Substrate and Layout Engineering to Suppress Self-heating in Floating Body Transistors

S. H. Shin[1], S. –H. Kim[2], S. Kim[2], H. Wu[1], P. D. Ye[1], and M. A. Alam[1]*
S. H. Shin and S. –H. Kim contributed equally to this study

*E-mail: {alam}@purdue.edu, Phone: (765) 494-5988, Fax: (765)-494-2706
[1]Department of ECE, Purdue University, West Lafayette, IN 47907, USA
[2]Korea Institute of Science and Technology (KIST), Hwarang-ro 14-gil 5, Seoul, 02792, Korea

I. Abstract

Self-heating (SH) has emerged as an important performance, variability, and reliability concern for floating body transistors (FB-FET), namely, extremely-thin-silicon-on-insulator (ETSOI), SOI-FinFET, gate-all-round NW-FET (GAA-FETs), etc. The floating body topology offers electrostatic control, but restricts heat outflow: apparently an intrinsic trade-off. In this paper, we trace the trajectory of heat flow in a broad range of transistors to show that the trade-off is not fundamental, and self-heating can be suppressed by novel device designs that ease thermal bottlenecks. Towards this goal, we (i) *characterize* SH in various FB-FETs with different channel materials (Si, Ge, InGaAs) by submicron thermo-reflectance imaging; (ii) *identify* universal features and common thermal bottlenecks across various transistor technologies, (iii) *offer* novel, technology-aware device design to ease the bottlenecks and reduce self-heating, and (iv) *experimentally* demonstrate the effectiveness of these strategies in suppressing self-heating. We conclude that thermal aware transistor design can suppress self-heating without compromising performance and electrostatic control of the transistor.

II. Introduction

Floating body transistors (FB-FET), such as ETSOI-FET, SOI-FinFET, GAA-FET, etc., use high-k dielectrics, surround-gate (SG) geometry, and new channel material to reduce leakage, control short channel effect, and improve drive-current/speed, respectively [1-3]. Unfortunately, drive-current increases power-dissipation (P), while floating-body geometry and high-k dielectrics increases thermal resistance, (R_{th}). Some of the channel materials (Ge, InGaAs) themselves are poor thermal conductors. Since $\Delta T = P \times R_{th}$, the significant self-heating in these transistors degrades performance, (run-time) variability, and reliability of the transistors [4-7]. Since self-heating has emerged as an important design consideration for the sustained scaling of transistors and adoption of novel channel materials, one must develop thermal-aware scaling strategies to improve heat transfer, without sacrificing short channel control or high drive current [8].

Towards this goal, in this paper, we carefully trace the trajectory of heat flow in the various FB-FET by optical (thermorefelectance) method with sub-micron resolution. Remarkably, we find that regardless the technology, few universal bottlenecks (gate oxide, channel, BOX, drain contact) impede heat transfer to the substrate and define the average channel temperature (ΔT_{Avg}). Equally important, the cross-talk among the fins makes $\Delta T_{Max} \gg \Delta T_{Avg}$ among the fins, suggesting spatially non-uniform degradation and dramatically reduced NBTI, HCI, and TDDB lifetime for the technology [9]. Our analysis suggests design strategies to relax the thermal bottlenecks/cross talk by substrate, device and fin-layout engineering. We use simple test-structures to demonstrate the effectiveness of the proposed concepts.

III. Experimental Observations of Self-Heating in FB-FET

A schematic of a FB-FET is shown in Fig. 1. Our previous experiments in GAA-FET have demonstrated that once heat is generated at the channel/drain edge (#1), it first diffuses to the S/D (#2) and the gate (#3) contacts, followed by diffusion through the BOX (#4) to the substrate. The CCD and high-resolution TR-images of Si ETSOI, Ge NW-FET, InGaAs GAA-FET in Fig. 2 show that the trajectory of heat diffusion is essentially universal across technologies [10-13]: First, in all the cases, heat diffuses asymmetrically to the drain (#1, 2) (Fig. 2b, top view; Fig. 2c, front view). Second, the heat diffusion through the gate stack to the gate contact (#3) is also clear from the side view, shown in Fig. 2d. Finally, the importance of heat-diffusion through the BOX (#4) is ascertained as follows: Consider two transistors with identical transistor geometry, but different BOX structures, as in Fig. 3. The gate surface maximum temperature ($\Delta T_{S,Max}$), as measured by TR-images is far more sensitive to the BOX properties (Fig. 3a) compared to that of the thickness of the gate oxides (Fig. 3c). Our measurements imply that the overall thermal-resistance, R_{th} (given by the slopes of the $\Delta T_{S,Max} - P$ in Fig. 3a) is strongly correlated to the properties of the substrate within 100nm below the transistors. We can also reach the same conclusion by comparing results of two SG-FETs with and without air-gap below the transistor (Fig. 4). A close analysis of the images suggest that, unlike classical transistors, *self-heating of SG-FET is dictated by the substrate and the S/D contacts, not by the channel material or gate oxide.* This remarkably important insight suggests the possibility of decoupling self-heating (by controlling the design of BOX) from that of suppressing short channel effects and suppress leakage (though careful, but independent design of the gate oxide). In Sec. IV and V, we will explore the insight conceptually through simulation and experimentally through new test-structures.

