Towards Reliability- & Variability-aware Design-Technology Co-optimization in Advanced Nodes: Defect Characterization, Industry-friendly Modelling and ML-assisted Prediction

Zhigang Ji, Yongkang Xue, Pengpeng Ren, Jinfeng Ye, Yu Li, Yishan Wu, Da Wang, Shuying Wang, Junjie Wu, Zirui Wang, Yichen Wen, Shiyu Xia, Lining Zhang, Jianfu Zhang, Junhua Liu, Junwei Luo, Huixiong Deng, Runsheng Wang, Lianfeng Yang, and Ru Huang

***Abstract*— Reliability- & variability-aware Design Technology co-optimization (RV-DTCO) becomes indispensable with advanced nodes. However, four key issues hinder its practical adoption: the lack of characterization technique that offer both accuracy and efficiency, the lack of defect model with long-term prediction capability, the lack of compact model compatible with most EDA platforms, and the low efficiency in circuit-level prediction to support frequent iterations during co-optimization. Demonstrating with 7nm technology, this work tackles these issues by developing an efficient characterization method for separating defects, introducing a comprehensive test-data-verified defect-centric physical-based model & an industry-friendly OMI-based compact model, and proposing a machine learning-assisted approach to accelerate circuit-level prediction. With these achievements, a RV-DTCO flow is established and demonstrated on 3nm GAA technology to bridge the material level to the circuit level. The work paves ways in boosting adoption of RV-DTCO in both circuit design & process development for ultimate nodes.**

Index Terms— Design Technology co-optimization (DTCO), FinFET, reliability, variability, Discharging-based multi-pulse technique (DMP), OMI, ST-GNN

I. Introduction

IN the era of post-Moore where performance growth becomes difficult from simply the technology scaling, Design-Technology co-optimization (DTCO) has been considered indispensable [1]. Comparing with the optimization for power/performance/area (PPA), the incorporation of reliability and variability considerations becomes essential for mission-critical applications, such as those found in the automotive [2] and healthcare sectors [3]. While DTCO with time-zero variability has been well-established [4], addressing time-dependent reliability and variability within the DTCO framework remains a challenging endeavor. Recent efforts have been made to develop a reliability- and variability-aware DTCO (RV-aware DTCO) flow to tackle these challenges [5]–[9]. However, for the practical adoption of the RV-aware DTCO flow, several critical issues need to be addressed and resolved, which include:

i. Efficient and accurate characterization technique is required. Existing characterization methods can be broadly categorized into bottom-up and top-down approaches. The bottom-up approach, such as random telegraph signal (RTS) [10], [11] and time-dependent defect spectroscopy (TDDS) technique [12], investigate the statistics of individual fluctuations in selected samples and thus are quite time-consuming. The top-down approach, such as extended measure-stress-measure (eMSM) [13], [14] and on-the-fly (OTF) [15], [16], captures the devices' macroscopic degradation and relies on mathematical fitting to decompose the contributions of different types of traps, however, too many fitting parameters may lead to uncontrollable errors in long-term prediction.

ii. Existing effects (reliability, variability) lack test-proven capability for long-term prediction, which is critical for analog & mixed-signal designs. Due to the stochastic nature of oxide defects, time-dependent variability induced by device degradation becomes critical, potentially impacting the end-of-life performance of circuits. Consequently, there is an urgent need for a test-proven prediction methodology that can accurately assess long-term variability, in addition to reliability. To the best of our knowledge, a comprehensive and reliable solution that addresses these challenges has yet to be developed.

iii. Solutions based on the standard model interface are needed. As an interface for the modification of SPICE model parameters that is supported by mainstream simulators, the CMC open model interface (OMI) [17]–[19] has gradually become an industry standard platform for evaluating circuit reliability. By developing reliability models that are compatible with the standard interfaces such as OMI, the proposed solutions can be applied to various circuit simulators, thus reducing the support costs and improving accessibility for both suppliers and end-users alike.

iv. A fast assessment methodology that applicable for (large-scale) circuit-level reliability is needed. Since RV-aware DTCO necessitates numerous iterations between

Yongkang Xue, Pengpeng Ren, Jinfeng Ye, Yishan Wu, Da Wang, Shuying Wang, Junjie Wu, Yichen Wen, Shiyu Xia and Zhigang Ji are with the National Key Laboratory of Science and Technology on Micro/Nano Fabrication, Shanghai Jiaotong University, Shanghai 200240, China (e-mail: zhigangji@sjtu.edu.cn).

Yu Li, Zirui Wang, Lining Zhang, Runsheng Wang and Ru Huang are with the School of Integrated Circuits, Peking University, Beijing 100871, China (e-mail: r.wang@pku.edu.cn).

Junwei Luo and Huixiong Deng are with Center of Materials Science and Optoelectronics Engineering, UCAS, Beijing 100871, China (e-mail: hxdeng@semi.ac.cn).

Lianfeng Yang is from Primarus Technologies, Shanghai, China.

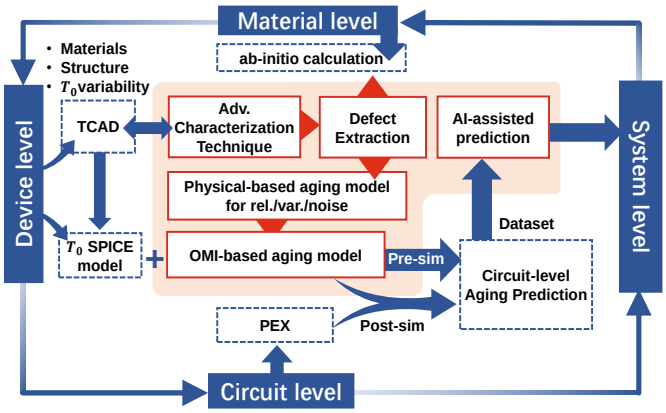


Fig.1. Reliability- & Variability-aware Design-Technology Co-optimization (RV-aware DTCO) methodology from material to circuit level proposed in this work. Based on the proposed defect-centric characterization and extraction technique, Physical-based and OMI-based aging model is established. ML-assisted prediction method is proposed to accelerate aging prediction of large-scale circuits.

process and circuit design, efficient evaluation techniques are crucial for minimizing development time and costs. Traditional approaches employed in commercial EDA tools often require substantial computational resources for circuit-level reliability prediction, resulting in slow and time-consuming analyses. Therefore, developing a fast and efficient (large-scale) circuit-level reliability evaluation methodology would not only reduce costs but also significantly shorten the development period of DTCO, paving the way for rapid innovation and the implementation of reliable electronic devices in various applications.

In this work, we proposed a novel RV-aware DTCO framework aiming to resolve above issues and accelerate the practical adoption of RV-aware co-optimization from material to circuit (**Fig.1**). This is made possible by achieving the following advances in this work:

1) An analytical method is proposed to separate different types of defects directly from the measured degradation, enabling accurate characterization of each type of defect while eliminating the need for time-consuming and laborious statistical analysis.

2) Proposed a defect-centric physical-based model that enables accurate long-term reliability and variability predictions, as well as an OMI-based compact model for effective simulation in the circuit level.

3) The proposed physical model is not only verified by

TABLE I

	This work	2022 IRPS [5]	2021 VLSI [6]	2019 TED [7]	2013 IEDM [8]
Experimental extraction of defect	✓	✓	✗	✗	✓
Ab-initio verified	✓	✓	✗	✗	✗
Independent-verified long-term physical model Rel. / Var.	✓	✓	✗	✗	✗
Compact model	✓	✗	✓	✓	✓
Industry standard Interface(OMI)	✓	✗	✗	✗	✗
Acceleration for fast iteration in DTCO flow	✓	✗	✗	✗	✗

Advances of the existing solutions to address the challenges of RV-aware DTCO. Our proposed methodology includes all the listed features from Ab-initio calculation, experimental extraction of defect, reliability & variability simulation to acceleration technique for large-scale circuits assessment.

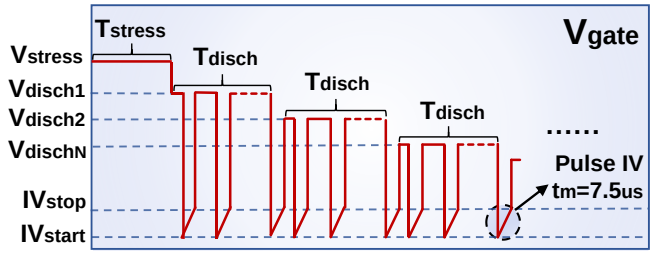


Fig.2. Illustration of the discharging-based multi-pulse (DMP) technique. In the charging phase, NBTI stress is applied to the DUT under desired stress level (V_{stress}) for pre-set time (T_{stress}). And in the discharging phase, gate voltage is lowered down step-by-step to gradually discharge the traps. Drain, source and substrate terminals should be connected to ground all the time except for Id-Vg measurement.

independent test results but the extracted trap properties can be well correlated to ab-initio calculation, which can be used in investigating the physical origins of traps and helpful for process optimization.

4) Proposed a Machine learning (ML)-based approach for the fast circuit-level RV prediction, providing a feasible solution for accelerating iterative processes in the DTCO framework.

Table.I summarizes the key advancements of our proposed framework over existing solutions in the field.

