Critical Role of Quantum Confinement on Transfer Length in Achieving High-Performance In_2O_3 Transistors with Ultra-Scaled Contacted Gate Pitch

Jian-Yu Lin[1], Chang Niu[1], Zehao Lin[1], Chang Liu[2], Juanjuan Lu[2], Haiyan Wang[2], Peide D. Ye[1, *]
[1]Elmore Family School of Electrical and Computer Engineering and [2]School of Materials Engineering,
Purdue University, West Lafayette, Indiana, USA, *Email: yep@purdue.edu

Abstract

In this work, for the first time, we study the contact length (L_C) and contacted gate pitch (CGP) scaling in ultrathin In_2O_3 field-effect transistors (FETs). A large 53 % decrease in transfer length (L_T) from 76 to 36 nm can be observed by increasing the In_2O_3 channel thickness (T_{ch}) from 1.2 nm to 2.0 nm, which can be understood by the positive-to-negative Schottky barrier height (Φ_{SB}) transition modulated by the quantum confinement (QC) effect in In_2O_3 channel. Leveraging the record-low L_T of 36 nm optimized by the QC, 2.0 nm In_2O_3 FETs demonstrate a record-low contact resistance (R_C) of 140 Ω·μm and a record-high maximum drain current ($I_{D,max}$) of 1.57 mA/μm at ultra-scaled CGP of 80 nm among all reported oxide semiconductor FETs with CGP scaling.

Introduction

Oxide semiconductor (OS) FETs have garnered renewed attention in recent years due to their potential application as back-end-of-line (BEOL) compatible transistors in monolithic 3D integration, a promising approach for advancing both "More Moore" and "More-than-Moore" technologies [1-7]. Among various OS FETs, atomic-layer-deposited (ALD) In_2O_3 FETs have exhibited significant potential owing to their high electron mobility of 152 $cm^2/V·s$ [2], remarkable on-current performance reaching nearly 20 mA/μm in gate-all-around structure [3], excellent reliability [4], and ultralow R_C of 23.4 Ω·μm (measured at 10 K) benefiting from the negative Φ_{SB} at the metal/In_2O_3 contact [5]. Although many works have focused on channel length (L_{ch}) scaling to achieve high-performance In_2O_3 FETs [2-7], L_C and CGP (= L_{ch} + L_C) scaling of In_2O_3 transistors has not been studied yet. CGP scaling is of great importance for reducing the transistor footprint and increasing the device density per area. In this work, the L_C and CGP scaling of ultrathin In_2O_3 FETs are investigated for the first time. Through increasing the T_{ch} of In_2O_3 from 1.2 nm to 2.0 nm, the L_T of the FETs can be reduced from 76 to 36 nm, which can be understood by the positive-to-negative Φ_{SB} transition governed by the QC effect. With the record-low L_T of 36 nm, 2.0 nm In_2O_3 FETs demonstrate a record-low R_C of 140 Ω·μm and a record-high $I_{D,max}$ of 1.57 mA/μm at ultra-scaled CGP of 80 nm among all OS FETs with scaled CGP. **Fig. 1** summarizes the highlights of this research.

Experiments

Fig. 2 (a) illustrates the schematic device structure of a back-gate In_2O_3 transistor with L_C scaling. Note that the CGP is defined as L_C plus L_{ch}. **Fig. 2 (b)** describes the fabrication process flow of the back-gate In_2O_3 transistors. As the start, 40 nm Pt was e-beam evaporated on $Al_2O_3/SiO_2/Si$ substrate as back-gate. 6 nm HfO_2 was then deposited by plasma-enhanced ALD (PE-ALD) at 200 °C to serve as gate dielectric. Next, ultrathin In_2O_3 with T_{ch} between 1.2 to 2.0 nm was grown by ALD at 225 °C. Isolation of the In_2O_3 channel was performed by Ar dry etching. 40 nm Ni was e-beam evaporated as the source/drain (S/D). Finally, post O_2 annealing at 270 °C for 30 minutes was used to improve the device performance. **Fig. 3** shows the cross-sectional scanning transmission electron microscopy (STEM) image with energy dispersive X-ray spectroscopy (EDS) elemental mappings of an In_2O_3 FET with an ultra-scaled CGP of 80 nm (L_{ch} = L_C =40 nm).

