Annealing of defect density and excess currents in Si-based tunnel diodes grown by low-temperature molecular-beam epitaxy

Sung-Yong Chung, Niu Jin, Ryan E. Pavlovicz, and Paul R. Berger

Department of Electrical and Computer Engineering, The Ohio State University, Columbus, Ohio 43210-1272

Ronghua Yu
Department of Physics, The Ohio State University, Columbus, Ohio 43210-1106

Zhaoqiang Fang
Semiconductor Research Center, Wright State University, Dayton, Ohio 45435

Phillip E. Thompson
Naval Research Laboratory, Washington, DC 20375-5347

(Received 16 January 2004; accepted 5 April 2004)

Deep-level transient spectroscopy (DLTS) measurements were performed in order to investigate the effects of post-growth heat treatment on deep level defects in Si layers grown by low-temperature molecular-beam epitaxy (LT-MBE) at 320 °C. In the LT-MBE as-grown samples, two dominant divacancy-related complex defects, of which the possible origins are suggested as \( P - V(E \text{ center}) + V - V(0/-) \) and \( V - V(-2/-) \) and others, were observed in P-doped \( n \) layers. When the as-grown samples were annealed at 700, 800, and 900 °C for 60 s by rapid thermal annealing, the total density of defects were decreased without generating other defects and most defects were annihilated at 900 °C. This study also compared the DLTS trends with performance of Si-based resonant interband tunnel diodes (RITDs) in terms of peak current density, valley current density, and peak-to-valley current ratio, which are closely related to the deep-level defects. The active regions of the RITDs were grown at the same substrate growth temperature and annealed at similar temperatures used in this DLTS study. © 2004 American Institute of Physics.

[I. INTRODUCTION]

Si-based epitaxial growth of mainstream SiGe heterojunction bipolar transistor technology has become commonplace, and many emerging Si-based three-dimensional transistors for complementary metal oxide semiconductors, nanoelectronic switches based on quantum dots and Si-based tunnel diodes also utilize Si-based epitaxy. To realize many of these emerging nanoscale devices, low-temperature growth is required to minimize segregation and diffusion that concurrently create point defects. As device sizes are aggressively scaled down to nanometer dimensions and power consumption requirements shrink, greater effort is required to understand the behavior and nature of defects created during low-temperature molecular-beam epitaxy (LT-MBE), because their presence directly affects many important electronic properties. A variety of device parameters that are altered by defects are minority-carrier lifetime, junction-leakage currents, and enhanced tunneling currents through deep-level defects. From a technological point of view, it is very important to understand the behavior of defects and their influence on the thermal budget of the device processing, especially considering a post-growth anneal to annihilate the point defects.

The key strategy in the development of Si-based resonant interband tunnel diodes (RITDs), which is a motive of this study and a prime example of point defect influence on device performance, was to define quantum wells (QWs) in both the conduction and valence bands by utilizing \( \delta \)-doped sheet carrier concentrations to achieve the degeneracy \( \left( \sim 10^{20} \text{ cm}^{-3} \right) \) needed for carriers to tunnel through the junction. However, the overall dopant concentration achievable using epitaxial growth techniques is limited under equilibrium growth conditions because, at high growth temperatures, dopants segregate and interdiffuse. LT-MBE is a far-from-equilibrium growth technique that has been developed to achieve degenerate \( \delta \)-doping by minimizing segregation and diffusion. However, LT-MBE concurrently creates point defects, leading to an elevated valley current, which are caused by tunneling through defects in the forbidden bandgap. Consequently, the peak-to-valley current ratio (PVCR), a figure-of-merit for the RITD, is suppressed. An as-grown Si RITD fabricated using LT-MBE shows a suppressed negative differential resistance (NDR) region, a unique property of tunnel diodes. The most dominant defects in Si LT-MBE growth are point defects created by the limited adatom mobility on the epitaxial surface leading to vacancies and by the strains and stresses stemming from the different atomic radii between Si and the dopants chosen. Defect complexes can arise from a combination of these defects.
A second key strategy adopted for Si-based RITDs\textsuperscript{5,6} was a short post-growth anneal to reduce point defects created during LT-MBE growth. A short 1 min post-growth rapid thermal annealing (RTA) ranging between 600 and 900 °C is adequate. In general, a higher temperature and/or a longer period of post-growth heat treatment leads to a lower defect concentration in the epitaxial films. However, annealing with a higher temperature or longer time, in contrast, causes the δ-doping spikes to broaden, resulting in performance degradation or failure of the RTTD by a loss in quantum confinement. Therefore, an optimum annealing temperature and time is sought for the best PVCR and current density.

