

Influence of post-annealing on the electrical properties of metal/oxide/silicon nitride/oxide/silicon capacitors for flash memories

Hee Dong Kim¹, Ho-Myoung An^{1,2}, Kyoung Chan Kim¹, Yu Jeong Seo¹
and Tae Geun Kim¹

¹ School of Electrical Engineering, Korea University, Anam-dong 5-ga, Seongbuk-ky, Seoul 136-701, Korea

² Department of Electronic Materials Engineering, Kwangwoon University, Wolgye-dong, Nowon-gu, Seoul 139-701, Korea

E-mail: tgkim1@korea.ac.kr

Received 10 January 2008, in final form 7 May 2008

Published 4 June 2008

Online at stacks.iop.org/SST/23/075046

Abstract

We report the effect of post-annealing on the electrical properties of metal/oxide/silicon nitride/oxide/silicon (MONOS) capacitors. Four samples, namely as-deposited and annealed at 750, 850 and 950 °C for 30 s in nitrogen ambient by a rapid thermal process, were prepared and characterized for comparison. The best performance with the largest memory window of 4.4 V and the fastest program speed of 10 ms was observed for the sample annealed at 850 °C. In addition, the highest traps density of $6.84 \times 10^{18} \text{ cm}^{-3}$ was observed with ideal trap distributions for the same sample by capacitance–voltage ($C-V$) measurement. These results indicate that the memory traps in the ONO structure can be engineered by post-annealing to improve the electrical properties of the MONOS device.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

The rapidly increasing market for portable electronic products has promoted the development of nonvolatile semiconductor memories (NVSM). Of the various NVSM devices, floating-gate (FG) flash memories outperform their competitors due to their full process compatibility with complementary metal-oxide-semiconductor (CMOS) technology and superior performance [1]. However, reducing the tunnel oxide thickness in the conventional FG flash technology is becoming increasingly difficult as the device size continues to shrink down.

Polysilicon-oxide-silicon nitride-oxide-silicon (SONOS) devices have been considered as a potential replacement for the FG structures in the next-generation NVSM because they have advantages in terms of their ability to be scaled down. SONOS devices store charges in discrete traps within the silicon nitride layer of the oxide-silicon nitride-oxide (ONO) layer, while the

FG structures store charges in a conductor (i.e., polysilicon) in the form of free carriers in conduction bands [2, 3]. SONOS technology has progressed to allow for two-bit operation in a single device as well as usage of high- k dielectrics in the ONO structure [4].

However, SONOS suffers from increasing reliability problems, such as a large leakage current, slow program/erase speed, poor retention due to physical and chemical reactions during the growth of the ONO layers [5, 6], due to the continual shrinkage of the ONO layers required for high-density memories. It has been reported that silicon nitride layers grown by low pressure chemical vapor deposition (LPCVD) generally give rise to high leakage currents and low film qualities, due to the interfacial traps (between the silicon substrate and tunnel oxide) induced during the deposition. Therefore, many researchers working in this area have attempted to improve the film quality by optimizing the

ONO structure, the thin-film growth method and the thermal annealing conditions [7–9].

In this work, we investigate the impact of the post-annealing process on the electrical characteristics such as programming, erasing and retention characteristics of metal/oxide/silicon nitride/oxide/silicon (MONOS) capacitors using deep level transient spectroscopy (DLTS).

2. Experiments

For the sample preparation, first, (100)-oriented p-type Si wafers were used as the starting substrates. A 20 Å thick tunneling oxide layer was then thermally grown by dry oxidation at 900 °C. A 65 Å thick silicon nitride layer was subsequently deposited by LPCVD at 750 °C by the reaction of dichlorosilane (SiCl_2H_2) and ammonia (NH_3) gas. An *in situ* 110 Å thick blocking oxide layer was then deposited by CVD. Prior to the post-annealing, the ONO structures were cleaned with a standard sulfuric acid and hydro-peroxide mixture (SPM) for 10 min. After drying in a nitrogen ambient atmosphere, the ONO samples were annealed by the rapid thermal annealing (RTA) process in nitrogen (N_2) ambient for 30 s [6]. The ONO films that were investigated were either as-deposited, or annealed at 750 °C, 850 °C or 950 °C. After cleaning with buffered oxide etchant (BOE) for 1 s, a 400 nm thick top aluminum (Al) electrode was deposited by an e-beam evaporation system and a bottom indium (In) contact was made on the substrate as an Ohmic contact. The electrical properties of the MONOS samples were measured using a Keithley 4200 semiconductor parameter analyzer, capacitance–voltage (C – V) meter (HP 4280A) and DLTS.

