Junction Engineering of 1T-DRAMs
Gino Giusi and Giuseppe Iannaccone

Abstract—One-transistor dynamic random access memories (DRAMs) (1T-DRAMs) are considered a promising candidate to overcome the limits of scalability of conventional one-transistor/one-capacitor DRAMs. Robust and reproducible operation has been demonstrated by experiments in MOSFET devices with a gate length ($L$) down to $\sim$50 nm, which prevents their use in future technological nodes. The main factors limiting the retention time of 1T-DRAMs are the Shockley–Read–Hall recombination in the channel and the band-to-band tunneling between channel and source/drain junctions, both enhanced by the relatively high field at both junctions. In this letter, we show through statistical and device simulations that, by introducing an underlap of $\sim$16 nm between the drain (source) junction and the gate, it is possible to reduce both the electric field at the junction and the impact of process variability, achieving 1T-DRAMs with $L = 10$ nm with a retention time in excess of 100 ms. We also show that field plates at the source and drain contacts do not provide additional advantages and that the junctionless transistor operation as 1T-DRAM is totally undermined by the impact of random dopants.

Index Terms—Band-to-band tunneling (BTBT), junctionless (JL), one-transistor dynamic random access memory (1T-DRAM), random dopant fluctuation, trap-assisted tunneling (TAT).

I. INTRODUCTION

One-transistor dynamic random access memories (1T-DRAMs) have recently gained attention as a replacement of conventional one-transistor–one-capacitor memory cells [1]–[3]. Variations of the single-gate or multigate MOSFET structure have been proposed via device simulation as options to increase retention, but often performance was overestimated because authors neglected the process of generation/recombination enabled by trap-assisted tunneling (TAT), which leads to reduced Shockley–Read–Hall (SRH) lifetime with increasing electric field. In 1T-DRAMs, TAT becomes the main factor limiting the retention time because of the high junction field. This aspect is very pronounced in small device structures because the junctions are very close to the middle of the channel, where the excess charge is stored. A similar degradation of the retention time is due to field-enhanced generation, i.e., band-to-band tunneling (BTBT) of charge carriers, which is usually included in device simulations. Some aspects of the influence of the junction field on the retention time have been already discussed [4], [5]. In this letter, we show that, by a proper junction engineering, it is possible to reduce the junction field, reducing both TAT degradation and BTBT degradation, increasing the retention time, and allowing 1T-DRAMs with a gate length of 10 nm to be used. Due to the small number of dopants in the device volume, we expect device variability to have a prominent effect on 1T-DRAM performance [6]. Therefore, we assess the effect of random dopant distribution on device operation by means of statistical device simulations.

Fig. 1. Simulated devices are DG nMOSFETs with two independent gates. (b)–(d) show drain-side engineering. Source side is symmetric.

II. SIMULATION DETAILS AND DEVICE STRUCTURE

We have performed MEDICI/DAVINCI (2-D/3-D) device simulations [7] using the drift–diffusion transport model with Fermi–Dirac statistics and concentration-dependent mobility. Recombination and generation mechanisms include SRH generation/recombination, Auger generation/recombination, impact ionization, and generation/recombination including tunneling [8]. The SRH lifetime is modeled as $\tau = \tau_0 / [1 + \Gamma(E)]$, where $\tau_0$ is a concentration-dependent lifetime modeled as in [9] with parameters as in [10], $E$ is the electric field, and $\Gamma(E) > 1$ is an electric-field-enhanced factor to be taken into account for generation/recombination including tunneling (TAT) [8]. All other model parameters are left to their default value as in the device simulator [7]. The simulated device structures are double-gate (DG) nMOSFETs with two independent gates, a gate length $L = 10$ nm [Fig. 1(a)], and a fin height of 10 nm (relevant for 3-D simulation only). Three different architectural options are investigated to reduce the electric field at the junction: 1) underlap $d_j$ between the gate and junctions [Fig. 1(b)]; 2) underlap with source/drain (S/D) field plates (workfunction of 4.16 eV) [Fig. 1(c)]; and 3) junctionless (JL) transistor [Fig. 1(d)]. Inversion-mode transistors [Fig. 1(a)–(c)] have gate WF = 4.8 eV, device width

