Transient Thermal and Electrical Co-Optimization of BEOL Top-Gated ALD In$_2$O$_3$ FETs Toward Monolithic 3-D Integration

Pai-Ying Liao, Dongqi Zheng, Sami Alajlouni, Zhuocheng Zhang, Mengwei Si, Member, IEEE, Jie Zhang, Member, IEEE, Jian-Yu Lin, Tatyana I. Feygelson, Marko J. Tadjar, Senior Member, IEEE, Ali Shakouri, and Peide D. Ye, Fellow, IEEE

Abstract—In this work, the transient thermal and electrical characteristics of top-gated (TG), ultrathin, atomic-layer-deposited (ALD), back-end-of-line (BEOL) compatible indium oxide (In$_2$O$_3$) transistors on various thermally conductive substrates are co-optimized by visualization of the self-heating effect (SHE) utilizing an ultrafast high-resolution (HR) thermo-reflectance (TR) imaging system and overcome the thermal challenges through substrate thermal management and short-pulse measurement. At the steady-state, the temperature increase ($\Delta T$) of the devices on highly resistive silicon (HR Si) and diamond substrates are roughly 6 and 13 times lower than that on a SiO$_2$/Si substrate, due to the much higher thermal conductivities ($\kappa$) of HR Si and diamond. Consequently, the ultrahigh drain current ($I_D$) of 3.7 mA/$\mu$m at drain voltage ($V_D$) of 1.4 V with direct current (dc) measurement is achieved with TG ALD In$_2$O$_3$ FETs on a diamond substrate. Furthermore, transient thermal study shows that it takes roughly 350 and 300 ns for the devices to heat-up and cool-down to the steady-states, being independent of the substrate. The extracted thermal time constants of heat-up ($\tau_h$) and cool-down ($\tau_c$) processes are 137 and 109 ns, respectively. By employing electrical short-pulse measurement with a pulsewidth ($t_{\text{pulse}}$) shorter than $\tau_h$, the SHE can be significantly reduced. Accordingly, a higher $I_D$ of 4.3 mA/$\mu$m is realized with a 1.9-nm-thick In$_2$O$_3$ FET on HR Si substrate after co-optimization. Besides, to integrate BEOL-compatible ALD In$_2$O$_3$ transistors on the front-end-of-line (FEOL) devices with the maintenance of the satisfactory heat dissipation capability, a FEOL-interlayer-BEOL structure is proposed where the interlayer not only electrically isolates the FEOL and BEOL devices but also serves as a thermally conductive layer to alleviate the SHE.

Index Terms—Atomic layer deposition, indium oxide (In$_2$O$_3$), self-heating effect (SHE), thermal engineering, thermo-reflectance (TR).

I. INTRODUCTION

D
epite the wide application in the display industry [1], oxide semiconductors attract more interest recently as back-end-of-line (BEOL)-compatible channel materials for thin-film transistor (TFT) applications even toward monolithic 3-D (M3D) integration [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17]. Among them, atomic-layer-deposited (ALD) indium oxide (In$_2$O$_3$) is of great interest due to its outstanding properties such as wafer-scale uniformity, high carrier mobility, BEOL compatibility, ambient stability, and capability of conducting ultrahigh current. Especially, the ALD growth method provides not only the angstrom-scale thickness control but also the atomically smooth surface conformality on 3-D surface structures such as deep trenches and side walls [2], [3], [4], [5], [6], [7], [8]. It has been demonstrated that back-gated (BG) ALD In$_2$O$_3$ transistors can be realized with an ultra-thin body of 0.5 nm, which is only a couple of atoms thick [6]. Besides, a high drain current ($I_D$) up to 2.2 mA/$\mu$m at drain bias ($V_D$) of 0.7 V is reached by 1.5-nm-thick BG In$_2$O$_3$ devices operated at enhancement mode (E-mode) [14].