3. Results and discussion

Figure 1 shows the leakage currents versus applied voltages curves for the as-deposited samples and those annealed at 750, 850 and 950 °C. The measured leakage current of the as-deposited MONOS sample was 10 pA and those of the samples annealed at 750, 850 and 950 °C were 100, 84 and 450 pA at the same voltage of 4 V, respectively. In a previous study [5], it was found that the gate leakage current increased with increasing annealing temperature due to thermal damage; however, in our experiment, the lowest gate leakage current was observed for the sample annealed at 850 °C. It can be considered that the optimum curing condition for the lowest interface trap in the ONO layer is met at the temperature of 850 °C [7–9]. In order to confirm whether 850 °C is the optimum annealing temperature for memory applications, we investigated the electrical properties such as the C – V memory window, the program/erase (P/E) characteristics, and the retention and the memory trap information for the MONOS samples annealed at various temperatures.

First of all, we measured the C – V hysteresis for the four samples to compare their memory abilities, as shown in figure 2. The voltage sweep was done in the range from –8 to 8 V at room temperature. The memory windows were measured to be 1.2, 0.8, 4.4 and 1.8 V for the as-deposited sample and those annealed at 750, 850 and 950 °C,

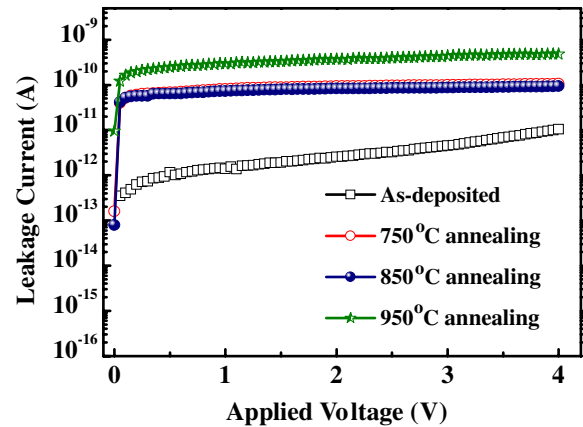


Figure 1. The leakage currents versus applied voltages for the as-deposited sample and those annealed at 750, 850 and 950 °C.

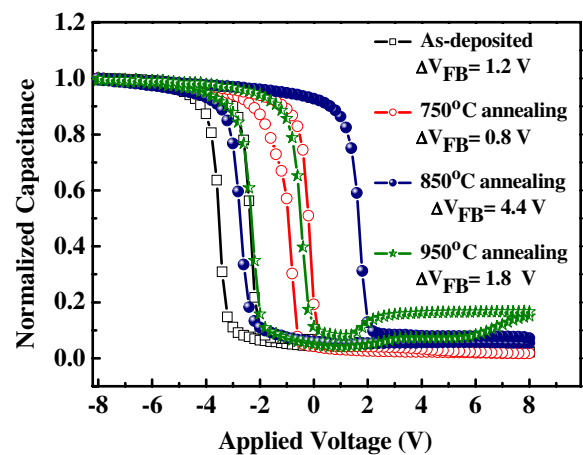


Figure 2. The C – V memory windows of the MONOS as-deposited sample and those annealed at 750, 850 and 950 °C.

