RC22934 (W0310-073) October 13, 2003
Electrical Engineering

IBM Research Report

Characterization of Breakdown Sites in Ultrathin Oxides by Optical Emission Imaging

J. C. Tsang, B. P. Linder
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

Research Division
Almaden - Austin - Beijing - Haifa - India - T. J. Watson - Tokyo - Zurich

CHARACTERIZATION OF BREAKDOWN SITES IN ULTRATHIN OXIDES BY OPTICAL EMISSION IMAGING

J. C. Tsang, and B. P. Linder
IBM T. J. Watson Research Center
PO Box 218
Yorktown Heights, NY 10598

ABSTRACT

Hot carrier luminescence excited by direct tunneling at low voltages through fresh, ultra-thin gate oxides in silicon FETs, and by localized transport processes during progressive breakdown of these devices, has been studied. The optical emission provides information on the number of breakdown spots, and the nature of the charge transport through a breakdown spot in nFETs. The transition from stress induced leakage currents involving isolated atomic defects to reversible breakdown events to final breakdown was accompanied by changes in the efficiency of the light emission. These changes show the evolution from highly inelastic transport processes to elastic transport through breakdown spots.

INTRODUCTION

The breakdown of ultra-thin insulators at low voltages is different from the abrupt, catastrophic breakdown of a thick insulator at high voltages.[1-3] The low voltage breakdown occurs gradually, involving modest currents. The gate leakage, I_G, generates defects in the oxide. At higher defect densities, a cluster of these defects provides a current path that does not locally destroy the oxide.[1,2] In this paper, we describe new experimental data about the leakage induced oxide states and how they contribute to the low voltage breakdown. Understanding this process helps answer the question of the reliability of devices using these oxides.

Silicon hot carrier luminescence (HCL) excited in the gate or channel of an FET was used to study the defects responsible for the initial breakdown of 1.5 nm thick gate oxides. Previous studies showed that breakdown in thicker oxides produces observable HCL.[4] We measured the HCL as the breakdown progresses,[1] imaging the HCL to spatially resolve I_G across the device. The growth of the breakdown was controlled by restricting the size of current jumps during constant voltage stress.[5] Measurable properties of the HCL are affected by the initial growth of Stress Induced Leakage Current (SILC), the onset of noisy discrete reversible discontinuities in I_G, the appearance of soft-breakdown currents with digital noise, and breakdown accompanied by analog noise. Our HCL images support the hypothesis that SILC is associated with inelastic tunneling

(IET) processes.[6] The HCL images show that reversible I_G jumps are directly related to the irreversible breakdown for n- and pFETs. The intensity, L_L, of the localized HCL after soft breakdown can be consistent with its excitation by elastic transport (ET). In contrast, I_L is relatively weak at the initial breakdown, and as the breakdown progresses and I_G becomes large. These results connect the IET of the SILC defects, with the ET characteristic of soft breakdown,[7] and suggest how the properties of the defect cluster responsible for the breakdown evolve from those of the SILC defect.

We first summarize the process responsible for HCL. The samples, experimental measurements, and the steps which occur in the breakdown of our samples are described. Our principle results are illustrated and we show how they complement existing knowledge about the initial stages of the breakdown in ultra-thin gate oxides at low voltages.

EXPERIMENTAL DETAILS

The gate oxides in state-of-the-art FETs are less than 2nm thick and I_G in fresh devices for the direct tunneling regieme is due to ET. Carriers elastically tunneling through ultra-thin gate oxides have a maximum excess energy of eV_G. V_G is the gate voltage. For nFETs, because of the heavily doped gate, the hot carriers in the gate rapidly thermalize with energy distributions characterized by effective temperatures of several thousand degrees. Such "hot" carriers produce weak but detectable light. This process is identical to what occurs at the drains of saturated FETs and has been used for timing analyses in CMOS circuits.[8,9] The HCL can be observed through thin gates, and from the backside of an FET. To first order, the intensity of the emission from the gate, L_G, excited by uniform elastic tunneling through the gate oxide, depends linearly on I_G, and exponentially on $1/V_G$. The energy possessed by hot carriers as they enter the gate and relax depends on E_G. Carriers with less excess energy produce exponentially less light than "hotter" carriers.[9] For fresh devices, the efficiency of the HCL, $E_G=L_G/I_G$, increases by 1000X as V_G goes from 1.5 to 2.8V.

