Solid State Electronics 151 (2019) 6–10

Contents lists available at ScienceDirect

Solid State Electronics

journal homepage: www.elsevier.com/locate/sse

Power reduction for recovery of a FinFET by electrothermal annealing

Joon-Kyu Han[1], Jun-Young Park[1], Yang-Kyu Choi*

School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea

ARTICLE INFO

Keywords:
Electrothermal annealing (ETA)
FinFET
Joule heat
Oxide damage
Reliability
Thermal conductivity

ABSTRACT

A strategy of reducing the power consumption to cure gate dielectric damage by electrothermal annealing (ETA) is proposed. A tri-gate FinFET was fabricated to demonstrate the damage curing by the ETA. Localized Joule heat induced by high current flowing through dual gate electrodes successfully annealed the damaged gate dielectric. Furthermore, a design methodology to save power consumption during the ETA was explored. Electrical measurements and simulations were performed considering scaling-down and material engineering points of view. This work contributes to improving the reliability of the FinFET by developing the ETA approach with reduced power consumption.

1. Introduction

The tri-gate FinFET has been widely used in logic circuits because of its enhanced gate-to-channel controllability and immunity against short-channel effects (SCEs) [1]. However, during iterative device operation, aging of the gate dielectric by hot-carrier injection (HCI) or Fowler-Nordheim (FN) stress is inevitable. Typically, the gate dielectric damage causes reliability issues such as degradation of subthreshold slope (*SS*), a shift in threshold voltage (V_T), and decrement of the ON-state current (I_{ON}) [2].

To cure the abovementioned gate dielectric damage, wafer scale annealing, such as forming gas annealing (FGA) and deuterium (D_2) annealing under hydrogen and deuterium gas ambient has been widely used [3,4]. However, such approaches require long curing times and bulky as well as expensive equipment, such as a furnace. Most importantly, the conventional wafer scale annealing cannot be applied to a packaged chip because it can lead to the melting and decomposition of the metal interconnections.

To overcome these limitations of conventional annealing, an electrothermal annealing (ETA) method has been suggested which uses Joule heat generated by the high current flowing in the device itself to cure the damaged gate dielectric [5]. Compared to conventional FGA or D_2 annealing, the ETA has a faster annealing speed (< 1 s), lower thermal budget, and excellent annealing selectivity.

However, despite these advantages, the ETA can be a concern in terms of power consumption because the fundamental mechanism of the process involves high current density through the conducting materials. For instance, in prior studies, 12 mW was used to cure the damaged gate dielectrics of flash memory and 3.2 mW was consumed to repair radiation damage [5,6]. From an energy efficiency point of view, the ETA power consumption should be reduced, however, few such attempt have been made.

In this work, we demonstrate a strategy of low power consumption ETA to anneal the damaged dielectric of a tri-gate FinFET, fabricated on a silicon-on-oxide (SOI) substrate. The results are characterized by electrical measurements. In addition, a supportive simulation study was performed with the aid of a 3-dimensional thermal simulator (COMSOL) to investigate the effect of design parameters such as device dimension and structural material, to construct a FET which further reduces the power consumption needed for the ETA [7]. First, the effect of representative geometric device dimensions such as gate length (L_G), fin width (W_{Fin}), oxide thickness (t_{OX}), buried oxide thickness (t_{BOX}), and substrate thickness (t_{Sub}) were analyzed. Second, the reduction in power consumption based on the underlying substrate material and gate material was examined.

2. Experimental details

2.1. Device fabrication

A tri-gate FinFET was fabricated on a p-type (1 0 0) SOI wafer, which contains a buried oxide (BOX) with a thickness of 400 nm. The nominal height (H_{Fin}) and width (W_{Fin}) of the silicon fin were 50 nm and 94 nm, respectively. The length of the sidewall spacer (L_{Spacer}) was 25 nm and gate length (L_G) was 80 nm. A thermal oxide (SiO_2) layer of 5 nm was grown as the gate oxide and 100 nm of n^+ poly-Si was

* Corresponding author.
E-mail address: ykchoi@ee.kaist.ac.kr (Y.-K. Choi).
[1] These authors equally contributed to this work.