respectively. The largest memory window was observed for the sample annealed at 850 °C while the smallest one was observed for the sample annealed at 750 °C. The memory window decreased from 4.4 to 1.8 V as the annealing temperature increased from 850 to 950 °C. This might be due to the increase in the number of interfacial traps between the silicon substrate and tunnel oxide generated by thermal damage at high temperatures [7] judging from the C – V curve, where low-frequency components generally caused by interface traps are observed in the inversion region (>2 V) [10]. This result indicates that the memory window size is closely related to the annealing temperature since the trap density and its space distribution of traps are affected by it. The trapped charge density (N_T) can be evaluated by measuring the flat band voltage shift, which is given by [11]

$$N_T = \frac{\Delta V_{\text{FB}} \epsilon_{\text{Si}_3\text{N}_4}}{q [(\epsilon_{\text{Si}_3\text{N}_4} / \epsilon_{\text{SiO}_2}) X_{\text{SiO}_2} + X_{\text{Si}_3\text{N}_4}]} \quad (1)$$

where $X_{\text{Si}_3\text{N}_4}$ and X_{SiO_2} represent the thicknesses of the charge storage (Si_3N_4) layer and blocking oxide (SiO_2) layer, respectively. In our case, $X_{\text{Si}_3\text{N}_4} = 60$ Å and $X_{\text{SiO}_2} = 110$ Å. The terms $\epsilon_{\text{Si}_3\text{N}_4}$ and ϵ_{SiO_2} are the dielectric constants of silicon nitride and the blocking oxide. The trapped charge densities

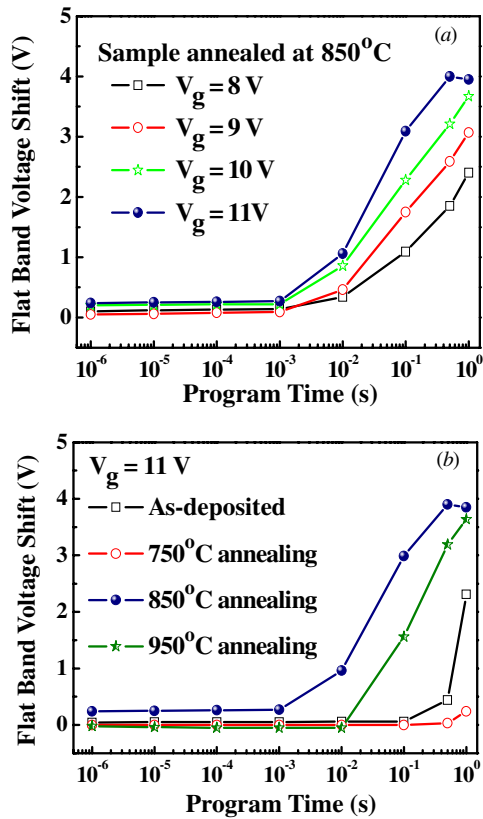


Figure 3. The program efficiency of the samples (a) annealed at 850 °C with different gate voltages ranging from 8 to 11 V, and the as-deposited sample and those annealed at 750, 850, and 950 °C at a gate voltage of 11 V.

(N_T) calculated from the as-deposited ONO structure and those annealed at 750, 850 and 950 °C are 1.86×10^{18} , 1.24×10^{18} , 6.84×10^{18} and $2.79 \times 10^{18} \text{ cm}^{-3}$, respectively.

Then, we investigated the program efficiency of the four samples by measuring the flat band voltage shifts against the gate voltages applied to the samples in the range from 8 to 11 V. The result for the MONOS sample annealed at 850 °C is shown in figure 3(a) as an example, where the saturation voltage—i.e., the point at which the program characteristics are saturated—turned out to be 11 V. However, the flat band voltage shift at 11 V seems to reverse from increasing to decreasing for 1 s pulses. This may be due to a discharge of the electrons through the gate electrode at a strong electric field. Similar observation has been reported in [12]. Then we compared the program characteristics of the four samples at a gate voltage of 11 V, as shown in figure 3(b). It is clearly observed that the sample annealed at 850 °C shows the fastest program speed, namely 10 ms at $V_g = 11$ V and the largest flat band voltage shift, namely 3.9 V at $V_g = 11$ V among the four samples. We think that this largest flat band voltage shift is associated with the largest trap density predicted by the C - V hysteresis of figure 2.

