electronics

MDPI

Review

Bias Temperature Instability of MOSFETs: Physical Processes, Models, and Prediction

Jian Fu Zhang *, Rui Gao †, Meng Duan, Zhigang Ji ‡, Weidong Zhang and John Marsland

School of Engineering, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK; r.gao90@ceprei.com (R.G.); meng.duan@synopsys.com (M.D.); zhigangji@sjtu.edu.cn (Z.J.); w.zhang@ljmu.ac.uk (W.Z.); j.s.marsland@ljmu.ac.uk (J.M.)

* Correspondence: j.f.zhang@ljmu.ac.uk

† Current address: Science and Technology on Reliability Physics and Application of Electronic Component Laboratory, The No. 5 Electronics Research Institute of the Ministry of Industry and Information Technology, Guangzhou 510610, China.

‡ Current address: National Key Laboratory of Science and Technology on Micro/Nano Fabrication, Shanghai Jiaotong University, Shanghai 200240, China.

Abstract: CMOS technology dominates the semiconductor industry, and the reliability of MOSFETs is a key issue. To optimize chip design, trade-offs between reliability, speed, power consumption, and cost must be carried out. This requires modeling and prediction of device instability, and a major source of instability is device aging, where defects gradually build up and eventually cause malfunction of circuits. This paper first gives an overview of the major aging processes and discusses their relative importance as CMOS technology developed. Attentions are then focused on the negative and positive bias temperature instabilities (NBTI and PBTI), mainly based on the early works of the authors. The aim is to present the As-grown-Generation (AG) model, which can be used not only to fit the test data but also to predict long-term BTI at low biases. The model is based on an in-depth understanding of the different types of defects and the experimental separation of their contributions to BTI. The new measurement techniques developed to enable this separation are reviewed. The physical processes responsible for BTI are explored, and the reasons for the failure of the early models in predicting BTI are discussed.

Keywords: reliability; instability; aging; degradation; bias temperature instability (BTI); NBTI; PBTI; yield; device variations; lifetime prediction

Citation: Zhang, J.F.; Gao, R.; Duan, M.; Ji, Z.; Zhang, W.; Marsland, J. Bias Temperature Instability of MOSFETs: Physical Processes, Models, and Prediction. *Electronics* **2022**, *11*, 1420. https://doi.org/10.3390/electronics11091420

Academic Editor: Yahya M. Meziani

Received: 15 March 2022
Accepted: 25 April 2022
Published: 28 April 2022

1. Introduction

Failures of electronic products have been commonly encountered in our daily life, and there is no way that a product can be made with a zero failure rate. Different products require different levels of reliability. For example, the failure rate in auto-electronic components must be lower than that in electronic toys. Failure can be broadly divided into two types: permanent breakdown and temporary malfunction, where a product can recover by, for example, restarting it. Failure rate can be improved by sacrificing performance such as speed and power consumption. A successful commercial product should have a well-balanced trade-off between the failure rate, performance, and costs. This requires modeling of instabilities to predict the failure rate based on an understanding of the underlying physical processes—the subject of this review. As CMOS technologies dominate the chip industry with a market share well over 90%, this work focuses on CMOS.

Ever since CMOS was invented in 1963, its reliability has always been a key issue [1–39]. Failure occurs at both the frontend [1–4], where MOSFETs are located, and the backend [5,6]. The backend includes multiple metal layers for connections and low-k dielectrics between these layers. The metal wires suffer from electromigration [5], where metal atoms gradually migrate from their initial locations, eventually resulting in open or short circuits. To

minimize time delay, the dielectric between these metal wires must have low dielectric constants, making them vulnerable to both mechanical and electrical breakdown [6]. The failure of the backend is not covered in this review, as the authors have been specializing on instability of the frontend.

There are different types of instability for MOSFETs at the frontend including mobile ions [7], hot carriers [8,9], time-dependent dielectric breakdown (TDDB) [10], bias temperature instabilities (BTIs) [1–4], and random telegraph noise (RTN) [11–13]. Their relative importance changes with time. To show this change, we named one type as the dominant instability for each decade after the invention of CMOS in Figure 1. In the 1960s and 1970s, mobile ions in gate oxide, such as sodium ions (Na^+), were a major concern [7]. They originate from contamination and have been effectively eliminated by using cleanroom technologies.

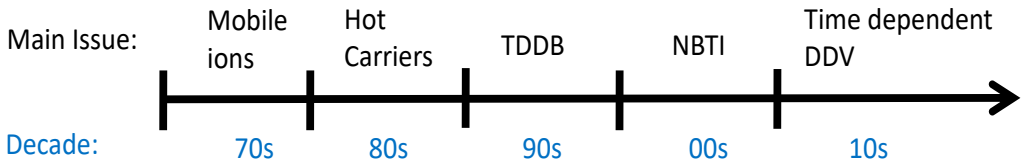

Figure 1. The main reliability issue in each decade for CMOS technologies. DDV: device-to-device variation.

In 1980s, the operation voltage was kept at 5 V when device sizes were downscaled as shown in Figure 2. This leads to an increase in the electrical field, which accelerates charge carriers in the conduction channel and makes them "hot". Through impact and ionization, hot carriers can cause damage to devices by generating interface states and forming space charges in gate oxides, which reduce the driving current, Id, under a given bias [8,9]. A typical definition for device lifetime is the time for an Id reduction of 10% [8,9], and hot-carrier-induced aging limited device lifetime in the 1980s.

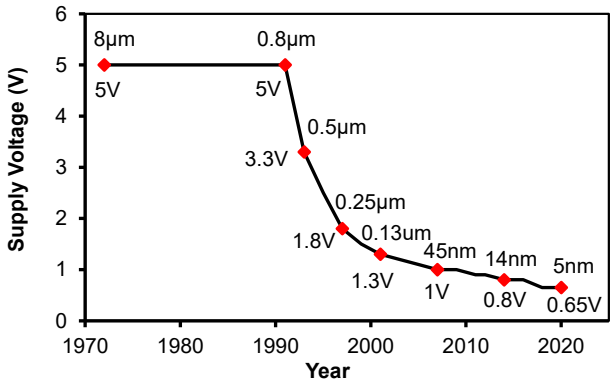

Figure 2. The supply voltage of different generations of CMOS technologies. It was fixed at 5 V before 1990s but has been reduced with the downscaling of device sizes since 1990.

In 1990s, the continuous reduction in the operation voltage, as shown in Figure 2, reduced the relative importance of hot carrier aging. Gate oxides became sufficiently thin, and direct tunneling occurred through them. This causes damage to gate oxides. The damage accumulates and eventually triggers oxide breakdown. As the damage accumulation takes some time, the breakdown is referred to as time-dependent dielectric breakdown (TDDB). TDDB was the dominant reliability issue in the 1990s [10].

In 2000s, negative bias temperature instability (NBTI) of pMOSFETs became the lifetime-limiting instability and attracted much attention [14–22]. The difference between hot carrier aging (HCA) and NBTI is that HCA occurs mainly near the drain junction, while NBTI happens uniformly. Before high-k dielectrics were used, silicon dioxides were nitrided to form oxynitrides, which increases NBTI [14]. BTI was also a major barrier to overcome in developing high-k processes [23–25].

Since 2010s, devices have become so small that device-to-device variations (DDVs) are a major challenge to circuit design. In addition to the static as-fabricated DDVs [26], device aging is stochastic and introduces time-dependent variation (TDV) [27–31]. Both HCA and BTI contribute to TDV. Moreover, for nanoscale devices, a single trap in the gate dielectric can capture a carrier from the conduction channel and induce considerable change of Id in the form of random telegraph noise (RTN) [11–13]. RTN is different from aging: aging shifts device parameters in one direction, while RTN causes their fluctuation in both directions. This further complicates the characterization and modeling of TDV.

This review focuses on BTI, first on NBTI in Section 2 and then on positive bias temperature instability (PBTI) in Section 3. We start by briefly reviewing their history and how advances in CMOS processes have affected them. For example, increasing nitrogen concentration can raise NBTI [14], and using a high-k/SiON stack can increase PBTI [23–25]. The shortcomings of the early works are reviewed, and it is shown that early models cannot predict long-term BTI at low biases, although they can fit test data well [32–34]. The recent progress made by the authors are then reviewed. Based on an in-depth understanding of the defect properties, a framework is proposed for the different types of defects, and new measurement techniques were developed to experimentally separate them. This lays the foundation for proposing a new as-grown-generation (AG) model that can predict the long-term BTI at low biases [35–39]. To support the statements made in this review, we included results and figures reported by early works. The sources of these figures are given at the end of their captions, where readers can find the detailed test procedures and conditions, if needed.

In addition to the instabilities mentioned above, there are other sources of instabilities, such as RF interferences on devices such as thermal sensors [40–42], which is beyond the scope of this review.

2. Negative Bias Temperature Instability (NBTI)

NBTI and PBTI mainly degrade the performance of pMOSFETs and nMOSFETs, respectively. Early attention was focused on NBTI, as it is generally higher than PBTI and limits the lifetime of pMOSFETs. We review NBTI first, followed by PBTI in Section 3.

2.1. Pre-2000 NBTI

NBTI was reported only a few years after the invention of CMOS technology and is one of the earliest instabilities observed for MOSFETs [1]. Under a negative gate bias, positive charges build up both in the gate oxide and from the generated interface states as shown in Figure 3a. This results in a negative shift in the capacitance–voltage (CV) characteristics, as shown in Figure 3b [1], and an increase in the magnitude of the threshold voltage of pMOSFETs [17].

Before 2000, the gate oxide used to study NBTI was relatively thick (e.g., >4 nm) and the oxide field applied was generally too low for electrons to tunnel through the oxide [1–3]. These NBTI works are hereafter referred to as "pre-2000 NBTI". Under these conditions, NBTI has the following features:

- The amount of positive charges formed in the gate oxide equal that originating from the generated interface states as shown in Figure 3b [1,3];
- Figure 4 shows that NBTI follows power law against stress time. The power exponent is insensitive to gate bias and temperature [14];
- Its recovery is insignificant;
- It also follows power law against gate bias;
- It is thermally activated.

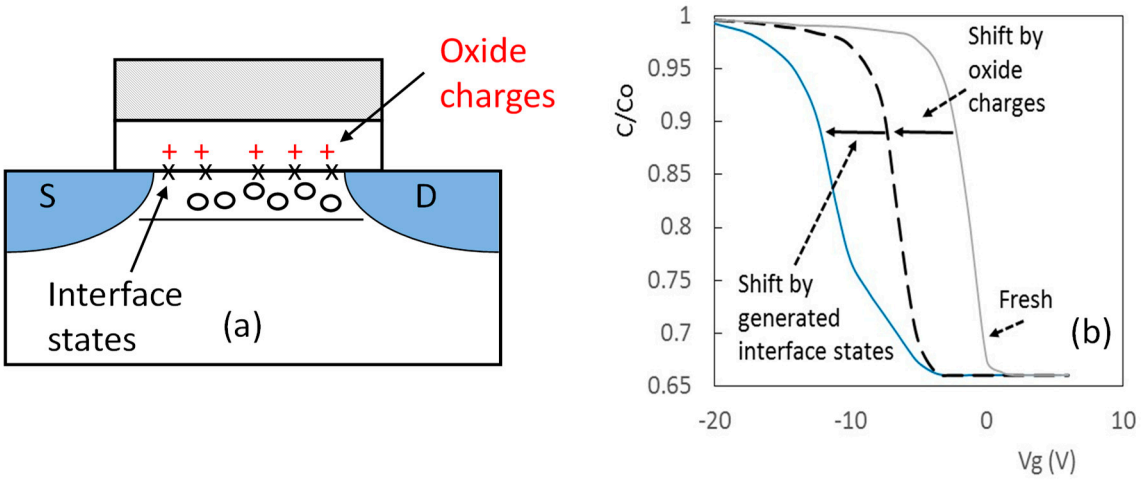

Figure 3. (**a**) A schematic illustration of oxide charges and the generated interface states; (**b**) the effect of NBTI on capacitance versus gate voltage characteristics: positive charges in gate oxide lead to a parallel shift and the generated interface states result in a nonparallel shift [1,17].

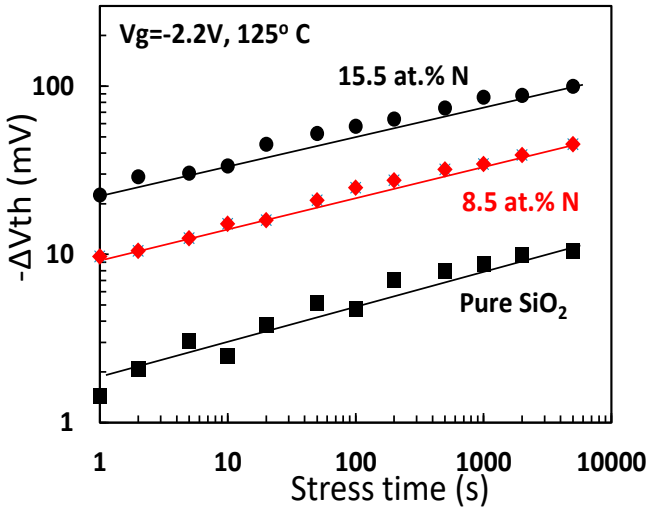

Figure 4. Increase of nitrogen density in SiON raises NBTI. The NBTI follows a power law [14].

On the physical process, the Si–H bond at the interface is generally believed to be the precursor for generated interface states [17]. During NBTI stress, Si–H is ruptured, and the released hydrogenous species then migrate into the gate oxide, leaving behind the Si dangling bond, also commonly known as the Pb-center, as the interface state [2,4,17]. The generated interface states in the lower half of Si bandgap are donor-like. As gate bias is swept in the negative direction, there are an increasing amount of them moving above the Fermi-level at the interface, Ef, and becoming positively charged. This leads to the nonparallel shift in CV as shown in Figure 3b.

In addition to the nonparallel shift, there is also a parallel negative shift, as shown in Figure 3b, which is caused by positive charges formed in the oxide. The magnitude of parallel shift is similar to that of nonparallel shift in Figure 3b, indicating a one-to-one correlation between the oxide charges and the positive charges from the generated interface states. This, however, does not mean that the oxide charge density is equal to the density of the generated interface states. As each Pb-center has two states, one acceptor-like in the upper-half of Si bandgap and one donor-like in the lower half of bandgap, the interface state density measured using a popular technique, such as charge pumping, should double the oxide charge density.

On the aging kinetics, it was proposed that the power law results from the diffusion of hydrogenous species through gate oxides as the rate limiting process [2,4]. This hypothesis was challenged as the gate oxide of modern MOSFETs is too thin and the time for hydrogenous species diffusing through it is too short to limit the aging process [17].

2.2. Post-2000 NBTI

2.2.1. Difference from Pre-2000 NBTI

In the three decades between the 1960s and 1990s, there were only a few papers on NBTI [1–4], since it was not the lifetime-limiting process. Since 2000, however, NBTI has been limiting the lifetime of pMOSFETs and has attracted much attention [14–22,31–39]. We refer to the NBTI works after the year 2000 hereafter as "post-2000 NBTI". The reasons for this increased importance of NBTI are given below.

As the channel length downscales, a thinner gate oxide is needed to control the drain-induced barrier lowering leakage (DIBL) current between the source and drain. On the other hand, when the gate oxide is thinner than 3 nm, direct tunneling current through it becomes considerable, increasing power consumption. To mitigate the gate leakage, SiO_2 is nitrided to form oxynitrides [14], which has a higher dielectric constant, so that thicker layers can be used for the same gate oxide capacitance. Oxynitrides also impede boron diffusion from the p+-poly-Si gate into the substrate. Figure 4 shows that an increase in nitrogen density leads to higher NBTI. Moreover, the operation voltage cannot be downscaled proportionally with device size now, since the silicon bandgap is a constant. This increases the electrical field within the device. The higher packing density of transistors also results in higher temperatures within a chip. These factors make NBTI high enough to limit the lifetime of pMOSFETs.