Next, let us discuss the second characteristic features of self-heating, namely the importance of cross-talk in a creating a non-uniform temperature profile, so that $\Delta T_{S,Max} \gg \Delta T_{S,Avg}$. Fig. 2d shows the temperature distribution across the fins, perpendicular to the channel. The middle fins are 30-50% hotter than the rest. This is generally true for all the technologies considered, see Fig. 2d and Fig. 4. This non-uniformity of temperature distribution leads to run-time variation in channel current (speed), broader Weibull slope for TDDB, and accelerated HCI, NBTI, and TDDB damage of middle fins compared to those at the edge [14]. As expected, the difference between $\Delta T_{S,Max}$ and $\Delta T_{S,Avg}$increases with number of Fins/NWs as well as with the input power density, as summarized

978-1-5090-3902-9/16/$31.00 ©2016 IEEE

in Fig. 5. Interestingly, this result implies that at same power, transistors with fewer NWs will have more predictable and uniform characteristics, because the ratio of $\Delta T_{Max}/\Delta T_{Avg} \rightarrow 1$.

IV. Substrate Engineering to reduce Self-heating

Given the universality and insights of the trajectory of heat-flow and various components that determine R_{th}, we can explore by 3D computational modeling various alternate structures that may reduce self-heating of SG-FET. For generality, we demonstrate the concept for three technologically-relevant transistors: SOI-Fin, SOI Planar, and GAA NWs structures. We assume that the transistors operate at the same power level, i.e., at comparable voltage and currents with same power density over channel width. The basic geometry and thermal properties are summarized in Fig. 6a (top) and Tables 1. Our goal is to sequentially modify various elements that dictate the heat-flow (the source/drain metal contact (#2), and the substrate (#4), and/or spacing the Fins, to reduce ΔT_{Avg}, and make $\Delta T_{Max}/\Delta T_{Avg} \rightarrow 1$.

First, we focus on the substrate-engineering by changing the BOX thickness and the substrate material. The results show that channel temperature (ΔT_C) decreases by 50-70%, depending on the relative thermal conductivity of the BOX materials, starting from SiO_2 (#1) to Al_2O_3 (#5), see Fig. 6a (bottom). The substrate with the highest thermal conductivity (SiC) substrate offers the maximum benefit, (#6 and #7 in Fig. 6a). The advantage is preserved even for GAA transistors, with 4-sided channel that support higher current density per NW than other Fin structures. Interestingly, Fig. 6b also shows that ΔT_c saturates relatively quickly with T_{BOX}, indicating that the R_{th} is dictated not by the entire T_{BOX}, but by the region close($\sim 50-100nm$) to the transistors, consistent with the physics of 3D heat diffusion from a point source. Inspired by this observation, we propose the following composite BOX designs, see Fig. 6c. Four different BOX designs with $T_{BOX} = 50nm$ are considered. The results confirm the importance of putting the high thermal conductivity substrate close to transistors, as in design #4. In contrast, the thermal properties and thickness of the gate oxides have relatively little impact of ΔT_C, see Fig. 7. Incidentally, these results are fully consistent those in Figs. 3 and 4.

Second, the observation of asymmetrical heat flow through S/D region suggests that we may either shorten the S/D extension ($L_{SD,Ext}$) as in Fig. 8, or increase the size of the drain pad. Fig. 9 and 10 show that the benefit is small, but significant, especially given the exponential sensitivity of reliability on ΔT_C. Finally, Fig. 11 shows that the idea of reducing the parasitic capacitances by creating an airgap next to S/D may actually be counterproductive, because increased self-heating may erode any gain in performance.

