Cathodoluminescence spectroscopy (CL) is a well-known technique for the characterization of semiconductor materials [1–6]. In CL, a high-energy electron beam in a scanning electron microscope (SEM) excites a material and generates luminescence that is collected and analyzed. CL emission gives valuable spatially resolved information about the band gap [7,8], carrier generation [9], defects [10,11], diffusion and carrier transport [12–14], recombination [15,16], and other optoelectronic properties of semiconductors that are used, e.g., in light-emitting diodes (LEDs) [17,18], lasers [19], solar cells [20], and more.

Understanding the interaction of a single electron with a semiconductor material and quantifying the resulting light emission process is essential in order to fully exploit the potential of CL. In semiconductors, the primary excitations of high-energy electrons are mostly bulk plasmons [21] that decay though an incoherent cascade creating electron-hole pairs that can then decay by radiative emission. In this way, depending on the sample, hundreds of photons can be created by a single electron [22,23]. The measured CL signal is then an average over multiple single-electron/semiconductor excitation events [24]. One key question in CL SEM experiments is whether each primary electron creates a similar number of photons, or whether the amount of light emission varies strongly from electron to electron. Answering this question requires understanding of the interaction and emission process at the single-electron and single-photon level.

Here, a time-resolved CL technique is introduced that answers this question by investigating the bunching in the photon statistics of CL [25] as a function of electron-beam current using a pulsed CL geometry. Photon bunching is the result of a strong correlation in the emission of photons due to the fact that multiple excitations are generated by a single electron. From the photon correlation measurements, the fraction of electrons that create photons is identified in a quantitative way. As a model system, we investigate CL emission from InGaN quantum wells (QWs) embedded in GaN [26,27]. By combining bunching measurements in pulsed and continuous mode with Monte Carlo simulations we derive a quantitative model from which the single-electron interaction probability is deduced. In the 6–30-keV energy range studied here, the interaction probability with the QWs varies from 1% to 100%. From the data, we identify the effect of a barrier layer in the semiconductor sample and find the effective carrier diffusion length in the bottom GaN substrate to be ~200 nm.

InGaN multiple QW structures embedded in GaN were grown by metalorganic chemical vapor deposition (MOCVD). Figures 1(a) and 1(b) show a high-angle-annular dark-field scanning transmission electron microscopy (HAADF STEM) image taken on a cross section of the sample prepared using focused-ion-beam (FIB) milling. The image shows 10 In_{0.18}Ga_{0.82}N QWs with a thickness of 2 nm spaced by 15-nm GaN layers buried under a 250-nm-thick p-doped GaN layer.
FIG. 1. (a) HAADF STEM image of a cross section of the InGaN/GaN QW sample. The white box in (a) is magnified in (b). (c) Casino simulation of the electron trajectories in the sample for a primary energy of 10 keV. (d) CL spectrum collected at 10 keV ($I = 10.5 \text{ pA}$). The spectral range within the dashed lines is used in the $g^{(2)}$ measurements. (e) CL spectra for electron energies in the range 6–30 keV. (f) Average number of photons produced per incident electron as a function of electron energy, integrated over the spectral band indicated in (d).

GaN capping layer. The chemical composition of the sample was analyzed using energy-dispersive x-ray spectroscopy (EDX) [28]. In the STEM image a dark line is observed in the capping layer 20 nm above the QWs [see Fig. 1(b)] which corresponds to a 2-nm-thick AlGaN barrier layer [29].

Figure 1(c) shows the overlap between the electron cascade and the QW layers at a primary electron energy of 10 keV calculated using Casino simulations [30]. Figure 1(d) shows the corresponding CL spectrum at an electron current $I = 10.5 \text{ pA}$.

The data are corrected for the system response by comparing measurements and calculations of the transition radiation spectrum of a single-crystalline Al reference sample [28,31]. All CL experiments presented here were conducted at room temperature. The spectrum clearly shows the InGaN peak emission at 2.77 eV. We changed the overlap with the QWs by varying the electron energy between 6 and 30 keV which effectively changes the depth and size of the interaction volume [6,16,24]. The CL spectra are shown in Fig. 1(e). For low energy (<10 keV) only the InGaN QW emission is observed. For increasing energy the GaN band-edge emission appears in the near UV (3.37 eV) and a broad peak appears at 2.14 eV corresponding to the well-known “yellow band” defect emission in the $n$-type GaN substrate underneath the QWs [32]. These contributions become more pronounced due to the increased overlap between the electron cascade and the GaN substrate.