Nevertheless, even though top-gated (TG) devices are particularly desired for practical applications, the explorations of In$_2$O$_3$ FETs mostly focus on BG structure. This is mainly due to the two challenges of defect induction during the formation of the high-$\kappa$ TG dielectric and severe self-heating...
effect (SHE) with high power density (PD) [7], [8]. The root cause of the former issue is that the formation of the high-\(k\) dielectric stack, ALD HfO\(_2\), likely fetches oxygen not only from the oxygen precursor but also the In\(_2\)O\(_3\) channel underneath, which induces oxygen vacancies at the channel. Fortunately, this can be partly resolved by lowering the growth temperature of ALD HfO\(_2\) from 200 °C down to 120 °C, followed by rapid thermal annealing (RTA) treatment in the O\(_2\) environment [5], [8]. On the other hand, the latter challenge remains a bottleneck of TG ALD In\(_2\)O\(_3\) devices.

In this work, an ultrafast high-resolution (HR) thermo-reflectance (TR) imaging system is introduced to visualize the transient and steady-state characteristic \(\Delta T\) of TG In\(_2\)O\(_3\) devices on different substrates to address the thermal issues. The substrate substitution method [7] is first applied to alleviate SHE. At the steady states, the SHE is mitigated by a factor of 6 and 13 with HR Si and diamond substrate, respectively. Compared to SiO\(_2\)/Si substrate, the SHE happens due to low-\(k\) of SiO\(_2\) with high power density (PD) [7], [8]. The substrate substitution method [7] is first applied to understand the transient thermodynamics of the self-heating process of TG ALD In\(_2\)O\(_3\) transistors, transient TR measurement is performed. The extracted time constants \(\tau_h\) and \(\tau_c\) are roughly 137 and 109 ns, respectively. By resolving SHE, an ultrahigh \(I_D\) of 3.7 mA/\(\mu\)m is realized with 2.5-nm-thick In\(_2\)O\(_3\) devices on the diamond substrate under direct current (dc) measurement. Furthermore, by pulse measurement with pulse widths shorter than \(\tau_h\), even higher \(I_D\) of 4.3 mA/\(\mu\)m is achieved with 1.9-nm-thick In\(_2\)O\(_3\) transistors on HR Si substrate under pulse measurement. Besides, BEOL-compatible TG ALD In\(_2\)O\(_3\) transistors are built on top of front-end-of-line (FEOL) Si devices with a thick plasma-enhanced chemical-vapor-deposited (PECVD) silicon nitride (Si\(_3\)N\(_4\)) in between, showing potentials of combining FEOL devices with BEOL devices together toward M3D integration. This article is an extended version of an accepted 2022 IEEE International Electron Devices Meeting (IEDM 2022) paper with more details and FEOL-BEOL integration included.

**II. Device Fabrication and Performance**

Fig. 1(a) exhibits the schematic of TG ALD In\(_2\)O\(_3\) transistors. Different \(\kappa\)-value substrates including SiO\(_2\)/Si, sapphire, HR Si, and diamond from 1.5 to 2200 W m\(^{-1}\) K\(^{-1}\), are used. After solvent cleaning, a 1.6–2.5 nm In\(_2\)O\(_3\) layer was conformally grown by ALD at 225 °C with trimethylindium (TMIn) and H\(_2\)O as the In and O precursors, respectively, followed by an Ar/BCl\(_3\) dry-etching step for device isolation. Then, 45 nm Ni was deposited as source/drain (S/D) contacts by e-beam lithography (EBL), e-beam evaporation, and a lift-off step. Next, a 7 nm HfO\(_2\) top dielectric layer was formed by ALD at 120 °C with [(CH\(_3\))\(_3\)]\(_2\)Hf (TDMAHf) and H\(_2\)O as the Hf and O precursors, respectively. Finally, the top-gate metal of 30/20 nm Au/Ni was deposited with the same process as the S/D metal, followed by an RTA treatment at 250 °C–300 °C in an O\(_2\) environment for 2–4 min. The overall thermal budget is as low as 300 °C, making it BEOL-compatible.