https://doi.org/10.1016/j.sse.2018.10.008

Available online 19 October 2018

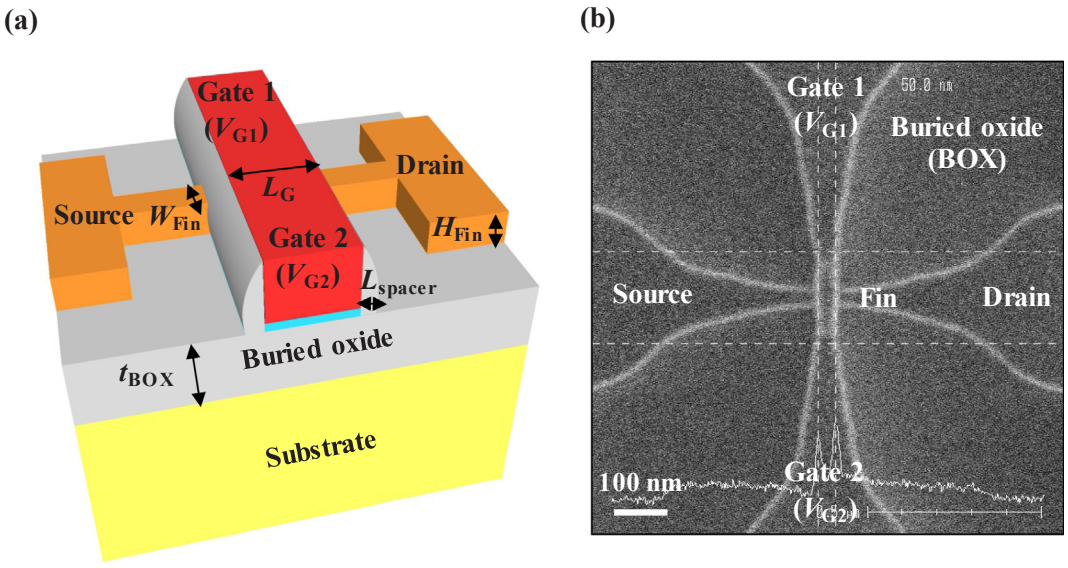

Fig. 1. (a) Schematic of the fabricated tri-gate FinFET on SOI substrate. (b) SEM image of the fabricated device.

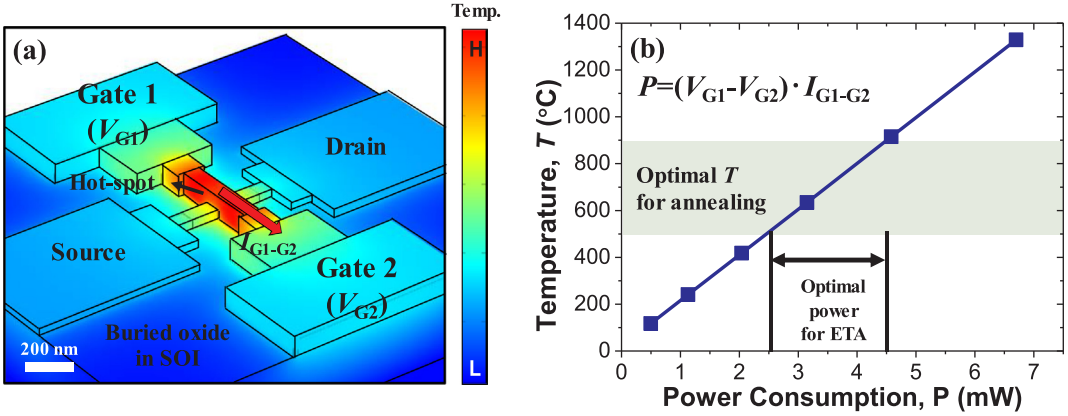

Fig. 2. (a) Simulated heat distribution profile in the poly-Si gate during the ETA. (b) Extracted temperature in Fig. 2(a) versus power consumption (P) during the ETA.

deposited as the gate electrode. A hard mask (SiO_2) of 30 nm was deposited on the poly-Si. Then, dual gate pads (Gate 1 for V_{G1} and Gate 2 for V_{G2}) with a dog bone shape, as shown in Fig. 2(a), were patterned to apply bias for the ETA. The detailed process flow was previously reported in [8]. Fig. 1 shows a schematic and a scanning electron microscope (SEM) image of the fabricated SOI tri-gate FinFET.

