
This is the author’s final accepted version.

There may be differences between this version and the published version. You are advised to consult the publisher’s version if you wish to cite from it.

http://eprints.gla.ac.uk/173658/

Deposited on: 20 November 2018
Thorough Understanding of Retention Time of Z2FET Memory Operation

M. Duan, C. Navarro, B. Cheng, F. Adamu-Lema, X. Wang, V. P. Georgiev, F. Gamiz, C. Millar, A. Asenov

Abstract—A recently reported zero impact ionization and zero subthreshold swing device Z2FET is a promising candidate for capacitor-less DRAM memory cell. In memory operation, data retention time determines refresh frequency, and is one of the most important memory merits. In this paper, we have systematically investigated the Z2FET retention time based on a newly proposed characterization methodology. It is found that the degradation of HOLD '0' retention time originates from the Gated-SOI portion rather than the Intrinsic-SOI region of the Z2FET. Electrons accumulate under Front Gate (FG) and finally collapse the potential barrier turning logic '0' to '1'. It appears that Shockley-Read-Hall (SRH) generation is the main source for electrons accumulation. Z2FET scalability has been investigated in terms of retention time. As the Z2FET is downscaled, the mechanism dominating electrons accumulation switches from SRH to parasitic injection of electrons from the cathode. Results show that downsampling of Lg has little effect on data '0' retention, but Lin is limited to ~125nm. An optimization method of the fabrication process is proposed based on this new understanding, and Lin can be further scaled down to 75nm. We have demonstrated by 2D TCAD simulation that Z2FET is a promising DRAM cells candidate particularly for IoT applications.

Index Terms — Z2FET, TCAD, DRAM, Transistor, Volatile Memory, Retention Time, Shockley-Read-Hall, Generation, Recombination, Injection.

I. INTRODUCTION

TRADITIONAL memory devices are facing increasing down-scaling challenges. 6T-SRAM suffers from creeping variability [1-2] and reliability [3-5] issues, which results in cell instability problems. One transistor and one capacitor DRAM cells (1T1C) struggle to maintain reasonable refresh time [6-7]. Efforts have been made to find new memory solutions, such as capacitor-less cells [8-10]. Floating body based memory structures are among the potential candidates, but impact ionization or band-to-band tunnelling (B2BT) limits their performance [11]. A recently proposed zero impact ionization and zero subthreshold swing device named Z2FET [10, 12], has shown significant technology advantages including CMOS technology compatibility, novel capacitor-less memory action and sharp switching characteristics, becomes a promising candidate for capacitor-less DRAM memory cell.

II. DEVICE AND TCAD SIMULATION DECK

The Z2FET (Fig. 1) has p-i-n structure on SOI substrate, with partial front gate (FG) and back gate (BG) regions controlling lateral potential barriers. The fabrication process is fully compatible with STMicroelectronics 28 nm FDSOI technology [13]. All simulations were carried out using Sentaurus Sdevice [14]. For memory operation, complementary potential barriers are established in the gated and intrinsic regions which prevents/allows electrons and holes to flows through the channel, depending on the electrons amounts stored under the front gate. For detailed memory operation please refer to the previously published papers [15-17]. Unless otherwise specified, devices with Lin=LG=200nm are used for simulation. The physical models [17] used in the simulation are summarized in Table 1:

<table>
<thead>
<tr>
<th>Models</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carrier statistics</td>
<td>Fermi Dirac Distribution</td>
</tr>
<tr>
<td>Temperature</td>
<td>300 K</td>
</tr>
<tr>
<td>Carrier Mobility</td>
<td>Doping dependence, Philips Model, High Field Saturation and Exponential mobility model</td>
</tr>
<tr>
<td>Effective intrinsic density</td>
<td>NoBand gap narrowing</td>
</tr>
<tr>
<td>Generation and recombination</td>
<td>Shockley Read Hall (Doping and Temperature Dependancy)</td>
</tr>
<tr>
<td>Impact ionization</td>
<td>Not considered</td>
</tr>
<tr>
<td>Band to Band tunnelling</td>
<td>NonlocalPath, Phonon assisted-110 and Direct model</td>
</tr>
<tr>
<td>Quantum Mechanics</td>
<td>Density Gradient (LocalModel=SchenkBGN)</td>
</tr>
</tbody>
</table>

Fig. 1 Z2FET cross section structure used for simulation. Three p-n junctions (J1, J2 and J3) are established by '+' VFG and '-' VBG bias. Both Lin and Lg are 200 nm for nominal device.

Table 1: Physical models and details used in the simulation
To examine the memory viability between data ‘0’ and ‘1’, the anode voltage (VA) is swept from 0 to 1.5V (Fig. 2a). A higher VA (SW1) is required to switch the device on if the initial state is ‘0’ owing to the higher established potential barrier in Gated-SOI region. Consequently, a lower VA is needed to switch the device off [15-16]. The simulations of all specific memory operation such as program ‘0’ (P0), program ‘1’ (P1), hold (H), and read (R) are illustrated in Fig. 2.