Results and Discussion

Fig. 4 (a) demonstrates the transfer characteristics of In_2O_3 FETs with different T_{ch} at L_{ch} = L_C = 1 μm. **Fig. 4 (b)** plots the T_{ch}-dependent field-effect mobility (μ_{FE}) extracted from the transconductance (g_m) of **Fig. 4 (a)**. In **Fig. 4**, two T_{ch}-dependent properties, threshold voltage (V_T) and μ_{FE}, in In_2O_3 FETs can be observed. When T_{ch} increases from 1.2 to 2.0 nm, the increase of carrier density in the In_2O_3 channel results in the negative shift of V_T [6, 8]. As for the higher μ_{FE} in thicker In_2O_3 film (e.g. μ_{FE} = 62.2 $cm^2/(V·s)$ when T_{ch} = 2.0 nm), this is due to the reduction of surface scattering effect on the transport at the channel center in thicker In_2O_3 film [6, 8]. **Fig. 5** studies the L_C scaling effect on the $I_D - V_{GS}$ curves of 2.0 nm In_2O_3 FETs at short L_{ch} = 80 nm. As the L_C scales down to 40 nm, some degradations of $I_{D,max}$ and g_m are caused by the increase of R_C due to the current crowding effect and the rise of S/D metal line resistance (R_{metal}). To further investigate the impact of L_C scaling, the transmission line method (TLM) was adopted to extract the R_C at different T_{ch} and L_C, as shown in **Fig. 6**. **Fig. 7 (a)** presents the dependency of R_C (extracted by TLM) on L_C in In_2O_3 FETs with different T_{ch}. Solid lines in the figure represent the fitting results of the experimental data using equation (1) [9]:

$$R_C = R_{C0}\coth(L_C/L_T)\,;\, L_T = \sqrt{\rho_C/R_{sh}} \qquad (1)$$

where R_{C0}, ρ_C, and R_{sh} denote the R_C when $L_C \gg L_T$, contact resistivity, and sheet resistance, respectively. When the T_{ch} changes from 1.2 to 2.0 nm, one can notice the reduction of R_{C0} from 233 to 146 Ω·μm [**Fig. 7 (a)**], L_T from 76 to 36 nm [**Fig. 7 (a)** and **Fig. 7 (b)**], and ρ_C from 1.2×10^{-6} to 7.7×10^{-8} $\Omega\cdot cm^2$ [**Fig. 7 (b)**]. All the above observations indicate that the Ni/In_2O_3 contact becomes better when the In_2O_3 film is thickened to T_{ch} = 2.0 nm. With good S/D contact in 2.0 nm In_2O_3 FETs, a record-low R_C = 140 Ω·μm at ultra-scaled L_C of 40 nm (**Fig. 8**) is achieved.