2.2. Electrical measurements

Electrical measurements of the initial state, damaged state, and repaired state were carried out with a semiconductor parameter analyzer (HP 4156C) under air ambient. HCI stress was intentionally applied to the fabricated device to provoke gate oxide damage. After that, the voltages for the ETA were applied to the dual gate pads with a pulse time (τ_{pulse}) of 100 ms. During the ETA process, the source voltage (V_S) and the drain voltage (V_D) were floated. The detailed bias conditions for the HCI and ETA are summarized in Table 1. It should be noted that an identical magnitude of voltage with the opposite sign was applied to each gate pad to prevent unwanted gate oxide breakdown during the ETA. For example, the voltage of gate 1 (V_{G1}) was the same as gate 2 (V_{G2}) with a minus sign, i.e., $V_{G1} = -V_{G2}$.

Table 1
Biasing conditions for the hci and repair.

	Hot-carrier injection (HCI)	Repair
V_{G1}	6 V	+3 V
V_{G2}	Floating	−3 V
V_S	GND	Floating
V_D	6 V	Floating
t_{pulse}	1000 s	100 ms

2.3. Device simulation

To estimate temperature in a nanoscale domain, a simulation study is necessary because the high temperature of the localized heat cannot be directly measured by equipment such as an infrared (IR) thermal detector, scanning thermal microscope (SthM), or thermal reflectance microscope (TRM) [9–11]. The IR thermal detector has a resolution of only a few micro-meters, which is not suitable for a small sized FinFET. Although SThM and TRM have much greater resolution, only relative heat distribution can be observed, and high temperature above 300 °C cannot be measured due to the limitations of the calibration method.

However, by employing the heat transfer module of the COMSOL simulator, we were able to obtain the temperature distribution profile when applying the measured current (I_{G1-G2}) through G1 and G2. As a