Finally, we observed that thermal crosstalk among the NWs increases the self-heating of the middle fins, which may lead to unequal current distribution among the NWs and overall reduced reliability of the transistors. Clearly, optimized layout with unequal spacing among the fins (while keeping the total width the same) would reduce temperature non-uniformity, see Fig. 12 and 13. The uniformity of the temperature should increase NBTI, HCI, and TDDB lifetimes significantly.

V. Experimental Verification

To verify our conceptual designs proposed in Sec. IV to reduce self-heating in SG-FET, we have used simple test structures (composed of metal lines terminated by asymmetrical contacts) to mimic the layout of the transistors, see Fig. 14a. The BOX stacks were deposited by material specific methods such as PECVD, ALD, and evaporator, as appropriate. Remarkably, the advantages suggested by simulation are all confirmed by experiments.

Regarding the importance of BOX material, we find that the slope of the $\Delta T_S - P$ ($i.e., R_{th}$) reduces with higher κ BOX material (Fig. 14b), as predicted by the theoretical modeling (Fig. 6a). The reduced R_{th} should translate to reduced self-heating. Fortunately, this improvement in self-heating does not compromise the transient response of the transistor, see Fig. 14c.

The structures and thermal images shown in Fig. 15a verifies the roles of drain pad and spacer engineering. As predicted in Fig. 9-11, the S/D engineering does offer small, but unambiguous advantage. We also confirmed that the dummy pad structures (mimicking Si_3N_4 spacers and dummy via) help to quickly distribute power from hot spot and reduce the peak temperature in the channel region (Fig. 16a). Interestingly, Fig. 16b shows that larger drain pads also improve the uniformity of heating among the fins by providing a parasitic pathway to redistribute heating through the pad. Finally, Fig. 17 demonstrates the importance of non-uniform Fin placement in reducing the peak temperature and reducing the temperature non-uniformity among the Fins, consistent with the theoretical model in Fig. 12, 13.

VI. Conclusions

The high spatio-temporal resolution of the TR imaging, coupled with 3D electro-thermal modeling of self-heating in floating-body transistors identifies the dominant and universal thermal bottleneck (e.g., drain side, internal channel, and substrate) for modern transistor topologies. With the bottlenecks identified, the conceptual re-design and verification though test-structure demonstrate the effectiveness of various strategies of self-heating. Self-heating will always be an important concern for floating body transistors, but our detailed analysis shows that it is possible to decouple the control of self-heating from the electrostatic control of the channel, and thereby design transistors with excellent short-channel control as well as reduced self-heating.

Acknowledgement

We acknowledge Birck Nanotechnology Center for the fabrication and characterization facilities and Prof. A. Shakouri's group, Z. Choi, and D. Back for the technical support. S. –H. Kim would like to thank the future SDTDP and KIST flagship program.

References

[1] A D. Hisamoto et al., IEEE TED, 47(12), 2320-2325, 2000.
[2] A. Majumdar et al., IEEE TED, 56(10), 2270-2276, 2009.
[3] N. Singh et al., IEEE EDL, 27(5), 383-386, 2006.
[4] C. Prasad et al., IEEE IRPS, 5D.1.1-5D.1.4., 2013
[5] S. W. Pae et al., IEEE IEDM, 557-560, 2015.
[6] D. Jang et al., IEEE IEDM, 289-292, 2015.
[7] R. Wang et al., IEEE IEDM, 753-765, 2008.
[8] T. Takahashi et al., IEEE IEDM, 809-812, 2011.
[9] S. E. Liu et al., IEEE IRPS, 4A.4.1-4A.4.4, 2014.
[10] J. J. Gu et al., IEEE IEDM, 27.6.1-27.6.4, 2012.
[11] H. Wu et al., IEEE IEDM, 2.1.1-2.1.4, 2015.
[12] S. H. Shin et al., IEEE TED, 62(11), 3516-3523, 2015.
[13] S. H. Shin et al., IEEE IEDM, 20.3.1-20.3.4, 2015.
[14] B. Weir et al., IEEE IEDM, 437-440, 1999.