Fig. 1(b) and (c) presents the electrical characteristics of a TG ALD In\(_2\)O\(_3\) FET with a channel length \(L_{ch}\) of 600 nm and a thin channel thickness \(T_{ch}\) of 1.6 nm on a SiO\(_2\)/Si substrate operated at enhancement-mode (E-mode). ON-OFF ratio of 10 orders of magnitude and SS value of 150 mV/dec are performed. (d) Output characteristics of a TG ALD In\(_2\)O\(_3\) FET with short \(L_{ch}\) of 80 nm on a SiO\(_2\)/Si substrate. (e) Severe SHE deteriorates the device performance of a TG ALD In\(_2\)O\(_3\) FET with high PD.

**III. Steady-State Characterization**

The ultrafast HR TR imaging equipment employed to systematically investigate and resolve the SHE is illustrated in Fig. 2(a). During the measurement, the device under test is applied by periodic \(V_{DS}\) pulses and a constant \(V_{GS}\) bias, where the source is common, to induce the SHE. A high-speed green light LED (wavelength of 530 nm) is equipped to illuminate the device surface for the synchronized charge-coupled device (CCD) camera to capture the surface reflectance. Fig. 2(b) demonstrates its working mechanism. As \(V_{DS}\) pulses start,
the device is turned on and accordingly self-heats up, and TR signals are captured as an active image after the steady-state is reached as shown in red. Similarly, when \( V_{DS} \) pulses end, the device is turned off and hence naturally cools down to ambient temperature, and TR signals are captured as a passive image after the steady-state is reached as shown in blue. This process is repeated numerous times, and the difference between the active and passive images is averaged in time domain. (c) Transformation from TR signal to a temperature scale.

Fig. 3 presents the observed \( \Delta T \) distribution of the TG In\(_2\)O\(_3\) transistors around the channel region. All the devices under test have identical structures and dimensions of 400 nm \( L_{ch} \) and 2 \( \mu m \) \( W_{ch} \), and the only difference is the substrate in use. Clearly, the one on a SiO\(_2\)/Si substrate is the most self-heated while the one on a diamond substrate is the least.

The cross sections of the \( \Delta T \) along the channel width (\( W_{ch} \)) direction normalized by PD is demonstrated in Fig. 4(a) where the PD is calculated by \( (I_D \times V_{DS})/(L_{ch} \times W_{ch}) \). Similar bell-like shapes are observed in all the cases. The maximum \( \Delta T \) of the devices with individual substrates and different PD is plotted in Fig. 4(b) where a linear relationship is obtained despite the substrate. The inversed slopes of the regression lines indicate the capability of the substrate to dissipate the generated Joule heat in the channel. To compare the thermal dissipation capabilities of the substrates, the \( \Delta T \) is extrapolated to a constant PD of 16 kW/mm\(^2\) in each case as shown in Fig. 4(c). Fig. 4(c) implies that the higher \( \kappa \) of the substrate is, the lower the maximum normalized \( \Delta T \) will be. Sapphire (\( \kappa = 40 \text{ W} \cdot \text{m}^{-1} \cdot \text{K}^{-1} \)) [21]), HR Si (\( \kappa = 140 \text{ W} \cdot \text{m}^{-1} \cdot \text{K}^{-1} \)) [22]), and diamond (\( \kappa = 2200 \text{ W} \cdot \text{m}^{-1} \cdot \text{K}^{-1} \)) [23]) substrates have \( \Delta T \) reduction by factors of 2.5, 6, and 13, respectively, compared with SiO\(_2\)/Si substrate with thick SiO\(_2\) (\( \kappa = 1.5 \text{ W} \cdot \text{m}^{-1} \cdot \text{K}^{-1} \)) [24]).