II. ADVANCED DEFECT SEPARATION AND CHARACTERIZATION TECHNIQUE

A. Discharging-based multi-pulse technique

By trap filling and the following gradual discharging, the discharging-based multi-pulse technique (DMP) [20] has been shown as a powerful tool in understanding oxide traps in transistors with different structures [21], channel materials [22], and dielectrics [23]. The test pattern of DMP is shown in **Fig.2**. In the charging phase, the device under test (DUT) is prepared by biasing the gate to the desired stress level ($V_{gstress}$) for the pre-set time (T_{stress}) while connecting the other terminals (source, drain & substrate) to ground. In this work, the time duration of 10s, 100s, 1000s, and 10000s are used during stress phase. And then in the discharging phase, the gate voltage is lowered down by ΔV to the first discharging level, $V_{disch1} = V_{gstress} - \Delta V$. This discharging phase will last for T_{disch}, during which several I_d-V_g measurements will be carried out. To reduce trap's recovery during the measurement, I_d-V_g curve is captured at the pulse edge within 7.5us. Once reaching T_{disch}, the discharge voltage lowers down to $V_{disch2} = V_{gstress} - 2*\Delta V$ for another T_{disch}. By repeating this procedure until V_{dischN}

TABLE II

Detailed information of DMP test in this work	
V_{stress} (or $V_{gstress}$)	-0.8V \ -1.0V \ -1.2V \ -1.4V \ -1.5V \ -1.6V \ -1.7V \ -1.8V
T_{stress}	10s \ 100s \ 1000s \ 10000s
V_{disch1}	$V_{stress} - \Delta V$
ΔV	-0.1V
V_{dischN}	+1V
T_{disch}	5s

1. V_{dischN} is the last discharge voltage in DMP procedure.
2. At each V_{dischN} , the measure-stress-measure (MSM) sequence was carried out and lasts for 5s (T_{disch}).

Detailed information of DMP test in this work.

reaches +1.0V in the N_{th} step. This V_{dischN} is properly selected to ensure the probing of deep traps while not introducing extra e-trapping. As mentioned above, we call the test pattern of **Fig.2** as one DMP test, if we repeat the DMP test on the same device, and each time the stress voltage ($V_{gstress}$) is greater than the last time, we call this test procedure as multi-DMP (m-DMP) test [23]–[25].

In this work, we chose the industrial-grade 7nm pFinFETs for the demonstration. The test is carried out by using the Keysight B1530A equipped with waveform generator / fast measurement unit (WGFMU). The pulse IV is performed in linear region (I_{dlin}) with V_d = -0.05V, and the threshold voltage V_{th} is monitored at a constant drain current of 1 μA × W/L. The testing temperature is 125 °C , and other detailed information of DMP test is shown in **Table.II.**

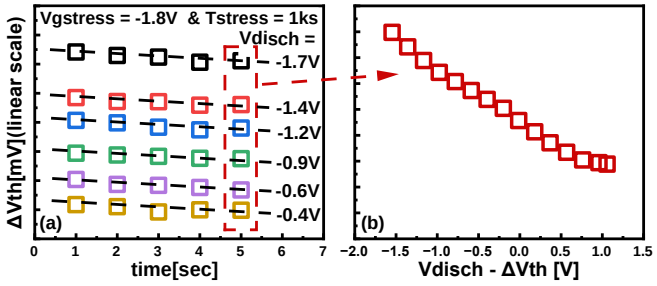


Fig.3. (a) The discharging trace under each V_{disch} in DMP test. (b)By extracting the last point under each V_{disch}, the relationship between the discharging voltage and the corresponding degradation can be obtained.

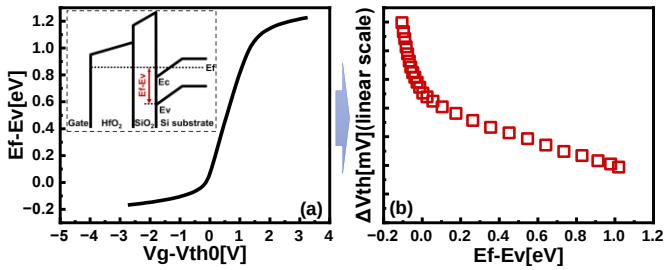


Fig.4. (a) The E_f-E_v against V_g-V_{th0} from calibrated TCAD simulation. V_{th0} is the threshold voltage of a fresh device without suffering degradation. (b) By converting V_{disch} -V_{th0} -ΔV_{th} to the E_f -E_v, the trap's equivalent energy level can be obtained. (V_{disch} is equivalent to V_g)

B. DMP test result preprocessing

After NBTI stress for preset stress voltage (V_{stress}) and stress time (T_{stress}), the traps are gradually detrapping under each discharge voltage (V_{disch}) as shown in **Fig.3a**. By extracting the last point under each V_{disch}, the relationship between the discharging voltage and the corresponding degradation can be obtained as shown in **Fig.3b**. Through the relationship between E_f -E_v and V_g -V_{th} obtained by calibrated TCAD simulation (**Fig.4a**), we can convert V_{disch} -V_{th} -ΔV_{th} to the energy level of E_f with respect to E_v at the Si/interfacial layer (IL) interface, i.e. (E_f -E_v) [26], and then the trap's density distribution against different energy levels can be obtained (**Fig.4b**). It is worth noting that such an energy level is the equivalent energy level that assumes the traps are located at the interface between the Si/IL, but not the true trap level in the system. Such transformation aims at the separation of different traps based on their energy profiles in the next section.

C. Equivalent energy profile of different traps

The separation flow of different traps based on their equivalent energy profiles is shown in **Fig.5**: When the

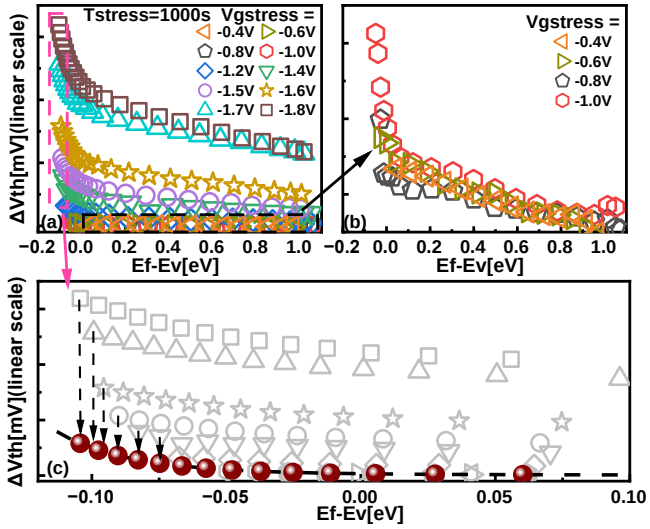


Fig.6. (a) The trap's energy profiles under different charging bias. (b) When charging bias is low, the energy profiles at different $V_{gstress}$ overlap well, , indicating that the trap energy level does not change after charging. (c) By shifting down the energy profiles of the higher bias and aligning them with profiles under lower bias, the energy profile of Type-A trap can be obtained.

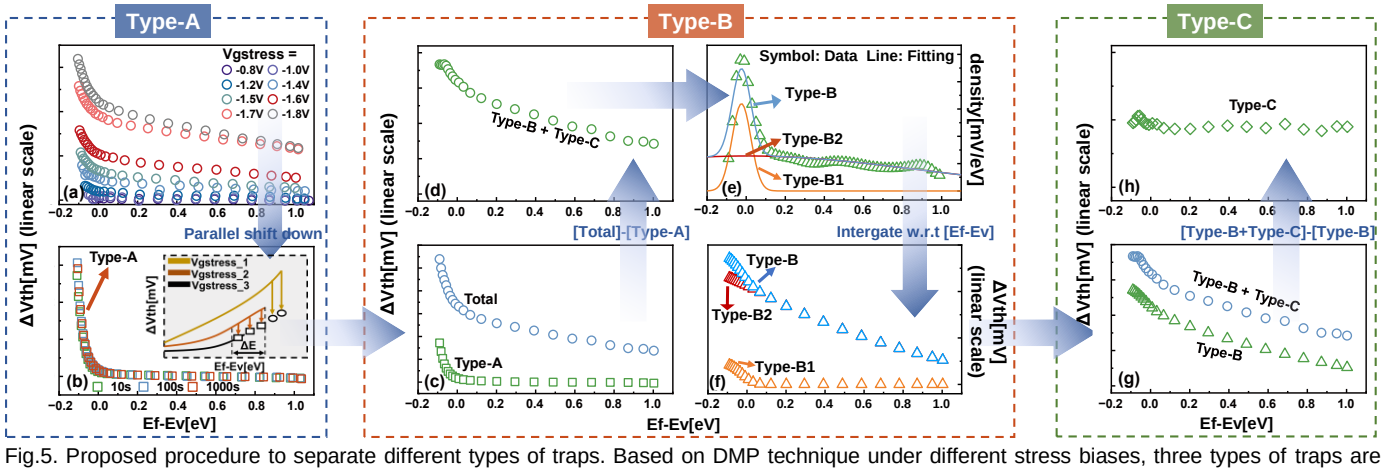


Fig.5. Proposed procedure to separate different types of traps. Based on DMP technique under different stress biases, three types of traps are clarified and separated. Type-A captures holes without changing energy levels while type-B shifts energy levels after hole capturing. Type-C presents constant energy levels and agree with interface states.

charging bias ($V_{gstress}$) is low, the energy profiles extracted from discharging after filling at different $V_{gstress}$ overlap well (**Fig.6b**), while they deviate from each other for higher $V_{gstress}$ (**Fig.5a & Fig.6a**), suggesting that there exist two types of traps in the oxide for such phenomenon [22]. For Type-A traps, if filled at a certain voltage, it can also be released at the same voltage when the bias is lowered, indicating that the trap energy level does not change after charging. In contrast, Type-B traps need to be released at an energy level lower than where they are filled, implying that Type-B may shift energy levels to a deeper level after hole capturing and exhibit the nature of switching traps. Because Type-B can only be discharged at a much lower energy level, for every two neighboring curves, by shifting down the energy profiles of the higher bias and aligning them with profiles under lower bias, Type-B traps will be removed. As illustrated by the red points in **Fig.6c**, the aligned curve represents the energy profile of Type-A trap.