This study extends earlier work of the authors\textsuperscript{9} in which growth temperature and dopant species effects on deep levels were investigated with LT-MBE as-grown silicon layers. Samples were studied by deep-level transient spectroscopy (DLTS).\textsuperscript{10} The possible origin of the defects was suggested as vacancy-related complexes. In this study, we investigate post-growth RTA heat treatment effects on these defects. This study ends by correlating the effect of RTA heat treatment and defect annihilation on the performance of RITDs grown by LT-MBE.

II. EXPERIMENT

Two sets of \( p^+ - n \) one-sided step-junction diodes (SJDs) were studied, in which the key difference was the growth temperatures, either 600 or 320 °C, of the Si:P-doped \( n \) layer described later, and examined by DLTS. Each data set consists of four samples taken from each wafer grown at two different growth temperatures, and each sample was then annealed at different temperatures. The epitaxial growth was carried out in a Vacuum Generators V-80 MBE system using solid sources. The basic layer structure, shown in Fig. 1, has an \( n \)-on-\( p \) growth sequence. For the first set of samples shown in Fig. 1(a), a 20 nm Si buffer layer was grown at 650 °C on the \( p^+ \)-Si substrate. Following the buffer layer, an 18 nm B-doped layer was grown during an appropriate substrate temperature adjustment, and the substrate temperature was then held at 600 °C for the growth of a 50 nm B-doped \( p^+ \)-layer \( (1 \times 10^{18} \text{ cm}^{-3}) \). For the phosphorus-doped \( n \) layer \( (1 \times 10^{17} \text{ cm}^{-3}) \), a thickness of 200 nm was determined to be large enough to allow examination of the growth and post-growth heat treatment effect upon deep-level defects and their densities within the \( n \) layer. The substrate temperature was 600 °C during Si MBE growth for this \( n \) layer. After a P cell temperature adjustment, the 30-nm P-doped \( n \) contact layer \( (1 \times 10^{20} \text{ cm}^{-3}) \) was grown at a substrate growth temperature of 320 °C. All growth was carried out at a rate of 0.1 nm/s.

The basic growth sequence for the second set of samples was the same as the first set of samples, except that the substrate growth temperature was decreased from 600 to 320 °C for the 200 nm phosphorus doped \( n \) layer. A substrate growth temperature of 320 °C has been widely used for Si-based RITDs.\textsuperscript{5,6}

Prior to device fabrication, three samples taken from each wafer were annealed under a forming gas of 95% \( N_2 \), 5% \( H_2 \) ambient in a Modular Process Technology Corporation RTP-600S furnace. The annealing time of all samples was fixed at 60 s. Anneal temperatures of 700, 800, or 900 °C were employed in both sets of as-grown samples. A fourth sample from each wafer was not heat treated and served as a comparison.

Ti/Au ohmic contacts (150 μm in diameter) were patterned photolithographically using lift-off and electron-beam evaporation, while backside ohmic contacts of Ti/Au were made by deposition across the full surface. Using the top contact metal as a self-aligned etch mask, all diodes were formed into a mesa structure approximately 300 nm tall. The wet etchant for mesa etching was HF: \( \text{H}_2\text{O}:\text{HNO}_3 \) (1:100:100) by volume ratio and etch rates varied between 100 and 150 nm/min.