The key reason behind the T_{ch}-dependent R_C in In_2O_3 FETs is the change of signs of Φ_{SB} controlled by the QC effect [6]. **Fig. 9** illustrates the schematic band diagrams at the interface between S/D metal and In_2O_3 channel with various T_{ch}. Due to the strong Fermi-level pinning at metal/In_2O_3 interface [6], the metal Fermi-level (E_{FM}) will be pinned at the charge neutrality level (CNL) of In_2O_3. When T_{ch} reduces from 2.0 nm [**Fig. 9 (b)**] to 1.2 nm [**Fig. 9 (a)**], the QC effect will widen the bandgap of In_2O_3 and move the energy of the conduction band minimum (E_C) upwards [6]. Considering the energy position of CNL is independent of T_{ch} [6, 8, 10, 11], the shift of E_C by QC will result in different band alignments among E_C, CNL, and E_{FM} at the metal/In_2O_3 interface. For instance, the E_C is "above" the CNL and the E_{FM} for 1.2 nm In_2O_3 [**Fig. 9 (a)**] [6], which creates a positive Φ_{SB} contact with larger R_C, ρ_C, and L_T (**Fig. 7**). On the other hand, the E_C is "below" the CNL and the E_{FM} for 2.0 nm In_2O_3 [**Fig. 9 (b)**] [6], which contributes to a negative Φ_{SB} contact with smaller R_C, ρ_C, and L_T (**Fig. 7**). This positive-to-negative shift of Φ_{SB} in In_2O_3 as the T_{ch} increases has been studied and verified in our previous work [6]. To sum up, the QC effect can modulate the following three things: 1. band alignments among E_C, CNL, and E_{FM} at the metal/In_2O_3 interface; 2. Φ_{SB} at the metal/In_2O_3 interface; 3. R_C, ρ_C, and L_T of In_2O_3 FETs. Based on the current crowding model described by eq. (1), R_C will elevate exponentially as $L_C < L_T$. Therefore, the understanding of QC effect to obtain $\Phi_{SB} < 0$ and minimize the L_T of In_2O_3 FETs is critical for achieving low R_C and high-performance transistors at ultra-scaled L_C and CGP.

Fig. 10 shows the $I_D - V_{GS}$ curves of 2.0 nm In_2O_3 FETs with ultra-scaled L_C = 80 or 40 nm (CGP = 120 or 80 nm), L_{ch} = 40 nm, and enhancement-mode operations (V_T > 0 V). **Fig. 11** presents the output characteristics of these ultra-scaled CGP devices using pulse I-V measurements to alleviate the self-heating effect [12]. Benefiting from the negative Φ_{SB} contact, record-low R_C = 140 Ω·μm, and record-low L_T = 36 nm in 2.0 nm In_2O_3 FETs, record-high $I_{D,max}$ of 2.16 and 1.57 mA/μm are achieved in ultra-scaled CGP = 120 and 80 nm, respectively, compared to other reported OS FETs with CGP scaling.

Conclusion

This work offers the first study of L_C and CGP scaling in ultrathin In_2O_3 FETs. It is found that using the QC effect in In_2O_3 to modulate the Φ_{SB} and minimize the L_T is essential to attain high-performance In_2O_3 devices in ultra-scaled CGP. **Table I** presents the benchmarking of OS FETs with sub-100 nm L_{ch}. Among all OS FETs with scaled CGP, the 2.0 nm In_2O_3 FETs in this work stand out with record-low R_C, record-low L_T, record-high $I_{D,max}$ in ultra-scaled CGP. Note that when reporting the L_T value, one should use eq. (1) to extract the L_T from $R_C - L_C$ dependency rather than using the x-intercept of TLM, which might give inaccurate and underestimated L_T values. This work is supported by NSF, SRC, and Samsung Electronic.

Highlights of this work

1. First investigation of L_C and CGP scaling in ultrathin In_2O_3 FETs
2. First study of QC effect on L_T in ultrathin In_2O_3 FETs with various T_{ch} and L_C
3. High performance ultrathin In_2O_3 FETs in ultra-scaled CGP of 80 nm (L_{ch} = L_C = 40 nm) achieved by smaller L_T modulated by QC effect:
   - ❑ Record-low R_C of 140 Ω·μm and L_T = 36 nm [among OS transistors with scaled CGP]
   - ❑ Record-high maximum drain current ($I_{D, max}$) = 1570 μA/μm [among OS transistors with scaled CGP]

Fig. 1. Highlights of this work, including the first investigation of L_C and CGP scaling in ultrathin In_2O_3 FETs, the first study of QC effect on L_T of ultrathin In_2O_3 FETs with T_{ch} = 1.2 – 2.0 nm, and record-high performance 2.0 nm In_2O_3 FETs (CGP = 80 nm) among reported OS transistors with scaled CGP.