Since the 45 nm CMOS process, a high-k/SiON stack replaced oxynitrides in 2007. The NBTI of the high-k/SiON stack is still important and dominated by the interfacial SiON layer as shown in Figure 5 [43]. The knowledge gained from the oxynitrides is applicable to the stack, and there is no new physical processes or defects identified for the NBTI of high-k/SiON stack.

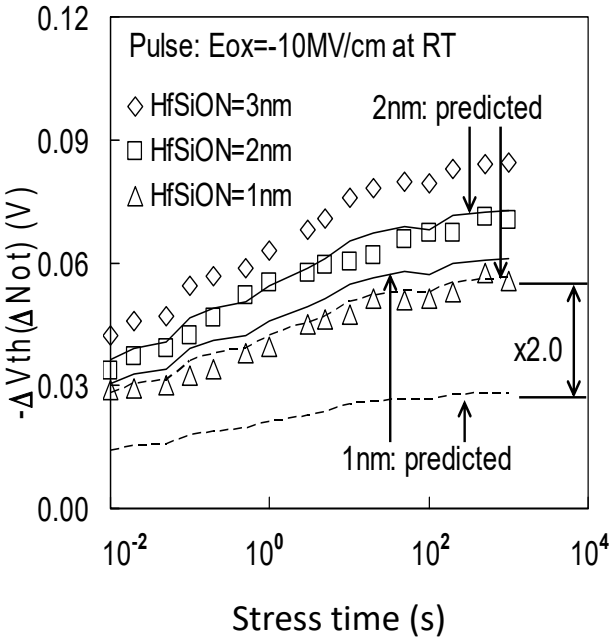

Figure 5. NBTI-induced positive charges with different HfSiON thicknesses. Symbols are test data. The dashed lines are predictions made by assuming charges at the Hf-k/SiON interface, which did not agree with test data. The solid lines are predictions made by assuming charges at SiON/Si interface, which agreed well with test data [43].

Unlike the pre-2000 NBTI, where tunneling through oxide is negligible, the thin oxide (e.g., <3 nm) and high field (e.g., >8 MV/cm) used post-2000 leads to considerable tunneling current during NBTI tests [17]. This changes the characters of NBTI in several ways. First, the one-to-one correlation of positive oxide charges with positive charges from the generated interface states, as shown in Figure 3b, often does not hold, and oxide charges can be substantially higher [44,45]. Second, Figure 6a shows that the kinetics can have a "hump" and no longer follow a power law [46]. Third, Figure 7 shows that NBTI now has a substantial recoverable component [47]. In the followings, we review the efforts made to understand and to model this complex behavior.

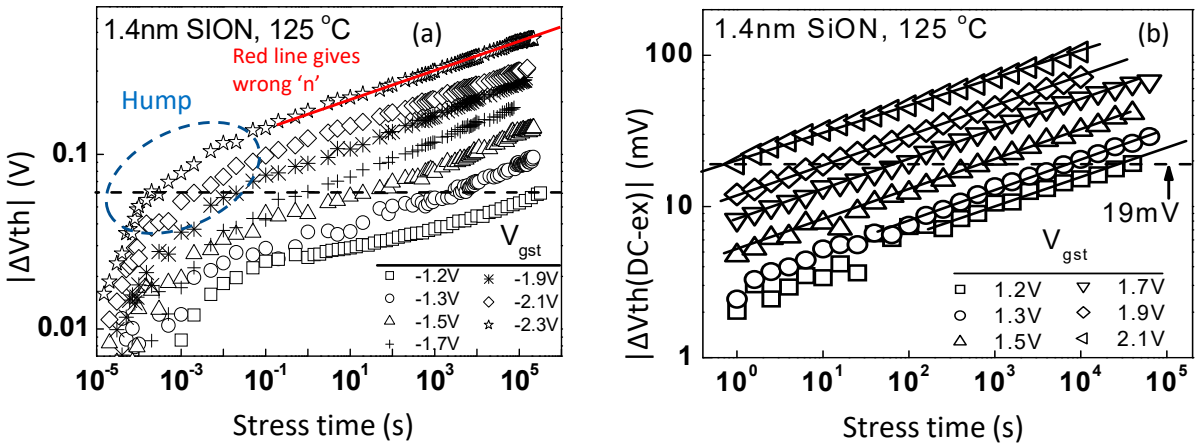

Figure 6. (**a**) The NBTI kinetics post-2000 does not follow a power law and has a "hump" when measured from the pulse (5 µs) Id-Vg. Although the data after the "hump" appear to follow the power law, the fitted red line underestimated the power exponent. (**b**) The kinetics appear to follow the power law when measured at a slow speed (10 ms per point) [46].

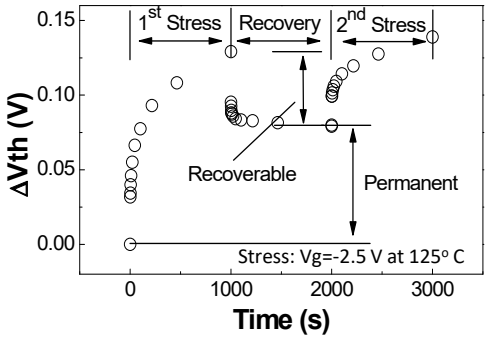

Figure 7. Substantial recovery occurs for post-2000 NBTI [47].

2.2.2. Failure of Early Models in Prediction

The presence of a substantial recoverable component in post-2000 NBTI, as shown in Figure 7, requires introducing new defects to explain the difference from the pre-2000 NBTI. It is well known that there are hole traps near the dielectric/substrate interface [48]. As holes can tunnel through the oxide under Vg < 0, they can fill these traps [49–51]. The filled traps are close to the substrate interface and are unstable. Once the stress negative bias is removed, they can be readily neutralized, resulting in the recovery observed in Figure 7.

Early models, such as the reaction–diffusion model [32] and the composite model [33,34], included this recoverable component by adding new kinetics, in addition to the power law. These models can fit test data well and examples are given for both SiON and high-k/SiON stack in Figure 8. The mission for modeling NBTI, however, is to predict it for cases where test data have not been used for fitting or do not exist. As a result, good fitting with test data is not sufficient to validate a model. When these models were used to predict NBTI at

lower biases, Figure 8 shows that there were large discrepancies between the prediction and the test data [17,36].

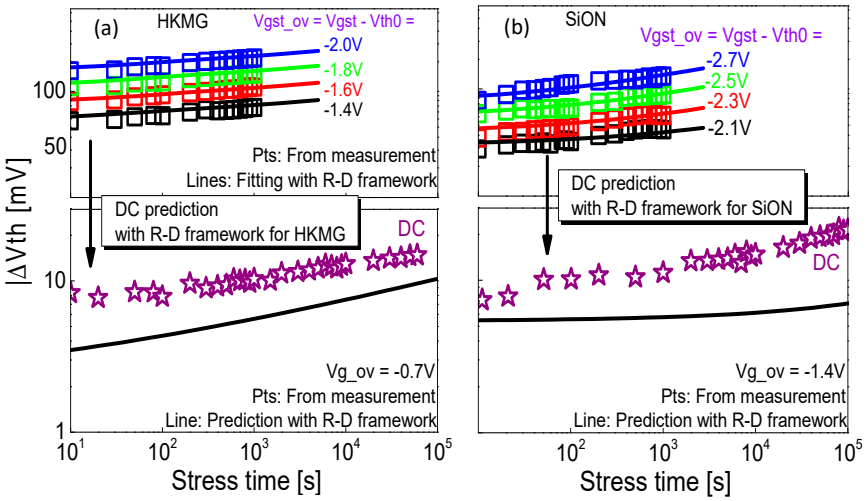

Figure 8. The reaction–diffusion (R–D) model could fit the test data well for both the high-k/SiON stack (**a**) and SiON (**b**). The model extracted from this fitting could not predict the NBTI at lower gate bias [36].

2.2.3. To Generate or Not to Generate, That Is the Question

To model the post-2000 NBTI successfully, an in-depth understanding of the defects is needed. Before proposing a kinetics, one needs to know the rate-limiting process. A key issue is whether the defects are generated or as-grown. If the defects are as-grown, filling them will limit the NBTI rate. On the other hand, if defects must be generated first before they can be filled, the generation process can limit the rate. The generation can follow different kinetics from that of trap-filling. There is some confusion, however, on how to separate these two and on the definition of generation. This will be clarified below.

As-grown traps, also known as pre-existing traps, exist in as-fabricated devices before electrical stresses. After filling and then neutralizing them, they can be refilled. Figure 9 shows that refilling follows the same kinetics as the 1st filling [48]. This indicates that they have not changed after its first filling-neutralization. We define as-grown traps as the traps whose refilling is the same as their 1st filling and is not affected by stresses.

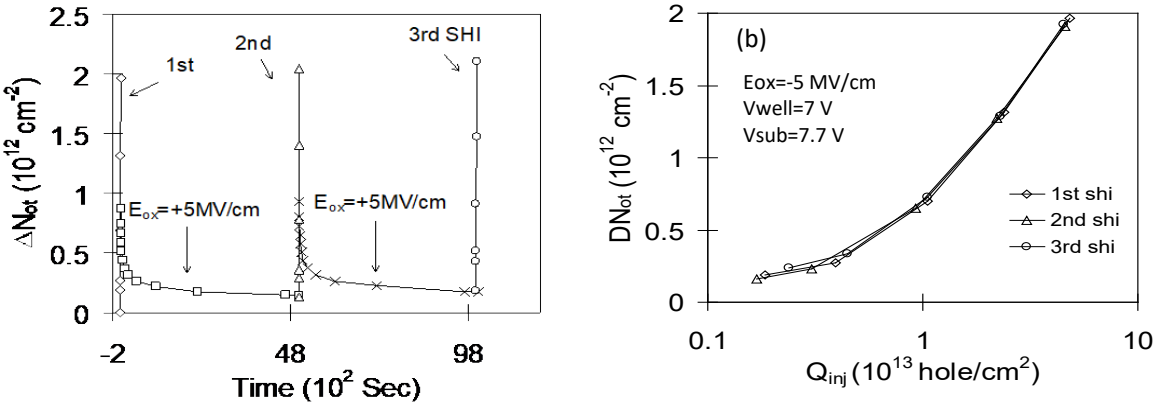

Figure 9. (**a**) As-grown traps can be repeatedly filled and then discharged; (**b**) the first filling and subsequent re-filling agreed well [48].

The concept of generation was used to describe the increase in the interface states following stresses [8,17]. In an as-fabricated device, the number of interface states is low. After stresses, it clearly rises under the same measurement conditions, because new

interface states are generated. As discussed earlier, the precursors of interface states in an as-fabricated device, Si–H, are not electrically active. The Si–H must be ruptured first to become an interface state [17]. We define the generation as the process for converting a precursor to a defect. This conversion may only happen once. After Si–H is broken, the resultant Pb-center can be repeatedly charged and discharged. Charging this interface state is much faster than the generation process; thus, the generation is the rate-limiting process for the build-up of interface states.

Trap generation in gate dielectric is also proposed during the time-dependent dielectric breakdown (TDDB) tests, where the gate current increases with the stress time [24]. It is proposed that this stress-induced leakage current (SILC) is caused by the generated defects, which can act as stepping-stones for the charge carrier to pass through the dielectric. Once the generated defects overlap and form a conduction path between gate and substrate, it triggers oxide breakdown. It has been proposed that the defects responsible for the breakdown are the generated electron traps, and there are other types of traps in the dielectric that do not contribute to the breakdown path [52].

Trap generation in gate dielectric was less clearly observed during NBTI tests. The gate current was not typically monitored here, and aging builds up continuously without saturation as shown in Figure 6. There are two potential explanations for this non-saturation behavior. One is that new traps are continuously generated as the stress time increases. The other is that there are traps with small capture cross sections or large capture times, which requires long stress time to fill them. In the following, we present results to support the case of trap generation.

Figure 10a shows the non-saturation of aging during the stress. To investigate it, a device was stressed, followed by neutralization, and then the 2nd stress as shown in Figure 10b [48]. Figure 10c compares the kinetics during the 1st and 2nd stresses. The initial trapping during the 2nd stress was clearly higher than that during the 1st, supporting that there are more traps available after stresses. These extra traps must be generated by the stress. Once a trap was generated, Figure 10d shows that they could be filled rapidly. The generated traps measured in this way were compared with the aging during the 1st stress in Figure 10a. The good agreement indicates that trap generation can play a dominating role in the non-saturation of aging.

Further support for trap generation in gate dielectric can be found from the energy profile of traps. Figure 11 shows that as-grown traps are located below the top edge of the Si valence band, Ev. As stress time increased, the traps above Ev increased, but the traps below Ev remained the same. This supports that new traps are generated above Ev. The as-grown and generated traps are energetically different, indicating that they are different types of traps.

2.2.4. A New Modeling Approach: Separating As-Grown Traps from Generated Defects

Although trap generation can dominate the long-term aging, as-grown traps may also contribute to it, since it is well known that there are slow traps. As stress time increases, slow traps are filled and contribute to the build-up of trapped charges. As filling as-grown traps is a different physical process from trap generation, it is expected that they follow different kinetics. The threshold voltage shift, ΔVth, measured during typical NBTI tests is the combined effect of all defects. The question is whether one can reliably extract these different kinetics by fitting them with this combined ΔVth simultaneously. Early efforts followed this approach, but the extracted models cannot be used to predict NBTI, as shown in Figure 8.

The failure of early models calls for new approach to make a breakthrough for establishing a model that can predict NBTI. Ideally, one should separate the contribution of generation from that of filling as-grown traps. This makes it possible to fit each kinetics separately with its own contribution only. A lot of efforts [35–39] have been made recently to separate as-grown traps from the generated ones experimentally and to model them,

based on an in-depth understanding of these defects. In the following, we review recent progresses on as-grown traps first and then the generated traps.

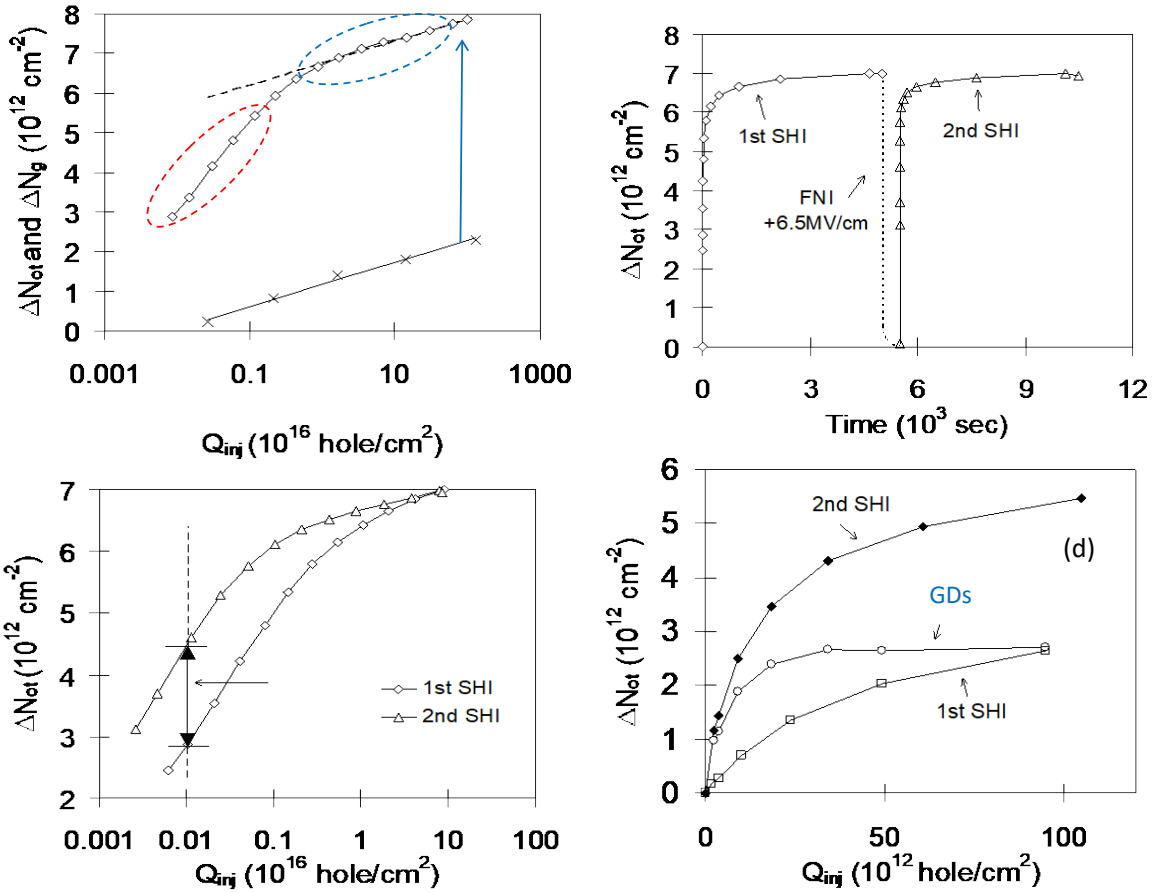

Figure 10. (**a**) The symbol "◊" represents defect build-up during stress. The symbol "×" represents generated defects extracted as shown in (**c**,**d**). (**b**) The test sequence: 1st stress, neutralization, and then 2nd stress. (**c**) The difference between the 1st and 2nd filling. The generated traps resulted in higher trapping during the 2nd filling. (**d**) The generated traps could be filled rapidly, and their filling was saturated [48].