The \( I–V \) characteristics were first checked at room temperature by an Agilent 4156C parameter analyzer. A Bio-Rad DL4600 system with a 100 mV test signal at 1 MHz was used to measure the \( C–V \) and DLTS characteristics. The \( C–V \) data, used to calculate the carrier profiles, was taken in the temperature interval of 100 to 350 K to establish if the carrier concentration varied as a function of temperature.

In the DLTS measurements, the temperature was varied from 100 to 350 K. Typical bias voltages (\( V_B \)) of −4.5 to −1.0 V were used with a filling pulse height (\( V_F \)) of −2.0 V to +0.5 V, a pulse width (\( W_F \)) of 1 ms and a rate window of 50/s. The rate window is set by the DLTS instrumentation and determined by the two sampling periods at \( t_1 \) and \( t_2 \) between which the change in capacitance due to the capture/emission processes is measured. The rate window (i.e., 1/\( \tau \)) is defined as ln(\( t_2/t_1 \))/\( (t_2-t_1) \). For a given rate window, there

\[ t \]
may be maxima (for majority traps) and minima (for minority traps) in the DLTS spectrum which are related to traps due to defects. If the DLTS spectrum versus the various rate windows is measured, Arrhenius plots can be obtained because the temperatures at which the minima and maxima occur are a function of the rate window. We can use Arrhenius analyses\(^9,11\) to determine activation energies and capture cross sections for the traps.

III. RESULTS

Figure 2(a) shows the DLTS spectra for the as-grown sample at 600 °C as well as the samples that experienced post-growth RTA at 700, 800, and 900 °C for 60 seconds. Shown in Fig. 2(b) are the DLTS spectra for the samples annealed at the same temperatures as in Fig. 2(a), but grown at the reduced substrate temperature of 320 °C. As will be discussed, the annealing temperatures and times were commensurate to those used for Si-based RITDs to achieve high PVCRs with the goal of annealing defects that contribute to an elevated valley current.

During the DLTS measurement, the majority carrier pulses were set such that maximum peak signals could be obtained. \(V_b\) of −4.0 V and \(V_f\) of +0.5 V were applied for the as-grown sample in group A (denoted for samples grown at 600 °C of substrate growth temperature), \(V_b\) of −2.0 V, and \(V_f\) of 0.0 V for the annealed samples in group A, and for the all samples in group B (denoted for the samples grown at 320 °C of substrate growth temperature), \(V_b\) of −3.0 V and \(V_f\) of 0.0 V, respectively. Pulse width \(W_f\) of 1.0 ms and rate window of 50/s were commonly employed in both sample group measurements.

As shown in the Fig. 2(a) for the as-grown sample at 600 °C, a dominant peak, labeled E1, emerged at around 220 K and separated into E1 and another peak, denoted as Ex, which emerged at around 240 K, when the sample was annealed at 700 °C, implying that E1 consists of at least two peaks. Several minor peaks, labeled E2, E3, and E4, were observed at around 160, 100, and 300 K in the spectra, respectively. The intensity of E1, including Ex, decreased as the annealing temperature increased, as did E3 and E4. The minor peak E2 became stronger for samples annealed at 700 °C and then the intensity of E2 decreased with higher temperature annealing. The increase in the intensity of E2 at 700 °C is approximately equal to the decrease in the intensity of E3. While peak intensities of E3 and E2 decrease at the higher temperatures simultaneously, the annihilation rate of the E3 peak is faster than that of E2, which might imply that the origin of E2 is related to the removal of E3. Some observed peaks appeared to be annihilated as their magnitude was below the detection limit when annealed at 900 °C for 60 s.

For the as-grown samples at 320 °C, shown in Fig. 2(b), the position of the peaks in the DLTS spectra slightly shifts to lower measurement temperatures with respect to the peaks observed in the as-grown sample at 600 °C. The dominant peak, E11, clearly shows that it consists of at least two more peaks, indicated as Exx and Eyy. Unlike Ex, which is a hidden shoulder of the E1 peak and emerged when the sample was annealed at 700 °C, Exx and Eyy were clearly observed prior to annealing in the as-grown sample, but were later greatly reduced when the sample was annealed above 700 °C for 60 s. The signal peaks, E33 and E22, were observed over a broad range of lower measurement temperatures (<−160 K) in the DLTS spectra. This indicates that the defects are more complex and not well defined.