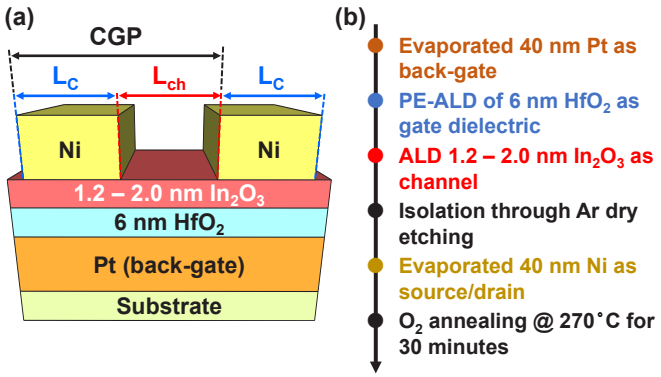

Fig. 2. (a) Device schematic of a back-gate In_2O_3 transistor. CGP is defined as L_{ch} plus L_C of source or drain metals. The minimum CGP in this work is 80 nm, which is L_{ch} + L_C = 40 + 40 nm = 80 nm. **(b)** Fabrication process flow of back-gate In_2O_3 transistors. Different T_{ch} from 1.2 – 2.0 nm of the In_2O_3 channel was grown by ALD at 225 °C with $(CH_3)_3In$ (TMIn) and H_2O as precursors.

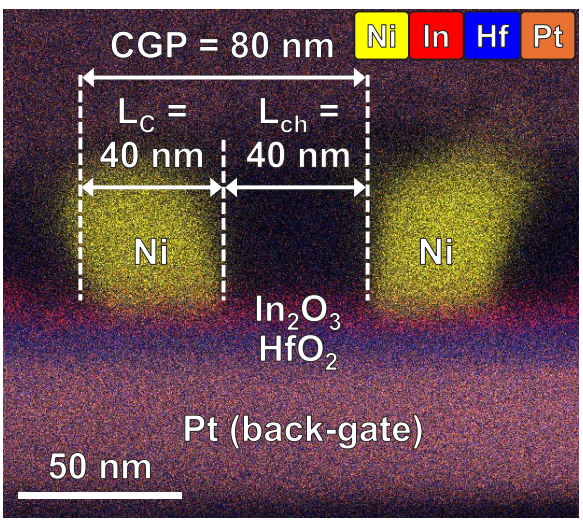

Fig. 3. Cross-sectional STEM image with EDS elemental mapping (Ni, In, Hf, and Pt) of an In_2O_3 FET with ultra-scaled CGP of 80 nm with L_{ch}=40 nm and L_C = 40 nm.

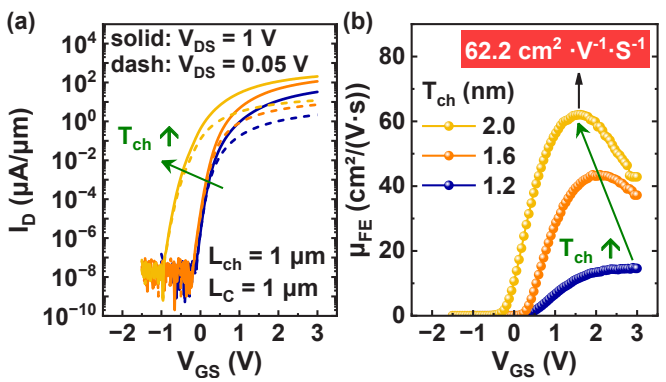

Fig. 4. (a) Transfer characteristics of In_2O_3 FETs with L_{ch} = L_C = 1 μm, and T_{ch} = 1.2, 1.6, and 2.0 nm. When T_{ch} increases from 1.2 to 2.0 nm, the increase of carrier density in In_2O_3 channel results in the negative shift of V_T [6, 8]. **(b)** Extracted μ_{FE} from the g_m of In_2O_3 FETs with different T_{ch}. Measured gate oxide capacitance (C_{OX}) of 1.6×10^{-6} F/cm^2 was used for the μ_{FE} extraction. The larger mobility in the thicker In_2O_3 channel comes from the reduction of surface scattering inside the transistor channel [6, 8].