On the other hand, the flat band voltage shift was very small, namely 0.24 V at $V_g = 11$ V, for both the as-deposited sample and the one annealed at 750 °C, which might be due to the insufficient number of traps of silicon nitride layer predicted by the C - V hysteresis of figure 2. In general, the flat

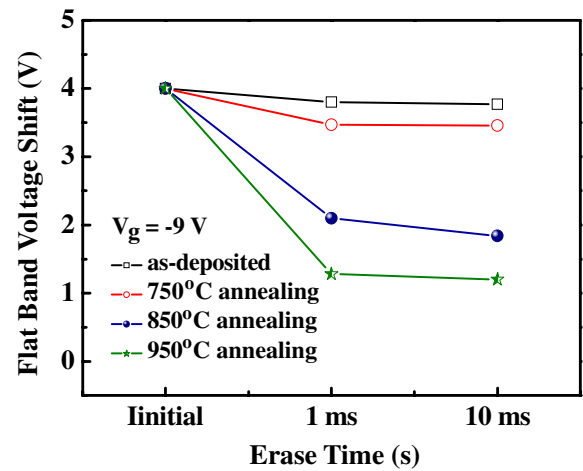


Figure 4. The erase efficiency of the as-deposited sample and those annealed at 750, 850 and 950 °C.

band voltage shift is a function of the trapped charge density, and therefore, these results are consistent with the flat band voltage shift of C - V results discussed in figure 2. Figure 4 shows the erase characteristics measured for the four samples at a gate voltage of -9 V. The flat band voltage shifts were large enough for the erase operation, namely -2.2 V at -9 V for the samples annealed at 850 °C and -2.7 V at -9 V for the samples annealed at 950 °C; however, as indicated in figure 4, the flat band voltage shifts of the as-deposited sample and the sample annealed at 750 °C are less than -0.6 V when the gate voltage with 1 ms at -9 V is applied. These program and erase characteristics would be expected to be closely related to the changes in the trap density or its energy distribution of traps across the ONO structures for different annealing temperatures, as described below in the discussion of the use of the DLTS technique. The program/erase characteristics for ONO samples are inferior compared to already published data in a SONOS device, as shown in figures 3 and 4. This is thought to be due to a low work function of the aluminum metal electrode. However, it is expected that they could be improved greatly by using high work function metals such as Pt as the gate electrode [13].

We also investigated the effect of post-annealing on the retention of the samples. Figure 5 shows the relation between the flat band voltage shift and read delay time for the as-deposited sample and those annealed at 850 and 950 °C. Since little shift of flat band voltage after program or erase operation is observed and is not appropriate to the retention experiment, the sample annealed at 750 °C was excluded for the retention test. In this experiment, the sample annealed at 850 °C also showed the best result among the three samples. The average charge decay rate for the as-deposited sample and those annealed at 850 and 950 °C are measured to be 17.58, 28.75 and 32.10 mV/decade, respectively, in the program state. In addition, the flat band voltage shifts were extrapolated against the read delay time to predict the long-term retention of the as-deposited sample and those annealed at 850 and 950 °C, and the values observed at 3×10^8 s (10 years) were

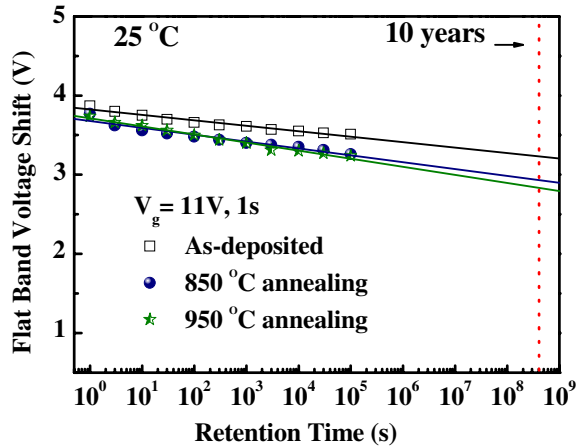


Figure 5. The relation between the flat band voltage shift and read delay time for the samples annealed at 850 and 950 °C, indicating the retention characteristics of the MONOS samples.