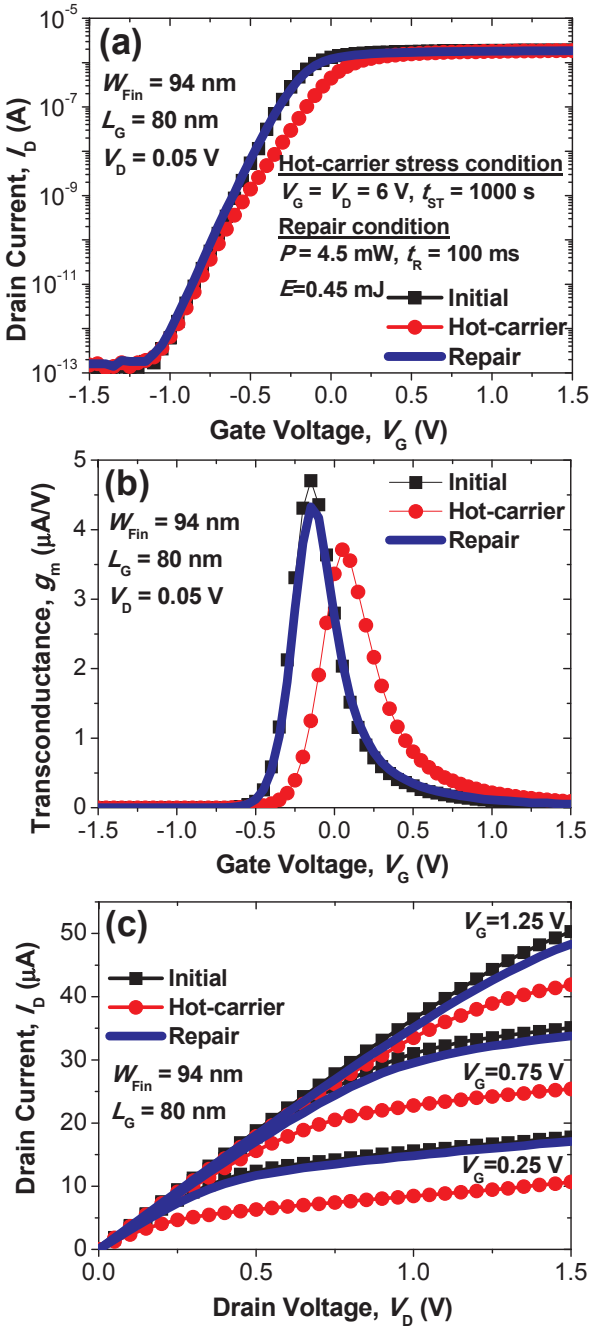

Fig. 3. (a) $I_D - V_G$ characteristics of the fabricated tri-gate FinFET for the fresh, hot-carrier degraded state, and repaired state. (b) Transconductance (g_m) of the fresh, hot-carrier degraded, and repaired state. (c) $I_D - V_D$ characteristics of the fresh, hot-carrier degraded, and repaired state.

reference, the simulated temperature was previously verified by observing the dopant activation characteristics of poly-Si under high temperature [12]. In the simulation, the thermal conductivity (κ) of Si, heavily doped poly-Si, SiO_2, HfO_2 and Ge were set to 150, 31.2, 1.25, 1.1 and 60 W/m K [13–15].

3. Results and discussion

Fig. 2(a) shows the heat distribution profile of the FinFET obtained from the simulation. The heat generated between the dual gate pads is localized at the center of the gate, that is, where the channel is located in a cross-bar structure, and this heat can be used to cure gate oxide damage. Fig. 2(b) indicates the extracted temperature (T) at the gate dielectric, which is linearly proportional to the power consumption (P). The P is simply calculated by multiplying the voltage across the dual gate pads times measured current through the gate ($P = (V_{G1} - V_{G2})$ I_{G1-G2}). The simulation result revealed that the appropriate range of P for the ETA is between 2.5 mW and 4.5 mW, because poly-Si breakdown can occur at a higher temperature [12].

Table 2
Electrothermal annealing effects on the soi tri-gate finfet.

	Incremental/decremental percentage of the changed device parameter compared with the initial state	
	Hot-carrier injection	Repair
g_m (μA/V)	−21%	−8%
SS (mV/dec)	−35%	+2%
I_{ON} (μA)	−11%	−5%

Fig. 3 shows the measured electrical characteristics of the FinFET. At P = 4.5 mW, which is the proper value from Fig. 2(b), the degraded device is successfully repaired and nearly returns to the initial state. Duration time of the ETA bias is 100 ms and the total energy consumption is 0.45 mJ. Extracted device parameters such as transconductance (g_m), subthreshold swing (SS) and ON-state current (I_{ON}) are summarized in Table 2. The mechanism of the damage curing is that pre-existing H^+ present in the gate oxide is diffused toward the interface and passivates the trap sites at high temperature by Joule heat. It should be noted that the degradation of I_{ON} due to the thermal oxidation was not found [16]. This is because the channel and the gate oxide were completely surrounded by n^+ poly-Si gate and gate spacers. Moreover, the ETA time of 100 ms is very short to grow unwanted additional gate oxide.

After electrical measurements, the design methodology was investigated by changing various geometric design parameters, including L_G, W_{Fin}, t_{OX}, t_{BOX}, and t_{Sub} in order to reduce the ETA power consumption. Fig. 4(a)–(c) shows the extracted temperature of the gate dielectric for various L_G, W_{Fin}, t_{OX}, t_{BOX}, and t_{Sub}. As the gate length (L_G) becomes shorter and the fin width (W_{Fin}) becomes narrower, the T is increased under the fixed P condition, as shown in Fig. 4(a). This is because heat dissipation into the atmosphere and substrate is reduced as L_G and/or W_{Fin} decreases. In addition, as the oxide thickness (t_{OX}) is reduced, the T becomes hotter. However, the level of temperature increase induced by the downscaled t_{OX} is negligible compared with that induced by the other downscaled design parameters in Fig. 4(a). These results indicate that as the device is miniaturized, the power consumption of the ETA can also be reduced. Fig. 4(c) shows the extracted T according to the thickness of the BOX (t_{BOX}) and silicon substrate (t_{Sub}) at the fixed P. As the t_{BOX} becomes thicker, the T becomes hotter due to the increased thermal isolation of the BOX layer. In contrast to t_{BOX}, t_{Sub} does not significantly affect the temperature because the thermal conductivity of Si is sufficiently high to dissipate the Joule heat during the ETA. As devices become more integrated, t_{Sub} decreases, due to the development of chip assembly technology such as through-silicon via (TSV) or wafer level packaging. However, the simulation confirmed that high power efficiency can be obtained regardless of the reduction in t_{Sub}. As a result, the important design parameters for reducing the P of the ETA are L_G, W_{Fin}, and t_{BOX}. The P can be scaled by 35% by reducing L_G, W_{Fin}, and t_{OX} in half, as shown in Fig. 4(d). These results indicate the proposed ETA is still a promising technology for future processes such as device down-scaling from the front-end process point of view and wafer back-grinding from the back-end process point of view.