Both n- and pFETs with 1.5 nm gate oxides were studied. The gate widths were $1<W<20\mu$m while the gate lengths were $0.125<D<10\mu$m. Images of the HCL were obtained by thermoelectrically and liquid nitrogen cooled CCD cameras. The test devices were viewed through a microscope with a 50X objective with the cameras attached to the camera port. The exposure times needed to image the HCL varied from 15sec to 30min, depending on V_G. All

measurements were made on the front side of the FET so that the imaged HCL was attenuated by the overlying gate. Constant voltages were applied to the gate to stress the gate oxides, and excite the HCL. I_G was measured independently. I_G transients due to breakdown events were limited by setting the compliance of the voltage source.

The V_G stress curves agreed with published results.[1] $I_G(t)$, where the source, drain and substrate were grounded, had an initial rapid rise, and a more gradual increase due to SILC. Reversible jumps in I_G, the pre-breakdown noise, were observed before the irreversible breakdown events. The final breakdown current showed analog noise, and the total increase in I_G was between 10-20μA at V_G=1.6V.

LIGHT EMISSION in FETs BEFORE and DURING GRADUAL BREAKDOWN

Figs. 1a-f show images of the HCL for V_G=1.6V from a 20x10(WxD)μm nFET. [I_G] (the average current during exposure), [I_G^{RMS}] (the root-mean-square deviation of I_G), L_G, and L_L, for each image are given in Table 1. At 1.6V, the nFET was stable during image acquisition since the rate at which breakdown occurs is ~10^6X slower than at 2.8V. Fig. 1a shows the emission from the gate due to I_G=16.21μA of direct tunneling current in the unstressed device. [I_G] agreed with the predicted current for this oxide at 1.6V.[10] The HCL came from the exposed portion of the gate of the FET. The square region in Fig. 1a which shows no emission was due to an optically opaque metal pad which covered part of the gate and facilitated the planarization of the device during processing. Elsewhere, where the gate can be observed, the HCL was uniform due to the uniform direct tunneling current. The source and drain of the FET were at the bottom and top of the image.

The device in Fig. 1a was stressed at 2.8V with a compliance of 5μA. After about 4100 sec, a 5μA jump occurred, the stress shut off, and the HCL at 1.6V measured. Fig. 1b shows the difference between this new measurement and Fig. 1a. There is a new, weak, emission spot in the upper right hand part of the gate (circled). Application of a second 2.8V stress with the same 5μA compliance produced a 1.6V HCL image whose difference with respect to Fig. 1a is shown in Fig. 1c. The weak breakdown spot in Fig. 1b disappeared. [I_G] and [I_G^{RMS}] were both reduced for Fig. 1c from Fig. 1b but were larger than for Fig. 1a. This device was then stressed again at 2.8V with a 5μA compliance. Fig. 1d shows extra emission with respect to Fig. 1a at the same spot as the localized emission in Fig. 1b. [I_G] for Fig. 1d was 440nA greater than [I_G] for Fig. 1c. [I_G^{RMS}] was also larger. After Fig. 1d was obtained, the device was held at V_G=1.6V for about 70000sec.