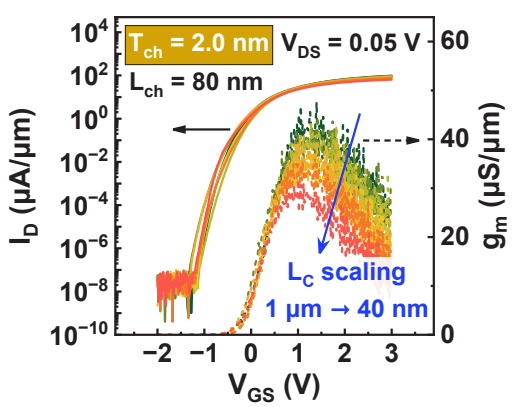

Fig. 5. Transfer characteristics of In_2O_3 FETs with short L_{ch} = 80 nm, T_{ch} = 2.0 nm, and different L_C = 1, 0.4, 0.2, 0.1, 0.08, 0.06, and 0.04 μm. The degradation of I_D and maximum g_m are caused by the rise of R_C due to the current crowding effect and the increase of metal line resistance when L_C becomes smaller.

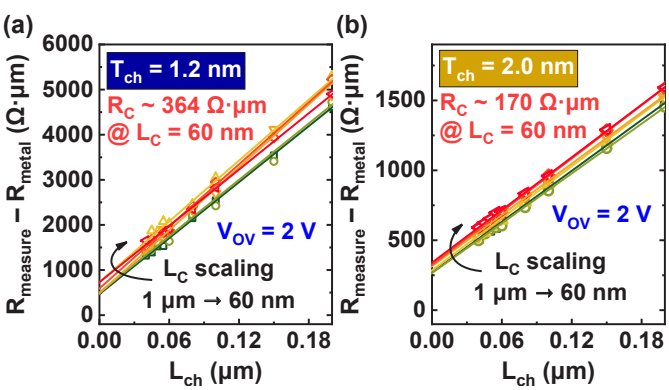

Fig. 6. TLM measurements at $V_{GS} - V_T = V_{OV}$ = 2 V on **(a)** 1.2 nm and **(b)** 2.0 nm In_2O_3 transistors with L_C = 1, 0.4, 0.2, 0.1, 0.08, and 0.06 μm. In the TLM analysis, to extract the correct R_C at different L_C, R_{metal} is removed from the measured total resistance ($R_{measure}$) before extracting R_C and sheet resistance (R_{sh}). For 2.0 nm In_2O_3 FETs, a low R_C of 170 Ω·μm can be extracted at scaled L_C of 60 nm.

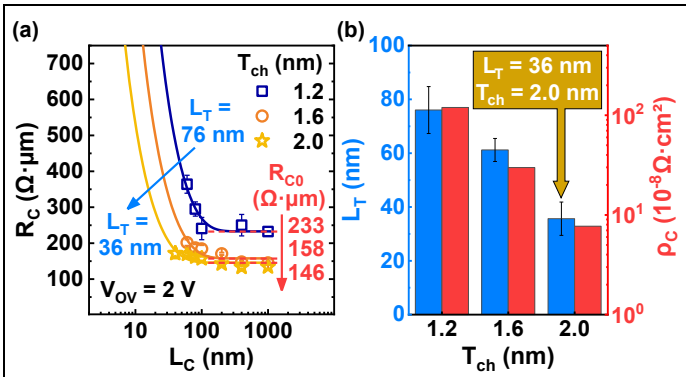

Fig. 7. **(a)** R_C extracted by TLM as a function of L_C in In_2O_3 FETs with different T_{ch} = 1.2, 1.6, and 2.0 nm at V_{OV} = 2 V. Symbols represent the experimental data. Solid lines are the fitting results using eq. (1), which can be used to extract L_T and the R_C when $L_C >> L_T$ (denoted as R_{C0}). **(b)** T_{ch}-dependent L_T and ρ_C extracted from **(a)**. In_2O_3 FETs with T_{ch} = 2.0 nm demonstrate a record-low L_T of 36 nm among OS transistors with scaled CGP. For 2.0 nm In_2O_3, ρ_C of 7.7×10^{-8} Ω·cm² can be extracted from $L_T = (\rho_C / R_{sh})^{0.5}$ = 36 nm using eq. (1).