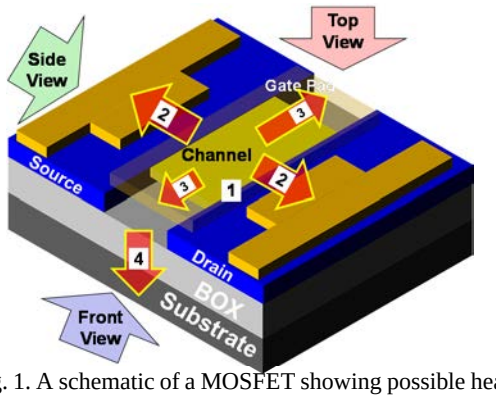


Fig. 1. A schematic of a MOSFET showing possible heat diffusion paths for various floating body transistors with SOI, SOI-Fin, and GAA NWs structures. The thermal images in Fig. 2 will be arranged from different angles such as top, front (perpendicular to channel), and side (along channel) views.

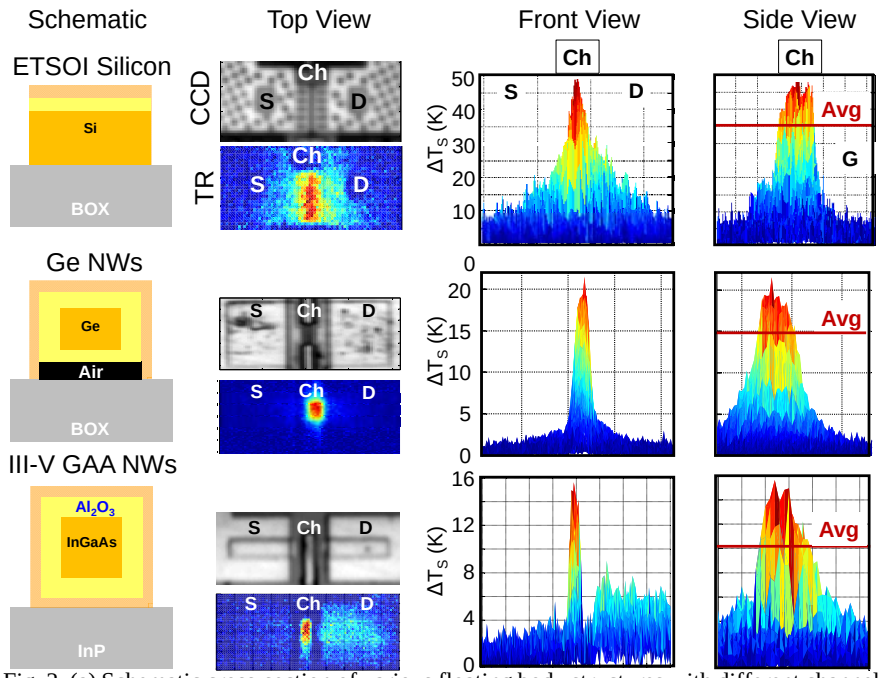
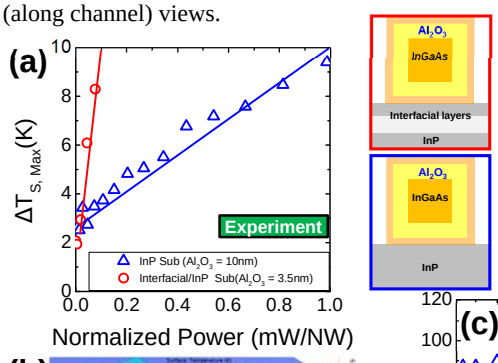
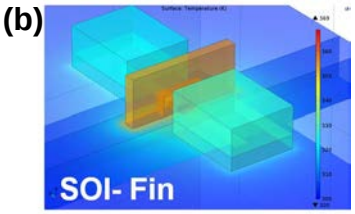


Fig. 2. (a) Schematic cross-section of various floating body structures with different channel materials. (b) CCD and thermal images (top view) show asymmetric heat diffusion through S/D. (c) Thermal images (front view, perpendicular to channel). (d) Thermal images (side view, perpendicular to the channel) show thermal cross talk in channel. Average temperature over the active region is shown for each structure [10-13].





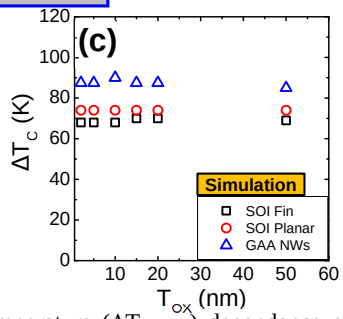


Fig. 3. (a) Gate surface maximum temperature ($\Delta T_{S,\,Max}$) dependence on normalized power (mW/NW) for InGaAs GAA NWs structure and side images are schematics of the InGaAs NWs W/ and W/O interfacial substrate on InP [10]. (b) An example of SOI-Fin structure simulation. (c) Channel temperature (ΔT_C) dependence for different structure using Al_2O_3 as gate oxide.