Motivated by the thermal studies, an In\(_2\)O\(_3\) transistor with \( T_{ch} \) of 2.5 nm and \( L_{ch} \) of 100 nm is built up on a diamond substrate to alleviate SHE. Fig. 5 presents its dc output and transfer characteristics where an ultrahigh \( I_D \) of 3.7 mA/\( \mu \text{m} \) is realized at \( V_{DS} \) of 1.4 V without observable SHE even with high PD, in great contrast to Fig. 1(e). The extracted field-effect mobility (\( \mu_{FE} \)) and SS are 55.6 cm\(^2\) \( \text{V}^{-1} \cdot \text{s}^{-1} \) and 185 mV/dec, respectively. Note that the degradation of the transfer characteristics from Figs. 1(b) to 5(b) is due to the much thicker \( T_{ch} \) as the electrical properties such as \( V_I \),
ON–OFF ratio, and SS value of ALD In$_2$O$_3$ devices are sensitively $T_{\text{ch}}$-dependent [2], [3], [4], [5], [6], [7].

IV. TRANSIENT CHARACTERIZATION

Although diamond with ultrahigh $\kappa$ is an excellent heat sinker, it is not economically affordable for mass production. TG transistors on Si substrate are much more practically desired. To further mitigate the SHE on a Si substrate, detailed transient thermal characteristics are studied. Fig. 6(a) illustrates the working mechanism for transient thermal characteristics. Similar to steady-state measurements, an active and a passive image are captured by the synchronized CCD camera in each period. The difference is that steady-state investigations only take active images after the steady-state is reached, while transient-state explorations take them throughout the whole process. By setting the active image to be captured at a specific time, the $1T$ distribution of that moment can be obtained through the same process shown in Fig. 2(c).

To experimentally observe the time-resolved self-heat-up and cool-down processes, the $V_{DS}$ pulses are set to start at 0 ns and end at 600 ns. Besides, the image capture time window is decreased to 50 ns, which is the minimum of the equipment capability and will be the resolution of the time-resolved measurement. From $t = 0$–1200 ns, TR images are taken every 50 ns. The heat-up process is illustrated in Fig. 6(b)–6(e) where the $V_{DS}$ pulses begin at $t = 0$ ns. The $\Delta T$ distribution around the channel region gradually increases after then due to SHE, and the transient process is observed. Similarly, the $V_{DS}$ pulses end at $t = 600$ ns, and the $\Delta T$ peak introduced by SHE gradually decreases as shown in Fig. 6(f)–6(i). The maximum $\Delta T$ of each moment of the devices with SiO$_2$/Si and HR Si substrates are summarized in Fig. 7(a). Two devices with the same structure and dimensions of 400 nm $L_{\text{ch}}$ and 2 $\mu$m $W_{\text{ch}}$ but different substrates are measured. It takes roughly 350 and 300 ns to reach steady-states for heat-up and cool-down, no matter whether SiO$_2$/Si or HR Si serves as the substrate.

By assuming the processes satisfy Newton’s law of cooling [25] (the rate of heat transfer is proportional to the temperature difference between the body and its surroundings) and utilizing their individual initial conditions, the following equations can be derived [26]:

$$\Delta T(t) = \Delta T_{\text{st}} \times (1 - e^{-t/\tau})$$

for heating-up and

$$\Delta T(t) = \Delta T_{\text{st}} \times e^{-t/\tau}$$

for cooling down where $\Delta T(t)$ represents the $\Delta T$ at time $t$, $\Delta T_{\text{st}}$ represents the steady-state $\Delta T$, and $\tau$ represents the thermal time constant. Accordingly, the following equations are obtained:

$$\frac{t}{\tau_h} = -\ln\left(1 - \frac{\Delta T(t)}{\Delta T_{\text{st}}}\right)$$

Fig. 5. (a) Output and (b) transfer characteristics of an In$_2$O$_3$ FET with $L_{\text{ch}}$ of 100 nm and $T_{\text{ch}}$ of 2.5 nm on a diamond substrate, performing ultrahigh $I_D$ of 3.7 mA/$\mu$m under dc condition.