MOSFET structure have been proposed via device simulation as options to increase retention, but often performance was overestimated because authors neglected the process of generation/recombination enabled by trap-assisted tunneling (TAT), which leads to reduced Shockley–Read–Hall (SRH) lifetime with increasing electric field. In 1T-DRAMs, TAT becomes the main factor limiting the retention time because of the high junction field. This aspect is very pronounced in small device structures because the junctions are very close to the middle of the channel, where the excess charge is stored. A similar degradation of the retention time is due to field-enhanced generation, i.e., band-to-band tunneling (BTBT) of charge carriers, which is usually included in device simulations. Some aspects of the influence of the junction field on the retention time have been already discussed [4], [5]. In this letter, we show that, by a proper junction engineering, it is possible to reduce the junction field, reducing both TAT degradation and BTBT degradation, increasing the retention time, and allowing 1T-DRAMs with a gate length of 10 nm to be used. Due to the small number of dopants in the device volume, we expect device variability to have a prominent effect on 1T-DRAM performance [6]. Therefore, we assess the effect of random dopant distribution on device operation by means of statistical device simulations.

II. SIMULATION DETAILS AND DEVICE STRUCTURE

We have performed MEDICI/DAVINCI (2-D/3-D) device simulations [7] using the drift–diffusion transport model with Fermi–Dirac statistics and concentration-dependent mobility. Recombination and generation mechanisms include SRH generation/recombination, Auger generation/recombination, impact ionization, and generation/recombination including tunneling [8]. The SRH lifetime is modeled as $\tau = \tau_0 / [1 + \Gamma(E)]$, where $\tau_0$ is a concentration-dependent lifetime modeled as in [9] with parameters as in [10], $E$ is the electric field, and $\Gamma(E) > 1$ is an electric-field-enhanced factor to be taken into account for generation/recombination including tunneling (TAT) [8]. All other model parameters are left to their default value as in the device simulator [7]. The simulated device structures are double-gate (DG) nMOSFETs with two independent gates, a gate length $L = 10$ nm [Fig. 1(a)], and a fin height of 10 nm (relevant for 3-D simulation only). Three different architectural options are investigated to reduce the electric field at the junction: 1) underlap $d_j$ between the gate and junctions [Fig. 1(b)]; 2) underlap with source/drain (S/D) field plates (workfunction of 4.16 eV) [Fig. 1(c)]; and 3) junctionless (JL) transistor [Fig. 1(d)]. Inversion-mode transistors [Fig. 1(a)–(c)] have gate WF = 4.8 eV, device width
BIAS USED DURING DEVICE OPERATION FOR ALL SIMULATED ARCHITECTURES. *ACTUAL BIAS REPORTED IN THE CORRESPONDING FIGURE CAPTION

<table>
<thead>
<tr>
<th>V</th>
<th>WRITE</th>
<th>HOLD</th>
<th>READ</th>
<th>ERASE</th>
</tr>
</thead>
<tbody>
<tr>
<td>TG</td>
<td>2.5V</td>
<td>*</td>
<td>*</td>
<td>2V</td>
</tr>
<tr>
<td>BG</td>
<td>-2.5V</td>
<td>VBG</td>
<td>*</td>
<td>2V</td>
</tr>
<tr>
<td>D</td>
<td>2V</td>
<td>0.5V</td>
<td>0.5V</td>
<td>0V</td>
</tr>
<tr>
<td>S</td>
<td>0V</td>
<td>0.5V</td>
<td>0V</td>
<td>0V</td>
</tr>
</tbody>
</table>

Table I

III. JUNCTION ENGINEERING OF 1T-DRAM

We start our analysis by considering a standard DG nMOSFET with L = 10 nm and an S/D doping of 10^{20} cm^{-3} [Fig. 1(a)] as a potential candidate for 1T-DRAM. Fig. 2 shows the READ after WRITE (RAW) and READ after ERASE (RAE) drain currents as functions of the HOLD time (t_H). Although the device shows a reasonable memory effect at low HOLD times (meaning high body effect), the fast degradation of state “1” and state “0,” as the HOLD time is increased, makes the device unsuitable as 1T-DRAM. In the same figure, the RAW and RAE currents are shown with TAT and BTBT turned off. The remaining leakage mechanism is band-to-band SRH recombination (S/D leakage), and the retention time greatly increases, showing that TAT and BTBT are the main limiting factors of retention.