During this period, V_G was switched on and off several times and $17\mu A < [I_G] < 17.8\mu A$ and $8.8nA < [I_G^{RMS}] < 109nA$. L_L varied by an order of magnitude while $I_G(t)$ away from the HCL spot decreased slightly. Fig. 1e shows the difference between the HCL image then obtained after a brief additional 2.8V stress and Fig. 1a. Fig. 1e shows no evidence for the localized HCL seen in the previous images. $[I_G]$ for Fig. 1e was close to that expected before breakdown when only the SILC is added to $I_G(0)$. $[I_G^{RMS}]$ for Fig. 1e was equal to $[I_G^{RMS}]$ for the unstressed device. Fig. 1f shows the reappearance of the original localized emission spot after further operation when $[I_G]$ increased by $1.27\mu A$. Row G in Table 1 describes an HCL image from this nFET obtained after 10^6sec of further operation at 1.6V. I_G as high as $25\mu A$ was measured during this operation. The HCL image for Row G was similar to Fig. 1f . It showed however, a reduction in L_G from its fresh oxide, value. I_G in row G was equal to the original gate current plus the SILC at breakdown. Since the SILC increased due to the operation of the device for 10^6 sec, and there was localized emission due to a breakdown event, the $16.89\mu A$ of I_G in row G was consistent with a reduction of the direct tunneling I_G, which explains some of the decrease in L_G. This is discussed in the next section.

Figs. 1a-f show breakdown at a single spot in the nFET. There were two reversible jumps in I_G but only a single emission location as breakdown progressed. This agrees with the statistical analysis of such breakdown in these nFETs by Linder et al.[5,11] They defined two time scales describing the gradual breakdown including the time to breakdown, and the times required for the breakdown current to increase by given amounts. They derived how these quantities are related for different types of breakdown, and how they should scale with device area. The area of the nFET under test was consistent with a single breakdown spot. Also, we find that increases in SILC produced no increases in $L_G(t)$, when there were no localized emission spots, or away from the localized emission spots.

Fig. 2 shows HCL obtained from a pFET for a 1.5nm gate oxide in a $20x10(WxD)\mu m$ pFET under V_G=-2.9V stress. All other contacts were grounded. I_G was $138\mu A$ for the fresh device. A compliance limit of $5\mu A$ was applied to the stress and triggered several times during the stress. The images in Fig. 2 were all obtained at V_G=-2.9V, with 15 sec exposures. This exposure time was at least an order of magnitude smaller than the time between the turn on of the stress and the first breakdown discontinuity in I_G. Fig. 2a shows the HCL from this pFET when V_G=-2.9V is first applied. The emission was largely uniform over the device. The notch in the upper left hand corner of Fig. 2a was due to the overlap of an opaque metal pad with part of the gate. Fig. 2b was

obtained after a 2.5μA jump in $I_g(t)$ at about 270sec and shows one weak and one bright emission spot near the lower gate edge. Excess current with digital noise appeared and disappeared during this stress. About 2700sec later, Fig. 2c was obtained showing negligible localized emission at the spots previously identified in Fig. 2b. A very weak emission spot near the left, upper corner of the gate is seen. 1200sec later, this weak emission spot is clearly seen in Fig. 2d. About 3000sec later, the light emission image shows only weak localized emission at the three spots which earlier showed strong localized emission. Continued stressing of this device produced Fig. 2e which again shows the strong localized HCL in Figs. 2b and 2d. After over 12000 sec of V_G=-2.9V, a very strong breakdown spot in the location of the localized emission first seen 8000sec earlier is observed in Fig. 2f, as well as several other emission spots. The gate current increased to 193μA during these measurements.

The results in Fig.2 show there are multiple breakdown spots in this pFET. They agree with earlier statistical analysis of $I_G(t)$ near breakdown in our pFETs by Linder et al.[11] This allowed them to infer that there were multiple breakdowns spots. Interestingly, we observe reversible behavior, and the final breakdown occurred at the location of one of the earlier spots.

ANALYSIS

Our experimental results provide three insights into the breakdown of our ultra-thin gate oxide system at low voltages.