According to the above method of separating Type-A, the Type-A traps extracted from multi-DMP measurements under different charging time is shown in **Fig.5b**, they overlap each other, suggesting the nature of fast saturation.

We then subtract Type-A from the total degradation, as shown in **Fig.5c&d**, the remaining part contains both oxide traps and interface traps. What is worth noting is that even during discharging under different E_f-E_v, the threshold voltage (V_{th}) is always sensed at the constant current, which can be approximately considered as the same surface potential. Therefore, the contribution from interface traps (labeled as Type-C traps) in the measured V_{th} shift is the same for different E_f -E_v. The impact of interface traps can thus be removed through the mathematical differentiation with respect to the energy levels. As shown in **Fig.5e**, Type-B shows a broad distribution in energy levels and it cannot be fitted by a single Gaussian distribution. Therefore, we introduce a dual Gaussian distribution and a good fitting can be achieved, which indicates the existence of two types of oxide traps. We label the narrow one as Type-B1 and the wider one as Type-B2. By fitting with two Gaussian distributions and integrating along the energy levels, the total Type-B traps (Type-B1&Type-B2) can be extracted, as shown in **Fig.5f**. Finally, by subtracting Type-A, Type-B1, and Type-B2 from the total degradation in **Fig.5g**, Type-C can be obtained, exhibiting a constant value against energy level in **Fig.5h**, which seems that it is hard to be "discharged". However, just as mentioned above, its independence of energy level is just because of the same sensing level.

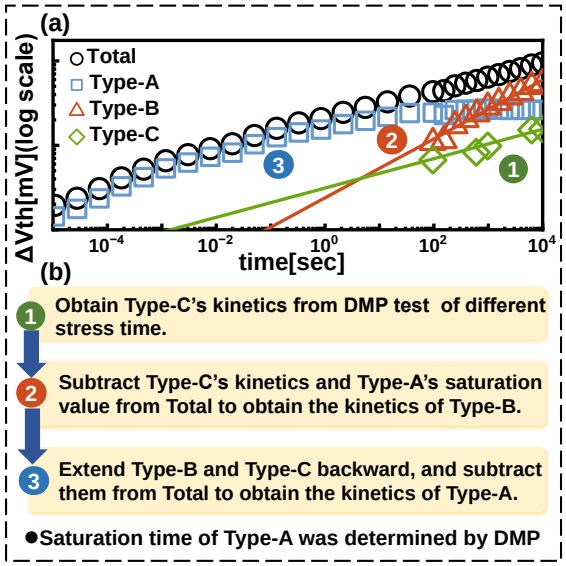


Fig.7. (a) The extracted charging kinetics of three types of traps under specific stress condition. (b) Charging kinetics extraction procedure of three types of traps.

D. Charging kinetics of different traps

Understanding the charging kinetics of each type of traps can be critical for modelling long-term reliability and variability. To separate different traps from their charging kinetics, the extended measure-stress-measure (eMSM) sequence is adopted. The kinetics of Type-C trap (i.e. interface traps) should be extracted first from DMP test of different stress time as shown as '◇' in **Fig.7a**. Since Type-A can quickly saturate (**Fig.5b**), the kinetics of Type-B at longer time can be derived by subtracting Type-C and the saturated Type-A ('Δ' in **Fig.7a**). By assuming both Type-B and Type-C traps exhibit a power law relationship, and back-extrapolating them, the kinetics of the Type-A trap can be obtained ('□' in **Fig.7a**) via subtracting

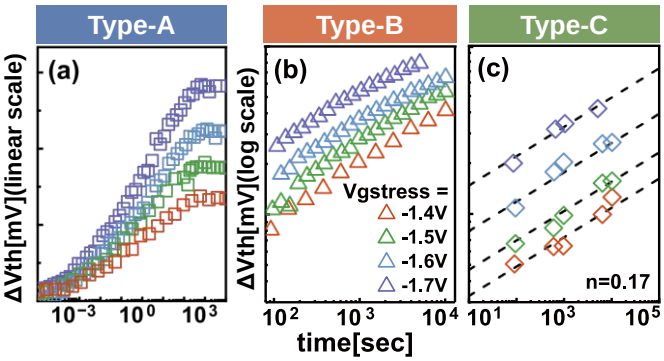
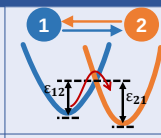
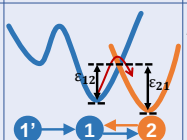


Fig.8. The extracted charging kinetics of (a) Type-A traps, (b) Type-B traps and (c) Type-C traps with stress time under varying stress biases. The charging kinetics can be well fitted by power law relationship with the time exponent of 0.17.

TABLE III

	Model		Fitting parameters
Type-A (2 state NMP)		$k_{12} = pv_{th}\sigma\vartheta e^{-\beta\varepsilon_{12}}$ $k_{21} = N_v v_{th}\sigma\vartheta e^{-\beta\varepsilon_{21}}$	$(<E_T>, \sigma_{E_T}, <S>, \sigma_S, \sigma_{ccs}, N_T)$
Type-B (3 state NMP)		$k_{11'} = A * exp(B * E_{ox}) * [1 - exp(-(t/\tau)^{\beta})]$ $k_{12} = pv_{th}\sigma\vartheta e^{-\beta\varepsilon_{12}}$ $k_{21} = N_v v_{th}\sigma\vartheta e^{-\beta\varepsilon_{21}}$	$(<E_T>, \sigma_{E_T}, <S>, \sigma_S, \sigma_{ccs}, N_T, A, B, \tau, \beta)$
Type-C (power-law)	——	$\Delta V_{th} = A * V_g{}^m * t^n$	(A, m, n)

Established Physical-based defect model for three types of traps and their corresponding parameters used for fitting. In these equations, p is the concentration of holes in the channel, N_v is effective density of states valance band, v_{th} is the thermal velocity of the carriers within the channel, σ represents capture cross section, and ϑ denotes Wentzel-Kramers–Brillouin (WKB) tunneling factor, ε_{12} and ε_{21} denote the barrier heights for transitions between these two wells, respectively. Other parameters such as A, B, τ, β, m, n are fitting parameters.

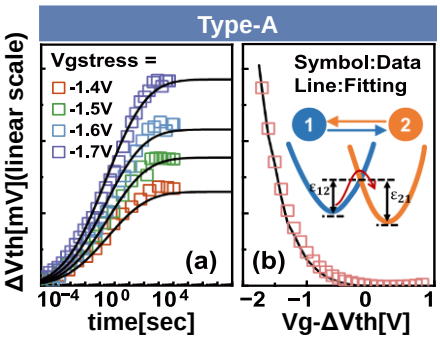


Fig.9. (a) The charging kinetics and (b) equivalent energy level distribution of Type-A traps can be well modelled by the two-state non-radiative multiphonon(NMP) theory.

Type-B and Type-C from the total degradation. **Fig.7** summarizes the entire extraction procedure. By repeating the procedure, the kinetics of different types of traps under different voltages can be obtained (**Fig.8**).

E. Physical-based Defect Model

After directly separating different types of traps from the experimental data following the characterization technique above, we can extract physical parameters of each trap according to their respective physical-based model by fitting. In this work, we assumed that, for each type of trap, its physical property including the energy level and the relaxation energy follows the Gaussian distribution. Their mean values and variances are obtained by fitting each trap's charging kinetics and their equivalent energy level distributions (**Fig.9&10**) with their respective model in **Table III**. What is worth noting is that due to the incomplete discharging at each discharge phase in DMP test, we also put discharging time, discharging voltage of each discharge phase and the corresponding ΔV_{th} into energy profile's fitting.

Type-A traps: Type-A traps capture a hole without altering their energy levels, which suggests that the microscopic structure undergoes negligible rearrangements. This behavior is consistent with the properties of structural defects that are widely observed in amorphous systems, where the disordered nature of these materials leads to a variety of defect states [27]. To better understand the trapping/de-trapping process of Type-A traps, a simplified two-state model based on the non-radiative multiphonon (NMP) theory is employed [28], [29]. This model effectively captures the fundamental mechanisms governing the transitions between the two states.

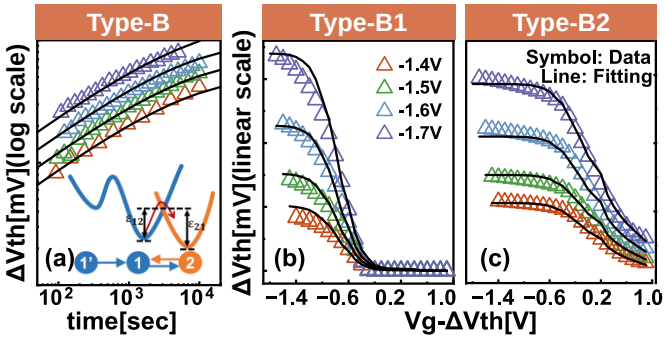


Fig.10. (a) The charging kinetics of Type-B (B1&B2) traps, (b) equivalent energy level distribution of Type-B1 traps, and (c) equivalent energy level distribution of Type-B2 traps can be modelled by incorporating the activation state into the two-state NMP theory.

The schematic of the two-state model is shown in **Table.III**, in which energy wells 1 and 2 represent the energy of two states before and after trapping, respectively. ε_{12} and ε_{21} denote the barrier heights for transitions between these two wells, respectively. Once the negative stress voltage is applied, the barrier height ε_{12} is lowered, which facilitates more holes jump over the barrier through thermal emission. To account for the distribution, a Gaussian distribution for the trap energy (E_T) and relaxation energy (S) is assumed. As shown in **Fig.9**, the charging kinetics and equivalent energy level distribution of Type-A traps can be well described by the 2-state NMP model.