In Fig. 5, the level of heat dissipation via the substrate is simulated for representative substrates used for mass production: SiO_2 for SOI (κ = 1.25 W/m K), Ge for strained technology (κ = 60 W/m K), and bulk-Si (κ = 130 W/m K). For a fair simulation, all of the geometric

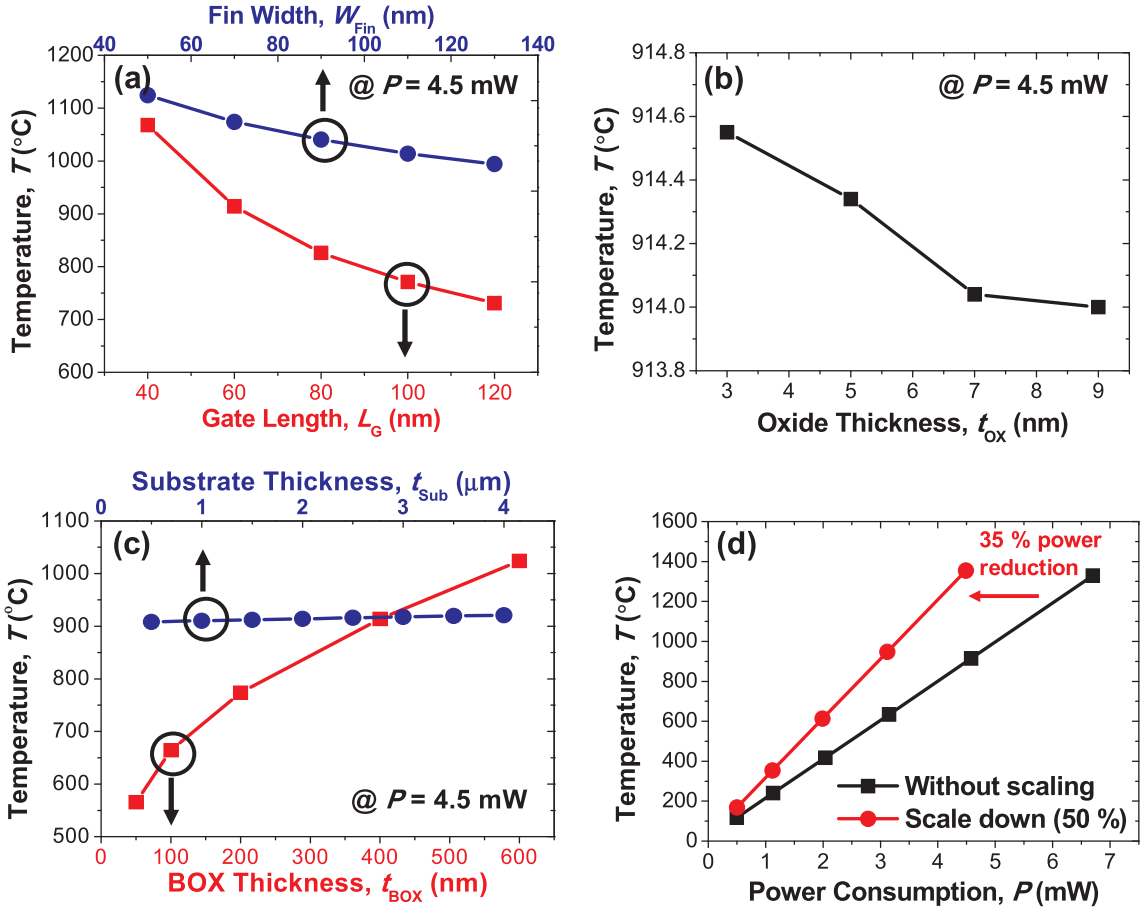

Fig. 4. (a) Extracted temperature versus L_G and W_{Fin} at $P = 4.5$ mW. (b) Extracted temperature versus t_{OX}. (c) Extracted temperature versus t_{BOX} and t_{Sub} at $P = 4.5$ mW. P can be reduced by increasing the BOX thickness. (d) Extracted temperature versus P before scaling ($L_G = 80$ nm, $W_{Fin} = 90$ nm, $t_{OX} = 5$ nm) and after scaling ($L_G = 40$ nm, $W_{Fin} = 45$ nm, $t_{OX} = 2.5$ nm). P can be reduced by scaling down the device.

design parameters and power consumption were fixed. As the thermal conductivity (κ) of the substrate is lowered, the heat generated from the gate electrode becomes less likely to dissipate into the substrate. Hence, the T rises as κ is decreased, as shown in Fig. 5(a). Consequently, a SOI device is advantageous in terms of the ETA as well as its traditionally reported advantages against short-channel effects (SCEs) with the aid of gate-to-channel controllability. In other words, a small magnitude of κ underneath the channel is