First, the HCL images clearly show that the reversible and early breakdown events foreshadowed the eventual irreversible breakdown. The HCL associated with reversible and irreversible changes in I_G occurred at a single spatial location for nFETs and a limited group of spatial locations for our pFETs. Breakdown studies of thicker (2.2nm) oxides at higher stresses (4V) have suggested, in contrast, that the final breakdown event can occur at a different location than the earlier stage breakdown events.[4]

Secondly, our data show that the defect induced SILC cannot efficiently excite HLC. We find no additional light emission in Fig. 1e after two reversible breakdown events and where I_{SILC} is estimated to be 660nA. (I_G(Fig. 1e)-I_G(0)), and [I_G^{RMS}] were both negligible here. The change in I_G due to SILC before breakdown can also be measured from $I_G(t)$ during the V_G=2.8V stress. For Fig. 1, after 4100 seconds of V_G=2.8V, before breakdown, $\Delta J/J_o$= 3%. This corresponds to a 500nA SILC at 1.6V. It could also be correlated with results from other studies of SILC. Our 3%

value was in reasonable agreement with the published value of 5% for our 200μm^2 device,[1] independently confirming that the SILC in Fig. 1b is ~660nA. The post breakdown emission, $L_G(t)$ away from the localized HCL spots in Figs. 1b-f showed no increase over L_G in Fig. 1a, in fact, a small, 0.5% decrease, demonstrating that SILC does not produce HCL. A 1-5meV change in V_G can produce a 1% decrease in L_G. Such changes can be obtained in the device by modifications of the surface potential with constant voltage stress. This result is consistent with previous work suggesting that SILC involves inelastic transport processes which cannot efficiently excite HCL.[6]

Thirdly, our results show how the carrier transport through the gate oxide changes as the localized breakdown progresses. The intensity of the HCL emission per carrier going through the gate, or the efficiency of the HCL, depends strongly on the excess energy of the hot carriers.[8] The efficiency of HCL varies significantly as the breakdown event progresses. The localized emission can have an efficiency, E_L, comparable to the efficiency of the fresh device emission. Since the initial $E_G(0)$ is due to ET, if $E_L=E_G(0)$, then the localized HCL is also excited by ET. This is consistent with previous assignments of soft breakdown transport to ET.[7] However, early in the breakdown, E_L was lower than $E_G(0)$. This connects the defect induce IET SILC to the soft breakdown events involving multiple defects and ET. At higher breakdown currents where hard breakdown characteristics appear, we also find that $E_L<E_G(0)$.

At a measurement time t, the efficiency of the uniform gate leakage excited HCL, E_G, and the efficiency of the localized breakdown excited HCL, E_L, are $E_G(t)=L_G(t)/I_G(t)$, and $E_L(t)=L_L(t)/I_L(t)$. L_L and the L_G excited by the ET before breakdown, and away from the breakdown spot can be measured from the HCL images. It is necessary to resolve the measured I_G, into its SILC, I_L, and direct tunneling away from the breakdown spot, components. The localized breakdown current is $I_G(t)$-I_{SILC}- $I_G(0)$. At breakdown, I_{SILC} at breakdown for Fig. 1 was about 660nA, $I_G(0)$=16.21μA and is assumed not to change since the breakdown affects only the properties of the gate oxide at the breakdown spot. Therefore, $L_L(t)$ immediately after breakdown can be correlated with $I_L(t)$. For Fig. 1b and d, $E_L=0.1xE_F(0)$. Figs. 1c and 1e have no localized emission and no localized emission efficiency. For Fig. 1f, $E_L=E_G(0)$. In all of these cases, L_G away from the breakdown spot was unchanged from L_G before breakdown showing that the gate current away from the breakdown spot was unchanged.

We observe no localized HCL in Fig. 1c, but have an excess localized current of 210nA. Given the low efficiency of the localized emission in Fig. 1d, if we consider that the HCL is the

sum of two currents, one which generate HCL with the efficiency of $E_G(0)$ through ET, and the remainder which is non-radiative like the SILC, then there are large, localized, non-radiative currents for Figs. 1b and c for I_{SILC}=660nA. These localized, non-radiative, currents are larger than the radiative currents.