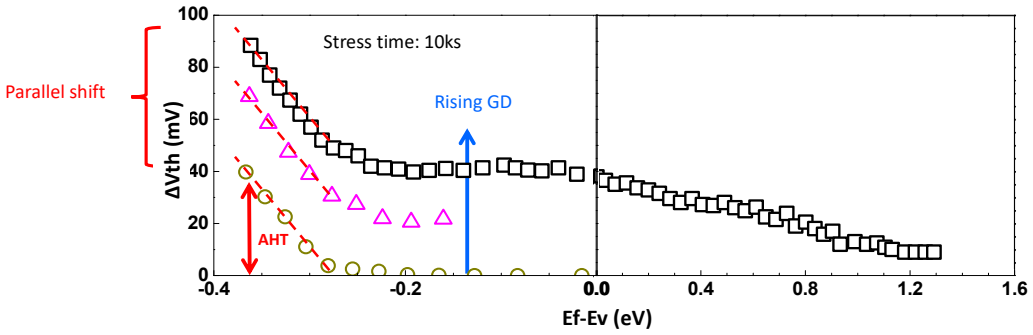

Figure 11. The as-grown defects "A" are located below silicon Ev. The generated defects are above Ev. The stress did not change the as-grown defects, as shown by the parallel upward shift of the dashed lines [35,36].

2.2.5. As-Grown Traps and Their Modeling

When biases are applied to a device, filling as-grown traps occurs simultaneously with generating new traps, which complicates their separation. To characterize as-grown traps, we must find an experimental condition under which new trap generation is negligible.

This can be achieved by stressing a device heavily first. Since aging follows a power law with a power exponent well less than one, Figure 12 shows that the aging rate slows down quickly as the defects build up. This makes it possible for new trap generation to become negligible during the characterization of as-grown traps on a heavily stressed device. By definition, the stress will not affect as-grown traps that can be repeatedly charged and then discharged. The negligible new trap generation was confirmed by the recyclability of the charging–discharging of as-grown traps in Figure 13 [53].

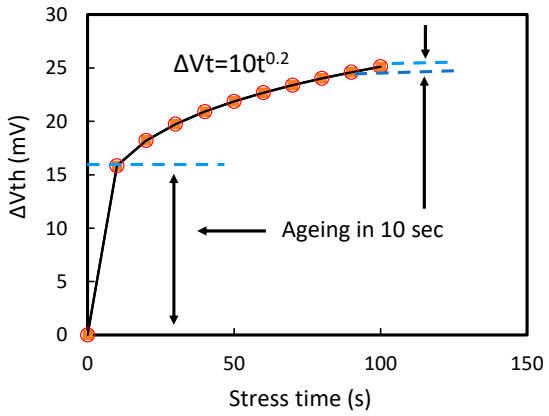

Figure 12. An illustration of the rapid reduction in the aging rate with stress time.

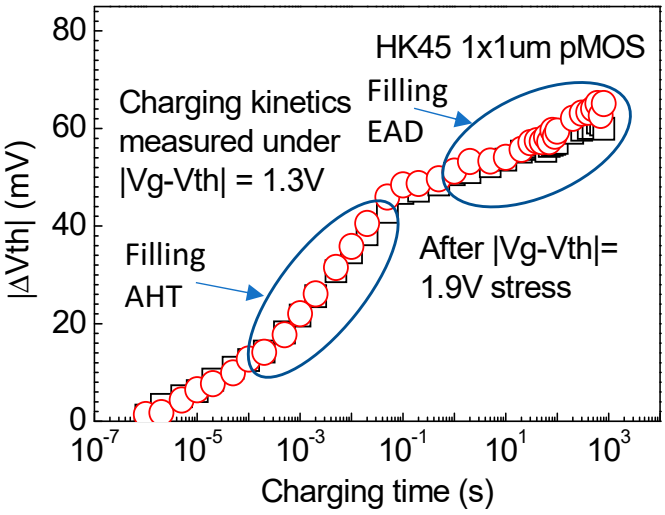

Figure 13. Recyclability of trapping–detrapping for as-grown defects. The symbol "□" represents first filling. After discharging, the traps were refilled (the symbols "o"). The trapping was dominated by filling as-grown hole traps (AHTs) initially, followed by filling the energy-alternating defects (EADs) [53].

Figure 13 shows that there are two types of as-grown traps: as-grown hole traps (AHTs) and energy-alternating defects (EADss). Filling AHTs is responsible for the rapid build-up of ΔVth initially, while EADs cause the subsequent slow and non-saturating increase. The filling process for AHTs is different from that for EADs. Figure 14a,b show that AHTs are filled by holes from the valence band of silicon, and their energy level does not change after charging, although the presence of an energy barrier makes the filling thermally activated. In contrast, EADs are filled by phonon-assisted process: during charging, holes must overcome an energy barrier before settling down in a lower energy well in Figure 15a. In the electron energy band diagram, the lower hole energy level corresponded to a higher electron energy level, i.e., an upward shift as shown in Figure 15b. The higher the energy

barrier, the longer the filling will take. As a result, the slow and non-saturating filling can originate from a spread of the energy barrier in Figure 15a.

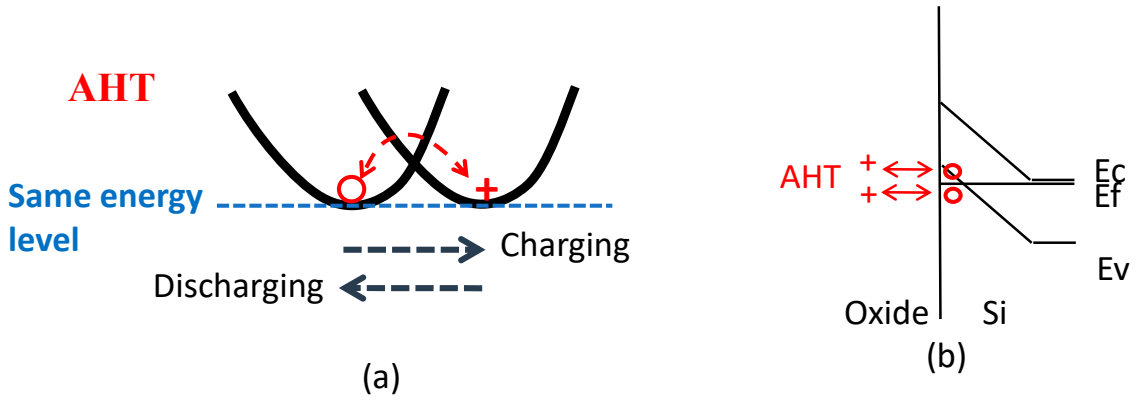

Figure 14. Charging and discharging of as-grown hole traps (AHTs): (**a**) the energy level of the hole trap did not change by charging–discharging; (**b**) the energy level of AHTs was below Si Ev. The symbol 'o' represents a hole.

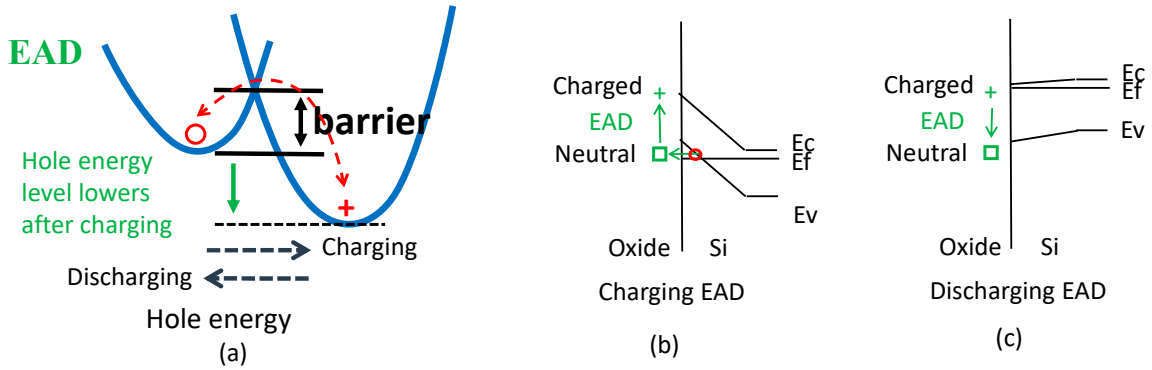

Figure 15. Charging and discharging EADs: (**a**) the hole energy level was lower following charging, which corresponded to a rise in its electron energy level in (**b**); (**c**) the electron energy level of the hole trap was lower after discharging. The symbol '□' represent a hole trap in its neutral state.

To support the above hypothesis on the presence of EADs, Figure 16 compares the dependence of ΔVth on $|Vg - Vth|$ during charging and discharging. During charging, the energy level is progressively swept in the negative direction by applying more negative Vg, allowing defects further below Ev to be filled, while the opposite occurs for discharging [37]. Figure 16 shows that the ΔVth during discharging was clearly higher than that during charging. As shown in Figure 15b,c, after an EAD is filled, its electron energy level rises, so that it cannot be discharged at the same energy level for filling. This is responsible for the higher ΔVth, i.e., the hysteresis observed during discharging in Figure 16.

Based on the different impacts of charging on the energy level of AHTs and EADs, a progressive charging–discharging technique was developed to separate them as shown in Figure 17a. The details of this technique can be found in Reference [39]. The orange curve in Figure 17b shows the measured saturation level of AHT, AHT_{sat}, at different overdrive voltages, Vgov = Vg − Vth. The dependence of AHT_{sat} on Vgov follows an exponential function, given in Table 1.

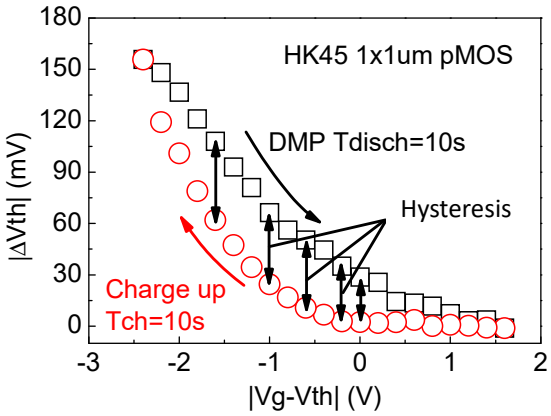

Figure 16. Presence of EADs: after charging, the electron energy level of the EADs rise, so that they cannot be discharged at the same energy level, resulting in the hysteresis [53].

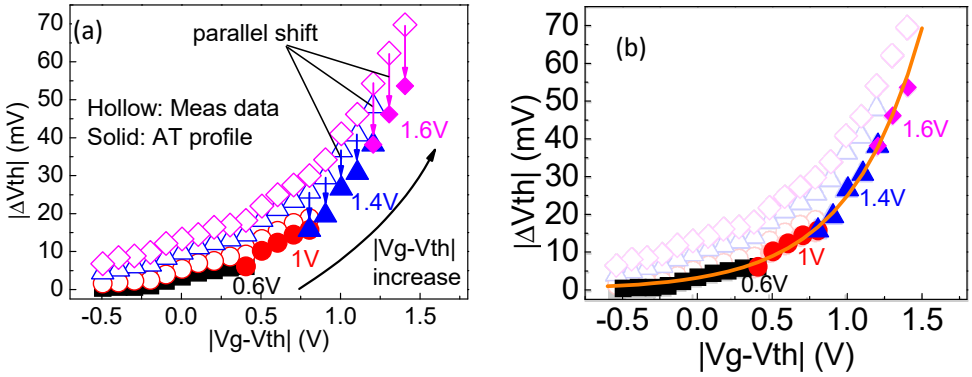

Figure 17. (**a**) The progressive charging–discharging technique for extracting the saturation level of AHTs; (**b**) the orange curve is the saturation level of AHTs at a given $|Vg - Vth|$ [37,53].

Table 1. The formula of the as-grown-generation (AG) model for BTI.

Defects	Formula
Saturation level of AHT/AET	$AT_{sat} = p_1 \cdot \exp(p_2 \cdot V_{gov})$
Filling AHT/AET	$AT = AT_{sat} \cdot \left[1 - \exp\left(-\frac{tch}{\tau}\right)^{\gamma}\right]$
EAD	$EAD = g_2 \cdot V_{gov}^{m_2} \cdot t^{n_2}$
GD	$GD = g_1 \cdot V_{gov}^{m_1} \cdot t^{n_1}$

At a given Vgov, the transient build-up of total ΔVth with time is represented by the black squares in Figure 18a. The green triangles represent EADs, which were obtained from $\Delta Vth - AHT_{sat}$. The EAD follows a power law. The EAD over the short term can be obtained by extrapolating this power law, as shown by the green dashed line. Once the EAD is known, AHT can be obtained from $\Delta Vth - EAD$, as shown by the red circles. The kinetics for filling AHTs is given in Table 1. Figure 18b shows AHT filling at different $Vg - Vth$. Although the saturation level, AHT_{sat}, increases with $|Vg - Vth|$, the normalized kinetics in Figure 18c were independent of $|Vg - Vth|$ and could be extracted by fitting the test data [37].

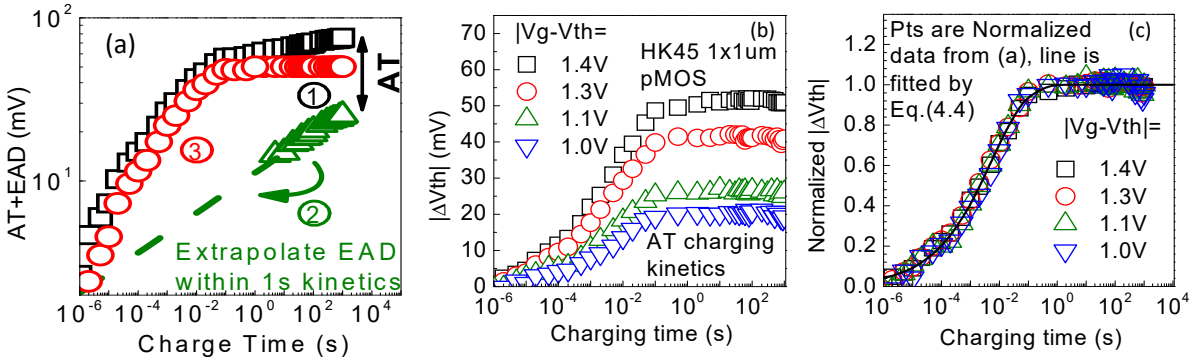

Figure 18. (**a**) Separating AHTs from EADs. The saturation level of AHT was first subtracted from the total ΔVth ("□") to give the EAD ("Δ"). The EAD was then fitted with a power law and extrapolated over the short term (see dashed line). The AHT ("o") was obtained by subtracting the green dashed line from the total ΔVth ("□"). (**b**) The AHT kinetics at different charging biases. A higher charging $|Vg - Vth|$ gives a higher AHT. (**c**) The AHT kinetics normalized against its saturation level [37].