preferred in order to reduce the power consumption in the ETA. Even though the SOI technology is not commonly used for mass production by industries, the proposed methodology is still applicable to a bulk-substrate device with device downsizing.

In order to improve mobility, strained technology, primarily using Ge as a stressor, has been widely used. Knowing whether the ETA power consumption can be reduced or not can be a concern in strained technology. For this reason, the temperature distribution, which depends on the ratio of Ge contents in $Si_{1-x}Ge_x$, was simulated, as shown in Fig. 5(b). The physical constants for $Si_{1-x}Ge_x$ such as electrical conductivity and thermal conductivity were adopted from the published book [17]. Unlike the improvement in mobility, the thermal conductivity of $Si_{1-x}Ge_x$ decreased and saturated as the Ge ratio was increased. This characteristic is caused by the alloy disorder scattering of

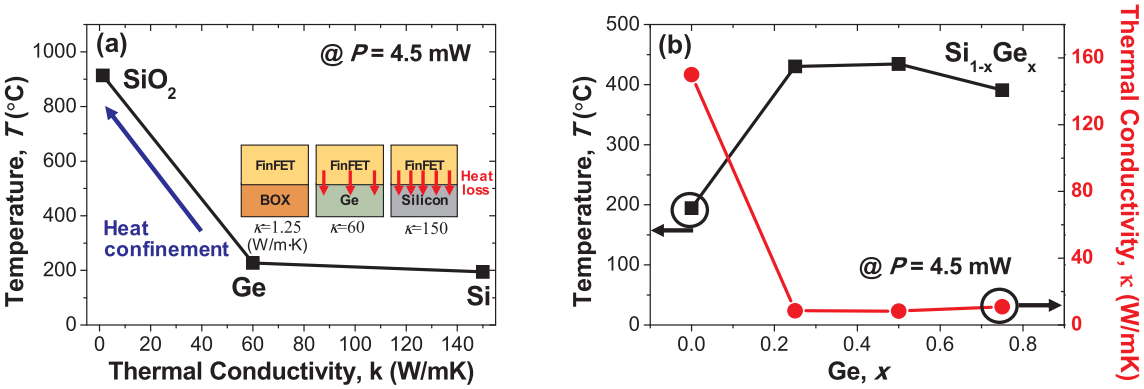

Fig. 5. (a) Induced temperature versus thermal conductivity of the substrate. A lower P can be obtained on the substrate which has lower thermal conductivity. (b) Induced temperature versus Ge portion in the $Si_{1-x}Ge_x$ substrate. The device on the $Si_{1-x}Ge_x$ substrate has a lower P than on the Si substrate.