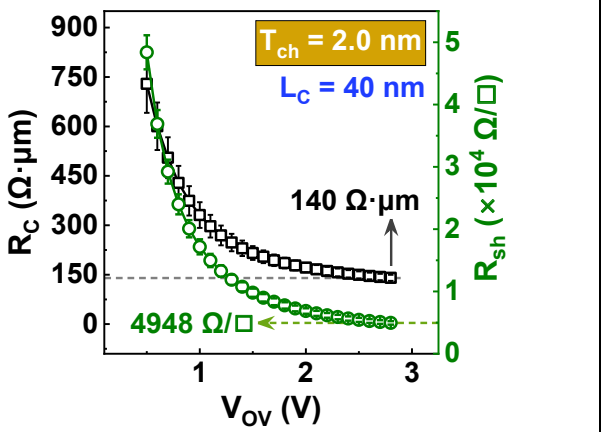

Fig. 8. R_C and R_{sh} extracted from TLM as a function of V_{OV} in In_2O_3 FETs with T_{ch} = 2.0 nm and L_C = 40 nm. This work shows the smallest R_C of 140 Ω·μm among OS transistors with scaled L_C and CGP. The ultralow R_C value under such a scaled L_C benefits from the negative Schottky barrier height at metal/In_2O_3 contact.

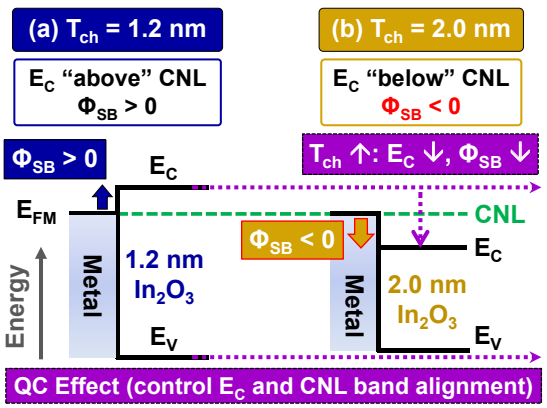

Fig. 9. Schematic band diagrams between metal/In_2O_3 contact with various In_2O_3 thickness of **(a)** T_{ch} = 1.2 nm and **(b)** T_{ch} = 2.0 nm. When T_{ch} decreases from **(b)** 2.0 nm to **(a)** 1.2 nm, QC effect will move the energy of E_C upward to above CNL and widen the bandgap of In_2O_3 [6]. Band alignments between E_C and CNL are essential in determining whether the Φ_{SB} is > 0 [when E_C is "above" CNL, like **(a)**] or < 0 [when E_C is "below" CNL, like **(b)**] [6]. Φ_{SB} < 0 means better contact between metal/In_2O_3 with lower R_C, ρ_C, and L_T.

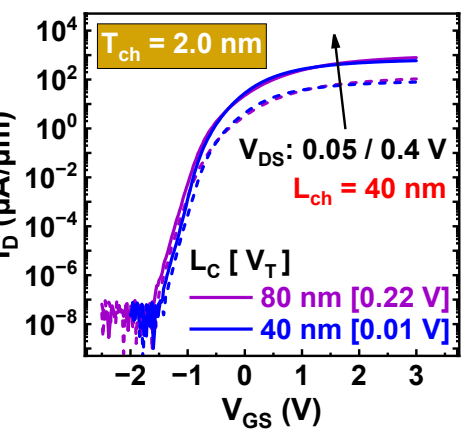

Fig. 10. Transfer characteristics of 2.0 nm In_2O_3 FETs with ultra-scaled L_C (= 80 / 40 nm) and CGP ($L_C + L_{ch}$ = 120 / 80 nm) at short L_{ch} = 40 nm. Both devices are enhancement-mode with linearly extrapolated V_T > 0.