Filling EADs at different Vgov is shown in Figure 19a. Figure 19b shows that the extracted time power exponent is insensitive to Vgov, and Figure 19c shows that EADs also follow a power law against Vgov.

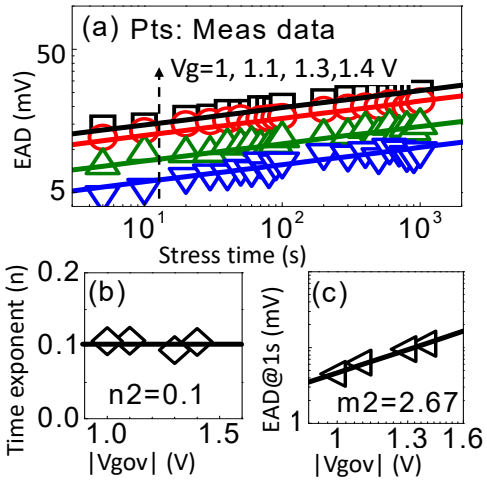

Figure 19. (**a**) The kinetics of an EAD follows the power law; (**b**) the time power exponent was insensitive to the stress bias; (**c**) an EAD also follows a power law against the stress overdrive voltage, Vgov [37].

2.2.6. Two Types of Generated Traps

After understanding and modeling as-grown defects, we now turn our attentions to the generated defects. Filling as-grown hole traps is a relatively fast process and is responsible for the observed "hump" in Figure 6a. As the stress time increased, however, more traps were generated, and they dominated the long-term aging. The presence of the "hump" made it difficult to extract the power exponent for the long-term aging, as we did not have a straight line in Figure 6a. One may focus on the region after the "hump" where the data were approximately in a straight line as shown in Figure 6a. The power exponent, n, extracted in this way, however, underestimates the real n. This is illustrated in Figure 20, where a constant, representing the saturation level of AHTs, was added to a power law. The resultant data can be fitted reasonably well with the power law, but giving a wrong power exponent, which is the slope of the fitted line. As a result, one should not use good fitting with the test data as a criterion to validate a model.

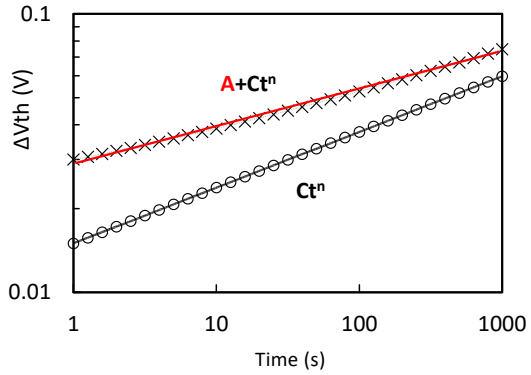

Figure 20. When a constant, A, was added to the power law, the data (symbol '×') could still be fitted with the power law reasonably well, but the power exponent (i.e., the slope of the red line) was underestimated, when compared with its real value (i.e., the slope of the black line).

As AHTs are located below Ev, they can be readily discharged. This explains why NBTI recovery was much more prominent post-2000 in Figure 7; AHTs were filled during post-2000 NBTI, since high field and thin oxide were used, while they were hardly filled during pre-2000 NBTI. In another word, the post-2000 NBTI has much higher AHT components than that of the pre-2000 NBTI. To minimize the contribution of AHTs to the measured ΔVth, one can introduce a delay during the measurement [23]. This allows AHTs to recover before ΔVth was measured and Figure 6b shows that the "hump" disappears, and the power law is restored [46]. The problem is that the extracted power exponent, n, now depends on the recovery time, as shown in Figure 21, and one does not know which n is correct. Figure 21 shows that when the power law is extrapolated to predict the device lifetime, these different n cause significant uncertainties.

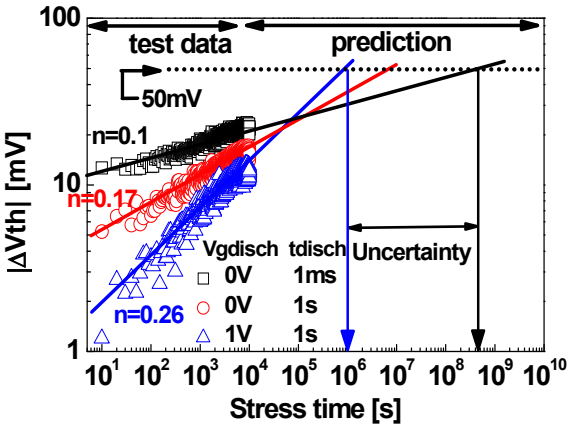

Figure 21. The power exponents extracted from test data depend on the delay between stress and measurement. The longer the delay, the larger the power exponent [38].

To overcome this challenge, an in-depth understanding of the generated defects is needed. In Figure 22a, a device was first stressed, and trapped charges were then neutralized [49]. This was followed by applying −5 and +5 MV/cm, alternatively. Part of the neutralized traps can be repeatedly recharged under −5 MV/cm and neutralized under +5 MV/cm, so that they are referred to as cyclic positive charges (CPCs) [49–51]. Part of recharged traps cannot be neutralized under −5 MV/cm, and they are called anti-neutralization positive charges (ANPCs). When the same ±5 MV/cm was applied on a fresh device, both the CPC and ANPC were absent, so that both were generated defects.

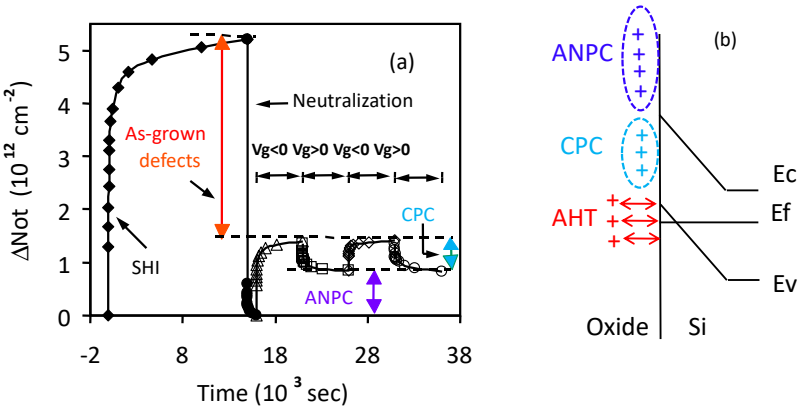

Figure 22. Different types of positive charges in a gate dielectric. (**a**) In addition to as-grown defects, some of the generated defects can be repeatedly charged and discharged by alternating the gate bias polarity, and they are called as cyclic positive charges (CPCs). The rest of generated positive charges are difficult to neutralize and are referred to as anti-neutralization positive charges (ANPCs). (**b**) The energy location of the different types of positive charges [49].

The different behavior of CPC and ANPC can be explained from their different energy levels. Figure 22b shows that CPC are within the bandgap of Si, so that they can be repeatedly charged–discharged, as their energy level moved above and below Ef under −5 and +5 MV/cm, respectively. In contrast, the higher energy levels of ANPC kept them above Ef under +5 MV/cm, so that they did not discharge.

As both CPC and ANPC are generated defects, it is natural to speculate that they have the same physical origin, and their difference is quantitative. For example, the same type of precursors with a spread of bonding strengths and/or angles can be responsible for their energy differences in Figure 22b. On the other hand, it is also possible that they originated from two different types of precursors. Both types of precursors can interact with hydrogenous species during NBTI tests: one results in CPC and the other is converted to ANPC [17]. This was supported by the following test results.

Figure 23a,b show the impact of pre-stress hydrogen exposure on the subsequent generation of CPC and ANPC, respectively [54]. Without the pre-stress hydrogen exposure, CPC increased gradually with the stress time initially and then saturated, as the precursors run out. With increasing hydrogen exposure, CPC can reach this saturation level already at the first stress point. In contrast, ANPC generation does not saturate. After the hydrogen exposure, the same amount of ANPC was created during the subsequent stress. These differences indicate that CPC and ANPC have different precursors.

2.2.7. Modeling the Generated Traps

Based on the above understanding, not only as-grown traps but also the generated CPC can recover. When compared with AHT, the relatively high energy level of CPC in Figure 22b led to a gradual discharge of CPC, which can be seen from the non-flat tail when plotted against the logarithmic discharge time in Figure 24. As a result, some CPCs were lost if there was a delay between stress and measurement. As illustrated in Figure 25, a loss of CPC led to an increase in the extracted n, which contributed to the uncertainty observed in Figure 21. To remove this uncertainty, we must minimize the loss of CPC.

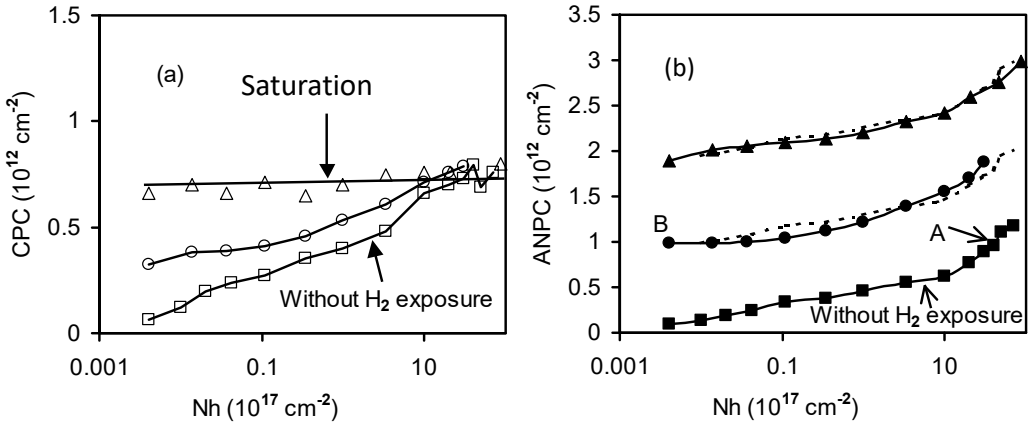

Figure 23. Impact of hydrogen exposure on CPC (**a**) and ANPC (**b**). The precursors for CPC are limited, and they can be converted to CPC by either electrical stress or hydrogen exposure. In contrast, the ANPC generated by electrical stress is not affected by the hydrogen exposure [54].

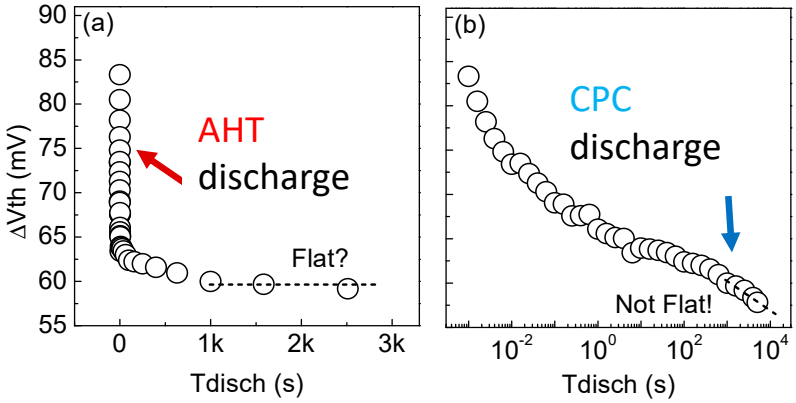

Figure 24. Discharge against linear (**a**) and logarithmic (**b**) time. The AHT can be neutralized rapidly due to its lower energy level. This is followed by a gradual discharge of CPC [53].

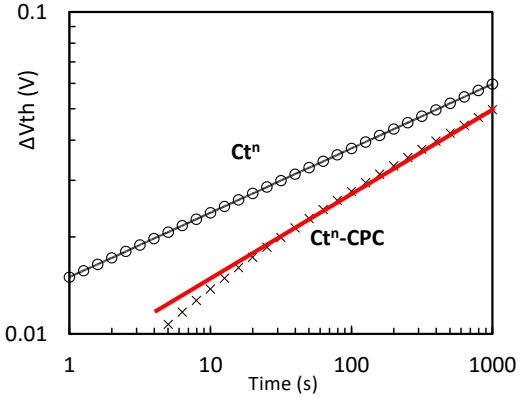

Figure 25. An illustration of the impact of a loss of CPC on the aging kinetics. Although the data after the loss (symbols "×") can be fitted reasonably well with a power law (red line), the power exponent (i.e., the slope of the red line) is overestimated, when compared with its true value (i.e., the slope of the black line).

The stress–discharge–recharge (SDR) technique in Figure 26a was designed to capture the generated defects in their entirety [38]. During stress, all defects are charged. To separate the as-grown defects from the generated ones, a discharge step was used. The conditions of this discharge step were set to neutralize all as-grown defects, but inevitably some CPC were also neutralized. A recharge phase was then used to recapture these lost

CPCs. During recharge, both as-grown defects and the lost CPC were refilled. The filling of as-grown defects can be determined by applying the same recharge step on a fresh device as shown in Figure 26b. After stress, the generated CPC increased the recharge level, so that the lost CPC could be obtained from the difference between the recharge levels pre- and post-stress in Figure 26b. Figure 27 shows that the lost CPC increased with both discharge time and discharge voltage. After adding them back to the measured ANPC, the total generated defects become independent of measurement conditions.

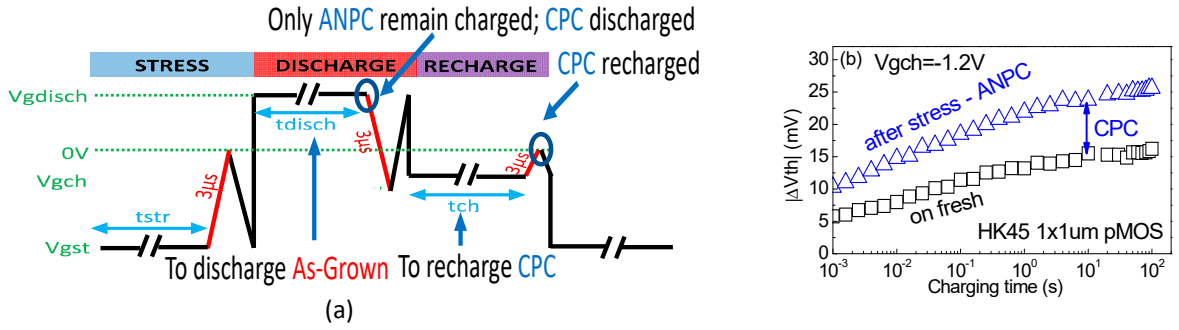

Figure 26. (**a**) The gate bias waveform for the stress–discharge–recharge (SDR) technique; (**b**) the recharge step: the increased |ΔVth| after stress from that of a fresh device was caused by the generated CPC [38].

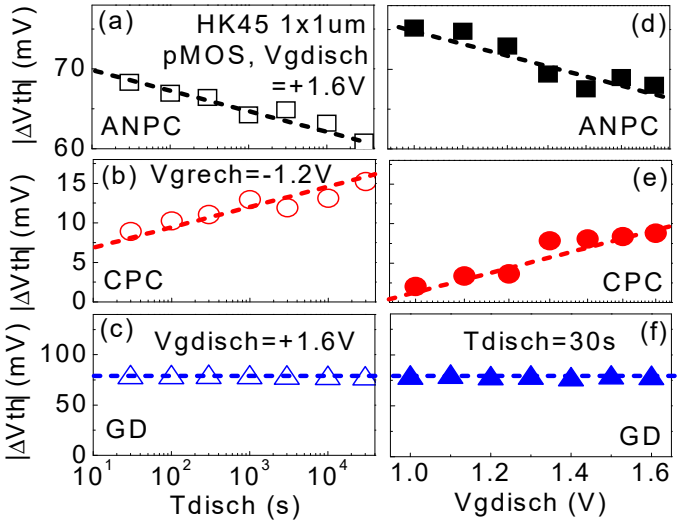

Figure 27. The ANPC and CPC measured by the SDR technique using different discharging times (**a**,**b**) and discharging voltages (**d**,**e**). The total generated defect (**c**,**f**) (i.e., the sum of ANPC and CPC) was independent of the measurement conditions [38].