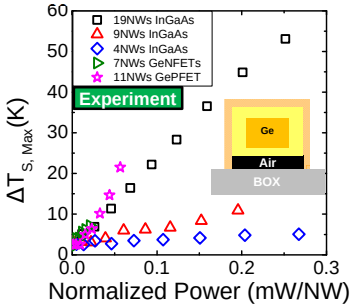


Fig. 4. $\Delta T_{S,\,Max}$ dependence on normalized power for different # of NWs with Ge with air gap suspended and InGaAs with GAA NWs transistors.

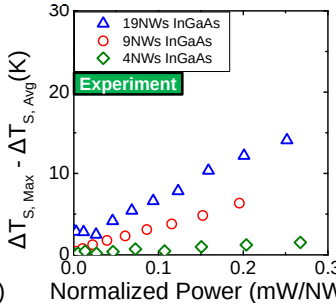
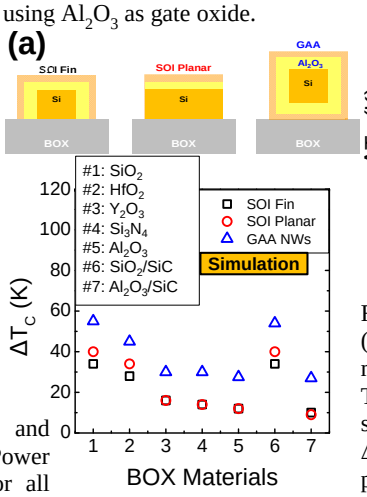


Fig. 5. Difference between max and average gate temperature over the channel area ($\Delta T_{S,\,Max}$ - $\Delta T_{S,\,Avg}$), depending on normalized power.

	Simulation Parameters
Geometries	SOI Planar, SOI Fin, GAA NW
Channel Material	Si
L_{ch} (nm)	40
W_{ch} (nm)	40
H_{ch} (nm)	10
$L_{SD,\,Ext}$ (nm)	10~100nm
T_{ox}	5 ~ 50nm Al_2O_3
T_{BOX}	10 ~ 300nm
Dielectrics	SiO_2, HfO_2, Y_2O_3, Si_3N_4, Al_2O_3
Substrate	Si, SiC

Materials	Cp (J/(kg-K)	Rho (kg/m^3)	k (W/(m-K))
SiO_2	730	2200	1.4
HfO_2	287	9678	2.3
Y_2O_3	455	5029	13.4
Si_3N_4	700	3100	20
Al_2O_3	730	3965	35
SiC	690	3216	490

Table. 1. Thermal properties (top) and geometries for simulations (bottom). Power density for channel area is same for all simulations, GAA NWs has higher power.



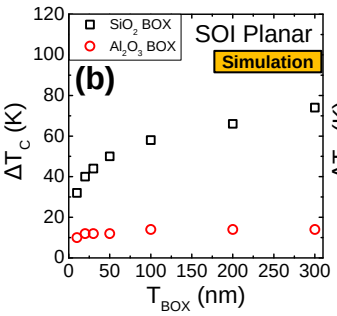
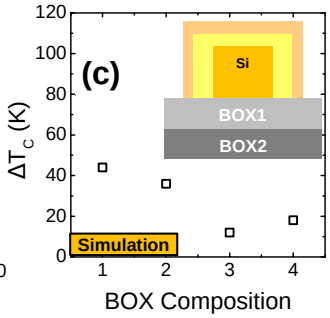


Fig. 6. (a) Channel temperature (ΔT_C) dependence on BOX materials and substrates (Fixed T_{BOX} = 20nm) for various structures (top-schematics). (b) ΔT_C dependence on T_{BOX} for SOI planar.



#1: 50nm SiO_2 only for BOX1&2
#2: 25nm SiO_2(BOX1)/25nm_Al2O3(BOX2)
#3: 50nm Al_2O_3 only for BOX1&2
#4: 25nm Al_2O_3 (BOX1)/25nm SiO_2 (BOX2)

(c) ΔT_C dependence on different composition of BOX materials with fixed T_{BOX} = 50nm. For #2 and #4, BOX1 and 2 are switched.