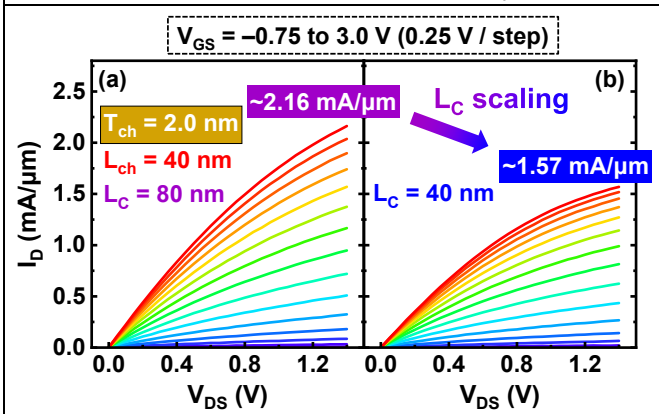

Fig. 11. Output characteristics of 2.0 nm In_2O_3 FETs with **(a)** L_C = 80 nm (CGP = 120 nm) and **(b)** L_C = 40 nm (CGP = 80 nm) at L_{ch} = 40 nm. Pulse I-V measurements were adopted to suppress the self-heating effect [12].

Table I. Benchmarking of sub-100 nm L_{ch} OS FETs with or without L_C and CGP scaling

Ref.	Material	V_T (V)	L_{ch} (nm)	L_C (nm)	CGP (nm)	$I_{D,max}$ (μA/μm)	R_C (Ω·μm)	L_T (nm)
[13]	ITO	0.3	50	/	/	1260	70	/
[14]	IWO	–0.1	50	/	/	815	300	/
[15]	IGZO	> 0	40	40	80	570	340	51*
[16]	IGZO	> 0	35	80	115	1930	227	55
		> 0	35	40	75	1330	325	
[5]	In_2O_3	–3.8	50	/	/	2620	43	10*
This work	In_2O_3	0.22	**40**	**80**	**120**	**2160**	**133**	**36**
		0.01	**40**	**40**	**80**	**1570**	**140**	(22*)

***L_T extracted by TLM, which might be inaccurate and underestimated**

References: [1] S. Datta *et al.*, *IEEE Micro*, pp. 8, 2019. [2] Z. Lin *et al.*, *2024 VLSI*, T4-3. [3] Z. Zhang *et al.*, *IEEE-EDL*, p. 1905, 2022. [4] Z. Zhang *et al.*, *2023 VLSI*, T11-3. [5] C. Niu *et al.*, *2023 IEDM*, 37-2. [6] Jian-Yu Lin *et al.*, *2024 IEDM*, 12-6. [7] M. Si *et al.*, *Nat Electron*, 164–170, 2022. [8] M. Si *et al.*, *Nano Letters*, pp. 500, 2021. [9] H. Murrmann and D. Widmann, *IEEE-TED*, pp. 1022, 1969. [10] Van de Walle *et al.*, *Nature*, p. 626, 2003. [11] J. Robertson, *J. Vac. Sci. Technol.*, pp. 1785, 2000. [12] P. -Y. Liao *et al.*, *2022 IEDM*, pp. 12.4.1. [13] Y. Kang *et al.*, *IEEE-TED*, pp. 4692, 2024. [14] Eknath Sarkar *et al.*, *2024 IEDM*, 12-1. [15] Z. Wu *et al.*, *IEEE-EDL*, pp. 408, 2024. [16] W. Zhao *et al.*, *2024 IEDM*, 36-2.