The localized emission in Fig. 1f has an efficiency equal to $E_G(0)$. The emission from the gate away from the breakdown spot was unchanged from the fresh device showing that the breakdown did not change the direct tunneling current, confirming our assumptions on obtaining I_L from I_G.

Since under stress, the surface potentials at the interfaces can change, the assumptions that the direct tunneling excited gate emission away from the breakdown spot, and the direct tunneling currents, do not depend on the applied stress, need not be true. From row G of Table 1, after the device was held at 1.6V over 10^6 seconds, if these assumptions were used, E_L=19x$E_G(0)$. This is unlikely, and arises from the fact that our calculated I_L is very small (20nA) because I_G=16.89μA, and the original values for the direct tunneling current (16.21μA) and SILC(660nA) sum to 16.87μA. In fact, for this measurement, there was a significant decrease in the intensity of the light emission away from the breakdown spot for row G, consistent with both a decrease in the efficiency of the direct tunneling excited HCL and the direct tunneling current. If the value of I_F+I_{SILC} were reduced from 16.89 to 16.5μA, then our results in Table 1 would be consistent with the localized emission in row G arising from ET. While our discussion of E_L is based on I_{SILC}=660nA, the qualitative features of our argument stand for a variety of different possible I_{SILC}. For example, if we use the I_{SILC}=500nA, all of the qualitative features of the above discussion remain valid.

The initial breakdown of FETs at low voltages is due to the creation of neutral traps in the oxide, and the eventual appearance of a percolation path through the oxide involving these defects.[1,12] Our observation of low efficiency localized HCL, or equivalently, large non-radiative currents at the initiation of breakdown, therefore is the physical connection between the IET processes which are responsible for the SILC and the ET processes which are characteristic of soft-breakdown in ultra-thin oxides at low voltages.[7]

Since $Si0_2$ is highly ionic, point defects in $Si0_2$ feature relaxation of the neighboring atoms as the charge state of the defect changes. Blochl and Stathis[13] associated the neutral trap state required by the SILC to a neutral hydrogen bridge defect with a relaxation energy of 1.7eV. Such

energy shifts are associated with defect relaxation times of the order of picoseconds. If the defect state is occupied by a trapped carrier for times long compared to the relaxation time, this state is a simple channel for IET. Simple sequential tunneling through such states in the percolation picture cannot be achieve without substantial energy loss. If several trap states and the contacts are in close proximity, the relaxation of the isolated point defect can be minimized. Blochl and Stathis pointed out that the size of the energy relaxation will depend on the time the carrier spends on a trap site.[13] The full energy relaxation requires the electron to be trapped for a time long compared to the relaxation time. For example, there are about 1500 SILC sites for our device in Fig. 1.[1] If I_{SILC}=660nA for the whole device, the SILC per defect is 0.4nA, Since each defect traps only one electron at a time, this implies a defect occupation time of 30psec, which is long compared to the lattice relaxation time. In contrast, at breakdown, $I_L \sim 1\mu A$. While the number of sites involved in the charge transport via a percolation model involving the sequential occupation of neighboring defect states, has increased by 2 or 3, the three order of magnitude increase in the current means that the amount of time available for a carrier to occupy a particular site will be much less than a picosecond.[13] Therefore, lattice relaxation effects can be weaker. The onset of breakdown with neutral hydrogen bridge defects in close proximity must be accompanied by changes in the details of the charge transport through the individual defects which make up the percolation path. These changes are reflected in the change in E_L which depends strongly on the excess energy of the hot carriers when they reach the gate.