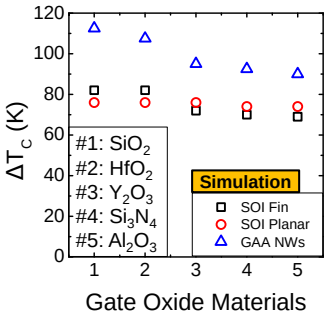


Fig. 7. ΔT_C dependence on various gate oxide (T_{ox} = 10nm, T_{BOX} (SiO_2) = 300nm).

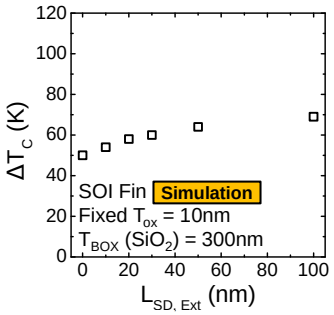


Fig. 8. ΔT_C dependence on length of S/D extension (T_{ox} = 10nm, T_{BOX} (SiO_2) = 300nm) with SOI Fin.

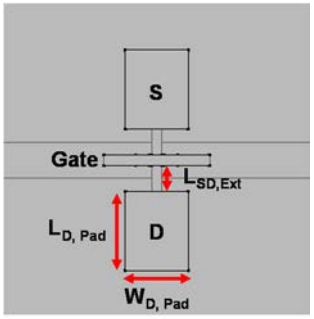


Fig. 9. Top view of simulated structure. $W_{D, Pad}$ is scaled to observe pad size dependence.

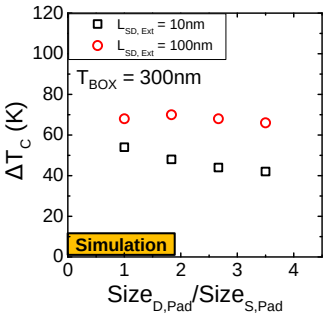


Fig. 10. ΔT_C dependence on scaling of drain pad size (T_{ox} = 10nm, T_{BOX} (SiO_2) = 300nm).

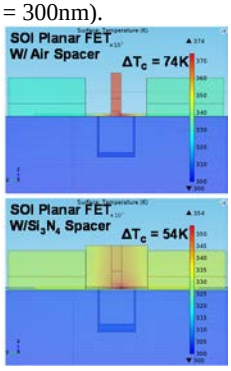


Fig. 11. Numerical simulations for with air and Si_3N_4 spacer.

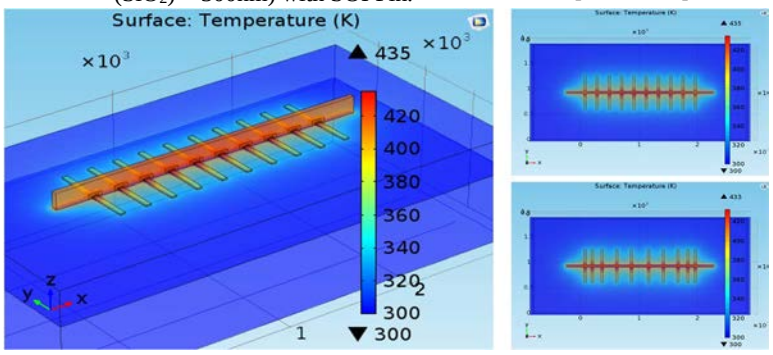


Fig. 12. (a) Simulation for 10 fin arrangement for SOI Fin structure. (b) Top view image for uniform gap in fins. (c) By keeping active region for Fig. 12. (b), gap in the fins in center region is widely distributed but, outer fins are closer to reduce thermal cross talks in the center region. S/D pads are intentionally removed for clarity.

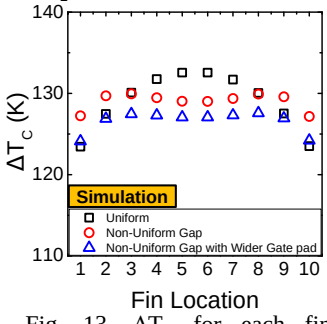
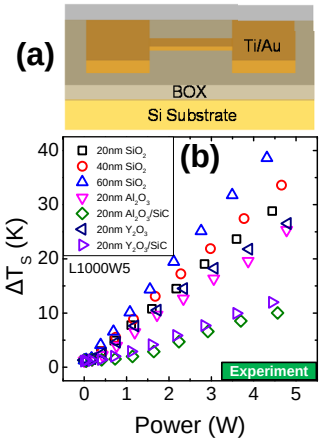


Fig. 13. ΔT_C for each fin by changing spacing among the fins. Temperature profile becomes more uniform with physically non-uniform gap and wider pad gate pad also contribute to the uniformity.