Since both groups of samples have been grown by the same MBE machine and similar doping concentrations, the peaks observed are expected to be closely related or to have the same defect origin. As the peaks in groups A and B are positive and the samples have a \(p^+−n\) one-sided SJ structure, the peaks observed in the DLTS spectra indicate majority electron traps in the phosphorus doped \(n\) layer. No minority peaks with concentrations above \(10^{13} \text{cm}^{-3}\) were revealed.

A reference sample was used to study the nature of defects by thermal emission studies, using the Arrhenius plots.
of $T^2/e_n$ versus $1000/T$ obtained by varying the rate windows from 4/s to 1000/s for as-grown sample at 600 °C in group A. The activation energy levels of the observed E1 and E2 are $E_C = 0.48$ eV and $E_C = 0.27$ eV, respectively.

The annealing effect observed in the DLTS spectra on the density of defects is as-illustrated in Fig. 3 as the defect density versus annealing temperature. Figure 3(a) is for the data set of samples grown at 600 °C and Fig. 3(b) is for the set of samples grown at 320 °C. The defect density can be calculated directly from the change in capacitance. Because the observed defects are the electron traps in a $p^+ - n$ diode, the equation used to calculate the defect density is

$$N = \frac{2\Delta C}{C_J} (N_D - N_A),$$

where $N$ is the trap concentration, $\Delta C$ is the maximum capacitance change in the DLTS spectra, $C_J$ is the junction capacitance under quiescent reverse-biased conditions, and $N_D - N_A$ is the net donor concentration on the $n$-side of the junction, where the trap is observed. $C_J$ and $N_D - N_A$ were obtained from prior $C-V$ measurements.

From the $C-V$ measurements, the carrier concentration in 320 °C as-grown sample is about 1.5 times higher than that in 600 °C as-grown sample. Based upon these values of $C_J$ and $N_D$, the E11 peak is approximately two times larger than E1. It should be noted that the measured trap density is only a lower boundary and represents a minimum density, because the trap levels may not be completely depopulated or filled by the emission and capture of carriers during the DLTS measurement.

Although overall defect densities decrease as the annealing temperatures increase in both groups of samples, the samples grown at 320 °C show that up to 700 °C, the rate of decrease for the dominant defect, E11, is small, and it is then accelerated for elevated annealing temperatures of 800 and 900 °C. The density of defects represented by E3 in the samples grown at 600 °C [Fig. 2(a)], slightly increased at 700 °C and then decreased at the upper temperatures. Most defects in both samples were annihilated by annealing at 900 °C, meaning their density was at or below the detection limit.

IV. DISCUSSION

A. Peak identification and suggested mechanism of defect formation

The dominant defects in the LT-MBE-grown samples are expected to be of intrinsic and/or dopant-related origin, because (1) the impurity concentration in the MBE-grown layers is low and (2) the defects found were annealed out without generating any other defects of comparable density. From the authors’ previous study of LT-MBE-grown Si, which compared results with defects in Si grown by chemical vapor deposition, the possible origin of the dominant defect, E1 and E11, was suggested as an E center ($V-P$) plus a singly negatively charged divacancy, $V-V(0/-)$, complex. It is well known that the $V-V(0/-)$ is located at $E_C = 0.42$ eV in Si$^{22,23}$ and strongly overlaps with the E center. Another commonly observed defect in Si$^{22,23}$ is a doubly negatively charged divacancy, $V-V(-2/-)$, at $E_C = 0.23$ eV. It corresponds to E2 and E22 including E3 and E33 in this study, and those are expected to be complexes combining this doubly negatively charged divacancy ($V-V(-2/-)$) with other defects. The activation enthalpies of E1 and E2 are $E_C = 0.48$ eV and $E_C = 0.27$ eV, respectively.