After removing the impact of measurement conditions on the generated defects and the contribution from as-grown defects, Figure 28 shows that the generated defects follow a power law with a power exponent that is not sensitive to either stress bias or temperature. Moreover, Figure 29 shows that the power exponent extracted using the SDR technique was also insensitive to the fabrication processes. This is in contrast with the wide spread of power exponents reported by early works in Figure 29.

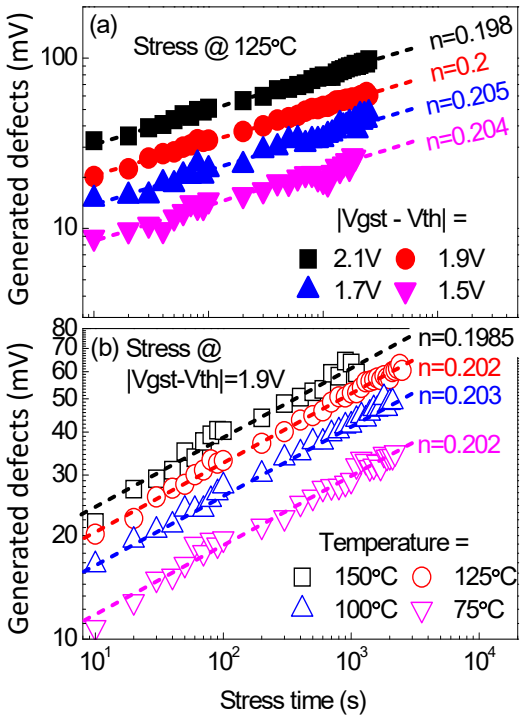

Figure 28. The kinetics of Generated Defects under different stress voltages (**a**) and different temperature (**b**). The power exponents are insensitive to either the stress bias or temperature [38].

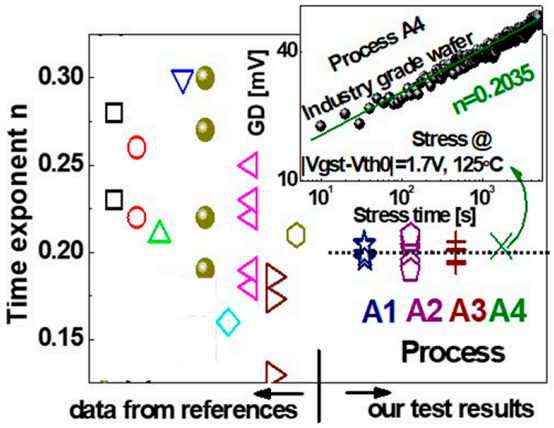

Figure 29. A comparison of the power exponents reported by our work with those by earlier works. The details of these early works can be found in Reference [38].

2.2.8. A Framework of Defects Responsible for NBTI

A framework of the defects responsible for NBTI is summarized in Figure 30. As-grown defects exist in fresh or as-fabricated devices pre-stress. By definition, they do not increase after stress and can be repeatedly charged–discharged. There are two types of as-grown defects: as-grown hole traps (AHTs) and energy-alternating defects (EADs). AHTs have energy levels below Si Ev, which do not change after charging. They dominate the initial build-up of ΔVth, and their saturation is responsible for the "hump" observed in the kinetics.

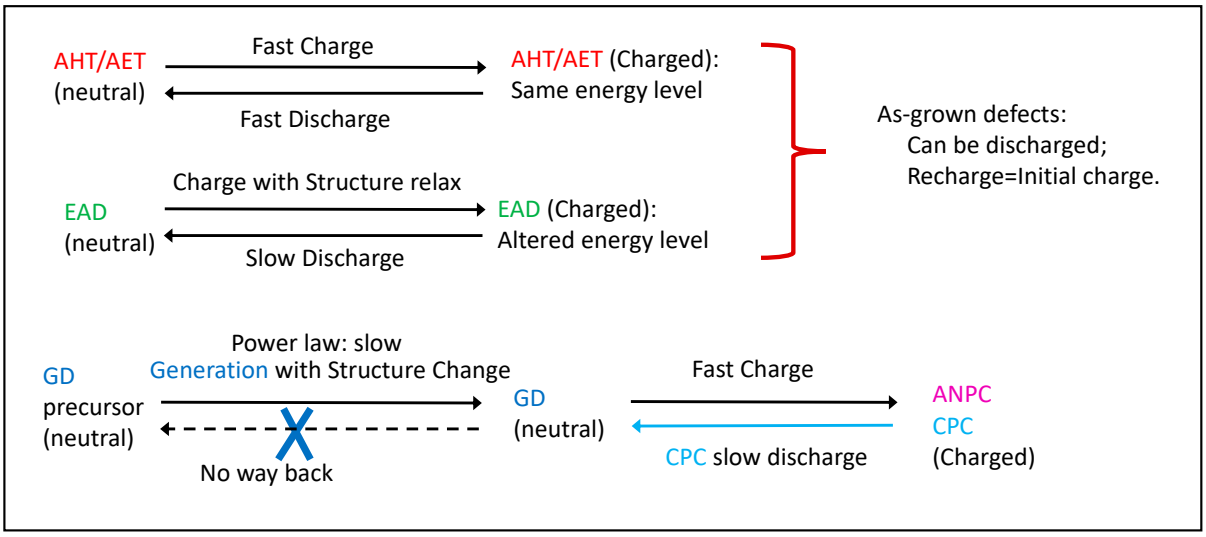

Figure 30. A framework for the defects responsible for bias temperature instabilities.

In contrast, after charging, the EAD structure relaxes and their hole energy level lowers, i.e., their electron energy level rises. This makes discharging EADs less efficient than discharging AHTs. EAD charging follows a power law against time without saturation. This wide spread of charging time may originate from a wide distribution of the energy barrier for charging, as the dielectric is amorphous. The charging time increases exponentially for higher energy barrier.

The precursors for the generated defects cannot be charged directly, and they must be converted to traps first. This conversion process from precursors to traps is referred to as trap generation. The generation results in the structural change of defects; it is a slow process that follows a power law and can be the rate-limiting process for long-term NBTI. Once generated, the trap-filling is relatively fast, and they will not return to their precursor status under typical NBTI test conditions.

Some of the generated traps are within Si bandgap. They can be repeatedly charged–discharged and are called cyclic positive charges (CPCs). The rest of the generated traps can have sufficiently high electron energy levels to remain charged after removing stress bias. They are called anti-neutralization positive charges (ANPC).

After separating the generated defects from the as-grown defects, not only the power law but also the one-to-one correlation between the charges from the generated interface states and those from generated oxide traps were restored as shown in Figure 31. As a result, the differences in the NBTI pre- and post-2000 originated from the contribution of as-grown defects to the post-2000 case. The generated defects were the same. This is not surprising, as the generated defects by NBTI are located close to the dielectric/Si interface [43,48], and they did not change when the dielectric became thinner post-2000.

The one-to-one correlation observed in Figure 31 may lead to the speculation on the electrochemical reaction responsible for the generation process in Figure 32. Holes are an essential reactant, as interface states are not generated in the absence of holes during positive bias temperature stresses [39,55]. At the interface, holes can react with either Si–H or GD-precursors near the interface and release hydrogenous species. For example, the Si–H can be ruptured, and the released hydrogen can then react with the GD-precursor to create a trap. This results in a pair of products: a Pb center as the generated interface state and a generated trap in the dielectric near the interface.

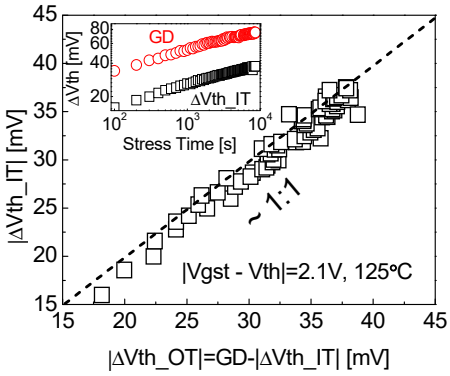

Figure 31. The one-to-one correlation between generated oxide charges and the charges from generated interface states [53].

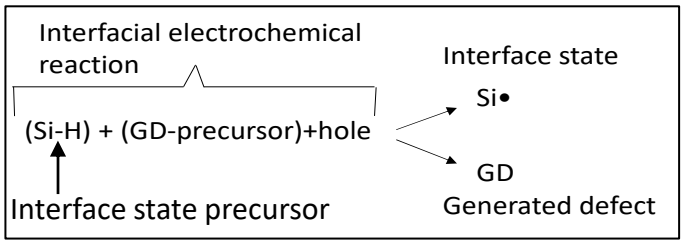

Figure 32. The speculated electrochemical reaction for NBTI. '•' represents a valence electron [17].

2.2.9. A Predictive As-Grown-Generation (AG) Model

Based on the defect framework in Figure 30, the as-grown-generation (AG) model is proposed and the kinetics of each type of defects are given in Table 1. Like the early models [32–34], the AG model fit the test data well as shown in Figure 33. Unlike the early models, the AG model, extracted from fitting the short accelerated test data, can predict long-term NBTI at low operation bias. It should be emphasized that the long-term test data at low bias were not used for the fitting. The samples used in Figure 33 come from different fabrication processes, and the good agreement between the prediction and test data demonstrates the general prediction capability of the AG model.

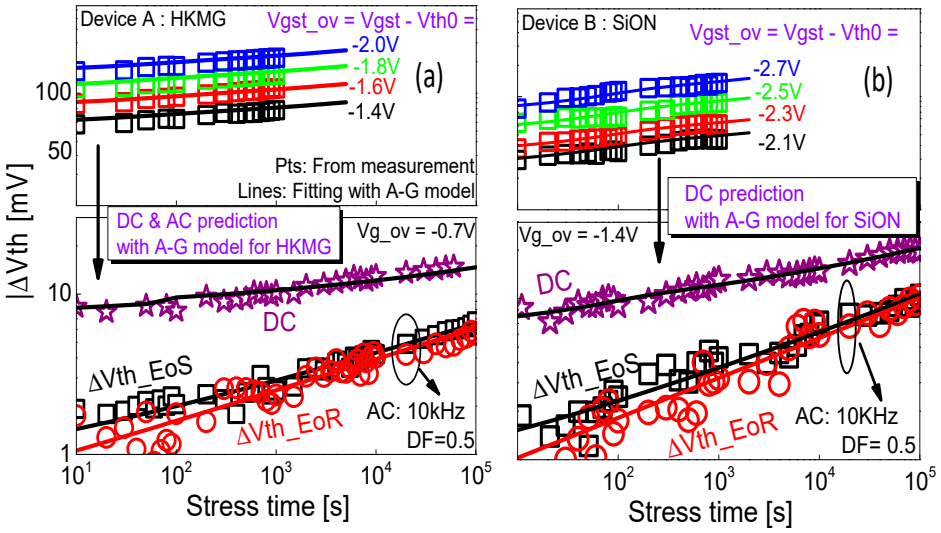

Figure 33. Prediction of NBTI by the AG model for a high-k/SiON stack (**a**) and SiON (**b**). The model parameters were extracted by fitting the accelerated test data in the top panels. They were then used to predict the NBTI at lower biases and for a longer time in the bottom panels [36].

The success of the AG model come from the experimental separation of as-grown traps from the generated ones. Without this separation, all parameters for both as-grown and generated defects have to be extracted together by fitting the measured total ΔVth against time. In this case, there are different kinetics and too many parameters to be reliably extracted from one set of measured data. By experimentally separating the contribution of different defects and kinetics, it becomes possible to fit one set of data with only one kinetics, allowing reliable extraction of its parameters.

3. Positive Bias Temperature Instability (PBTI)

As pMOSFETs and nMOSFETs are switched on by negative and positive gate bias, respectively, NBTI mainly affects pMOSFETs, while PBTI mainly affects nMOSFETs. The relative importance of PBTI against NBTI is process dependent [23–25], and their impact on circuits can be added together rather than cancelling each other out. For example, Figure 34a shows that for a SRAM cell, NBTI and PBTI stresses occur in different inverters, making one inverter different from the other. Both NBTI and PBTI contribute to the reduction of the static noise margin, which is proportional to the size of butterfly eyes in Figure 34b [29]. As a result, both require modeling to optimize circuit performance.

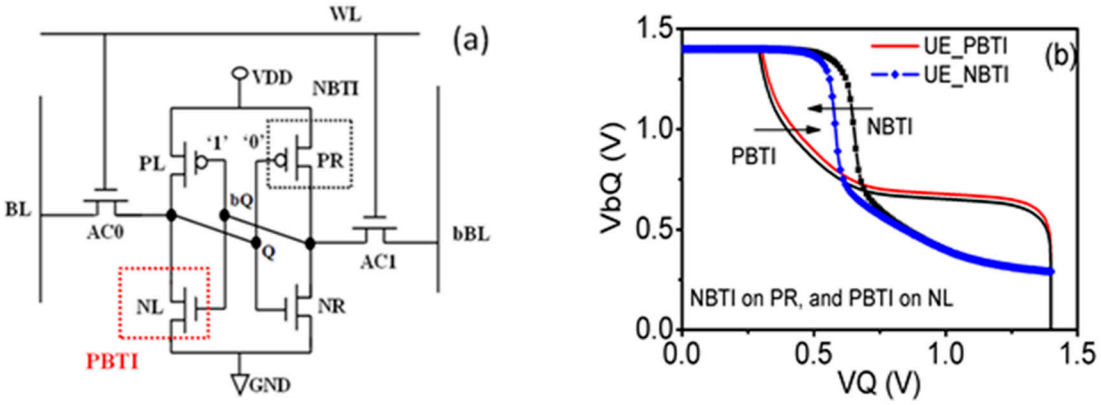

Figure 34. (**a**) When a SRAM cell holds a data bit, PBTI occurs in one of the pull-down nMOSFETs, while NBTI happens in the pull-up pMOSFET of the opposite inverters; (**b**) both PBTI and NBTI contribute to the reduction in the static noise margin (i.e., the size of butterfly eyes) by making the two inverters imbalance [29].

3.1. History of PBTI

Figure 31 shows that NBTI originates from both generated interface states and positive charges formed in the interfacial region of gate oxide. In contrast, Figure 35 shows that interface states are not created for PBTI, so that PBTI only originated from negative charge formation in the gate dielectric through filling acceptor-like electron traps [56–61]. Early works showed that, if arsenium, a common dopant for Si, was left in the gate oxides, they formed electron traps [56]. Water diffused into SiO_2 produces electron traps with a well-defined capture cross section of 10^{-17} cm^2 [57]. When aluminum was used as the gate in early generation CMOS technologies, hydrogenous species also induced smaller traps with capture cross-section on the order of 10^{-18} cm^2 [58–61].

When poly-si was used as the gate for the self-aligned CMOS processes, the high temperature anneal after gate implantation effectively drives these hydrogenous species out of SiO_2. Figure 36 shows that there were little as-grown electron traps for poly-si gated SiO_2, and electron traps must be generated by carrier tunneling through the oxide under a high oxide field [62]. When gate SiO_2 is relatively thick (e.g., >5 nm), electron tunneling through gate oxide during operation is negligible, so that PBTI is insignificant. For thinner SiO_2, tunneling carriers can create new electron traps. These electron traps can act as stepping-stones to form the gate-induced leakage current. They do not form stable space charges in the gate oxide and PBTI is again insignificant.