CONCLUSIONS

Hot carrier luminescence from carriers tunneling through ultra-thin gate oxides at lower voltages provides important information about the breakdown in these systems. The evolution of the breakdown has been traced optically. Direct spatial information about the breakdown including whether it involves progressive changes of a single location or multiple locations has been obtained. The strong dependence of the HCL on the excess energy of the hot carriers moving through the oxide, and the comparison of the efficiencies of the localized breakdown emission and the intrinsic ET emission means that the efficiency of the emission process can be used to probe the elastic or inelastic nature of the charge transport. Both large changes in the intensity of the hot carrier light emission due to rapid changes in I_G during gradual breakdown, and the efficiency of the emission, reflecting changes in the transport mechanism through the breakdown spot, are

observed. A transition between inelastic carrier transport through stress induced defects, and elastic tunneling through groups of stress induced defects is found.

Acknowledgments: We thank Dr. J. H. Stathis for many helpful suggestions and comments.

REFERENCES:

1. J. H. Stathis, IBM J. Res. & Dev. 46, 265-286 (2002).

2. F. Monsieur, E. Vincent, D. Roy, S. Bruyere, J. C. Vildeuil, G. Pananakakis, and G. Ghibaudo, Proc. 40th Annual International Reliability Physics Symposium, 2002, 45-54.

3. D. R. Wolters and J. F. Verwey, "Breakdown and Wearout Phenomena in Si02 Films,' Instabilities in Silicon Devices, G. Barbottin and A. Vapaille, Eds, North Holland, Amsterdam, 1986, 315-362.

4. T. Pompl, C. Engel, H. Wurzer and M. Kerber, Microelectronics Reliability 41, 543-551 (2001).

5. B. P. Linder, J. H. Stathis, R. A. Wachnik, E. Wu, S. Cohen, A. Ray, A. Wayshenkar, 2000 Symposium on VLSI Circuits, Digest of Technical Papers, p. 214.

6. S. Takagi, N. Yasuda, and A. Toriumi, IEEE Trans. Electron Dev., 46, 335-341 (1999).

7. E. Miranda, J. Sune, R. Rodriguez, M. Nafra, X. Aymerich, L. Fonseca, and F. Campabadal, IEEE Trans. Electron Dev., 47, 82-89 (2000).

8. J. C. Tsang and J. A. Kash, Appl. Phys. Lett., 70, 889-891 (1997).

9. J. C. Tsang, J. A. Kash and D. P. Vallett, Proc. IEEE, 88 1440-1459 (2000).

10. S. Lo, D. A. Buchanan, Y. Taur, and W. Wang, IEEE Elec. Dev. Lett. 18, 209-211 (1997).

11. B. P. Linder, and J. H. Stathis, INFOS (to be published)

12. R. Degrave, G. Groeseneken, R. Bellens, J. Ogier, M. Depas, P. Roussel and H. Maes, IEEE Trans. Electron Dev. 45, 904-910 (1998).