0.77, 1.0, and 1.16 V, respectively, at room temperature. This result indicates that the retentions of the three samples would all be more than 10 years at a program voltage of 11 V at room temperature [14, 15].

On the other hand, we measured the DLTS signals for the four samples, in order to investigate the effect of post-annealing on the trap density and its energy distribution of traps in the ONO structure and thereby find out why the sample annealed at 850 °C showed the best electrical characteristics such as programming, erasing and retention. We first measured the DLTS signals for both the silicon nitride/oxide and oxide samples, as shown in figures 6(a) and (b), respectively, to obtain a reference for identifying the origin of the DLTS signals measured for the ONO structure, namely whether they were induced by silicon nitride-related traps or by Si/SiO₂ interface-related traps. The peak positions of the DLTS signals for both samples were different, with the silicon nitride trap related peak A being observed near 295 K in the MIS structure and the Si/SiO₂ interface trap related peak B being observed near 335 K in the MOS structure.

Figure 7(a) shows the DLTS signals of the as-deposited MONOS sample and those annealed at 750, 850 and 950 °C. The peak positions of the DLTS signals ranged from 252 to 295 K in the vicinity of peak A, indicating that the silicon nitride-related traps (traps related to the silicon nitride layer) are much more dominant than oxide-related traps (traps related to the interface between silicon and silicon oxide) in these samples, regardless of the annealing temperature. A bias voltage with a pulse width of 10 ms and a height of 2 V (trap-filling pulse) was applied to the sample and the rate window was fixed at 13.86 s⁻¹.

In order to ensure that the DLTS signals observed in these samples originate from the silicon nitride-related traps, which are useful for the trap density or its energy distribution of traps [16], we also measured the small-pulse (SP) DLTS signals for the same samples in order to identify the interfacial traps between the silicon substrate and tunnel oxide layer, as shown in figure 7(b). In this figure, all of the maximum peaks, regardless of the annealing temperature, are located

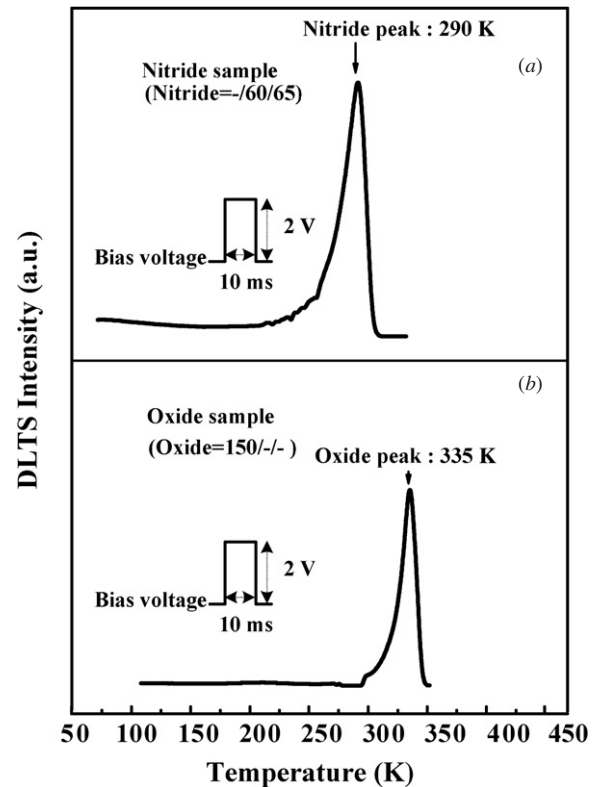


Figure 6. The DLTS signals measured from the (a) silicon nitride/oxide sample and (b) oxide sample.