Type-B traps: To account for the lattice relaxation process, the behaviors of both Type-B1 and Type-B2 traps are modelled by a three-state process as shown in **Table.III**, which captures the complex dynamics involved in the charging process and provides a more accurate representation of such type of traps. During charging, the trap needs to be initially activated from state 1' to state 1, after which the charging can proceed from state 1 to state 2. By fitting the charging kinetics of Type-B traps (i.e. Type-B1&Type-B2 obtained from **Fig.8(b)**) and the DMP results of both Type-B1 and Type-B2 (data from **Fig.5f**), the model parameters can be determined. As demonstrated in **Fig. 10a-c**, the lines generated from the model exhibit a good agreement with the experimental results. It is worth noting that without considering the activation process (1'->1), the kinetics of Type-B traps cannot be well fitted, which further confirms the switching nature of Type-B traps, highlighting the necessity of incorporating this additional state into the modeling process.

Type-C traps: As depicted in **Fig. 8c**, the degradation behaviour of Type-C traps can be described using the classical power law relationship. The time exponent is found to be 0.17, which is also broadly observed in recent years across multiple technologies by using the delay-corrected DCIV technique [30] or by probing V_{th} degradation under conditions where hole trapping is minimal. Historically, the time exponent was considered as a signature of interface state generation controlled by the Reaction-Diffusion (R-D) process. In this process, it is assumed that the Si-H bond breaks at the Si/IL interface and diffuses as a hydrogen molecule into the oxide [31]. However, theoretical studies have suggested that the Si-H bond can be rather stable [32], casting doubt on the validity of the R-D process as the primary mechanism. Recently, an alternative explanation has been proposed, suggesting that the breakage of Si-H bonds may be resulted from their reaction with atomic hydrogen originating from either the substrate [33] or the gate [34]. This reaction proceeds through an exothermic process: Si-H + H → Si- + H_2. This new hypothesis provides a more plausible mechanism for the observed behaviour, taking into account the stability of the Si-H bond.

III. AB-INITIO CALCULATION

A. Discussion on the physical origin

The trap level (E_T) and relaxation energy (S) can be used as the signature to pursue the physical origin of the traps. Therefore, we can compare the E_T and S values obtained by fitting the experimental data with the results calculated by ab-initio calculation for the defects of different configuration to explore the potential physical origin of the three types of traps.

The formation energy of defects can be evaluated as

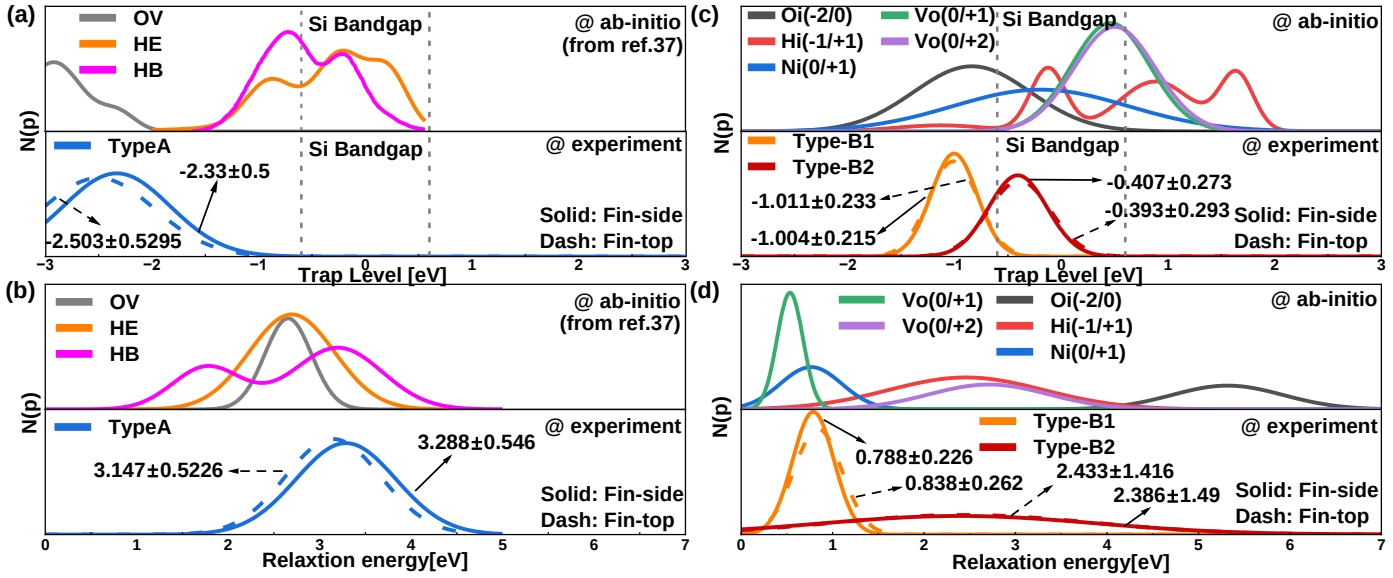


Fig.11. Comparison of the Ab-initio calculation and experiment for (a) energy level and (b) relaxation energy of Type-A in the IL; (c) energy level and (d) relaxation energy of Type-B (B1&B2) in the HK layer.

$$E_F = E(\partial, q) - E(host) + \sum_i n_i(E_i + \mu_i) + q[E_{VBM}(host) + E_f + \Delta V]$$

where ∂ is the type, q is the charge state, $E(\partial, q)$ is the system energy with defect and charge, *E(host)* is the energy of the perfect system, n_i is the number of element, μ_i is the chemical potential relative to the element material, E_{VBM} is the host valence band maximum (VBM). ΔV is the difference of electric potential between perfect and defect systems with a place far away from the defect. E_T between the charge q_1 and q_2 can be extracted from the Fermi level where the formation energy of q_1 is equal to that of q_2. S can be obtained through relaxing the configuration from the equilibrium for charge state q_1 to the equilibrium for another charge state q_2 [35].

For defects in SiO_2, several candidates have been suggested, including the oxygen vacancies (Vo) [36], the hydrogen bridge (HB) and the hydroxyl E' centers (HE) [37]. E_T and S have been systematically calculated in ref.37. The comparison between the theoretical and the experimental values is shown in **Fig.11a&b**. Both HB and HE defects have a considerable part of components within the Si bandgap. However, Type-A traps are situated below the Si valance band, and their energy levels are relatively deep. Consequently, these two defect types may not be suitable candidates for Type-A traps. Oxygen vacancies (Vo) are the only defects exhibiting deep energy levels. Although the energy level of Type-A traps is slightly shallower than that of Vo, considering the potential calculation error of 0.5eV in density functional theory (DFT) [37], a strong correlation between Type-A and Vo is plausible. Furthermore, the experimental and theoretical values of S are in good agreement. Therefore, Type-A can be considered as the pre-existing trap originating from oxygen vacancy. From this perspective, this finding aligns with the aforementioned assertion that Type-A traps are structural defects introduced during the fabrication process due to imperfections.

As depicted in the lower panel of **Fig.11c**, both Type-B1 and Type-B2 traps are shallower than Type-A traps. We constructed several potential defect structures commonly observed in high-k (HK) layers, including oxygen vacancy (Vo) [38], Interstitial hydrogen (Hi) [39], Interstitial Nitrogen (Ni) [40] and Oxygen interstitials (Oi) [38]. The amorphous nature of the HK layer is also taken into consideration. The calculated results are presented in the upper panels of **Fig.11c&d**. The comparison with the experimental data suggests that Type-B1&B2 traps may originate from Ni and Hi, respectively. Since the concentration of H can be much larger than that of N, Type-B2 traps are far more abundant than Type-B1 traps, which is confirmed by the experimental data (**Fig.5e-f**). Interstitial H could be stabilized as either the hydrogen bridge (HB) or the hydroxyl E', which undergoes a transition to the excited vibronic state after capturing a hole. By dissipating the excess energy through the multi-phonon emission process, they can relax to a lower energy level.

IV. MODEL VALIDATION FOR PREDICTIVE CAPABILITY

A. Validation for Reliability Predictive Capability

Irrespective of the device geometry, degradation is perceived as an ensemble of traps of different types with different filling statuses. Once the average trap number of each type of traps is determined by the extracted areal density, the average degradation can be estimated by calculating the filling occupancy through the corresponding physical processes and

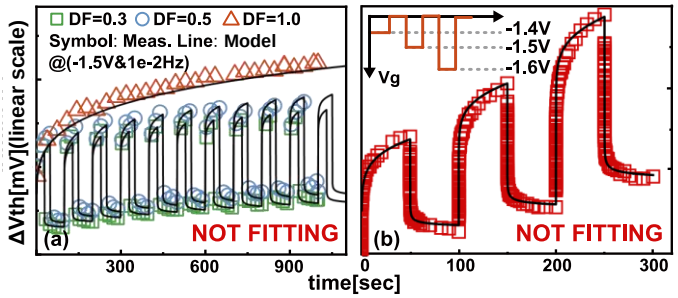


Fig.12. Comparisons of the proposed model and experiment data under (a) DC & AC conditions and (b) arbitrary waveform condition. All show good agreement between the prediction and the measurement.

summing up the results for all traps.

To demonstrate the prediction capability of the proposed model, we calculated degradation kinetics under direct current (DC) conditions, alternating current (AC) conditions with varying duty factors (**Fig.12a**), and arbitrary waveform conditions (**Fig.12b**). The results were then compared with experimental data. The comparison for duty factor and frequency dependences is also shown in **Fig.13**. Good agreement between model's predictions and the measured data serves as robust evidence supporting the validity of the proposed model. ("NOT FITTING" is used to highlight that good agreements are not from fitting data.)