### III. RETENTION TIME SIMULATION

In a conventional 1T1C DRAM cell, retention time indicates the capability that capacitor is able to retain sufficient charge to build up a readable differential voltage for the sense amplifier. To compensate for the capacitor leakage, the memory cell need to be refreshed periodically. In Z2FET, the stored ‘1’ state is an equilibrium state and there is no need to refresh. We will concentrate on state ‘0’ in this paper. Two types of retention time will be examined: HOLD ‘0’ (H0_Retention) and READ ‘0’ (R0_Retention).

#### A. Retention Time Extraction

The straightforward way of measuring H0 retention time is to apply a pulsed anode voltage and to monitor its current. The interval between P0 and the time output (IA) becomes ‘1’ determines the H0 retention time. This however cannot provide the intrinsic H0 retention time since the applied pulsed-VA accelerates the collapse of stored data and contaminates the obtained retention time [16]. Our new extraction procedure (Fig. 3) can be described as followings: 1) Ensure data ‘0’ is stored by P0. 2) Bias to Hold condition and monitor the potential (ψ) at point ‘P’ of Si channel (Fig. 1). 3) The time when potential drops to a certain level close to the saturation at point ‘N’ is defined as retention time. Reading before and after potential collapse in Fig. 3(c) justifies this methodology. The potential collapse under the FG is the basic evidence of data ‘0’ loss, and will be used hereafter to extract retention time.

#### B. Hold ‘0’ Retention time

Fig. 4 shows the retention time under hold ‘0’ condition (H0_Retention) for different VBG, VFG and different temperatures. Contrary to expectations, the retention time is constant with respect to applied voltages. Obviously, efforts to improve the H0_Retention by adjust VFG and VBG biases are counterproductive.

#### C. Read ‘0’ Retention time

It is also contrary to H0_Retention, Fig. 5 shows that read ‘0’ retention time (R0_Retention) has a strong dependency on both VBG and VFG. R0_Retention benefits from lower |VBG| and higher VFG. For a particular VFG value, the R0_Retention time can become infinitely long (>1000s in this simulation) for all simulated temperatures, as electron / hole cannot accumulate under the gate / intrinsic region. Therefore, R0_Retention is not a limiting factor for Z2FET operation, and a careful choose of VFG/VBG can locate the read operation in the ‘Safe’ region. Hereafter we will focus on the retention at hold condition to understand its degradation mechanism.

### IV. DISCUSSION OF RETENTION DEGRADATION MECHANISM

Under hold condition, VFG is biased positively and there are several possibilities to degrade the stored data ‘0’. The degradation is associated with the accumulation of electrons in the Gated-SOI region. These electrons might originate either from Shockley-Read-Hall (SRH) induced generation, or injection from cathode terminal. Either non-effective intrinsic potential barrier or holes accumulation induced barrier lowering in the Intrinsic-SOI can enhance cathode injection.

#### A. Carrier Concentration Evolution

Fig. 6 shows the evolution of the potential distribution along the Z2FET channel for the transition from non-equilibrated ‘0’ to equilibrated state ‘1’, which constitutes the degradation of data ‘0’. During degradation, the electron concentration in the...
Gated-SOI region rises by 8 order of magnitude, while the concentration of holes in the Intrinsic-SOI remain almost constant. Therefore, the collapse of the data ‘0’ is due to electron accumulation in Gated-SOI region.

B. SRH Generation-Recombination & B2BT

Shockley-Read-Hall (SRH) generation happens in a two-step process: 1) an electron in valence band is captured into a trap located in the bandgap. 2) the trapped electron gains additional energy to reach the conduction band. The completion of SRH process leaves a free hole in valence band and electron in conduction band. According to Eq. (1) [14], SRH rate is mainly determined by non-equilibrated electron/hole concentration and the corresponding lifetimes. The lifetimes are temperature sensitive Eq. (2) [14]. Since $\alpha$ is negative, higher temperature reduces the carrier lifetime and increases the SRH rate.

$$SRH = \frac{np - n_i^2}{\tau(p + p_i)}$$

(1)

$$\tau(T) = \tau(300K) \left( \frac{T}{300K} \right)^{\alpha}$$

(2)

Where $n$, $p$ and $n_i$ are electron, hole and intrinsic carrier concentration separately. $n_i = n_i \exp \left( \frac{E_{trap}}{kT} \right)$, $p_i = n_i \exp \left( -\frac{E_{trap}}{kT} \right)$, and $E_{trap}$ is the trap energy level relative to the intrinsic Fermi level. $\tau_e$, $\tau_h$ are electron, hole lifetimes correspondingly.