13. P. Blochl and J. H. Stathis, Phys. Rev. Lett. 83, 372-375 (1999).

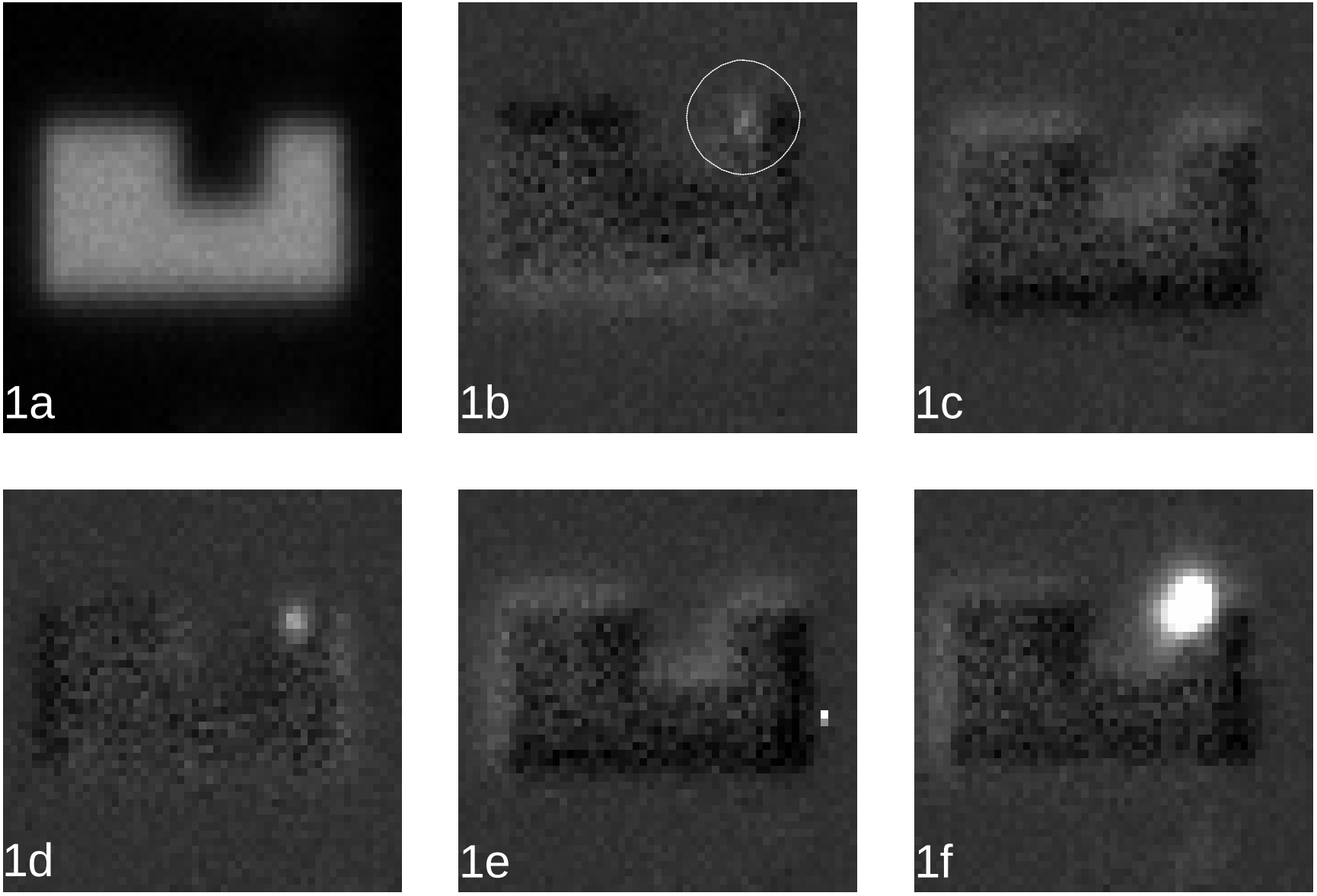

Figure 1. Light emission from a10x20um nFET with a1.5nm gate oxide excited by V_G=1.6V and stressed under a variety of conditions for voltages less than 2.8V. 1a is the image as recorded by the CCD camera. 1b-f are the difference images where Fig. 1a is subtracted from the acquired image. The outlines of the emission region in Figs. 1b-f are due to small misalignments in the pairs of images in the subtractions. A dotted circle indicates the weak emission in Fig. 1b. The large area of the emission in Fig. 1f is due to the saturation of of the grey scale by the emission.