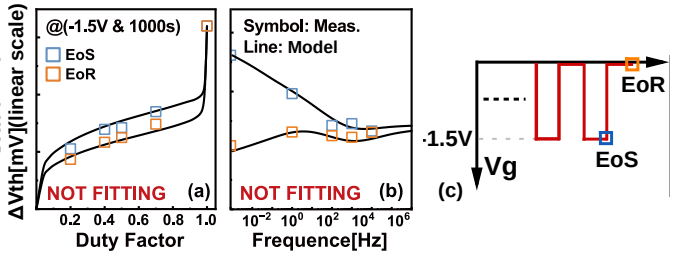


Fig.13. Comparisons of the proposed model and experiment data under AC conditions with different (a) duty factors and (b) frequencies. (c) The degradation is measured both under end of stress (EoS) and end of recovery (EoR).

Traditionally, long-term reliability under low stress voltage conditions can be predicted through extrapolation using simple power-law [41], log-law [42], or saturation power-law [43] relationships against voltage and time. We extracted the relevant model parameters by fitting experimental data under high stress voltages (ranging from -1.4V to -1.7V), and compared our model with traditional methods under V_g = -1.2V and T_{stress} = 50ks, as illustrated in **Fig.14**. The results demonstrate that our model has good long-term reliability prediction capability, whereas the power-law method tends to overestimate device degradation, and both log-law and

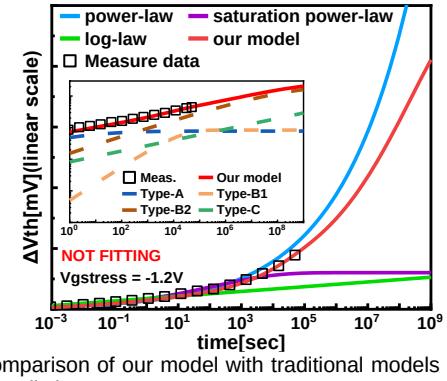


Fig.14. Comparison of our model with traditional models in long-term reliability prediction.

saturation power-law methods tend to underestimate it.

B. Validation for Variability Predictive Capability

When time-dependent variation (TDV) needs to be considered, the average threshold voltage shift (η) resulting from a single defect can be estimated through the charge sheet model. By assuming the Poisson-distributed trap number and the exponential-distributed ΔV_{th} fluctuation [44], [45], Monte-Carlo simulations can be performed to predict the temporal degradation for multiple devices over time. As shown in **Fig.15**, the red line represents the TDV measured on small-size devices under $V_{gstress}$ = -1.3V, while the gray line represents the simulation results obtained through the aforementioned method. The red line essentially falls within the range of the gray line, indicating that our model also has good variability predictive capability. However, to establish a more accurate TDV model, it is necessary to obtain the average number of defects (N) for small devices and the average threshold voltage shift (η) caused by a single defect through statistical experimentation. This will be addressed and improved in our subsequent work, further refining the model's

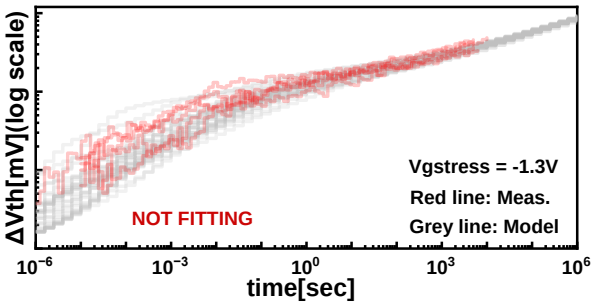


Fig.15. Monte-Carlo simulation results of the temporal device-to-device variation (DDV) induced by device degradation. The red line essentially falls within the range of the gray line, indicating that our model also has good variability predictive capability.

predictive capabilities.

We emphasize that all model parameters were extracted from DC stress data and energy profiles in **Fig.8-10** only. All the test data shown in **Fig.12-15** were not used for fitting. Moreover, all the predictions using the proposed model are based on the single set of model parameters.

V. OMI-BASED AGING MODEL

To facilitate the circuit-level prediction with good compatibility with various circuit simulators and EDA platforms, the CMC (Compact Model Council) open model interface (OMI) is adopted. By simplifying the proposed model and assuming that the waveforms of each node can be approximated as square waves, we can model different

TABLE IV

	Stress stage	Recovery stage
Type-A	$TypeA = AR_{1(A)} \cdot \left[1 - exp\left(-\left(\frac{t_s}{\tau}\right)^{\beta}\right)\right]$	$\frac{TypeA}{TypeA(0)} = AR_{3(A)}$
Type-B1	$TypeB_1 = AR_{1(B_1)} \cdot t_s^{n_1}$	$\frac{TypeB_1}{TypeB_1(0)} = AR_{3(B_1)}$
Type-B2	$TypeB_2 = AR_{1(B_2)} \cdot t_s^{n_2}$	$\frac{TypeB_2 - 0.4 \cdot TypeB_2(0)}{TypeB_2(0) - 0.4 \cdot TypeB_2(0)} = AR_{3(B_2)}$
Type-C	$TypeC = AR_{2(C)} \cdot t_s^{n_3}$	—
Age Rate	$AR_1 = p_1 \cdot exp(p_2 \cdot V_{gstress})$ $AR_2 = p_3 \cdot V_{gstress}^m$	$AR_3 = \frac{1}{1+B \cdot t_r^n} \cdot f(V_{gstress}, t_{eff\_stress})$
Total	$TD = TypeA + TypeB_1 + TypeB_2 + TypeC$	

Simplified OMI-based compact model for all types of traps, including the stress and recovery stages. In the stress stage, voltage is converted into the Age Rate (*AR*) with the exponential model and power-law model, the time kinetic of type-A was characterized by the stretch-exponential model, and time kinetics of other types of traps were characterized by the power-law model. In the recovery stage, recovery rate (AR_3) was characterized by an expression with stress time and stress voltage and recovery time. The total degradation (*TD*) is obtained by linear superposition of all types of traps.

operating voltages and fixed recovery voltages at 0V. The Simplified OMI-based compact model is presented in **Table. IV**. This compact model encompasses all characterized traps, and their kinetics can be divided into two stages: stress stage and recovery stage.

During the stress stage, operating voltage is converted into the Age Rate (AR) with the exponential model for Type-A & Type-B traps and the power-law model for Type-C traps. Time kinetics are characterized by the stretch-exponential model for Type-A traps and the power-law model for other trap types.

During the recovery stage, the recovery rate (AR_3) is characterized by an expression incorporating stress time, stress voltage, and recovery time. Degradation after the recovery stage is determined by the pre-recovery degradation multiplied by AR3. Recovery is only considered for oxide traps while not for interface traps.

The total degradation (TD) is obtained through the linear superposition of all trap types. This compact model enables handling of both stress and recovery stages with arbitrary workloads, making it a versatile tool for predicting and analyzing device degradation and circuit performance.

To validate the accuracy of the OMI-based compact model, we integrate the model into the EDA environment and perform simulations under various operating voltages over an extended period in arbitrary waveform. **Fig.16** shows the device degradation results calculated by both the defect-based physical model and the OMI-based compact model. It is evident from the figure that the results from both models are consistent under any working voltage over a long duration. Good agreement between the two methods demonstrates that the OMI-based compact model is well-suited for circuit-level reliability assessment.

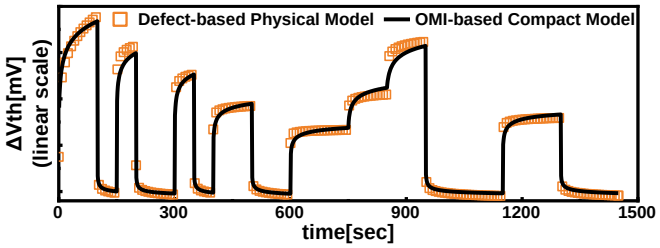


Fig.16. Good agreement between the data calculated by Defect based physical model and OMI-based compact model, making it ready to use for circuit-level reliability assessment.

VI. RV-AWARE DTCO DEMONSTRATION

By integrating the developed method for the trap property extraction, the physical and compact modelling into the proposed DTCO flow, the co-optimization from material level to circuit level become possible (**Fig.1**). The proposed DTCO framework flow is demonstrated on 3nm Gate-All-Around (GAA) technology, considering both process and aging induced device variation.

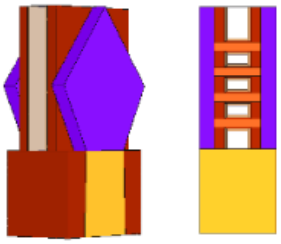


Fig.17. Illustration of the nanosheet structure simulated in this work

TABLE V

Parameter of GAA PMOSFET	Value
Length of Channel (L_{ch})	27nm
Width of Channel (W_{ch})	35nm
Thickness of Channel (H_{ch})	5nm
Thickness of Oxide (T_{ox}, SiO_2 + HfO_2)	0.8nm + 1.5nm
Doping Conc. of Substrate (N_{sub_D})	10^{15} cm^{-3}
Doping Conc. of Source & Drain (N_{sd_D})	10^{21} cm^{-3}
Doping Conc. of Channel (N_{ch_D})	10^{17} cm^{-3}

Related parameters of nanosheet structure in TCAD simulation.

Fig.17 illustrates the 3D structure of the GAAFET, comprising 4 nanosheets with a thickness of 5 nm individually. The length (L_{ch}) and width (W_{ch}) of each nanosheet is 27 nm and 35nm, respectively. The dielectric layer is composed of 0.8-nm-thick SiO2 and 1.5-nm-thick HfO2. Related device parameters are listed in **Table V**.