Figure 1(f) shows the QW emission intensity per incident electron (integrated over the 2.5–3.1-eV range in the spectrum) as a function of electron energy. When light is generated inside a high-index medium a large fraction remains trapped due to total internal reflection at the GaN/vacuum interface. To estimate the average number of photons produced per electron the data are corrected for the outcoupling efficiency ($\sim 3.5\%$) assuming orientation-averaged outcoupling, for a point dipole emitter beneath the GaN surface [28]. Figure 1(f) shows that the average number of photons generated per electron strongly increases from 0.4 to 180 as the energy is increased from 6 to 15 keV, which we attribute to the increased overlap of the electron cascade with the QWs and the increase in available energy. When the energy is further increased, the average photon generation within the filter range decreases, as the electron cascade penetrates beyond the QW layer stack. These data show that the electron trajectories play an important role in determining the CL response. A key question in order to fully understand the CL generation mechanism is whether the average emission probability derived in Fig. 1(f) is representative for each incoming electron or whether there is a subset of the electrons that each generates a larger number of photons.

To resolve this we measured the temporal distribution of the emitted photons. Figure 2(a) shows a schematic of the measurement technique. CL emitted from the sample is guided to a Hanbury-Brown-Twiss interferometer composed of a
beam splitter and two photon counting detectors [33]. This geometry allows us to measure the autocorrelation function \( g^{(2)}(\tau) \) that represents the probability to observe two photons separated in time by a delay \( \tau \). In quantum optics this technique is well known for the characterization of single-photon emitters which display characteristic antibunching behavior in optical spectroscopy [34] and CL-TEM experiments [35].

Figure 2(a) shows \( g^{(2)} \) measurements taken at low beam current \((I = 2.5 \text{ pA at 10 keV})\) using a short pass filter \((> 2.5 \text{ eV})\). A clear bunching behavior is observed \([g^{(2)}(0) \approx 11]\) reflecting a high probability to detect two photons within a short time delay. The shape of the peak can be fitted by a double-exponential decay with a main component corresponding to the luminescence decay of the QWs at 2.77 eV \((\tau_{\text{QW}} = 12 \text{ ns})\) and a small slower decay component that is assigned to the emission from the tail of the yellow band \((\tau_{\text{YN}} \sim 30 \text{ ns})\) [28,36]. Even if the long decay seems to agree with the lifetime found for the yellow band emission, this long decay could also be attributed to the complex dynamics of carrier diffusion and recombination in the InGaN/GaN system [37,38]. However, this component only has a minor contribution to the \( g^{(2)} \) function and can be disregarded for the \( g^{(2)} \) analysis [28]. The bunching effect is the consequence of multiple excitations by a single electron that lead to emission of a bunch of photons within the lifetime of the emitter. This was first demonstrated for CL excitation in a TEM [25].

Figure 2(b) shows \( g^{(2)} \) measurements for different currents in the range 0.13 e-/pulse (200 ns), 6 electrons/pulse (0.13 ns), 100 electrons/pulse (6.0 ns), and 1000 electrons/pulse (100 ns), corresponding to the longest pulse \((p = 200 \text{ ns})\) for high current \((I = 10.5 \text{ pA})\). For the shortest pulse \((p = 2 \text{ ns}, 0.13 \text{ electrons/pulse})\) \( g^{(2)}(0) \) increases to 35. Contrary to the continuous beam \( g^{(2)} \) measurement, the pulsed \( g^{(2)} \) function

![Figure 2](image_url)