in the vicinity of peak B, which is assumed to be related to the silicon-silicon dioxide interfacial trap, although the temperature of the maximum DLTS peaks ranges from 322 to 380 K for the different annealing temperatures. Because only interfacial traps can be detected by SP-DLTS, we can conclude that B originates from the Si-SiO₂ interfacial trap. The interface states can be investigated when the trap-filling pulse amplitude is small (<100 mV). In this case, the interfacial traps situated below the Fermi level are occupied by electrons and those above are empty, transition occurring over an energy range of ~2 kT at the Fermi level at the tunnel oxide/silicon substrate interface state, which can be scanned over one half of the band gap by applying different gate voltages [16].

Back in figure 7(a), in order to obtain more detailed information on the silicon nitride-related traps across the ONO structure, we calculated the activation energy of traps and trap density (N_T) of the ONO structure from the slope and the intercept with the y axis in the Arrhenius plot, obtained from the shift of the DLTS peaks at the maximum peak height for the different rate windows (0.4 to 15 s⁻¹). To calculate the trap density, the carrier density at 300 K was obtained from C-V data and then applied to the well-known equation $N_T = 2N_d(\Delta C/C_0)$, where N_d is the net doping concentration of Si, ΔC is the DLTS signals and C_0 is the capacitance at bias voltage used in the DLTS measurements [12, 18]. The calculated results are summarized in table 1. In fact, the trap density reported in the literature for the ONO structure is in the order of 10¹⁹ [11, 18], but the trap density in our experiment

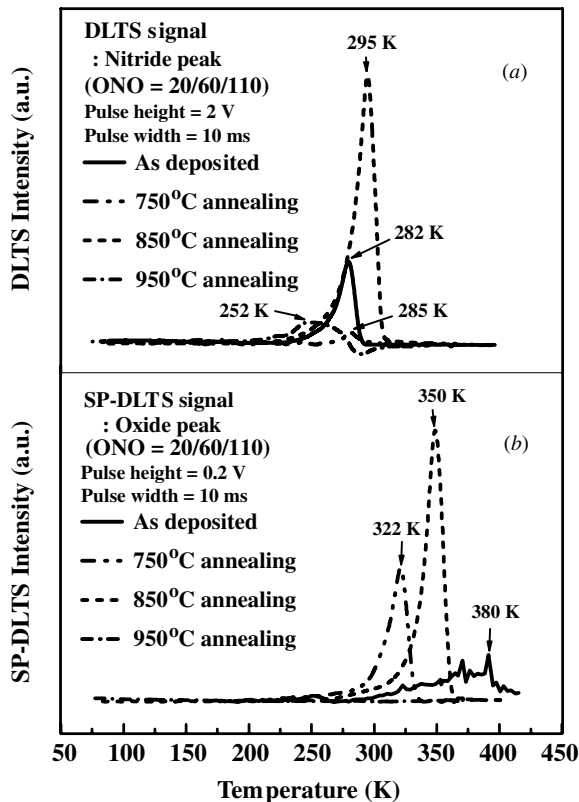


Figure 7. The DLTS signals measured from the as-deposited MONOS sample and those annealed at 750, 850 and 950 °C: (a) nominal mode and (b) small-pulse mode.

Table 1. E_a and N_T of the as-deposited MONOS sample and those annealed at 750, 850 and 950 °C.

	As-deposited	750 °C annealing	850 °C annealing	950 °C annealing
E_a (eV)	0.566	0.507	0.278	0.095
N_T (cm ⁻³)	8.75×10^{13}	1.43×10^{12}	2.42×10^{14}	3.61×10^{13}

is much lower, in the order of 10^{14} . This is because the volume taken for the observation period is much larger than that of the traps present in ONO for observation. The peak intensity of the DLTS signal was highest for the sample annealed at 850 °C, indicating that this sample has the largest silicon nitride-related trap density among the four samples. This phenomenon (silicon nitride-related trap density) might affect the C - V memory window and program/erase characteristics of the MONOS capacitor because the memory window and program/erase speed improved as the increased traps density of the silicon nitride layer in figures 2–4. On the other hand, the activation energy of the traps in silicon nitride layer in the ONO layers moved to the shallow energy levels within the silicon nitride layer as the annealing temperature increased, as shown in table 1. This is because the activation energy of the trap in a silicon nitride layer can be decreased by gaining the thermal energy during the annealing process [14]. The reason why the energy levels of traps become shallower when