Based on TCAD simulation, the variation induced by random dopant fluctuation (RDF) [46], metal gate granularity (MGG) [47] and oxide thickness variation (TOV) [48] are taken into account in device time-zero performances simulation using the impedance field method [49]. The impedance field method (IFM) offers a highly convenient, efficient, and accurate technique for statistical variability analysis. The core essence of IFM is to treat randomness as perturbation of reference device. Rather than solving the Poisson and drift-diffusion equations for numerous random device realizations, a 3D TCAD solution is required only once for the reference device. The simulation process can be outlined in the following two steps:

a. Conduct a perturbation simulation at each grid point, evaluating the variation of each relevant quantity;

b. Calculate the linear current response of the TCAD solution based on the precomputed quantities to obtain the corresponding I-V characteristics, and subsequently compute the statistics for all relevant quantities.

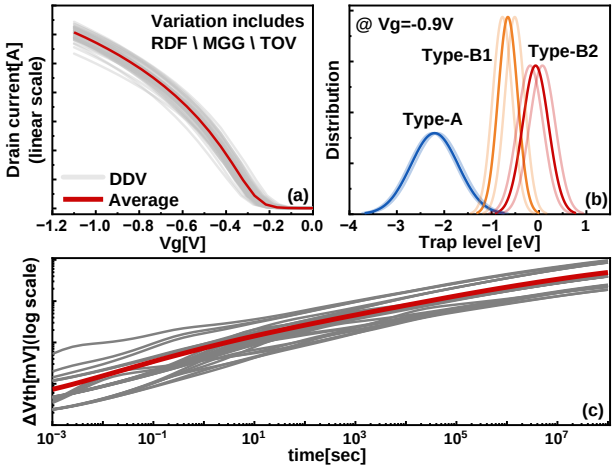


Fig.18. (a) Simulated time-zero performances considering RDF, MGG and TOV variation sources. (b) The modulation of process variation on energy levels of different types of traps. (c)Simulated temporal device variations considering both process variation and aging induced variation by stochastic nature of each type of trap.

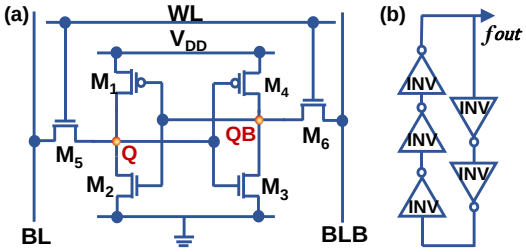


Fig.19. Schematics of the simulated (a) 6T-SRAM cell and (b) ring oscillator (RO). By analyzing the SRAM static noise curve (SNM) and the output frequency of the RO to assess circuit-level reliability.

As is shown in **Fig18a**, the variation of I_d-V_g curve induced by unsatisfactory process factors can be well simulated. Due to the modulation of process variation on the defect's local electrical field, the alignment of defect energy levels with the valence band will be changed, even under identical gate biases, indicating the coupling between process and aging-induced variation (**Fig.18b**). By combining these two sources of variation, the device's variation induced by aging & process induced variation can be predicted (**Fig.18c**).

Extending to circuit level, **Fig. 19** showcases circuits used to evaluate circuit-level reliability, including a 6T-SRAM cell (**Fig.19a**) and 5-stage ring oscillator (RO) (**Fig.19b**). To assess the reliability of the SRAM, we analyze the degradation of the static noise margin (SNM) over time. For the RO, we evaluate its reliability by examining the degradation of the output frequency with aging time and different operating voltages.

Fig. 20(a) shows the degradation of SRAM. The green curve represents the initial SNM with a fresh device, and the dashed rectangle indicates the maximum noise tolerance. As the two pull-up transistors (M1&M4) degrade over time (with only PMOS degradation considered in the simulation), it becomes increasingly difficult for nodes Q and QB to transition from GND to VDD, necessitating a larger negative bias to drive M1 and M4 to pull the nodes up to the VDD level. Consequently, the butterfly curve shifts from green to red, and the solid rectangle, which represents the maximum noise tolerance, continually decreases. **Fig.20(b)** demonstrates the degradation of output frequency in the RO. As the transistors degrade, the

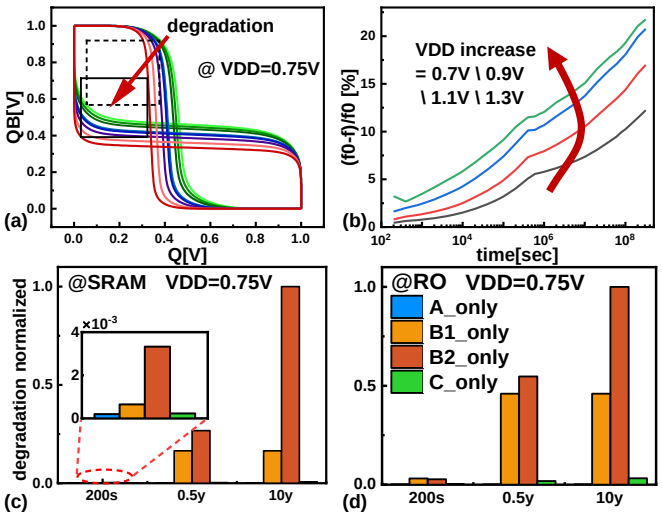


Fig.20. (a) SNM of SRAM and (b) Frequency of RO degrades with VDD and operation time. The contributions of each type of trap to (c) SRAM and (d) RO cell degradation.

inverter's driving ability weakens, and the propagation delay increases, resulting in a continuous decline in the RO's frequency. When the operating voltage (VDD) is 0.7 V, the degradation of output frequency reaches nearly 10% after 10 years. As VDD increases, the RO's lifetime correspondingly decreases.

The above results highlight the severity of circuit-level aging, underscoring the importance of identifying which type of trap causes such significant degradation. By pinpointing the source of this degradation, we can develop strategies to optimize the fabrication process and improve device reliability.

Fig. 20c&d exhibit the contributions of each type of traps to the SRAM and RO cells, respectively. In the initial stage of the SRAM cell, the degradation is relatively small, as shown in **Fig. 20c**. Type-B's contribution rapidly increases with aging time, with Type-B1 and Type-B2 accounting for approximately 40% and 60% after 0.5 years, respectively. Due to the saturation of Type-B1 and the growth of Type-B2, Type-B2's contribution rises to about 90% after 10 years. A similar situation occurs in the RO cell, where Type-B1 and Type-B2 are the primary contributors to circuit degradation, while the impacts of Type-A and Type-C can be largely disregarded.

Consequently, to enhance circuit performance and reliability, it is crucial to eliminate Type-B traps (especially Type-B2) during the process optimization stage.

VII. AI-ASSISTED PREDICTION

Traditional approach of circuit-level aging simulation adopted in commercial EDA tools requires high computation resources. Besides, changing simulation conditions requires re-analysis of the circuit, which takes a long iteration time and significantly increases the cost. The above shortcomings may hinder its use for the large-scale circuit reliability-aware design. To address this challenge, a fast assessment methodology based on spatial-temporal graph neural network (ST-GNN) is proposed. By taking both the structural topology and dynamic operation of the circuit into consideration, reliability prediction can be achieved with both high accuracy and efficiency.

In order to analyze circuits using GNN, it is necessary to convert circuits into graph. From circuit schematic diagram, the device can be treated as the node vector of the graph, and the connection between the device ports can be treated as the edge of the graph. The considered nodes include MOS transistors four terminals, without considering grounding, DC power supply, and AC sources. The impact factors on device reliability are included in the node feature. For example, transistor node X_g is composed of four features, namely effective channel length L_{eff}, effective width W_{eff}, gate source voltage V_{gs}, and drain source voltage V_{ds}. To simplify the

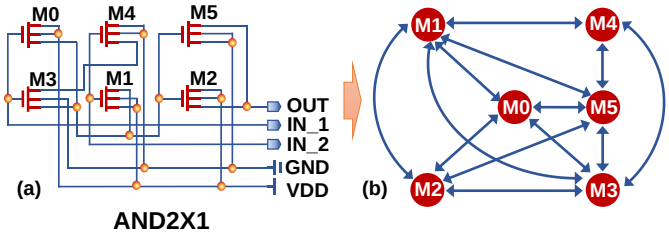


Fig.21. Illustration of how the topology of two-input AND is represented as a graph.

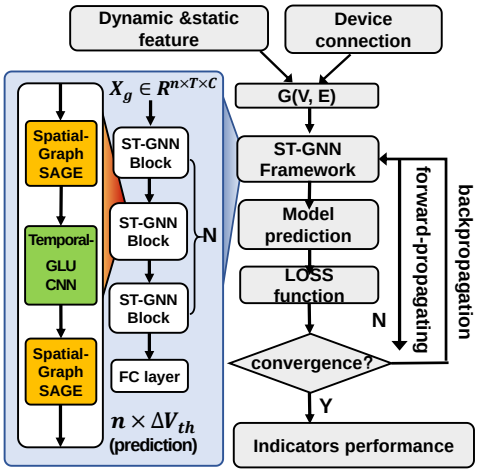


Fig.22. Training flow of the proposed ST-GNN framework. Both circuit topology and temporal information are considered.

procedure, the circuit diagram is treated as an unweighted graph, with weight values only associated with each node while not with edges. Regarding to the edge of the graph, the connections of MOS transistors: gate, source, drain, and substrate is considered as the same type of edge. **Fig.21** shows a dual input AND gate represented as an unweighted isomorphic undirected graph.

To consider both the structural topology and dynamic operation information of the circuit, the ST-GNN includes GNN of spatial domain and improved CNN (Convolutional Neural Networks) of temporal domain. GraphSAGE is adopted as the GNN model [50], which is a classical model that applies the message passing paradigm. In addition, it is a universal graph neural network that samples and aggregates neighboring nodes and generates target node embeddings in a graph. The improved CNN model of time domain processes time series composed of node features at different simulation steps, taking changes of circuit device parameters in dynamic stress simulation into account. Improved CNN adds gate linear units GLU and residual connections methods on the basis of two-dimensional convolution [51]. The GLU algorithm has a fixed length time window, in which the time series features can be compressed using convolution operations, and output control based on time information can be achieved through gating.