Fig. 6. (a) Working mechanism of transient thermal property study with an ultrafast HR TR imaging system. (b)–(i) Transient TR characteristics of a TG ALD In$_2$O$_3$ FET on HR Si substrate. The device was self-heated-up by applying a $V_{DS}$ pulse starting at 0 ns and ending at 600 ns. (b)–(e) Demonstrate the heat-up process while (f)–(i) demonstrate the cool-down process.
Fig. 7. (a) Transient $\Delta T$ results and comparison with different substrates. Roughly 325 ns is needed to reach steady-state in both cases. Time constant extraction of (b) heat-up and (c) cool-down processes from the transient $\Delta T$ measurement results with HR Si substrate of Fig. 7(a). The $t/\tau$ values are obtained by plugging the measured individual $\Delta T(t)$ values into the indicated formulas.

for heating-up and

$$ t = -\ln \left( \frac{\Delta T(t)}{\Delta T_{st}} \right) $$

for cooling-down where $\tau_h$ and $\tau_c$ represent the heat-up and cool-down thermal time constant, respectively. Fig. 7(b) and (c) demonstrates the calculated $t/\tau$ values by plugging the experimentally observed data points in Fig. 7(a) at time $t$ into the dedicated equations. The heat-up and cool-down thermal time constants, $\tau_h$ and $\tau_c$, can be consequently acquired by extracting the inversed slopes of the regression lines in Fig. 7(b) and (c), respectively. Finally, $\tau_h$ and $\tau_c$ are extracted to be 137 and 109 ns, respectively.

With the understanding of its transient thermal characteristics, short-pulse electrical co-optimization is proposed to alleviate SHE. Fig. 8(a) exhibits the working principles of pulse measurement in the time domain where both the drain and gate biases are pulses instead of constant. At the start of each pulse, $t_{rise}$ of time is used for setting the biases to the desired values. Then, they last for $t_{pulse}$ of time in which the first $t_{delay}$ of time is used for bias stabilization, and the following $t_{meas}$ of time is used for current measurement. The $t_{delay}$ time is required as the biases show some damping after being changed. After $t_{pulse}$, $t_{fall}$ of time is used for setting the biases back to zero, and they stay zero until the start of the next cycle. By applying pulse measurement, the device under test is only turned on for $t_{pulse}$ of time each cycle. As long as $t_{pulse}$ is short enough, the device under test is turned off before SHE degrades the device’s performance. Hence, the SHE can be mostly avoided.

Fig. 8(b) exhibits the $I_D-V_{DS}$ curves of a TG In$_2$O$_3$ FET with $T_{ch}$ of 1.9 nm and $L_{ch}$ of 80 nm on the HR Si substrate. The empty symbols represent the dc measurement results where $V_{DS}$ up to 1.8 V can be applied to achieve $I_D$ up to 2.6 mA/µm without observable SHE in its electrical performance. Nonetheless, SHE starts to take place with larger $V_{DS}$ with dc biases. On the other hand, the solid symbols represent the pulse-measurement results with $t_{pulse}$ of 120 ns. With short-pulse measurement, an ultrahigh $I_D$ of 4.3 mA/µm is achieved at a higher $V_{DS}$ of 3.2 V. In great contrast to Fig. 1(d), even with such high PD on an atomically thin channel, there is no observable SHE since the $t_{pulse}$ of 120 ns is even shorter than a heat-up thermal time constant, $\tau_h$, of 137 ns. The key fact is that the electrical response is much faster than the thermal response. Short-pulse electrical measurement can significantly reduce SHE and probe the potential of the material and device performance.

Besides, SiNx is able to serve as a promising interlayer dielectric for M3D integration due to its relatively high

Fig. 9. (a) Structure illustration of combining BEOL ALD In$_2$O$_3$ transistors and FEOL Si devices with 1-µm-thick PECVD SiNx. (b) Optical photograph of the structure in (a) where the indicated FEOL and BEOL devices are vertically separated by a 1-µm-thick PECVD SiNx layer.