**FIG. 2.** (a) Schematic of the CL \( g^{(2)} \) measurement set-up. Inset: \( g^{(2)} \) measurement at 10 keV (2.5 pA) showing clear bunching at 0 ns delay. (b) \( g^{(2)} \) measurements for different currents in the range 2.5–349 pA. The bunching effect decreases for increasing current. Inset: Amplitude \([g^{(2)}(0) - 1]\) of the bunching peak as a function of beam current. BS: beam splitter; APD: avalanche photodetector.
goes to 0 instead of 1 outside of the pulse window except during the time window corresponding to a subsequent pulse. In particular, the second peak observed at 1 μs, reflects the time between two photons created by electrons within two consecutive pulses. It measures the probability that two consecutive pulses contain electrons that generate CL, which is determined by the pulse width and the Poissonian distribution of the continuous electron beam. As this probability represents uncorrelated events, \( g^{(2)}(1 \mu s) \) serves as normalization of the \( g^{(2)} \) data in Fig. 3 [28]. Note that if the average number of electrons per pulse is less than 1, the width and shape of the peak centered around \( g^{(2)}(0) \) will only depend on the lifetime, as it is due to photons generated by the same electron, whereas the width and shape of the second peak are a convolution between the decay and the electron pulse shape.

Next, a quantitative model is derived to explain the data of Figs. 2 and 3. A Monte Carlo based model taking into account the electron interaction mechanisms and the Poissonian distribution of the beam is used to calculate the \( g^{(2)}(\tau) \) function for a given electron-beam condition [28]. As a starting point, we use a model previously derived for CL excitation on ultrathin TEM membranes [25]. These TEM experiments were performed in the weak interaction regime (high electron energy and thin samples) where the excitation efficiency is a simple Poissonian law given by the mean-free path and the sample thickness. Here, each electron will be fully stopped in the material and how this energy is distributed determines the response.

The closer the electron-hole pairs are generated to the QWs the higher the probability that they will excite them. The three main parameters that determine the amplitude of \( g^{(2)} \) in this case are the electron current \( I \), the lifetime \( \tau_{\text{QW}} \), and the probability \( \gamma \) for an incoming primary electron to excite the QWs. The current is measured using a Faraday cup and the lifetime is derived from the \( g^{(2)} \) function as described above, so that the only unknown parameter is the probability of excitation \( \gamma \). Because this model is based on correlation measurements, the result is neither sensitive to the collection efficiency nor the radiative quantum efficiency of the QWs, but only to the probability that excitations due to the primary electron reach the QWs. It therefore enables the decoupling of excitation efficiency and emission efficiency. This is not the case for other intensity-based detection experiments such as the one in Fig. 1(f).

Figure 4(a) shows a simulated \( g^{(2)} \) curve (blue) for pulsed excitation (\( p = 500 \text{ ps}, I = 10.5 \text{ pA}, \tau_{\text{QW}} = 12 \text{ ns}, \gamma = 0.5 \)), overlaid with the corresponding measurement (red curve). For these parameters, the experimental and simulated curves are very similar. The difference between the curves is ascribed to the fact that the long decay tail assigned to the yellow band, which is observed in the measurement, is not taken into account in the simulation. To confirm this, the dashed gray curve in Fig. 4(a) shows the short-lifetime component of the fit through the experimental data: very good agreement is observed with the simulated \( g^{(2)} \) curve.

Figure 4(b) shows the measured and simulated \( g^{(2)} \) amplitude \( [g^{(2)}(0) - 1] \) as a function of beam current for both continuous and pulsed configurations at 10 keV. All simulations were performed using \( \gamma = 0.5 \) which fits the experimental results best. The decreasing trend with current that is observed experimentally is very well reproduced for both continuous and pulsed measurements. The large difference in the \( g^{(2)} \) amplitude between pulsed and continuous excitation is also very well reproduced by the simulations using the same value of \( \gamma \), and a value of \( p = 0.5 \text{ ns} \) for the pulse width [28]. This analysis shows that 50% of the primary electrons cause excitation of the QWs without any \textit{a priori} knowledge of the geometry of the sample.