The ST-GNN framework consists of two blocks and one

TABLE VI

circuit	Freq[Hz]	max error [%]
DFFSR	5.00E+09	1.1752825
	2.00E+09	1.149435
OAI21X1	5.00E+09	0.3736537
	2.00E+09	0.5067706
FAX1	5.00E+09	2.2061551
	2.00E+09	1.119634
MUX2X1	5.00E+09	0.6838623
	2.00E+09	1.0061294
NOR2X1	5.00E+09	0.3887624
	2.00E+09	0.3836282
TBUFX2	5.00E+09	0.5663097
	2.00E+09	1.4122184
XNOR2X1	5.00E+09	0.8741502
	2.00E+09	0.3939159
XOR2X1	5.00E+09	0.6520148
	2.00E+09	0.6376211

The prediction error of the proposed ST-GNN for various standard cells.

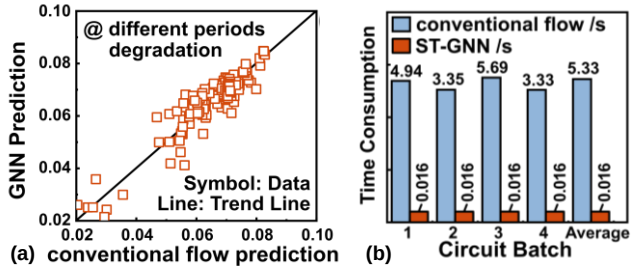


Fig.23. Comparison of the (a) prediction results and (b) time consumption between ST- GNN and conventional flow.

fully connected layer. Each block contains two GraphSAGE layers and one improved CNN layer. The GraphSAGE layer aggregates and updates nodes based on circuit diagram structural information. The improved CNN layer captures relevant temporal information in the temporal dimension. The fully connected layer is used to synthesize features and generate predicted values for device degradation. Besides, the standard library circuits with 45nm PTM are used for model training [52]. **Fig.22** shows the model framework and the process of model training.

From **Table VI**, it shows that the ST-GNN model achieves good performance in predicting aging values for different circuits and frequencies of AC source, with maximum percentage errors maintained within 2.3%. **Fig.23** (a) shows the fitting curve between predicted data and data from conventional approach, which presents small deviations compared to each other under AC source with different frequency. In addition, **Fig.23** (b) compares the calculation time of the traditional reliability model and the ST-GNN model, which indicates that the inference time of the ST-GNN framework model is significantly shorter than conventional approach for all batches of circuits, and the average acceleration ratio can reach over 200 times.

In conclusion, the ST-GNN framework can balance prediction accuracy and time overhead, laying the groundwork for future aging prediction tasks of large-scale circuits. Furthermore, applying the ST-GNN framework to EDA to predict transistor aging will contribute to improving designers' circuit design efficiency and reducing the DTCO time efforts.

VIII. Conclusion

Four key issues, including efficient and accurate characterization technique, long-term prediction capability, compatibility in most EDA platforms, and the prediction efficiency to enable fast iteration, hindered the practical adoption of RV-aware DTCO. This work tackles them by 1) proposing an analytical method to separate different types of traps directly from the measured degradation, 2) proposing a unified defect-based model for accurate long-term reliability and variability aging prediction, 3) developing an OMI-based compact model for the circuit-level aging assessment, 4) developing an ML-assisted approach based on ST-GNN that enables the efficiency improvement by over 200 times compared with conventional methods. With these key advances, a new RV-aware DTCO flow is established, which bridges material properties to circuit design.

REFERENCES

[1] A. Wei *et al.*, "Advanced Node DTCO in the EUV Era," in *2020 IEEE International Electron Devices Meeting (IEDM)*, 2020, vol. 2020-Decem, pp. 41.2.1-41.2.4, doi: 10.1109/IEDM13553.2020.9371921.

[2] J.-G. Ahn, I.-R. Chen, P.-C. Yeh, and J. Chang, "Design-For-Reliability Flow in 7nm Products with Data Center and Automotive Applications," in *2019 IEEE International Reliability Physics Symposium (IRPS)*, 2019, vol. 2019-March, pp. 1–5, doi: 10.1109/IRPS.2019.8720594.

[3] K. Takei, "High performance, flexible CMOS circuits and sensors toward wearable healthcare applications," in *2016 IEEE International Electron Devices Meeting (IEDM)*, 2016, pp. 6.1.1-6.1.4, doi: 10.1109/IEDM.2016.7838358.

[4] Liping Wang *et al.*, "Impact of Self-Heating on the Statistical Variability in Bulk and SOI FinFETs," *IEEE Trans. Electron Devices*, vol. 62, no. 7, pp. 2106–2112, Jul. 2015, doi: 10.1109/TED.2015.2436351.

[5] G. Rzepa *et al.*, "Performance and Variability-Aware SRAM Design for Gate-All-Around Nanosheets and Benchmark with FinFETs at 3nm Technology Node," in *2022 International Electron Devices Meeting (IEDM)*, 2022, vol. 2022-Decem, pp. 15.1.1-15.1.4, doi: 10.1109/IEDM45625.2022.10019528.

[6] Y. Zhao *et al.*, "A Unified Physical BTI Compact Model in Variability-Aware DTCO Flow: Device Characterization and Circuit Evaluation on Reliability of Scaling Technology Nodes," *Dig. Tech. Pap. - Symp. VLSI Technol.*, vol. 2021-June, pp. 2021–2022, 2021.

[7] S. Mishra *et al.*, "A Simulation Study of NBTI Impact on 14-nm Node FinFET Technology for Logic Applications: Device Degradation to Circuit-Level Interaction," *IEEE Trans. Electron Devices*, vol. 66, no. 1, pp. 271–278, Jan. 2019, doi: 10.1109/TED.2018.2875813.

[8] R. Wang *et al.*, "A unified approach for trap-aware device/circuit co-design in nanoscale CMOS technology," in *2013 IEEE International Electron Devices Meeting*, 2013, pp. 33.5.1-33.5.4, doi: 10.1109/IEDM.2013.6724745.

[9] R. Huang *et al.*, "Variability-and reliability-aware design for 16/14nm and beyond technology," in *2017 IEEE International Electron Devices Meeting (IEDM)*, 2017, vol. 7, pp. 12.4.1-12.4.4, doi: 10.1109/IEDM.2017.8268378.

[10] S. S. Chung and E. R. Hsieh, "Principles and Applications of Ig-RTN in Nano-scaled MOSFET," in *Noise in Nanoscale Semiconductor Devices*, Cham: Springer International Publishing, 2020, pp. 175–200.

[11] E. Simoen and C. Claeys, "Random Telegraph Signal: a local probe for single point defect studies in solid-state devices," *Mater. Sci. Eng. B*, vol. 91–92, pp. 136–143, Apr. 2002, doi: 10.1016/S0921-5107(01)00963-1.

[12] T. Grasser, H. Reisinger, P.-J. Wagner, F. Schanovsky, W. Goes, and B. Kaczer, "The time dependent defect spectroscopy (TDDS) for the characterization of the bias temperature instability," in *2010 IEEE International Reliability Physics Symposium*, 2010, pp. 16–25, doi: 10.1109/IRPS.2010.5488859.

[13] B. Kaczer *et al.*, "Ubiquitous relaxation in BTI stressing-new evaluation and insights," *IEEE Int. Reliab. Phys. Symp. Proc.*, pp. 20–27, 2008, doi: 10.1109/RELPHY.2008.4558858.

[14] B. Kaczer, V. Arkbipov, R. Degraeve, N. Collaert, G. Groeseneken, and M. Goodwin, "Disorder-controlled-kinetics model for negative bias temperature instability and its experimental verification," in *2005 IEEE International Reliability Physics Symposium, 2005. Proceedings. 43rd Annual.*, 2005, pp. 381–387, doi: 10.1109/RELPHY.2005.1493117.

[15] C. Liu, K. T. Lee, S. Pae, and J. Park, "New observations on hot carrier induced dynamic variation in nano-scaled SiON/poly, HK/MG and FinFET devices based on on-the-fly HCI technique: The role of single trap induced degradation," *Tech. Dig. - Int. Electron Devices Meet. IEDM*, vol. 2015-Febru, no. February, pp. 34.6.1-34.6.4, 2015, doi: 10.1109/IEDM.2014.7047170.

[16] M. Denais *et al.*, "On-the-fly characterization of NBTI in ultra-thin gate oxide PMOSFET's," in *Technical Digest - International Electron Devices Meeting, IEDM*, 2004, pp. 109–112, doi: 10.1109/iedm.2004.1419080.

[17] D. Dunlop and K. McKinley, "OMI-a standard model interface for IP delivery," in *Proceedings of Meeting on Verilog HDL (IVC/VIUF'97)*, 1997, pp. 83–90, doi: 10.1109/IVC.1997.588538.

[18] W. R. Davis, C. Shaw, and A. R. Hassan, "How to write a compact reliability model with the Open Model Interface (OMI)," in *2020 IEEE International Reliability Physics Symposium (IRPS)*, 2020, vol. 2020-April, pp. 1–2, doi: 10.1109/IRPS45951.2020.9128222.

[19] R. Wang *et al.*, "OMI/TMI-based modeling and fast simulation of random telegraph noise (RTN) in advanced logic devices and circuits," in *Proceedings of International Conference on ASIC*, 2019, pp. 7–10, doi: 10.1109/ASICON47005.2019.8983538.