Authorized licensed use limited to: Purdue University. Downloaded on April 26,2023 at 15:56:10 UTC from IEEE Xplore. Restrictions apply.
thermal conductivity (4.5 W·m⁻¹·K⁻¹ [27]), low leakage current [28], [29], great CMOS compatibility [30], and low growth temperature enabled by PECVD [31], [32]. To integrate BEOL-compatible TG ALD In₂O₃ TFTs on the FEOL devices, a 1-µm-thick layer of SiNx was grown by PECVD at a low temperature of 150 °C as an interlayer on SEMATECH chips with Si transistors to maintain decent heat dissipation capability. TG ALD In₂O₃ transistors were then fabricated on top of the thick SiNx interlayer as illustrated in the diagram sketch of Fig. 9(a). Fig. 9(b) exhibits an optical image of the fabricated devices where a FEOL Si device and a BEOL ALD In₂O₃ device are indicated in the white circle and blue square, respectively. The FEOL and BEOL transistors are separated by a 1-µm-thick SiNx layer in between for satisfactory electrical insulation.

Fig. 10(a) shows the transfer characteristics of the SEMATECH E-mode Si nMOS with variant Lch from 75 nm to 1 µm at VDS of 0.1 V where great switching behaviors are exhibited regardless of Lch. ON–OFF ratios range from 4 to 6 orders of magnitude, and SS values are all approximately 90–110 mV/dec. Fig. 10(b) illustrates the corresponding output characteristics of a long channel device with Lch of 1 µm where exceptional saturation properties are exhibited. Besides, Fig. 10(c) similarly reveals the ID–VDS curves of a device with a short Lch of 75 nm which reaches a high ID of 1.2 mA/µm at VDS of 1 V. The electrical characterization of the FEOL Si devices was measured before the deposition of the SiNx layer and the fabrication of the BEOL devices. Fig. 10(d) shows the TR measurement of a TG In₂O₃ transistor with Lch of 400 nm, Wch of 2 µm, and channel thickness of 1.8 nm in the FEOL–SiNx–BEOL structure. It is noticeable that the hottest region performs roughly wider ΔT distribution compared with previous cases shown in Fig. 3, which might be due to the reflected heat from the SiNx/FEOL interface to the surface with the presence of the 1-µm-thick SiNx layer. Fig. 10(e) illustrates the comparison of the thermal dissipation capability of the devices on PECVD SiNx with that on other substrates, indicating that the PD-normalized ΔT of In₂O₃ transistors in the FEOL–SiNx–BEOL structure is approximately 1.9 times lower than that on a SiO₂/Si substrate.

Nevertheless, the TG ALD In₂O₃ transistors fabricated on the SiNx–FEOL structure demonstrate poor switching behaviors with an ON–OFF ratio smaller than 1 order of magnitude. This is plausibly due to the defect-rich surface of the PECVD SiNx layer [33], [34], [35]. More investigations and efforts are needed to better realize the integration of FEOL and BEOL fabrications, but this work proposed the FEOL-interlayer-BEOL structure where the interlayer not only separates the FEOL and BEOL devices but also serves as a thermally conductive layer to alleviate the SHE of the BEOL ALD In₂O₃ transistors. The realization of this structure also benefits the applications of ALD In₂O₃ transistors toward M3D integrations at advanced technology nodes.

V. Conclusion

In summary, transient thermal and electrical properties of TG ALD In₂O₃ FETs on various thermally conductive substrates are co-optimized employing an ultrafast HR TR imaging technique to unveil the problematic SHE. By using HR Si substrate and electrical short-pulse measurement, an ultrahigh ID of 4.3 mA/µm is achieved on atomically thin In₂O₃ devices without observable SHE. This work demonstrates that the understanding of both thermal and electrical transient dynamics is important to resolve thermal bottleneck on atomically thin oxide semiconductor devices. Besides, a FEOL–interlayer-BEOL structure is proposed where the interlayer not only electrically isolates the FEOL and BEOL devices but also serves as a thermally conductive layer to alleviate the SHE and exhibit the potentials for the applications of BEOL ALD In₂O₃ TFTs toward M3D integration.

References