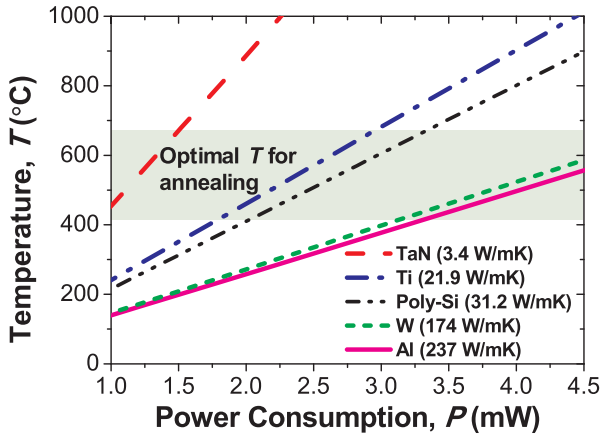

Fig. 6. Induced temperature versus P for various gate materials such as (TaN, Ti, n^+ poly-Si, W, and Al). The P can be further reduced by employing a lower κ material as a gate.

phonons due to the large mass difference between Si and Ge [18]. Thus, the decreased thermal conductivity of the $Si_{1-x}Ge_x$ substrate reduces thermal dissipation during the ETA, which favorably results in increased temperature at the same power consumption.

A high-k metal gate (HKMG) structure composed of a high-k gate dielectric and metal gate is widely used in logic devices to increase gate capacitance. Typically, a metal gate requires a smaller repair voltage ($V_{G1} - V_{G2}$) for the ETA than a poly-Si gate owing to the high electrical conductivity of the metal gate [19]. However, a study of the power consumption during the ETA has not been carried out in detail. Fig. 6 shows the extracted temperature with various gate materials according to the power consumption. For the simulations, HfO_2 (κ = 1.1 W/m K) was used as the high-k gate dielectric. TaN (κ = 3.4 W/m K), Ti (κ = 21.9 W/m K), W (κ = 174 W/m K), and Al (κ = 237 W/m K) were used for the metal gate. The results indicate that the larger P is required for the W gate and Al gate compared to the poly-Si gate, while the smaller P is needed for TaN gate and Ti gate. The reason is that the gate with a high κ suffers more from heat loss due to convection cooling than the gate with a low κ. Hence, to reduce the power consumption during the ETA, it is inferred that the low-κ gate material is preferred.

4. Conclusion

A tri-gate FinFET was fabricated and its oxide damage was cured nearly to the pristine state by the ETA. The power consumption of the ETA process was reduced by scaling-down the device dimensions. A 50% reduction in L_G, W_{Fin}, and t_{OX} reduced the ETA power consumption down to 35%. Thus, the reduction of power consumption is scalable. From a material standpoint, the low thermal conductivity of the substrate and the gate material was advantageous to reducing the power consumption of the ETA. The proposed strategy to reduce the ETA power consumption could be very useful with advanced fabrication technology such as down-scaling, thinned back-grinding of a wafer prior to packaging, and new material based engineering including strained technology and high-k metal gate process.

Acknowledgements

This work was supported by the Center for Integrated Smart Sensors funded by the Ministry of Science and ICT as Global Frontier Project (CISS-2011-0031848) in part. This work also partially sponsored by the Global Ph. D. Fellowship Program (2017H1A2A1042274) through the National Research Foundation (NRF) of Korea.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.sse.2018.10.008.