To answer the question of whether this emission probability is representative for each incoming electron or whether a subset of the electrons generates a larger number of photons, the electron energy was varied. As shown in Fig. 1(f), we can control the emission probability from the QWs by varying the electron energy. Figure 5(a) represents simulated electron trajectories inside GaN for 6, 8, and 10 keV. The position of the QWs and the thin AlGaN barrier are indicated in the figure. Only a selection of trajectories is shown here. Figure S11 in [28] shows a more complete overview of the number of interactions as a function of depth. There it is visible that a small number of interactions actually does occur below the AlGaN layer for 6 keV [not visible in Fig. 5(a)] which explains QW signal at that voltage. From these data it is clear that the QW excitation probability \( \gamma \) increases for increasing energy, decreasing the average time between two electrons that can generate CL emission.

Figure 5(b) shows \( g^{(2)} \) measurements for different electron-beam energies in the range 6–30 keV. The bunching effect strongly decreases as the energy increases from 6 to 15 keV. We attribute this to the fact that for increasing energy the time between two electrons responsible for emission decreases, similar to the trend observed in Fig. 2(b) for increasing current. For energies above 15 keV, no further decrease in \( g^{(2)} \) is observed, which we attribute to the fact that the QW interaction probability is close to unity.

Figure 5(c) shows the simulated \( g^{(2)} \) amplitude \( [g^{(2)}(0) - 1] \) (\( I = 10.5 \text{ pA}, \tau_{\text{QW}} = 12 \text{ ns} \)) as a function of \( \gamma \). The amplitude decreases as the fraction of electrons exciting the QWs increases. In the limit of \( \gamma = 1 \), \( g^{(2)} \) converges to the minimum value related to the average time separation between electrons which is determined the current \( (I = 10.5 \text{ pA}) \). At \( I = 10.5 \text{ pA} \) the average time between electrons is about 15 ns,
FIG. 5. (a) Electron trajectory simulations (casino) for 6, 8, and 10 keV. The color indicates the electron energy relative to the primary electron energy $E_{\text{max}}$. The InGaN QWs (white lines) and the AlGaN barrier (black line) are indicated. (b) $g^{(2)}$ measurements (continuous beam) for electron energies ranging from 6 to 30 keV ($I = 10.5$ pA). The curves for 15, 20, 25, and 30 keV are very similar and lie on top of each other. (c) Amplitude of the modeled $g^{(2)}$ curve as a function of $\gamma$ ($I = 10.5$ pA, $\tau = 12$ ns). The amplitudes derived from the $g^{(2)}$ measurement for each energy in (b) are overlaid (red dots), allowing us to retrieve $\gamma$ for each electron energy.

Close to the lifetime ($\tau_{\text{QW}} = 12$ ns) explaining why $g^{(2)}(0) \sim 1$ if all the electrons interact. The red dots in Fig. 5(c) represent the amplitude of the $g^{(2)}$ peaks from Fig. 5(b): for each energy the corresponding value of $\gamma$ can now be derived; it is plotted in Fig. 6(a). For the lowest energy (6 keV) $\gamma = 0.01$ which means that 1% of the primary electrons interact with the QWs. At 10 keV we find $\gamma = 0.5$, consistent with the data in Fig. 4. For energies above 10 keV $\gamma$ converges to 1, in correspondence with the convergence in the $g^{(2)}$ measurements of Fig. 5(b) for energies above 15 keV.

By combining the analysis shown above with the data in Fig. 1(f), which shows the number of photons generated per incident electron, we can now derive the number of photons per electron interacting with the QWs. For example, at 6 keV only 1% of the electrons interact ($\gamma = 0.01$) but an average of 0.4 photons/electron are emitted [see Fig. 1(f)], therefore each primary electron that interacts produces 40 photons on average. Similarly, at 10 keV (where $\gamma = 0.5$), each primary electron that interacts produces 120 photons on average. In order to explain the large variation in photon generation rate for each incident electron a model combining electron trajectory simulations (Casino [30]) and a carrier diffusion model is developed [28]. In the model, it is assumed that carriers between the surface and the AlGaN barrier cannot reach the QWs, while carriers produced below the barrier within a certain diffusion length from the QWs ($D_L$) can excite them.