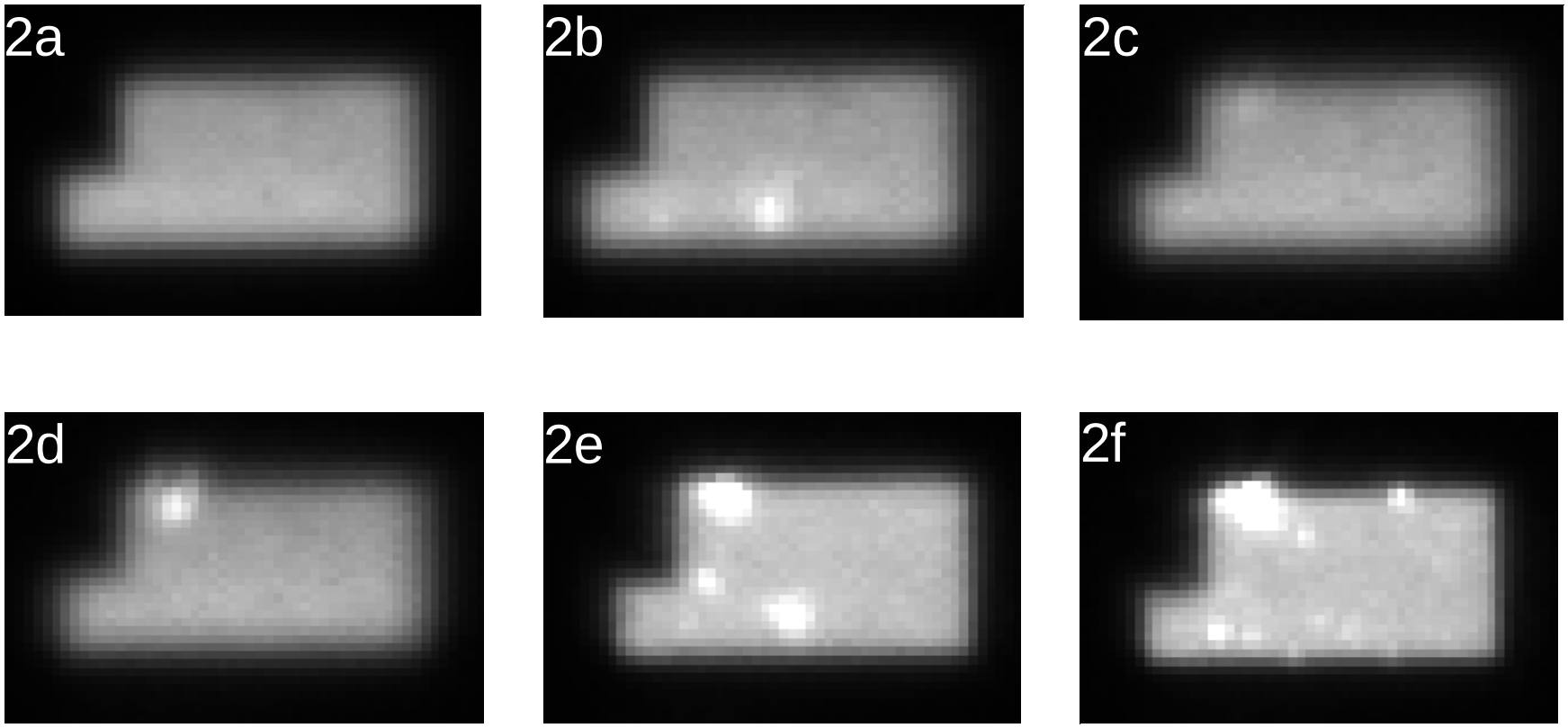

Figure 2. Light emission images from a 10x20um pFET with a 1.5nm gate oxide excited by -2.9V stress. Multipixel emissions are due to saturation of the grey scale by the emission.

Table 1

	$I_G(\mu A)$	$I_G^{RMS}(\mu A)$	L_G(cts)	L_L(cts)
1a	16.21	3.51	198894	
1b	17.22	22.7	195432	458
1c	17.08	6.87	195834	<100
1d	17.29	406.	197129	796
1e	16.87	3.84	191283	<100
1f	18.15	467.	197056	9619
1g	16.89	11.8	165212	3972

Gate currents, RMS deviation of the gate currents, large area emission intensities, and localized emission intensities for the images of the device in Figure 1, and one additional image not shown in Fig. 1.