the annealing temperature increases is now investigated and will be reported elsewhere. This phenomenon might affect the retention of the MONOS capacitor because the average decay rates increased (namely the retention time decreased) as the activation energy level of the trap decreased in figure 5. It is certain that the distributions of the silicon nitride-related traps are closely related to the electrical operation of the MONOS devices, which can be observed effectively by DLTS analysis. Again, optimizing the annealing condition is important to improve the electrical characteristics of ONO thin films for charge-trap memory devices.

4. Conclusion

We studied the effect of post-annealing on the electrical properties of MONOS capacitors using DLTS. Four samples, namely the as-deposited specimen and those annealed at 750, 850 and 950 °C for 30 s in nitrogen ambient by a rapid thermal process, were prepared and characterized for comparison. It was found that the sample annealed at 850 °C with the highest trap density in the silicon nitride layer showed the best electrical performance. This result indicates that DLTS can provide a useful means of optimizing ONO structures and thereby improving the electrical properties of charge-trap memory devices.

Acknowledgment

This work was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government (MEST) (Quantum Photonic Science Research Center).

References

- [1] Wu K H, Chien H C, Chan C C, Chen T S and Kao C H 2005 *IEEE Trans. Electron Devices* **52** 987–92
- [2] Gritsenko V A, Nasyrov K A, Novikov Y N, Aseec A L, Yoon S Y, Lee J W, Lee E H and Kim C W 2003 *Solid-State Electron.* **47** 1651–5
- [3] White M H, Wang Y, Wrazien S J and Zhao Y 2006 *Int. J. High Speed Electron. Syst.* **16** 479–502
- [4] Lee Y K, Sung S K and Sim J S 2002 *J. Korean Phys. Soc.* **41** 908–11
- [5] Mao L F 2007 *Semicond. Sci. Technol.* **22** 1203–8
- [6] Lin Y H, Chien C H, Chang C Y and Lie T F 2006 *J. Vac. Sci. Technol.* **24** 682–5
- [7] Zhan N, Ng K L, Wong H, Kok C W and Poon M C 2003 *IEEE Conf. Electron Devices Solid-State Circuits* **17** 431–4
- [8] Lucovsky G, Wu Y, Niimi H, Misra V and Phillips J C 1999 *Appl. Phys. Lett.* **74** 2005–7
- [9] Kim J K, Cheong H J, Park K H, Kim Y, Yi J Y, Bark H J, Bang S H and Cho J H *J. Korean Phys. Soc.* **42** 316–9
- [10] Fedorenko Y G, Truong L, Afanas'ev V V and Stesmans A 2004 *Mater. Sci. Semicond. Process.* **7** 1–6
- [11] Kim T H, Park I H and Lee J D 2006 *Appl. Phys. Lett.* **89** 063508
- [12] Lee C H, Hur S H, Shin Y C and Choi J H 2005 *Appl. Phys. Lett.* **86** 152908

- [13] Jeon S and Kim C W 2006 *Electrochem. Solid State Lett.* **9** G265–67
- [14] Kim J S, Kim E K, Park K, Yoon E, Han I K and Park Y J 2005 *Physica E* **26** 91–5
- [15] Wrazien S J, Zhao Y, Krayner J D and White M H 2003 *Solid-State Electron.* **47** 885–91
- [16] Seo Y J, Kim K C, Sung Y M, Cho H Y, Joo M S, Pyi S H and Kim T G 2008 *Appl. Phys. Lett.* **92** 132104
- [17] Johnson N M 1979 *Appl. Phys. Lett.* **34** 802–4
- [18] Yang Y L and White M H 2000 *Solid-State Electron.* **44** 949–58