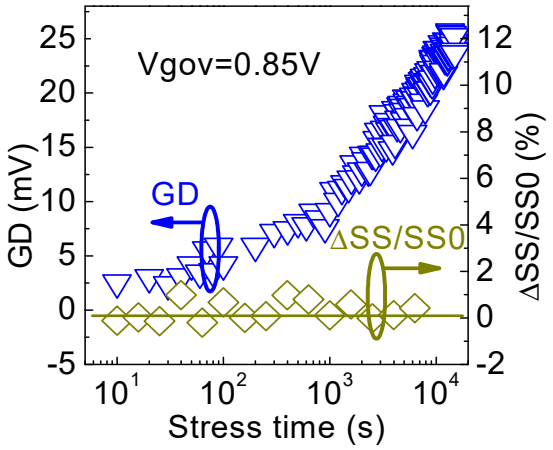

Figure 35. Under positive gate bias, generated defects (GDs) increased, but the negligible change in subthreshold swing (SS) indicates that interface states were not created [39].

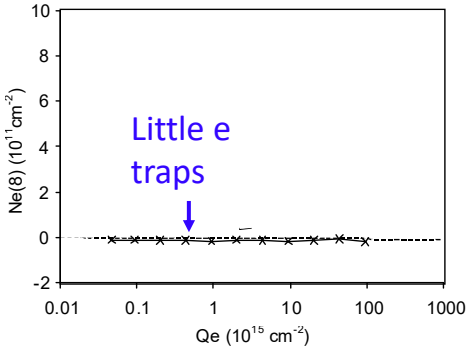

Figure 36. There are little as-grown electron traps in the poly-Si-gated SiO_2 [62].

When the high-k/SiON stack is used, PBTI becomes considerable. In the early stage of high-k process development, PBTI was so severe that it limited the commercial use of the process as detailed in the next section.

3.2. PBTI as the Limiting Instability during the Early Stage of High-k Process Development

Figure 37a,b show the PBTI of a HfO_2 (4 nm)/SiO_2 (1 nm) stack during the development of the high-k process [24,25]. The Id-Vg recorded for the rising and falling Vg pulse edges in Figure 37a is compared in Figure 37b. The Id-Vg recorded at the falling edge was shifted in the positive direction by over half a volt from the Id-Vg of the rising edge. This was caused by electron trapping under a positive Vg during the t_{op} period of several microseconds.

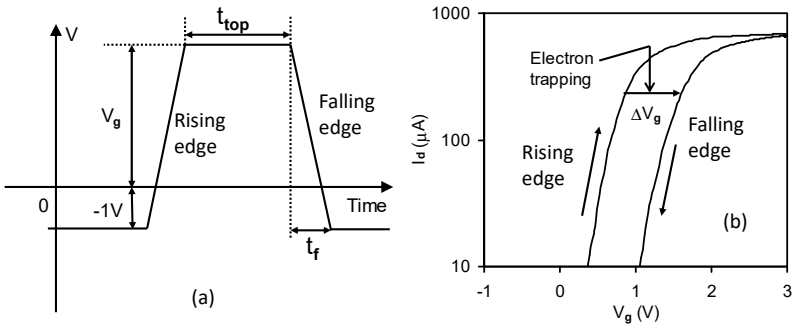

Figure 37. (**a**) The gate bias waveform; (**b**) the pulse Id-Vg recorded at the rising and falling edges of the gate bias [25].

Figure 38 shows that trapped electrons are not stable, and some of them can be lost when the falling edge time is longer than 30 μs [63]. As a result, the energy level of these electron traps is shallow and above the lower edge of silicon conduction band, Ec. These traps are as-grown and can be repeatedly charged and discharged [63].

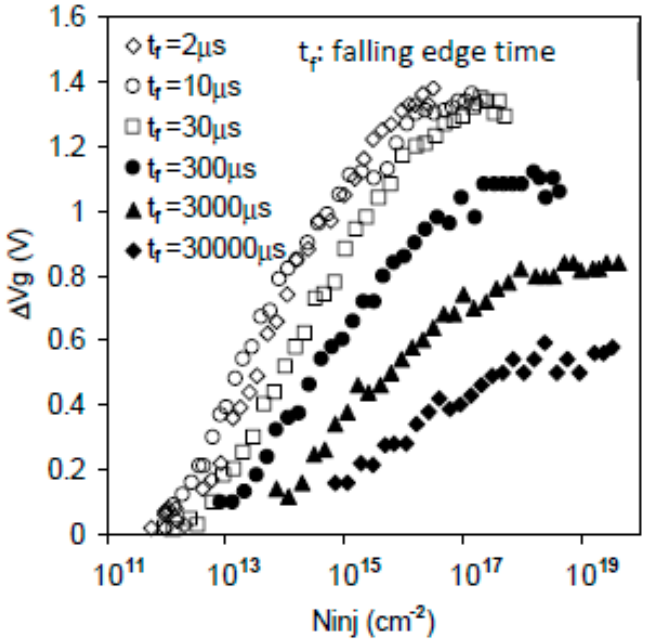

Figure 38. An increase in the falling edge time resulted in a lower trapping level because the trapped electrons can be detrapped before the measurement [63].

Significant efforts have been made to overcome this huge PBTI. To find the location of these as-grown electron traps, the PBTIs of different HfO_2 thicknesses were measured in Figure 39. The grey regions are the assumed trap locations. It can be seen that neither a pile-up of traps at the high-k/SiO_2 interface nor a uniform distribution of traps in the high-k layer agree with the test data. Good agreement was obtained by assuming there were no traps around 1.3~1.8 nm at one or both ends of the high-k layer [63]. Figure 40 shows that PBTI reduced rapidly as the high-k layer became thinner.

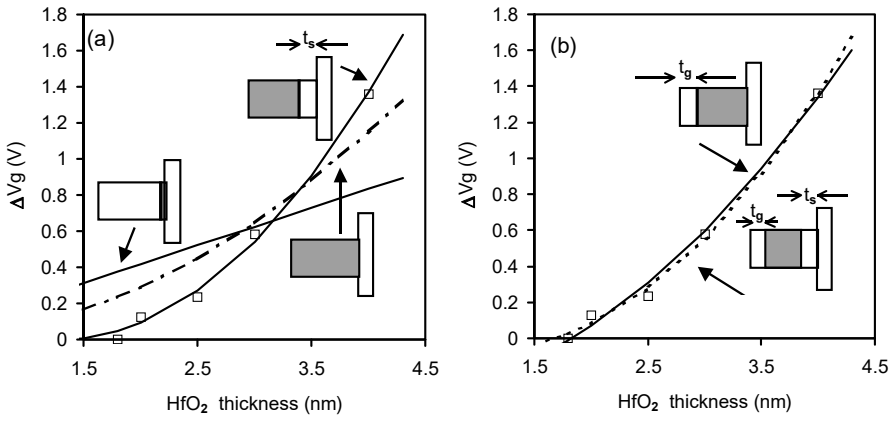

Figure 39. The location of as-grown electron traps in HfO_2. Symbols represent the test data, and the lines are fitted with traps located in the grey regions [63].

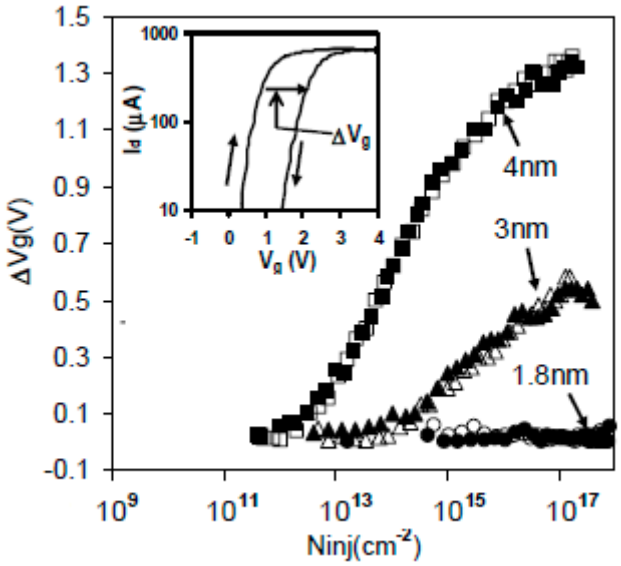

Figure 40. The rapid reduction of as-grown electron trapping with the downscaling of the thickness of HfO_2 [63].

The absence of electron trapping near the end of the high-k layer could be because electrons there can escape to the electrodes and will not form a steady space charge. It is also possible that thick high-k layer could be partially crystallized, resulting in the shallow traps. The suppression of these shallow traps by using thin high-k layers has allowed their commercial use since the advent of 45 nm CMOS technology in 2007.

3.3. PBTI of Modern High-k/SiON Stacks

Although the suppression of shallow as-grown electron traps has reduced PBTI significantly, PBTI still exists in modern commercial CMOS processes with high-k/SiON stacks [9,39,55]. One example is given in Figure 41a, which shows that PBTI is comparable with NBTI [9]. Moreover, Figure 41b shows that the recovery of PBTI is substantially less than that of NBTI. It confirms that these electron traps are energetically deeper than those responsible for the PBTI in the early stage of high-k process development as shown in Figure 37. When compared with hole traps for NBTI that pile up at the dielectric/substrate interface, the electron traps for PBTI were relatively distant from the dielectric/Si interface [43,63], which also contribute to the relative stability of PBTI.

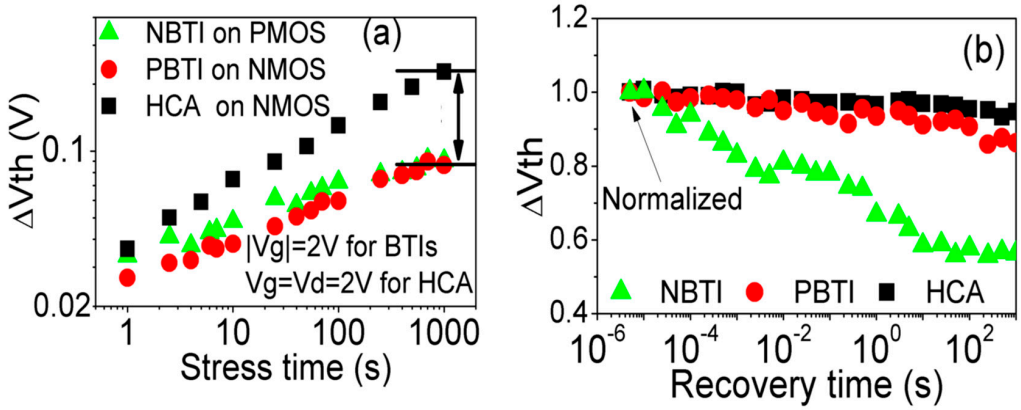

Figure 41. A comparison of PBTI with NBTI during stress (**a**) and recovery (**b**). (**a**) Shows that the PBTI was similar to NBTI during stress for this CMOS process but more stable during recovery [9].

To characterize the electron traps responsible for PBTI, their energy profile was probed. After charging them, as shown in Figure 42a, they were gradually lifted above the Si Ec to

allow them to discharge as shown in Figure 42b [55]. The discharging at different energy levels resulted in the energy profiles in Figure 42c. These electron traps were below Si Ec under flat band conditions and peaked around 1.4 eV below the conduction band edge of HfO_2.

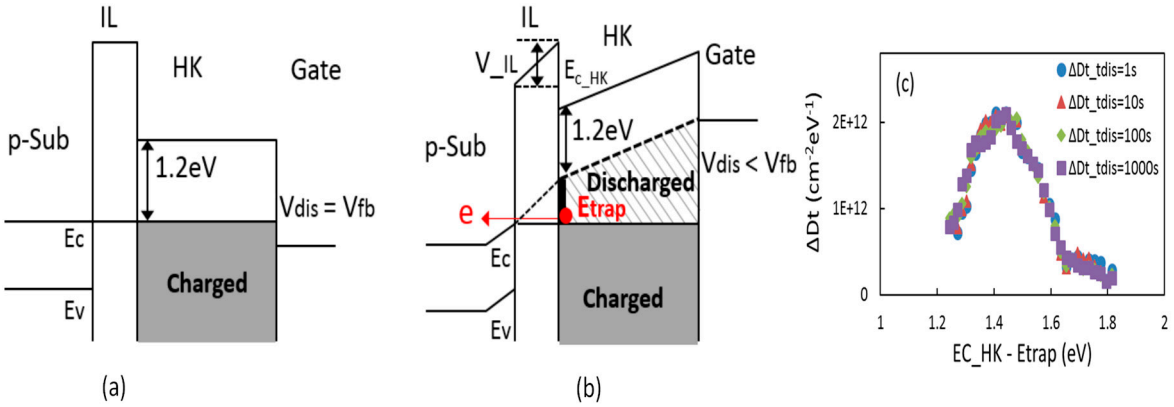

Figure 42. Probing the energy distribution profile of electron traps: (**a**) the electron traps below Si Ec were first charged; (**b**) applying a positive Vg will lift some charged traps above Ec for discharging, i.e., the striped region; (**c**) the extracted energy profile of electron traps by progressively increasing Vg for discharging [55].

3.4. As-Grown Defects for PBTI

The experience of modeling NBTI shows the importance of separating as-grown defects from the generated ones. The question is whether the electron traps observed in Figure 42 are as-grown or generated. To answer it, we charged and then discharged these electron traps by alternating gate bias polarity in the stage 1 of the test in Figure 43a [55]. It can be seen that the charging–discharging was recyclable, indicating that they were as-grown. To further support this, the device was heavily stressed in the stage 2. In the following stage 3, the same gate bias polarity alternation as that in the stage 1 was reapplied. Figure 43b shows that the charging–discharging of electron traps before and after the heavy PBTI stress agrees well, so that they were not affected by the stress, i.e., they are as-grown. After the heavy stress, there are electron traps that cannot be neutralized under Vg = −1.8 V at the end of stage 2. These anti-neutralization electron traps (ANET) did not exist before the heavy stress in the stage 1; thus, they were generated.

Like NBTI, the as-grown defects for PBTI can be divided into as-grown electron traps (AETs) and energy alternating defects (EADs). The energy levels of the AETs did not change with charging–discharging, while the energy levels of the EADs were lowered following charging as shown in Figure 44. This difference allows for their separation as shown in Figure 44 [39].

On filling kinetics, an AET can be filled rapidly, and it saturates with time. On the other hand, filling an EAD follows a power law. The saturation level of AET is determined from the measurement in Figure 44a. This saturation level is then subtracted to obtain the EAD after the AET saturation as shown by the green triangles in Figure 45. These EAD data were fitted with a power law. To obtain the AET before its saturation, the EAD power law was extrapolated to short time as shown by the green, dashed line. An AET over a short time was evaluated by subtracting the extrapolated EAD as shown by the circles in Figure 45.

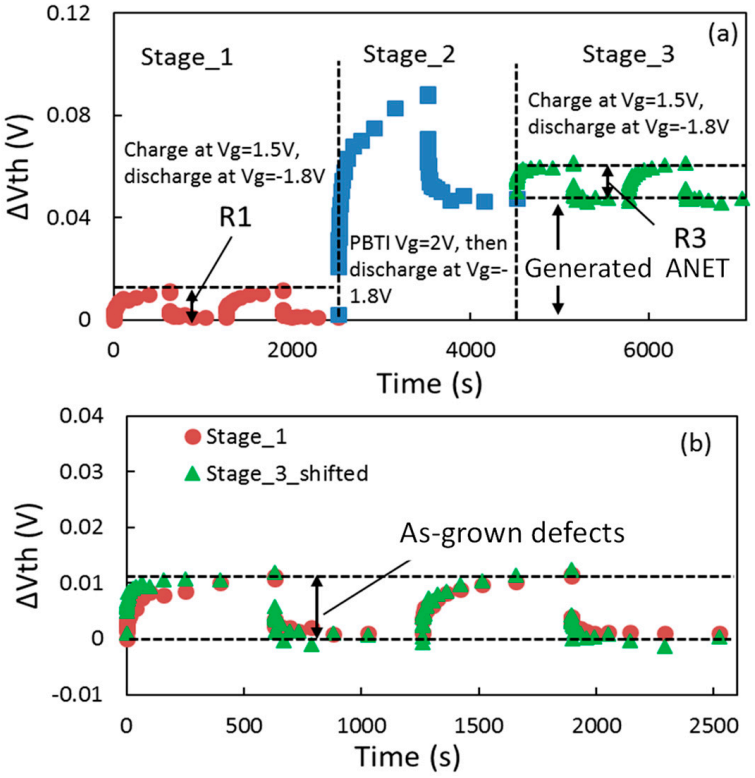

Figure 43. (**a**) Test sequence for confirming the presence of as-grown defects and the generated Anti-neutralization electron traps (ANETs) by PBTI; (**b**) a comparison of the as-grown defects pre- and post-heavy PBTI stress [55].