B2BT is insensitive to temperature and driven by the electric field according to Eq. (3) [14].

$$B2BT = A\phi^0 \exp \left( -\frac{R}{T} \right)$$

(3)

Due to the virtual bias induced doping, the Z2FET features three separate p-n junctions. The distribution of SHR generation-recombination along the channel is plotted in Fig. 7. The SHR mainly occurs in the region J1 and J3. Negative SHR of J1 indicates recombination and positive value of J3 indicates generation. B2BT generation is only significant in the J3 region, but quantitatively, it is negligible comparing to SRH generation.
The simulated time dependent SRH generation at J2, J3 and gated-SOI is illustrated in Fig. 8. The SRH at J3, being 300 times larger than that at J2, dominates the generation process during non-equilibrium transitional state. The SRH within Gated-SOI is the lowest one and drops from the level of SRH_J2, until reaching saturation. Overall SRH rate in J2 and Gated-SOI region are insignificant and affect little the electrons accumulation.

Fig. 9 (a) and (b) presents the simulated SRH_J3 generation rate under different VFG, VBG biases and temperatures. SRH_J3 changes little with gate biases but is drastically affected by temperature: higher thermal excitation enhances SRH. All these features are coherent with our previous analysis. To rule out the contribution from B2BT, we also simulated B2BT generation by monitoring the peak value in J3. Fig. 9 (c) and (d) shows that under various front/back gate biases and temperatures, B2BT generation is always much lower than SRH generation. Therefore, B2BT contribute little to the electronics accumulation in the Gated-SOI region.

Finally, we verify the cathode current leakage contribution to the electron accumulation in Gated-SOI. A negative cathode leakage current (inset in Fig. 10) is observed during hold ‘0’, which means electrons are flowing into Z2FET, giving a fake impression that this leakage contributes to the electrons accumulation.

Firstly, the current magnitude is quite small (~10^-14 A). Secondly, we have observed simultaneous SRH carrier recombination happening at J1 (Fig. 7). When comparing the evolution of this leakage current with SRH_J1 after normalization, we find they are exactly following each other. This strongly indicates that most of the leakage current is only contributing to the SRH recombination. Although some electrons can possibly diffuse towards the anode, the magnitude should be insignificant; otherwise, there would be a deviation between the leakage current and the SRH rate. Hence, we conclude that the accumulated electrons in the Gated-SOI region are mainly from SRH generation, which occurs in the J3 region. Both B2BT and the leakage current from cathode has a negligible impact on the electrons accumulation.

V. SCALABILITY

A. Scalability of Lin and Lg Length

The Z2FET scalability was analyzed for different VFG biases in terms of Lin and Lg in Fig. 11. Almost identical H0-Retention values (> 1ms) are obtained when shortening Lg from 200 nm to 30 nm regardless of the employed VFG (Fig. 11a). Contrarily, downsizing Lin below 125 nm severely disturbs the retention time (Fig 11b). Moreover, there is a clear VFG dependency where the retention time dramatically drops for higher VFG (inset of Fig 11b). This is not consistent with Fig. 4 and 9, where SRH generation dominates the degradation. So, a distinct mechanism differing from SRH must be involved in the data ‘0’ loss in scaled Z2FETs. This can be confirmed in Fig. 12 where the SRH generation rate at J3 is plotted as a function of Lin and no significant change is noticed.

B. Current Injection from Cathode

The cathode leakage current is monitored to investigate what accelerates the data ‘0’ degradation in shorter Lin device.
Unlike long device (Fig. 10) where the injected current was insignificant, in short Lin device the cathode current is much larger. More interestingly, the cathode current exhibits two distinct plateaus. A disagreement (Fig. 13) is observed when comparing again the normalized cathode leakage current and SRH recombination. Coincidentally, the SRH recombination presents good agreement with the lower plateau when shifted downwards. This suggests that the upper plateau current flow (shadow part) is effectively contributing to the electron accumulation in Gated-SOI, further degrading the data ‘0’ retention time. This can be also confirmed from the coincidence that the dropping down time of upper plateau (100ns) is similar with the retention time (Lin=75nm) in Fig. 11.

Fig. 14 summarizes the different ‘0’ state degradation mechanisms. In long Lin Z2FET, SRH generation dominates and is responsible of repopulating the Gated-SOI with electrons. On the other hand, in short Lin Z2FETs, the degradation mechanism swaps to the parasitic electron injection from the cathode due to the poor back-gate electrostatic controllability.

C. Optimization

Fig 11 demonstrated a minimum Lin of 125 nm to achieve at least 1 ms retention time. In order to improve the scalability, the cathode current injection needs to be limited. A thinner intrinsic-Si film (Fig. 15) would be able to enhance the back-gate electrostatic control and thus suppress cathode leakage current.