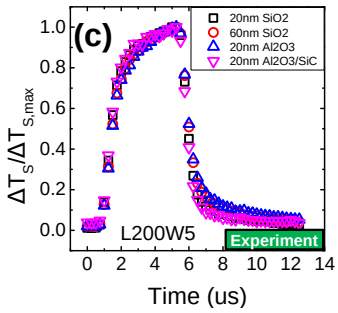
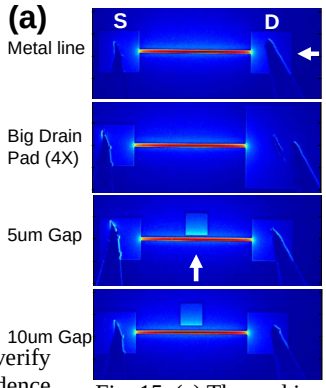


Fig. 14. (a) Simple test structure to verify thermal aware solutions. (b) ΔT_S dependence on BOX materials. (c) Transient curves of ΔT_S normalized by maximum ΔT_S for each material. Zoomed in cooling transient (inset).



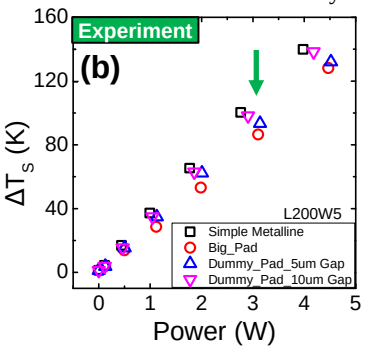
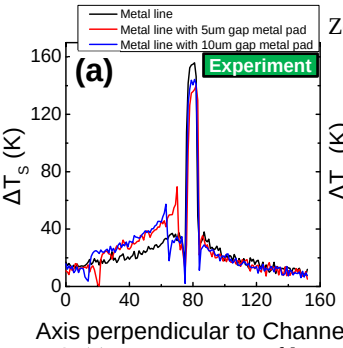


Fig. 15. (a) Thermal images for the test structure with big drain pad, 5 and 10 um gap away dummy pad to channel area respectively. (b) ΔT_S dependence on power for the various thermal aware structure.



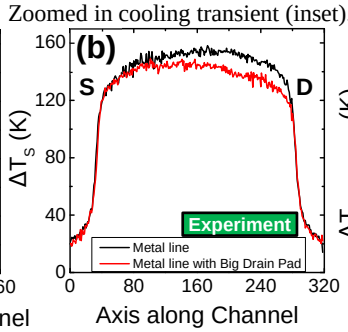
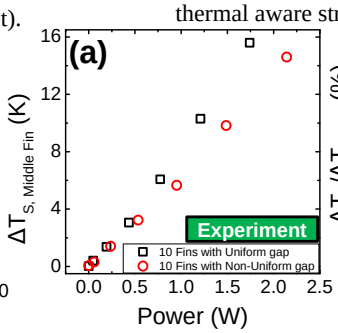
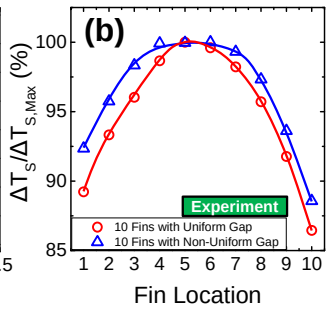


Fig. 16. (a) Temperature profile perpendicular to the channel with and without dummy pads. (b) Temperature profile along channel with and without big drain pad. Asymmetry temperature profile can be made intentionally.





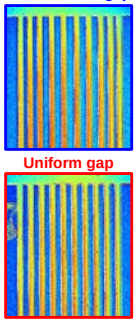


Fig. 17. (a) ΔT_S at middle fin depending on power for 10 fins uniform gap and 10 fins non-uniform gap structures. (b) Temperature profiles of for uniform gap and non-uniform gap (side) for each fin perpendicular to channel and corresponding thermal images (side).