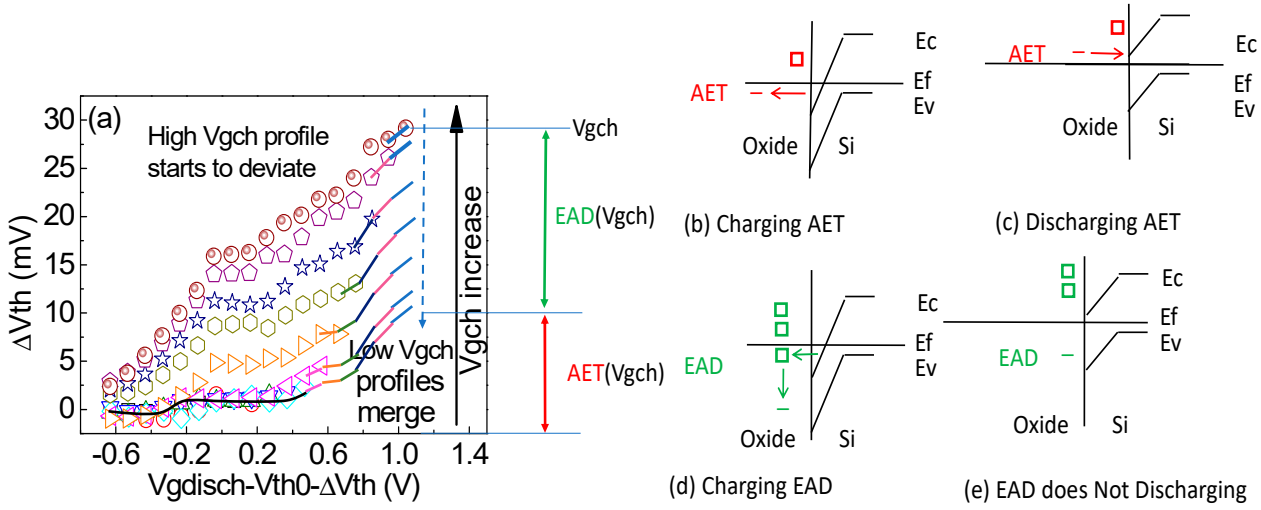

Figure 44. (**a**) Tests for separating as-grown electron traps (AETs) from the as-grown energy-alternating defects (EADs) [39]. (**b**) When an AET is below Ef, it is charged. (**c**) The energy levels of the AETs did not change after charging. It was discharged when above the same Ef for its charging. (**d**) When an EAD is below Ef, it is charged. After charging, the energy level of the EAD is lowered, so that it will not be discharged under the same Ef for its charging in (**e**).

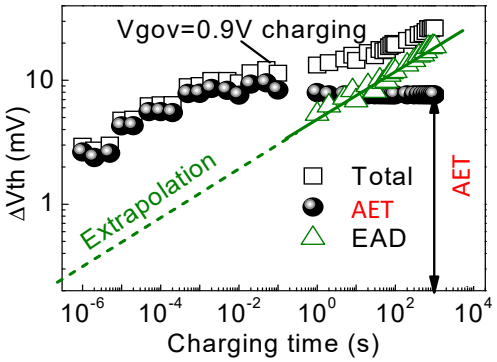

Figure 45. Extracting the kinetics of an EAD and an AET from the measured total ΔVth [39].

The separated AET and EAD at different Vgov are given in Figure 46a,b, respectively. The power exponent of the EAD was insensitive to Vgov, and the AET followed the same kinetics after normalizing against their saturation level.

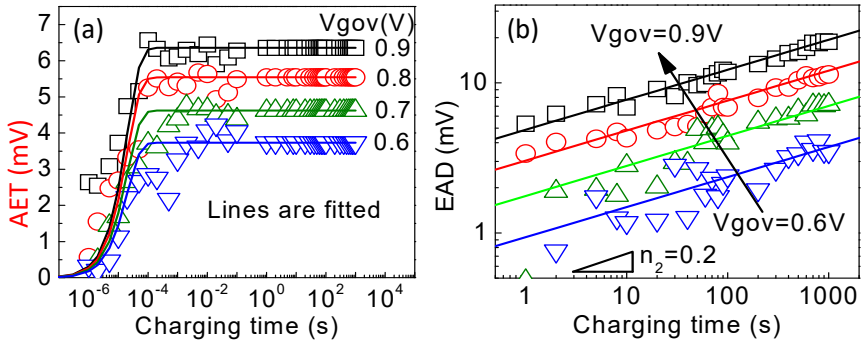

Figure 46. (**a**) AET kinetics under different filling Vgov; (**b**) EAD kinetics under different Vgov [39].

3.5. As-Grown-Generation (AG) Model of PBTI

The measured ΔVth during typical PBTI tests consists of both as-grown and generated defects. Although they could fit the power law well in Figure 47a, the extracted power exponent in Figure 47b depended on the measured delay [39]. For a delay of 1 ms, typically used in early works, the power exponent also changes with stress bias. When the extracted power law was used to predict PBTI at lower bias, Figure 48 shows that there were large discrepancies. As a result, the measured ΔVth must not be used to extract the power law directly, and it is essential to separate it into as-grown and generated defects.

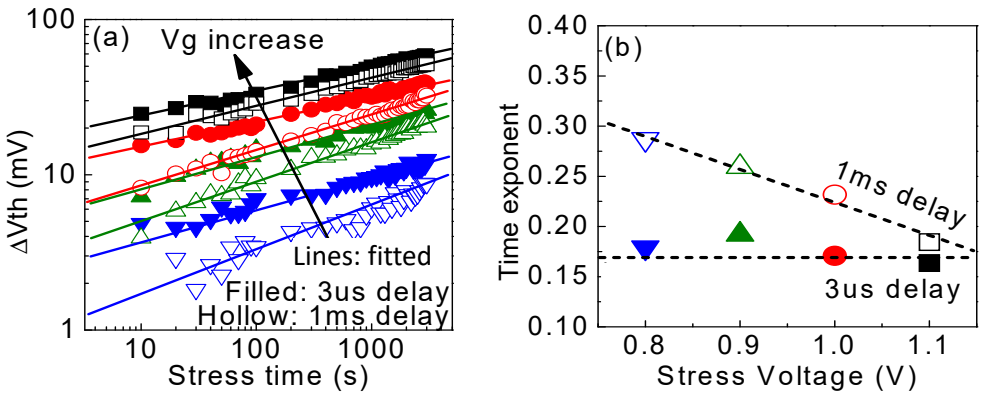

Figure 47. (**a**) Fitting power law with the measured total ΔVth. The lines are fitted, and the symbols are measured data. (**b**) The extracted power exponent depended on the delay time and stress bias [39].

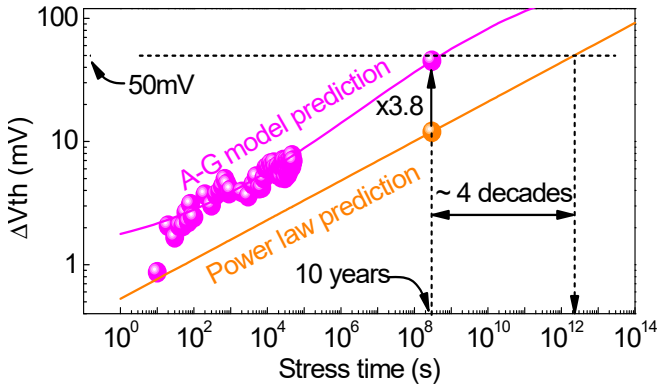

Figure 48. The AG model extracted from the accelerated PBTI tests can predict the PBTI at low biases, while the power law directly fitted with the same test data overestimates PBTI lifetime by 4 orders of magnitude [39].

The stress–discharge–recharge (SDR) technique in Figure 26 can also be applied to PBTI. After removing the contribution of as-grown defects, Figure 49a shows that the power exponent extracted from the generated defects measured by the SDR technique became independent of the measurement conditions. Moreover, Figure 49b shows that the power exponent was insensitive to the stress bias.

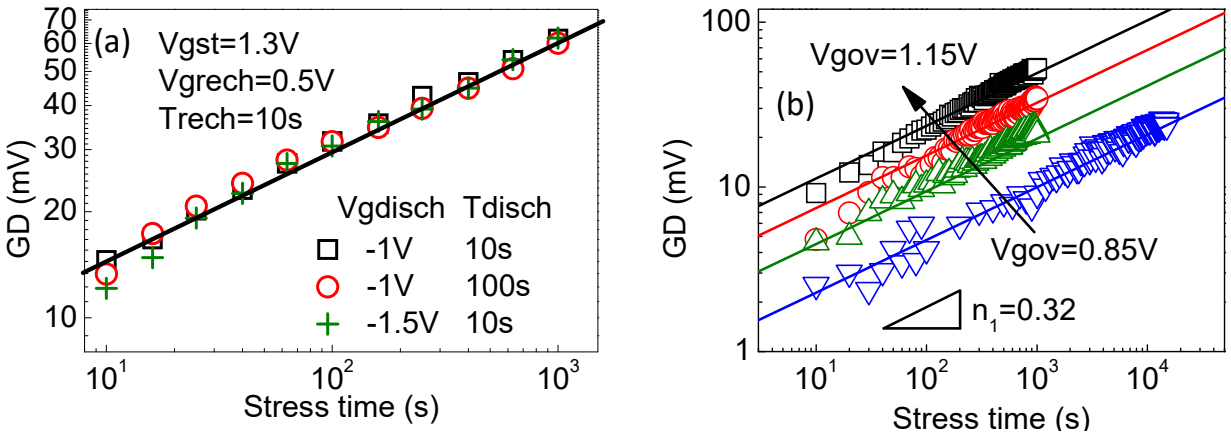

Figure 49. (**a**) The generated defects measured by the SDR technique were independent of the measurement conditions. (**b**) The GD kinetics under different stress Vgov. The power exponents were insensitive to Vgov [39].

By combining the modeling of as-grown defects with that of generated defects, the as-grown-generation model in Table 1 can also be applied to PBTI [39]. Figure 48 shows that the AG model can be used to predict the PBTI at low bias.

4. Conclusions

This work reviewed the frontend reliability issues of CMOS technology. After a brief discussion of the key sources of instability during different nodes of CMOS development, attention was focused on the bias temperature instability of MOSFETs. The as-grown-generation (AG) model was presented, which can predict the long-term BTI under operation biases. An in-depth understanding of underlying physical processes led to the proposal of a framework for the defects responsible for BTI.

There are as-grown defects in fresh devices before electrical stresses, which are defined as the traps whose first charging is the same as any subsequent charging after neutralization. Their density and properties will not be changed by stresses. They can be further divided into as-grown hole/electron traps (AHTs/AETs) and energy-alternating defects (EADs).

The AHTs/AETs are filled rapidly and can dominate the initial BTI. They are saturated with stress time and are responsible for the "hump" often observed in the BTI kinetics. They can be readily neutralized and make a major contribution to the recovery of BTI. Their energy levels do not change after charging. In contrast, the energy level of an EAD becomes deeper after charging, making them more stable than AHTs/AETs. Their charging follows a power law with a power exponent generally different from that of defect generation.

In addition to the as-grown defects, new defects can be generated by converting precursors to electrically active traps. This conversion process may only happen once and is slow and rate-limiting. Filling the generated traps is relatively fast, and the generated traps will not return to their precursor status under normal BTI stress conditions. Some of the generated traps can be repeatedly charged–discharged by alternating gate bias polarity, and they are called cyclic positive charges/electron traps. The rest of generated traps can survive the recovery and are referred to as anti-neutralization positive charges/electron traps.

Early works extract all model parameters by fitting the measured total threshold voltage shift. Although these models can fit the test data well, they cannot predict the long-term BTI at low biases. To make a breakthrough, the authors followed a new approach: the contributions of different defects were experimentally separated first by developing new measurement techniques and each set of data was used to fit the kinetics of only one type of defects. This led to the development of the as-grown-generation (AG) model, which can not only fit the test data but also predict the long-term BTI under low biases.

Author Contributions: This review article was first written by J.F.Z. and all authors contributed equally. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Engineering and Physical Science Research Council of UK under grant numbers GR/L28531/01, GR/R10387/01, EP/C003071/1, EP/I012966/1, EP/L010607/1, and EP/T026022/1.

Acknowledgments: The test samples were supplied by IMEC and Qualcomm Technologies International Ltd., Cambridge, UK. The authors thank their project partners for their support and valuable discussions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Deal, B.E.; Sklar, M.; Grove, A.S.; Snow, E.H. Characteristics of the Surface-State Charge (Qss) of Thermally Oxidized Silicon. *J. Electrochem. Soc.* **1967**, *114*, 114–266. [CrossRef]
2. Jepson, K.O.; Svensson, C.M. Negative bias stress of MOS devices at high electric fields and degradation of MNOS devices. *J. Appl. Phys.* **1977**, *48*, 2004–2014. [CrossRef]
3. Blat, C.E.; Nicollian, E.H.; Poindexter, E.H. Mechanism of negative-bias-temperature instability. *J. Appl. Phys.* **1991**, *69*, 1712–1720. [CrossRef]
4. Ogawa, S.; Shimaya, M.; Shiono, N. Interface-trap generation at ultrathin SiO_2 (4–6 nm)-Si interfaces during negative-bias temperature aging. *J. Appl. Phys.* **1995**, *77*, 1137–1148. [CrossRef]
5. Black, J.R. Electromigration Failure Modes in Aluminum Metallization for Semiconductor Devices. *Proc. IEEE* **1969**, *57*, 1587–1594. [CrossRef]
6. Chen, F.; McLaughlin, P.; Gambino, J.; Wu, E.; Demarest, J.; Meatyard, D.; Shinosky, M. The Effect of Metal Area and Line Spacing on TDDB Characteristics of 45nm Low-k SiCOH Dielectrics. In Proceedings of the 2007 IEEE International Reliability Physics Symposium Proceedings, Phoenix, AZ, USA, 15–19 April 2007; pp. 382–389.
7. Nauta, P.K.; Hillen, M.W. Investigation of mobile ions in MOS structures using the TSIC method. *J. Appl. Phys.* **1978**, *49*, 2862–2865. [CrossRef]
8. Hu, C.; Tam, S.C.; Hsu, F.-C.; Ko, P.-K.; Chan, T.-Y.; Terrill, K.W. Hot-Electron-Induced MOSFET Degradation—Model, Monitor, and Improvement. *IEEE J. Solid-State Circuits* **1985**, *20*, 295–305. [CrossRef]
9. Duan, M.; Zhang, J.F.; Manut, A.; Ji, Z.; Zhang, W.; Asenov, A.; Gerrer, L.; Reid, D.; Razaidi, H.; Vigar, D.; et al. Hot carrier aging and its variation under use-bias: Kinetics, prediction, impact on Vdd and SRAM. In Proceedings of the IEEE International Electron Devices Meeting (IEDM), Washington, DC, USA, 7–9 December 2015; pp. 547–550. [CrossRef]
10. Degraeve, R.; Ogier, J.L.; Bellens, R.; Roussel, P.J.; Groeseneken, G.; Maes, H.E. A New Model for the Field Dependence of Intrinsic and Extrinsic Time-Dependent Dielectric Breakdown. *IEEE Trans. Electron Devices* **1998**, *45*, 472–481. [CrossRef]
11. Kirton, M.J.; Uren, M.J. Noise in solid-state microstructures: A new perspective on individual defects, interface states and low-frequency ($1/f$) noise. *Adv. Phys.* **1989**, *38*, 367–468. [CrossRef]