It should be noted that, unlike other studies in which the as-grown samples experienced proton irradiation or ion implantation in order to intentionally induce defects, the samples studied here were found to naturally form from point defects whose possible origin is similar to that seen in previous studies. It was also found that two divacancy-related defects can be formed, regardless of which dopants were used.

In general, the mechanism for an E-center creation is explained by the fact that doping with P results in a contraction of the silicon lattice due to the different atomic radii of P in a Si matrix. The mechanism for the creation of divacancy defects assigned to the E1, E11, and E2, E22 traps
observed in the MBE as-grown samples is suggested subsequently. In Si implanted with heavy ions, the creation of $V-V(0/-)$ and $V-V(-2/-)$ defects is explained by the presence of a local lattice distortion and strain in the damaged peak region. The local distortion and strain formed during ion implantation create enough unstable high energy to initiate Jahn–Teller distortion for divacancy creation. In the MBE-grown layers studied, the local lattice distortion can be formed during MBE growth. Since MBE is a nonequilibrium growth technique at $600^\circ$C and even more so at $320^\circ$C, the Si adatom surface mobility is kinetically limited and creates vacancies and interstitials during growth. The lower the growth temperature, the more plentiful and potentially complex the defects may be, especially if the defects are located at shallow levels close to the band edge. However, it should be noted that doubly and singly negatively charged divacancies, themselves, are not strongly affected by dopants, although dopants can cause defect complexes by combining with existing divacancy defects.

Monakov et al. found similar peaks in SiGe samples, and suggested the possible origin of the peaks, $E_{xx}$, $E_{yy}$, and $E_x$, shown here, as comprised of the dominant $E$ center plus $V-V(0/-)$ due to (1) the strain by the Ge atoms in the Si matrix or (2) the transformed $V-V(0/-)$ or complex difference from both $V-V(0/-)$ and $V-P$ pair. Since in this study, the peaks $E_{xx}$ and $E_{yy}$, including $E_x$, were discovered in samples without any Ge, this suggests that (2) is a more likely reason.

B. The effect of the defect annihilation on Si-based RITD performance

As mentioned earlier, the annealing temperatures were chosen in the range where Si-based RITDs have shown optimum performance. A brief description of the working mechanism of these RITDs, conceptually combines the operating principles of resonant tunneling of resonant tunnel diodes (RTDs) and band-to-band tunneling of tunnel diodes. Si-based RITDs have been realized using $\delta$-doping and very thin intrinsic tunnel barriers. The $\delta$-doped regions define QWs in the conduction and valence bands. Resonant tunneling takes place from the filled state in the conduction band QW to the empty state in the valence band QW. Between the two $\delta$-doped QWs is an intrinsic Si/SiGe composite spacer layer inserted in order to minimize dopant interdiffusion. LT-MBE of the $\delta$-doping layers and the central spacer is required to minimize dopant segregation and diffusion that could broaden the $\delta$-doping QWs and reduce carrier confinement.

If the tunneling region is free of traps, the $I-V$ characteristics consist of tunnel current and thermal diffusion current only. The resulting PVCR, which is defined as the ratio of the peak current density (PCD) to the valley current density (VCD) would always be determined by the PCD for an RITD since the valley current is fixed by the thermal diffusion current after the applied forward bias makes the respective QW energies uncross. However, the LT-MBE technique introduces some imperfections and trap centers. In LT-MBE-grown Si, as shown in the DLTS spectra in Fig. 2, there exist more than two dominant defects with energy levels around $E_C \sim 0.48$ for $E_1$ and $E_C \sim 0.27$ for $E_2$. Interestingly, the energy level of $E_1$ is very close to that of the Au impurity introduced in the excess current study of early Ge Esaki diodes in 1961. Sah suggested that the existence of deep levels leads to the excess current in which several carrier transition processes beyond band-to-band tunneling are involved. These processes consist of (1) Shockley–Read–Hall-type process including trap-to-trap, trap-to-band, and band-to-band and (2) tunneling process between trap states to band states. Since this excess current manifests as a low PVCR and an increase in static power dissipation for tunnel diode circuits, it is desirable to suppress the valley current as low as possible. The remedy adapted in Si-based RITDs is the addition of a post-growth annealing heat treatment to remove point defects.