Fig. 13 Normalized current injection from cathode (IK) and SRH recombination rate at J1 region. IK has two plateaus. Higher plateau does not match SRH_J1 recombination but lower one does.

Fig. 14 Illustration of different mechanisms of data ‘0’ degradation. For Z2FET with long Lin, electrons accumulation is from SRH generation. However, for short Lin device, electron injection induced leakage current starts dominating.

Fig. 15 Z2FET structure optimization by removal of the raised Si along Intrinsic-SOI to improve back gate electrostatics.

Fig. 16 (a) compares the leakage current before and after optimization. For shorter Lin device (<125nm), the leakage current can be reduced by a factor of 1000. Consequently, with the epitaxy optimization the scaling of Lin can reach ~75 nm as shown in Fig. 16 (b). Although leakage current in Lin > 125 nm device is suppressed by optimization, the retention time is similar compared with the non-optimized one (Fig. 11b), confirming that cathode leakage current does not play an important role.
In this paper, we have systematically investigated the Z2FET retention time based on a novel characterization methodology. Extreme long retention time under read condition can be achieved by optimizing VFG/VBG bias. However, under hold condition the retention time almost independent on the biasing conditions. Hence, emphasis of this study was on the degradation mechanism under hold condition. We showed that SRH generation at J3 region dominates the retention degradation. The leakage current from cathode is negligible. However, when Lin is scaled down to less than 125nm, leakage current start to dominate the data '0' degradation. An optimization approach is proposed which reduces the leakage current by reducing the thickness of epitaxial layer in the Intrinsic-SOI. Simulations show that the Lin scalability can be achieved by optimizing VFG/VBG bias. However, under hold condition the retention time almost independent on the biasing conditions. Hence, emphasis of this study was on the degradation mechanism under hold condition. We showed that SRH generation at J3 region dominates the retention degradation. The leakage current from cathode is negligible. However, when Lin is scaled down to less than 125nm, leakage current start to dominate the data '0' degradation. An optimization approach is proposed which reduces the leakage current by reducing the thickness of epitaxial layer in the Intrinsic-SOI. Simulations show that the Lin scalability can be improved and down to 75nm, which in turn will be able to increase significantly the DRAM cells volume density.

The authors would like to acknowledge Prof. S. Cristoloveanu for some fruitful discussion, Prof. M. Bawedin, Dr. J. Lacord, and Dr. M.S. Parihar for the supply of simulation deck. Thanks Dr. P. Galy and Dr. K.H. Lee for some technical discussion. This work is supported financially by Horizon2020 European Project: REMINDER/687931.

VI. CONCLUSION

In this paper, we have systematically investigated the Z2FET retention time based on a novel characterization methodology. Extreme long retention time under read condition can be achieved by optimizing VFG/VBG bias. However, under hold condition the retention time almost independent on the biasing conditions. Hence, emphasis of this study was on the degradation mechanism under hold condition. We showed that SRH generation at J3 region dominates the retention degradation. The leakage current from cathode is negligible. However, when Lin is scaled down to less than 125nm, leakage current start to dominate the data '0' degradation. An optimization approach is proposed which reduces the leakage current by reducing the thickness of epitaxial layer in the Intrinsic-SOI. Simulations show that the Lin scalability can be improved and down to 75nm, which in turn will be able to increase significantly the DRAM cells volume density.

The authors would like to acknowledge Prof. S. Cristoloveanu for some fruitful discussion, Prof. M. Bawedin, Dr. J. Lacord, and Dr. M.S. Parihar for the supply of simulation deck. Thanks Dr. P. Galy and Dr. K.H. Lee for some technical discussion. This work is supported financially by Horizon2020 European Project: REMINDER/687931.

REFERENCES


[13] H. E. Dirani; M. Bawedin; K. Lee; M. Parihar; X. Mesicot; P. Fonteneau; Ph. Galy; F. Gamiz; Y.-T. Kim; P. Ferrari; S. Cristoloveanu, “Competitive 1T-DRAM in 28 nm FDSOI technology for low-power embedded memory”, 2016 IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S), 2016. DOI: 10.1109/S3S.2016.7804402


[17] C. Navarro; M. Duan; M. S. Parihar; F. Adamu-Lema; S. Coseman; J. Lacord; K. Lee; C. Sampedro; B. Cheng; H. Dirani; J. C. Barbe; P. Fonteneau; S. Kim; S. Cristoloveanu; M. Bawedin; C. Millar; P. Galy; C. L. Royer; S. Karg; H. Riel; P. Wells; Y. T. Kim; A. Asenov; F. Gamiz, “Z²-FET as Capacitor-Less eDRAM Cell For High-Density Integration”, IEEE Transactions on Electron Devices, Vol. 64, No. 12, pp. 4904, 2017. DOI: 10.1109/TED.2017.2759308