12. Mehedi, M.; Tok, K.H.; Zhang, J.F.; Ji, Z.; Ye, Z.; Zhang, W.; Marsland, J.S. An assessment of the statistical distribution of Random Telegraph Noise Time Constants. *IEEE Access* **2020**, *8*, 1496–1499. [CrossRef]
13. Mehedi, M.; Tok, K.H.; Ye, Z.; Zhang, J.F.; Ji, Z.; Zhang, W.; Marsland, J.S. On the Accuracy in Modeling the Statistical Distribution of Random Telegraph Noise Amplitude. *IEEE Access* **2021**, *9*, 43551–43561. [CrossRef]
14. Tan, S.S.; Chen, T.P.; Ang, C.H.; Chan, L. Mechanism of nitrogen-enhanced negative bias temperature instability in pMOSFET. *Microelectron. Reliab.* **2005**, *45*, 19–30. [CrossRef]
15. Kaczer, B.; Grasser, T.; Roussel, P.J.; Franco, J.; Degraeve, R.; Ragnarsson, L.-A.; Simoen, E.; Groeseneken, G.; Reisinger, H. Origin of NBTI variability in deeply scaled pFETs. In Proceedings of the 2010 IEEE International Reliability Physics Symposium, Anaheim, CA, USA, 2–6 May 2010; pp. 26–32. [CrossRef]
16. Grasser, T. Stochastic charge trapping in oxides: From random telegraph noise to bias temperature instabilities. *Microelectron. Rel.* **2012**, *52*, 39–70. [CrossRef]
17. Zhang, J.F.; Ji, Z.; Zhang, W. As-grown-generation (AG) model of NBTI: A shift from fitting test data to prediction. *Microelectron. Rel.* **2018**, *80*, 109–123. [CrossRef]
18. Waldhoer, D.; Schleich, C.; Michl, J.; Stampfer, B.; Tselios, K.; Ioannidis, E.G.; Enichlmair, H.; Waltl, M.; Grasser, T. Toward Automated Defect Extraction From Bias Temperature Instability Measurements. *IEEE Trans. Electron Devices* **2021**, *68*, 4057–4063. [CrossRef]
19. Zhang, J.; Wang, Z.; Wang, R.; Sun, Z.; Huang, R. Body Bias Dependence of Bias Temperature Instability(BTI) in Bulk FinFET Technology. *Energy Environ. Mater.* **2021**, 1–4. [CrossRef]
20. Lee, J. Physics-informed machine learning model for bias temperature instability. *AIP Adv.* **2021**, *11*, 025111. [CrossRef]
21. Kishida, R.; Suda, I.; Kobayashi, K. Bias Temperature Instability Depending on Body Bias through Buried Oxide (BOX) Layer in a 65 nm Fully-Depleted Silicon-On-Insulator Process. In Proceedings of the 2021 IEEE International Reliability Physics Symposium (IRPS), Monterey, CA, USA, 21–25 March 2021. [CrossRef]
22. Bhootda, N.; Yadav, A.; Neema, V.; Shah, A.P.; Vishvakarma, S.K. Series diode-connected current mirror based linear andsensitive negative bias temperature instability monitoringcircuit. *Int. J. Numer. Model.* **2022**, *35*, e2953. [CrossRef]
23. Hicks, J.; Bergstrom, D.; Hattendorf, M.; Jopling, J.; Maiz, J.; Pae, S.; Prasad, C.; Wiedemer, J. 45 nm Transistor Reliability. *Intel Technol. J.* **2008**, *12*, 131–144. [CrossRef]
24. Zhao, C.Z.; Zahid, M.B.; Zhang, J.F.; Groeseneken, G.; Degraeve, R.; De Gendt, S. Properties and dynamic behavior of electron traps in HfO_2/SiO_2 stacks. *Microelectron. Eng.* **2005**, *80*, 366–369. [CrossRef]
25. Zhao, C.Z.; Zhang, J.F.; Zahid, M.B.; Govoreanu, B.; Groeseneken, G.; De Gendt, S. Determination of capture cross sections for as-grown electron traps in HfO_2/HfSiO stacks. *J. Appl. Phys.* **2006**, *100*, 093716. [CrossRef]
26. Asenov, A.; Cheng, B.; Dideban, D.; Kovac, U.; Moezi, N.; Millar, C.; Roy, G.; Brown, A.R.; Roy, S. Modeling and simulation of transistor and circuit variability and reliability. In Proceedings of the IEEE Custom Integrated Circuits Conference 2010, San Jose, CA, USA, 19–22 September 2010; pp. 1–8. [CrossRef]
27. Duan, M.; Zhang, J.F.; Ji, Z.; Zhang, W.; Kaczer, B.; Schram, T.; Ritzenthaler, R.; Groeseneken, G.; Asenov, A. New analysis method for time-dependent device-to-device variation accounting for within-device fluctuation. *IEEE Trans. Electron Devices* **2013**, *60*, 2505–2511. [CrossRef]
28. Duan, M.; Zhang, J.F.; Ji, Z.; Ma, J.G.; Zhang, W.; Kaczer, B.; Schram, T.; Ritzenthaler, R.; Groeseneken, G.; Asenov, A. Key issues and Techniques for Characterizing Time-Dependent Device-to-Device Variation of SRAM. In Proceedings of the 2013 IEEE International Electron Devices Meeting, Washington, DC, USA, 9–11 December 2013; pp. 774–777. [CrossRef]
29. Duan, M.; Zhang, J.F.; Ji, Z.; Zhang, W.; Kaczer, B.; Schram, T.; Ritzenthaler, R.; Groeseneken, G.; Asenov, A. Development of a Technique for Characterizing Bias Temperature Instability-Induced Device-to-Device Variation at SRAM-Relevant Conditions. *IEEE Trans. Electron Devices* **2014**, *61*, 3081–3089. [CrossRef]
30. Duan, M.; Zhang, J.F.; Ji, Z.; Zhang, W.; Kaczer, B.; Schram, T.; Ritzenthaler, R.; Thean, A.; Groeseneken, G.; Asenov, A. Time-dependent variation: A new defect-based prediction methodology. In Proceedings of the 014 Symposium on VLSI Technology (VLSI-Technology): Digest of Technical Papers, Honolulu, HI, USA, 9–12 June 2014; pp. 74–75. [CrossRef]
31. Gao, R.; Ji, Z.; Manut, A.B.; Zhang, J.F.; Franco, J.; Hatta, S.W.M.; Zhang, W.; Kaczer, B.; Linten, D.; Groeseneken, G. NBTI-Generated Defects in Nanoscaled Devices: Fast Characterization Methodology and Modeling. *IEEE Trans. Electron Devices* **2017**, *64*, 4011–4017. [CrossRef]
32. Mahapatra, S.; Goel, N.; Desai, S.; Gupta, S.; Jose, B.; Mukhopadhyay, S.; Joshi, K.; Jain, A.; Islam, A.E.; Alam, M.A. A Comparative Study of Different Physics-Based NBTI Models. *IEEE Trans. Electron Devices* **2013**, *60*, 901–916. [CrossRef]
33. Huard, V. Two independent components modeling for Negative Bias Temperature Instability. In Proceedings of the 2010 IEEE International Reliability Physics Symposium, Anaheim, CA, USA, 2–6 May 2010; pp. 33–42. [CrossRef]
34. Huard, V.; Parthasarathy, C.R.; Bravaix, A.; Hugel, T.; Guerin, C.; Vincent, E. Design-in-Reliability Approach for NBTI and Hot-Carrier Degradations in Advanced Nodes. *IEEE Trans. Device Mater. Reliab.* **2007**, *7*, 558–570. [CrossRef]
35. Ji, Z.; Hatta, S.F.W.M.; Zhang, J.F.; Ma, J.G.; Zhang, W.; Soin, N.; Kaczer, B.; De Gendt, S.; Groeseneken, G. Negative Bias Temperature Instability Lifetime Prediction: Problems and Solutions. In Proceedings of the 2013 IEEE International Electron Devices Meeting, Washington, DC, USA, 9–11 December 2013; pp. 413–416. [CrossRef]

36. Ji, Z.; Zhang, J.F.; Lin, L.; Duan, M.; Zhang, W.; Zhang, X.; Gao, R.; Kaczer, B.; Franco, J.; Schram, T.; et al. A test-proven As-grown-Generation (A-G) model for predicting NBTI under use-bias. In Proceedings of the 2015 Symposium on VLSI Technology (VLSI Technology), Kyoto, Japan, 16–18 June 2015; pp. 36–37. [CrossRef]
37. Gao, R.; Ji, Z.; Hatta, S.M.; Zhang, J.F.; Franco, J.; Kaczer, B.; Zhang, W.; Duan, M.; De Gendt, S.; Linten, D.; et al. Predictive As-grown-Generation (A-G) model for BTI-induced device/circuit level variations in nanoscale technology nodes. In Proceedings of the 2016 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 3–7 December 2016; pp. 778–781. [CrossRef]
38. Gao, R.; Manut, A.B.; Ji, Z.; Ma, J.; Duan, M.; Zhang, J.F.; Franco, J.; Hatta, S.W.M.; Zhang, W.; Kaczer, B.; et al. Reliable time exponents for long term prediction of negative bias temperature instability by extrapolation. *IEEE Trans. Electron Devices* **2017**, *64*, 1467–1473. [CrossRef]
39. Gao, R.; Ji, Z.; Zhang, J.F.; Marsland, J.; Zhang, W.D. As-grown-Generation Model for Positive Bias Temperature Instability. *IEEE Trans. Electron Devices* **2018**, *65*, 3662–3668. [CrossRef]
40. Aiello, O.; Fiori, F. On the Susceptibility of Embedded Thermal Shutdown Circuit to Radio Frequency Interference. *IEEE Trans. Electromagn. Compat.* **2012**, *54*, 405–412. [CrossRef]
41. Xie, S. The Design Considerations and Challenges in MOS-Based Temperature Sensors: A Review. *Electronics* **2022**, *11*, 1019. [CrossRef]
42. Kim, J.; Koo, Y.; Song, W.; Hong, S.J. On-Wafer Temperature Monitoring Sensor for Condition Monitoring of Repaired Electrostatic Chuck. *Electronics* **2022**, *11*, 880. [CrossRef]
43. Zhang, J.F.; Chang, M.H.; Ji, Z.; Lin, L.; Ferain, I.; Groeseneken, G.; Pantisano, L.; De Gendt, S.; Heyns, M.M. Dominant layer for stress-induced positive charges in Hf-based gate stacks. *IEEE Electron Device Lett.* **2008**, *29*, 360–1363. [CrossRef]
44. Chang, M.H.; Zhang, J.F. On positive charges formed under negative bias temperature stresses. *J. Appl. Phys.* **2007**, *101*, 024516. [CrossRef]
45. Ji, Z.; Zhang, J.F.; Chang, M.H.; Kaczer, B.; Groeseneken, G. An analysis of the NBTI-induced threshold voltage shift evaluated by different techniques. *IEEE Trans. Electron Devices* **2009**, *56*, 1086–1093. [CrossRef]
46. Ji, Z.; Lin, L.; Zhang, J.F.; Kaczer, B.; Groeseneken, G. NBTI lifetime prediction and kinetics at operation bias based on ultrafast pulse measurement. *IEEE Trans. Electron Devices* **2010**, *57*, 228–237. [CrossRef]
47. Duan, M.; Zhang, J.F.; Ji, Z.; Zhang, W.; Kaczer, B.; De Gendt, S.; Groeseneken, G. Defect loss: A new concept for reliability of MOSFETs. *IEEE Electron Device Lett.* **2012**, *33*, 480–482. [CrossRef]
48. Zhang, J.F.; Sii, H.K.; Groeseneken, G.; Degraeve, R. Hole trapping and trap generation in the gate silicon dioxide. *IEEE Trans. Electron Devices* **2001**, *48*, 1127–1135. [CrossRef]
49. Zhang, J.F.; Zhao, C.Z.; Chen, A.H.; Groeseneken, G.; Degraeve, R. Hole traps in silicon dioxides—Part I: Properties. *IEEE Trans. Electron Devices* **2004**, *51*, 1267–1273. [CrossRef]
50. Zhao, C.Z.; Zhang, J.F.; Groeseneken, G.; Degraeve, R. Hole traps in silicon dioxides—Part II: Generation mechanism. *IEEE Trans. Electron Devices* **2004**, *51*, 1274–1280. [CrossRef]
51. Zhao, C.Z.; Zhang, J.F.; Chang, M.H.; Peaker, A.R.; Hall, S.; Groeseneken, G.; Pantisano, L.; De Gendt, S.; Heyns, M.M. Stress-induced positive charge in Hf-based gate dielectrics: Impact on device performance and a framework for the defect. *IEEE Trans. Electron Devices* **2008**, *55*, 1647–1656. [CrossRef]
52. Zhang, W.D.; Zhang, J.F.; Zhao, C.Z.; Chang, M.H.; Groeseneken, G.; Degraeve, R. Electrical signature of the defect associated with gate oxide breakdown. *IEEE Electron Device Lett.* **2006**, *27*, 393–395. [CrossRef]
53. Gao, R. Bias Temperature Instability Modelling and Lifetime Prediction on Nano-Scale MOSFETs. Ph.D. Thesis, Liverpool John Moores University, Merseyside, UK, 2018.
54. Zhao, C.Z.; Zhang, J.F. Effects of hydrogen on positive charges in gate oxides. *J. Appl. Phys.* **2005**, *97*, 073703. [CrossRef]
55. Duan, M.; Zhang, J.F.; Ji, Z.; Zhang, W.; Vigar, D.; Asenov, A.; Gerrer, L.; Chandra, V.; Aitken, R.; Kaczer, B. Insight into Electron Traps and Their Energy Distribution under Positive Bias Temperature Stress and Hot Carrier Aging. *IEEE Trans. Electron Devices* **2016**, *63*, 3642–3648. [CrossRef]
56. DeKeersmaecker, R.F.; DiMaria, D.J. Electron trapping and detrapping characteristics of arsenic-implanted SiO_2 layers. *J. Appl. Phys.* **1980**, *51*, 1085–1101. [CrossRef]
57. Nicollian, E.H.; Berglund, C.N.; Schmidt, P.F.; Andrews, J.M. Electrochemical Charging of Thermal SiO_2 Films by Injected Electron Currents. *J. Appl. Phys.* **1971**, *42*, 5654–5664. [CrossRef]
58. Zhang, J.F. Oxide defects. In *Bias Temperature Instabilities for Devices and Circuits*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 253–285. [CrossRef]
59. Zhang, J.F.; Taylor, S.; Eccleston, W. Electron Trap Generation in Thermally Grown Silicon Dioxide Under Fowler-Nordheim Stress. *J. Appl. Phys.* **1992**, *71*, 725–734. [CrossRef]
60. Zhang, J.F.; Taylor, S.; Eccleston, W. A Quantitative Investigation of Electron Detrapping in SiO2 Under Fowler-Nordheim Stress. *J. Appl. Phys.* **1992**, *71*, 5989–5996. [CrossRef]
61. Zhang, J.F.; Taylor, S.; Eccleston, W. A Comparative Study of The Electron Trapping and Thermal Detrapping in SiO_2 Prepared by Plasma and Thermal Oxidation. *J. Appl. Phys.* **1992**, *72*, 1429–1435. [CrossRef]

62. Zhang, W.D.; Zhang, J.F.; Lalor, M.; Burton, D.; Groeseneken, G.; Degraeve, R. Two types of neutral electron traps generated in the gate silicon dioxide. *IEEE Trans. Electron Devices* **2002**, *49*, 1868–1875. [CrossRef]
63. Zhang, J.F.; Zhao, C.Z.; Zahid, M.B.; Groeseneken, G.; Degraeve, R.; De Gendt, S. An assessment of the location of as-grown electron traps in HfO_2/HiSiO stacks. *IEEE Electron Device Lett.* **2006**, *27*, 817–820. [CrossRef]