The folded $I-V$ characteristics of a Si-based RITD with a 6 nm tunneling barrier layer between two $\delta$-doped layers are plotted with the annealing temperature as parameter.

![Image](https://example.com/fig4.png)

FIG. 4. $I-V$ characteristics of the RITD with 6 a nm tunneling barrier between two $\delta$-doped layers are plotted with the annealing temperature as parameter.

It is not unexpected that PCD decreases as the annealing temperature increases, because dopants localized in the $\delta$-doped layers diffuse which broadens the dopant profile. Consequently, the broadened $\delta$-doping profile results in lower tunnel current, as the confined states in the QW are lost. As discussed earlier, dopant segregation and interdiffusion are suppressed by the LT-MBE growth technique, but elevated thermal cycling can lead to similar undesirable results. On the contrary, VCD is highest at low annealing temperatures, because reduced annealing temperature are not as effective as higher annealing temperatures to remove point defects. A higher defect density causes larger excess currents...
due to the higher probability of tunneling through defects. This matches well with the previous defect density trends with respect to post-growth annealing temperature shown in Fig. 3. The plot in Fig. 3(b) is from the samples grown at 320 °C, which is the same growth temperature used for the RITD discussed here.

Although annealing at 900 °C for 60 s is shown to annihilate most point defects in the bulk layers studied here, as clearly shown in Fig. 3, it concurrently suppresses band-to-band tunneling currents in Si-based RITDs due to the loss of quantum confinement resulting in almost no NDR shown in Fig. 5. Therefore, an optimal annealing temperature for the best PVCR should be sought and lies between these two extremes. For an RITD with a 6 nm spacer, the highest PVCR is obtained when it was annealed at 800 °C. This implies that the annealing temperature up to 700 °C with a fixed duration of 60 s was not enough to effectively reduce defect densities.

V. CONCLUSION

DLTS measurements were performed on LT-MBE-grown bulk Si samples to determine the possible origin of the point defects and their influence by post-growth annealing heat treatments. The results are discussed in the context of the performance of Si-based RITDs which were grown under similar conditions at the same growth temperature, by the same MBE system, and at similar post-growth annealing temperatures. The samples used for DLTS measurement are $p^+ - n$ diodes with thick Si:P layers used as the lightly doped $n$ layers that are probed directly by the DLTS technique. Two samples were grown with the $n$ layer at 320 and 600 °C, respectively, and pieces of each were further annealed at 700, 800, and 900 °C, respectively.

From the DLTS measurements and thermal emission studies, two dominant complex vacancy-related defects, an E-center plus a singly negatively charged divacancy $V^- - V (0/-)$ and doubly $V^- - V (2/-)$ negatively charged divacancy and others were observed in the as-grown samples without intentionally introducing defects and/or impurities. After the samples were annealed, the defect densities gradually decreased and most of the defects were reduced to the detection limit at 900 °C for 60 s.

When Si-based RITDs were annealed at various annealing temperatures in the range of 700 to 900 °C, the valley current density, which is largely caused by defect-related tunneling through these defects in the tunnel barrier, was decreased, as was the peak tunneling current density. The best peak-to-valley current ratio was obtained with the annealing at 800 °C. When the RITDs are annealed at 900 °C, no PVCR is observed. There is an upper temperature limit imposed upon the annealing of RITD devices as dopant out-diffusion from the $\delta$-doped QWs deleteriously effects the QW confinement, levels of compensation within the intrinsic spacer between the QWs, and modifies the overall tunneling distance.

ACKNOWLEDGMENTS

The Ohio State University was supported by the National Science Foundation (ECS-0196208, ECS-0196054, and ECS-0323657). Support by the Center for Materials Research and the NSF contributed towards the participation of undergraduate researchers. The work at NRL was supported by the Office of Naval Research.