[20] Z. Ji *et al.*, "A test-proven As-grown-Generation (A-G) model for predicting NBTI under use-bias," in *2015 Symposium on VLSI Technology (VLSI Technology)*, 2015, vol. 2015-Augus, no. 1, pp. T36–T37, doi: 10.1109/VLSIT.2015.7223693.

[21] L. Zhou *et al.*, "Understanding Frequency Dependence of Trap Generation under AC Negative Bias Temperature Instability Stress in Si p-FinFETs," *IEEE Electron Device Lett.*, vol. 41, no. 7, pp. 965–968, 2020, doi: 10.1109/LED.2020.2992263.

[22] P. Ren *et al.*, "Understanding charge traps for optimizing Si-passivated Ge nMOSFETs," in *2016 IEEE Symposium on VLSI Technology*, 2016, vol. 2016-Septe, no. 1, pp. 1–2, doi: 10.1109/VLSIT.2016.7573367.

[23] Z. Ji *et al.*, "An Investigation on Border Traps in III–V MOSFETs With an In 0.53 Ga 0.47 As Channel," *IEEE Trans. Electron Devices*, vol. 62, no. 11, pp. 3633–3639, Nov. 2015, doi: 10.1109/TED.2015.2475604.

[24] H. Chang *et al.*, "Comparative Study on the Energy Distribution of Defects under HCD and NBTI in Short Channel p-FinFETs," in *2021 IEEE International Reliability Physics Symposium (IRPS)*, 2021, vol. 2021-March, pp. 1–5, doi: 10.1109/IRPS46558.2021.9405162.

[25] L. Zhou *et al.*, "Impact of Electron trapping on Energy Distribution Characterization of NBTI-Related Defects for Si p-FinFETs," in *2020 IEEE International Symposium on the Physical and Failure Analysis of Integrated Circuits (IPFA)*, 2020, vol. 2020-July, pp. 1–5, doi: 10.1109/IPFA49335.2020.9260885.

[26] S. W. M. Hatta *et al.*, "Energy distribution of positive charges in gate dielectric: Probing technique and impacts of different defects," *IEEE Trans. Electron Devices*, vol. 60, no. 5, pp. 1745–1753, 2013, doi: 10.1109/TED.2013.2255129.

[27] K. R. Farmer and R. A. Buhrman, "Defect dynamics and wear-out in thin silicon oxides," *Semicond. Sci. Technol.*, vol. 4, no. 12, pp. 1084–1105, 1989, doi: 10.1088/0268-1242/4/12/011.

[28] T. Grasser, "Stochastic charge trapping in oxides: From random telegraph noise to bias temperature instabilities," *Microelectron. Reliab.*, vol. 52, no. 1, pp. 39–70, Jan. 2012, doi: 10.1016/j.microrel.2011.09.002.

[29] G. Rzepa *et al.*, "Comphy — A compact-physics framework for unified modeling of BTI," *Microelectron. Reliab.*, vol. 85, no. April, pp. 49–65, Jun. 2018, doi: 10.1016/j.microrel.2018.04.002.

[30] A. Neugroschel, G. Bersuker, and R. Choi, "Applications of DCIV method to NBTI characterization," *Microelectron. Reliab.*, vol. 47, no. 9-11 SPEC. ISS., pp. 1366–1372, 2007, doi: 10.1016/j.microrel.2007.07.037.

[31] S. Mahapatra, P. BharathKumar, and M. A. Alam, "Investigation and Modeling of Interface and Bulk Trap Generation During Negative Bias Temperature Instability of p-MOSFETs," *IEEE Trans. Electron Devices*, vol. 51, no. 9, pp. 1371–1379, Sep. 2004, doi: 10.1109/TED.2004.833592.

[32] R. Wang *et al.*, "Understanding Hot Carrier Reliability in FinFET Technology from Trap-based Approach," *IEEE Int. Electron Devices Meet.*, pp. 661–664, 2021.

[33] L. Tsetseris, X. J. Zhou, D. M. Fleetwood, R. D. Schrimpf, and S. T. Pantelides, "Physical mechanisms of negative-bias temperature

instability," *Appl. Phys. Lett.*, vol. 86, no. 14, p. 142103, Apr. 2005, doi: 10.1063/1.1897075.

[34] T. Grasser *et al.*, "Gate-sided hydrogen release as the origin of 'permanent' NBTI degradation: From single defects to lifetimes," *Tech. Dig. - Int. Electron Devices Meet. IEDM*, vol. 2016-Febru, pp. 20.1.1-20.1.4, 2015, doi: 10.1109/IEDM.2015.7409739.

[35] C. Freysoldt *et al.*, "First-principles calculations for point defects in solids," *Rev. Mod. Phys.*, vol. 86, no. 1, pp. 253–305, Mar. 2014, doi: 10.1103/RevModPhys.86.253.

[36] T. Grasser *et al.*, "On the microscopic structure of hole traps in pMOSFETs," in *2014 IEEE International Electron Devices Meeting*, 2014, vol. 2015-Febru, no. February, pp. 21.1.1-21.1.4, doi: 10.1109/IEDM.2014.7047093.

[37] D. Waldhoer *et al.*, "Toward Automated Defect Extraction from Bias Temperature Instability Measurements," *IEEE Trans. Electron Devices*, vol. 68, no. 8, pp. 4057–4063, 2021, doi: 10.1109/TED.2021.3091966.

[38] J. L. Lyons, A. Janotti, and C. G. Van de Walle, "The role of oxygen-related defects and hydrogen impurities in HfO2 and ZrO2," *Microelectron. Eng.*, vol. 88, no. 7, pp. 1452–1456, Jul. 2011, doi: 10.1016/j.mee.2011.03.099.

[39] M. Kaviani, V. V. Afanas'ev, and A. L. Shluger, "Interactions of hydrogen with amorphous hafnium oxide," *Phys. Rev. B*, vol. 95, no. 7, p. 075117, Feb. 2017, doi: 10.1103/PhysRevB.95.075117.

[40] J. L. Gavartin, A. L. Shluger, A. S. Foster, and G. I. Bersuker, "The role of nitrogen-related defects in high- k dielectric oxides: Density-functional studies," *J. Appl. Phys.*, vol. 97, no. 5, 2005, doi: 10.1063/1.1854210.

[41] H. Aono *et al.*, "Modeling of NBTI degradation and its impact on electric field dependence of the lifetime," in *2004 IEEE International Reliability Physics Symposium. Proceedings*, 2004, no. m, pp. 23–27, doi: 10.1109/RELPHY.2004.1315296.

[42] J. Bhaskarr Velamala *et al.*, "Compact Modeling of Statistical BTI Under Trapping/Detrapping," *IEEE Trans. Electron Devices*, vol. 60, no. 11, pp. 3645–3654, Nov. 2013, doi: 10.1109/TED.2013.2281986.

[43] Z. Yu, J. Zhang, R. Wang, S. Guo, C. Liu, and R. Huang, "New insights into the hot carrier degradation (HCD) in FinFET: New observations, unified compact model, and impacts on circuit reliability," in *2017 IEEE International Electron Devices Meeting (IEDM)*, 2017, no. 978, pp. 7.2.1-7.2.4, doi: 10.1109/IEDM.2017.8268344.

[44] B. Kaczer *et al.*, "Origin of NBTI variability in deeply scaled pFETs," in *2010 IEEE International Reliability Physics Symposium*, 2010, pp. 26–32, doi: 10.1109/IRPS.2010.5488856.

[45] B. Kaczer, P. J. Roussel, T. Grasser, and G. Groeseneken, "Statistics of multiple trapped charges in the gate oxide of deeply scaled MOSFET devicesapplication to NBTI," *IEEE Electron Device Lett.*, vol. 31, no. 5, pp. 411–413, 2010, doi: 10.1109/LED.2010.2044014.

[46] C. Shin, X. Sun, and T.-J. K. Liu, "Study of Random-Dopant-Fluctuation (RDF) Effects for the Trigate Bulk MOSFET," *IEEE Trans. Electron Devices*, vol. 56, no. 7, pp. 1538–1542, Jul. 2009, doi: 10.1109/TED.2009.2020321.

[47] A. R. Brown, N. M. Idris, J. R. Watling, and A. Asenov, "Impact of Metal Gate Granularity on Threshold Voltage Variability: A Full-Scale Three-Dimensional Statistical Simulation Study," *IEEE Electron Device Lett.*, vol. 31, no. 11, pp. 1199–1201, Nov. 2010, doi: 10.1109/LED.2010.2069080.

[48] A. Asenov, S. Kaya, and J. H. Davies, "Intrinsic threshold voltage fluctuations in decanano MOSFETs due to local oxide thickness variations," *IEEE Trans. Electron Devices*, vol. 49, no. 1, pp. 112–119, 2002, doi: 10.1109/16.974757.

[49] K. El Sayed, E. Lyumkis, and A. Wettstein, "Modeling Statistical Variability with the Impedance Field Method A systematic comparison between the Impedance Field and the 'atomistic' Method," *Int. Conf. Simul. Semicond. Process. Devices, SISPAD*, pp. 205–208, 2012.

[50] J. Zhou *et al.*, "Graph neural networks: A review of methods and applications," *AI Open*, vol. 1, no. September 2020, pp. 57–81, 2020, doi: 10.1016/j.aiopen.2021.01.001.

[51] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive Representation Learning on Large Graphs," *Conf. Neural Inf. Process. Syst. (NIPS 2017)*, no. Nips, pp. 1–11, Jun. 2017, doi: https://doi.org/10.48550/arXiv.1706.02216.

[52] Wei Zhao and Yu Cao, "New Generation of Predictive Technology Model for Sub-45nm Design Exploration," in *7th International Symposium on Quality Electronic Design (ISQED'06)*, 2006, pp. 585–590, doi: 10.1109/ISQED.2006.91.