position of the sample, ranging from 3.7 to 7.3 in these three examples. Instead, β_{emitter} remains in the 0.61-0.64 range.

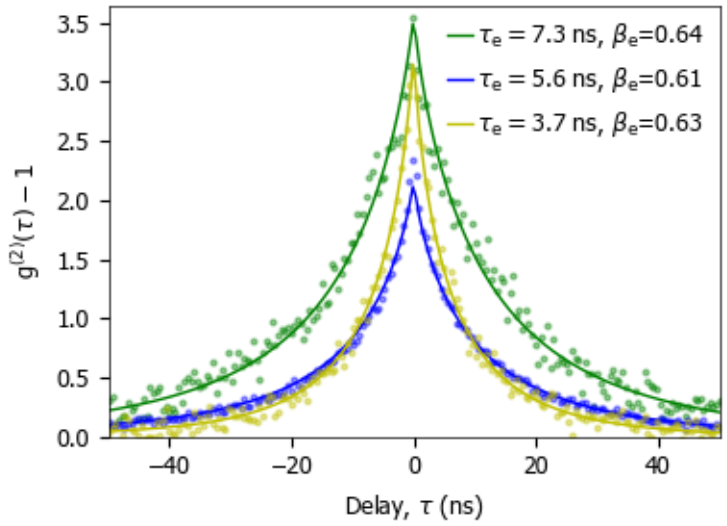

Figure S8: $g^{(2)}(\tau)$ measurements obtained with a 10 keV continuous electron beam at three different spots on the sample. The solid lines are the fits from Eq. (S36) when $y(t)$ is a stretched exponential, with fit parameters $\tau_e \equiv \tau_{\text{emitter}}$ and $\beta_e \equiv \beta_{\text{emitter}}$.

References

(1) Meuret, S.; Tizei, L. H. G.; Cazimajou, T.; Bourrellier, R.; Chang, H. C.; Treussart, F.; Kociak, M. Photon Bunching in Cathodoluminescence. *Phys. Rev. Lett.* **2015**, *114* (19), 1–5.

(2) Meuret, S.; Coenen, T.; Zeijlemaker, H.; Latzel, M.; Christiansen, S.; Conesa-Boj, S.; Polman, A. Photon Bunching Reveals Single-Electron Cathodoluminescence Excitation Efficiency in InGaN Quantum Wells. *Phys. Rev. B* **2017**, *96* (3), 1–8.

(3) Meuret, S.; Coenen, T.; Woo, S. Y.; Ra, Y. H.; Mi, Z.; Polman, A. Nanoscale Relative Emission Efficiency Mapping Using Cathodoluminescence g(2) Imaging. *Nano Lett.* **2018**, *18* (4), 2288–2293.

(4) Fox, M. *Quantum Optics: An Introduction (Oxford Master Series in Physics)*; OUP Oxford, 2006.

(5) Bóna, M. Who Knows What It Looks Like, But It Exists. The Probabilistic Method. In *A Walk Through Combinatorics*; WORLD SCIENTIFIC, 2016; pp 381–416.

(6) Solà-Garcia, M.; Meuret, S.; Coenen, T.; Polman, A. Electron-Induced State Conversion in Diamond NV Centers Measured with Pump–Probe Cathodoluminescence Spectroscopy. *ACS Photonics* **2020**, *7* (1), 232–240.

(7) Meuret, S.; Solà Garcia, M.; Coenen, T.; Kieft, E.; Zeijlemaker, H.; Lätzel, M.; Christiansen, S.; Woo, S. Y.; Ra, Y. H.; Mi, Z.; Polman, A. Complementary Cathodoluminescence Lifetime Imaging Configurations in a Scanning Electron Microscope. *Ultramicroscopy* **2019**, *197*, 28–38.

(8) Drouin, D.; Couture, A. R.; Joly, D.; Tastet, X.; Aimez, V.; Gauvin, R. CASINO V2.42 - A Fast and Easy-to-Use Modeling Tool for Scanning Electron Microscopy and Microanalysis Users. *Scanning* **2007**, *29* (3), 92–101.