References

[1] Hisamoto D, Lee WC, Kedzierski J, Takeuchi H, Asano K, Kuo C, et al. FinFET—A self-aligned double-gate MOSFET scalable to 20 nm. IEEE Trans Electron Devices 2000;47:2320–5. https://doi.org/10.1109/16.887014.
[2] Hu C, Tam SC, Hsu FC, Ko PK, Chan TY, Terrill KW. Hot-electron-induced MOSFET degradation-model, monitor, and improvement. IEEE J Solid-State Circuits 1985;20:295–305. https://doi.org/10.1109/JSSC.1985.1052306.
[3] Onishi K, Kang Chang Seok, Choi Rino, Cho Hag-Ju, Gopalan S, Nieh RE, et al. Improvement of surface carrier mobility of HfO2/MOSFETs by high-temperature forming gas annealing. IEEE Trans Electron Devices 2003;50:384–90. https://doi.org/10.1109/TED.2002.807447.
[4] Lyding JW, Hess K, Kizilyalli IC. Reduction of hot electron degradation in metal oxide semiconductor transistors by deuterium processing. Appl Phys Lett 1996;68:2526–8. https://doi.org/10.1063/1.116172.
[5] Lue HT, Du PY, Chen CP, Chen WC, Hsieh CC, Hsiao YH, et al. Radically extending the cycling endurance of Flash memory (to > 100M Cycles) by using built-in thermal annealing to self-heal the stress-induced damage. Tech Dig – Int Electron Devices Meet IEDM 2012:199–202. https://doi.org/10.1109/IEDM.2012.6479008.
[6] Moon DI, Park JY, Han JW, Jeon GJ, Kim JY, Moon J, et al. Sustainable electronics for nano-spacecraft in deep space missions. Tech Dig – Int Electron Devices Meet IEDM 2016:794–7. https://doi.org/10.1109/IEDM.2016.7838524.
[7] Comsol multiphysics user's guide, Comsol Inc., Burlington, MA, USA; 2011.
[8] Moon DI, Kim JY, Jang H, Hong HJ, Kim CK, Oh JS, et al. A novel FinFET with high-speed and prolonged retention for dynamic memory. IEEE Electron Device Lett 2014;35:1236–8. https://doi.org/10.1109/LED.2014.2365235.
[9] Urakawa S, Tomai S, Ueoka Y, Yamazaki H, Kasami M, Yano K, et al. Thermal analysis of amorphous oxide thin-film transistor degraded by combination of joule heating and hot carrier effect. Appl Phys Lett 2013;102. https://doi.org/10.1063/1.4790619.
[10] Fiege GBM, Feige V, Phang JCH, Maywald M, Gorlich S, Balk LJ. Failure analysis of integrated devices by scanning thermal microscopy (SThM). Microelectron Reliab 1998;38:957–61. https://doi.org/10.1016/S0026-2714(98)00086-9.
[11] Ryu SY, Kim DU, Kim JK, Choi HY, Kim GH, Chang KS. Surface-temperature measurement and submicron defect isolation for microelectronic devices using thermoreflectance microscopy. Int J Thermophys 2015;36:1217–25. https://doi.org/10.1007/s10765-014-1681-6.
[12] Park JY, Moon DI, Seol ML, Jeon CH, Jeon GJ, Han JW, et al. Controllable electrical and physical breakdown of poly-crystalline silicon nanowires by thermally assisted electromigration. Sci Rep 2016;6:1–7. https://doi.org/10.1038/srep19314.
[13] Ju YS, Goodson KE. Phonon scattering in silicon films with thickness of order 100 nm. Appl Phys Lett 1999;74:3005–7. https://doi.org/10.1063/1.123994.
[14] Tai YC, Mastrangelo CH, Muller RS. Thermal conductivity of heavily doped low-pressure chemical vapor deposited polycrystalline silicon films. J Appl Phys 1988;63:1442–7. https://doi.org/10.1063/1.339924.
[15] Maycock PD. Thermal conductivity of silicon, germanium, Iii-V xompounds and Iii-V alloys. Solid-State Electron Pergamon Press 1967;10:161–8. https://doi.org/10.1016/0038-1101(67)90069.
[16] Choi SJ, Moon DI, Duarte JP, Ahn JH, Choi YK. Physical observation of a thermomorphic transition in a silicon nanowire. ACS Nano 2012;6:2378–84. https://doi.org/10.1021/nn2046295.
[17] Kasper E. Properties of strained and relaxed silicon germanium. London: INSPEC; 1995.
[18] Bhandari CM, David MR. CRC handbook of thermoelectrics. CRC Press; 1995.
[19] Park JY, Moon DI, Bae H, Roh YT, Seol ML, Lee BH, et al. Local electro-thermal annealing for repair of total ionizing dose-induced damage in gate-all-around MOSFETs. IEEE Electron Device Lett 2016;37:843–6. https://doi.org/10.1109/LED.2016.2574341.

Yang-Kyu Choi received the B.S. and M.S. degrees from Seoul National University, Seoul, South Korea, in 1989 and 1991, respectively, and the Ph.D. degree from the University of California at Berkeley, Berkeley, CA, USA, in 2001. He is currently a professor with the School of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon, South Korea.