The simulated data are shown in Fig. 6(a) and match very well with the experiment. At low energy most of the interactions occur above the barrier layer and only a very small fraction is exciting the QWs ($\gamma < 0.1$). As the energy is increased more interactions occur below the barrier, resulting in an increase of the excitation rate of the QWs. At the highest energy almost all electrons are interacting; $\gamma$ converges to 1. Above 10 keV the $g^{(2)}$ amplitude is too small to derive data for $\gamma$. The sensitivity could be improved by performing experiments at lower beam current. The strong variation of $\gamma$
with voltage observed here results from the stratified nature of the material. It would disappear for a homogeneous medium where the current and lifetime would be the only relevant parameters determining $\gamma$.

Finally, using the diffusion model we can determine the average number of photons generated per primary electron as a function of electron energy. Figure 6(b) shows this data for three diffusion lengths (200, 300, and 400 nm). Above 15 keV, for the shortest diffusion length the curve decreases faster with increasing energy due to the fact that the carriers generated at larger depths cannot reach the QWs. The same trend is observed in the data from Fig. 1(f), also shown in Fig. 6(b) for reference. For these data an outcoupling efficiency of 3.5% is assumed. Comparing the experimental data and the model, the diffusion length is estimated to be around 200 nm. This diffusion length is in the same order of magnitude as values reported for $n$-doped GaN layers [40]. Differences with reported values can be caused by a large number of parameters including defect structure, built-in electric fields, and the presence of charge-repelling layers such as the AlGaN, making it difficult to model. In the Casino based approach, both the material geometry and the diffusion characteristics have to be known quite accurately to predict the electron interaction with the material. Although the method works rather well in this case and it verifies our approach it is more complex and less reliable than the $g^{(2)}$ analysis. In conclusion, the cathodoluminescence emission statistics of InGaN quantum wells embedded in GaN under 6–30 keV electron excitation were measured. Strong photon bunching is observed in the InGaN quantum well emission at 2.77 eV, due to the fact that a single primary electron can excite multiple photons. The bunching effect decreases with increasing beam current due to the increased temporal overlap of multiple electron excitations. For a continuous beam of $I = 10.5 \text{ pA}$, pulsing the electron beam ($p = 2\text{–}100 \text{ ns}$; 0.13–6 electrons per pulse) strongly enhances the bunching effect, in agreement with the Poissonian electron-beam statistics. A model based on Monte Carlo simulations is developed in which a fraction $\gamma$ of the primary electrons generates multiple photons per electron; it explains the $g^{(2)}$ measurements very well. The exciting electron fraction $\gamma$ strongly increases with electron energy due to the enhanced overlap of the electron cascades with the quantum wells. A decreasing trend for $\gamma$ is observed at higher energy and is ascribed to the finite diffusion length for carriers generated in the GaN substrate; the diffusion length is estimated to be around 200 nm. The combination of $g^{(2)}$ measurements in pulsed and continuous mode with CL spectral analysis provides a powerful tool to study optoelectronic properties of semiconductors and can find applications in many other optically active systems and devices.

We gratefully acknowledge the technical assistance of R. Buijs in the construction of the new CL microscope developed for this work. N. Schilder is acknowledged for the light outcoupling calculations. We gratefully acknowledge E. Kieft (Thermo Fisher-FEI), G. Weppelman, and J. Hoogenboom (TU Delft) for useful discussions and advice regarding beam blanking. We gratefully acknowledge S. Kolling for preparing the cross-section sample for TEM analysis. This work was part of the research program of the Netherlands organization of scientific research (NWO). It is also funded by the European Research Council (ERC). M.L. and S.C. gratefully acknowledge the German Research Foundation (DFG) for financial support through the cluster of excellence “Engineering of Advanced Materials” at the Friedrich-Alexander-Universität Erlangen-Nürnberg. Competing financial interests: T.C. is employee and A.P. is co-founder and co-owner of Delmic BV, a company that has brought a cathodoluminescence system on the market based on a design developed at AMOLF and that was used in this work.

PHOTON BUNCHING REVEALS SINGLE-ELECTRON . . PHYSICAL REVIEW B 96, 035308 (2017)


[41] F. J. García de Abajo, Optical excitations in electron microscopy, Rev. Mod. Phys. 82, 209 (2010).