Because both mechanisms are activated by the electric field at the S/D junctions, we investigate in this letter the three possible solutions to reduce the electric field at the junction illustrated in Fig. 1. The first solution is to separate the junction from the gate mask edge by an underlap distance d_j [Fig. 1(b)]. Fig. 3(a) shows a cutline of the electric field along the S/D direction (y = 0) during the HOLD phase, as a function of d_j: As d_j is increased, the junction field decreases, and TAT/BTBT currents decrease as well [11]. We take into account the variability induced by S/D random dopant fluctuations performing 3-D device simulations where impurities are randomly placed according to a Poisson distribution. As can be seen in Fig. 4, the higher the d_j, the lower the fresh current (the RAE current for t_H = 0) because of the higher series resistance.

Fig. 4 shows the RAW/RAE currents simulated for 15 different random dopant distributions and their average values for HOLD times (t_H) of 0 ns (filled symbols) and 100 ms (empty symbols). As d_j is increased, the lower RAE current produces, in turn, a lower RAW current. For d_j = 16 nm, a large READ sense margin (RSM = RAW – RAE) and a high current ratio (RAW/RAE) are found with an appreciable t_H = 100 ms. Of course, these results are obtained at the cost of an increased total device area. Because the total occupation area is proportional to 10 nm + 2d_j, it is questionable whether the proposed solution allows a higher RSM with respect to a case where the gate length is 10 nm + 2d_j and the underlap is null. We found that, at the considered scaled device dimension, the second solution is not suitable and the RSM is almost null for t_H = 100 ms. We also need to check whether a lower electric field due to the underlap could be paid by a low generation rate during the WRITE phase, requiring the use of higher V_DS. However, as shown in Fig. 3(b), this is not the case, and the WRITE time remains within 1 ns also for the higher value of d_j. This is due to the asymmetric top-/bottom-gate bias during WRITE [2]: the relatively high electron current due to the positive top-gate bias allows fast removal of generated excess electrons and allows fast charging of generated excess holes.

$W = 5 \text{ nm}, \text{S/D Gaussian doping profile with a peak value of} 10^{20} \text{ cm}^{-3} \text{and a standard deviation of} 2.8 \text{ nm}, \text{undoped body, and oxide (SiO}_2\text{) thickness} t_{\text{ox}} = 1.1 \text{ nm}. \text{JL FETs [Fig. 1(c)] have} t_{\text{ox}} = 3 \text{ nm}, W = 3 \text{ nm}, \text{a body doping of} 10^{19} \text{ cm}^{-3}, \text{and gate WF} = 5.5 \text{ eV}. \text{Table I shows the bias used for all simulated architectures during WRITE, HOLD, READ, and ERASE operations. Device operation during the HOLD, READ, and ERASE phases has been discussed in detail in [3], while the WRITE operation mode has been already used in [2].}

![Fig. 2](image_url1)

![Fig. 3](image_url2)

![Fig. 4](image_url3)
The dispersion of RAW/RAE currents, which is due to random dopant fluctuations, could also benefit by the presence of “field plates” on S/D contacts, shown in Fig. 1(c). Fig. 5 presents the RAW and RAE currents as functions of the separation between the field plates and the gate ($d_{fp}$) for a fixed junction separation ($d_{j} = 16$ nm).

It is evident from Fig. 5 that, as $d_{fp}$ decreases, the dispersion of the RAE current also decreases until $d_{fp} = 8$ nm. This is due to the fact that the peak electric field is concentrated in a region where the doping is lower. However, the drawback of this solution is that, as $d_{fp}$ is reduced, the junction field increases so that the RAE current increases too, as shown in Fig. 5, reducing the RSM. The optimum value is $d_{fp} = 8$ nm, corresponding to minimum dispersion and good RSM ($≈10^{-6}$ A) and current ratio ($≈10^{2}$). However, the limited gain with respect to a solution without field plates does not justify their use. As a further radical option, the junctions can be totally removed using JL devices [Fig. 1(d)] [12]. 1T-DRAM potential operation has been demonstrated for JL transistors 1 µm long by drain current hysteresis characteristics [13]. In order to keep the electric field low, we have used $t_{ox} = 3$ nm, $W = 3$ nm, and a bulk doping of $10^{19}$ cm$^{-3}$. Fig. 6 shows the RAW and RAE currents as functions of the HOLD time for deterministic and random dopant distributions in a device with $L = 10$ nm.

Simulations with nominal continuous doping show that JL devices could be used as 1T-DRAMs, showing a reasonable RSM. However, statistical simulations with discrete dopants reveal that, at these extremely scaled device dimensions, the fresh (e.g., the RAE) current is subject to a large dispersion, which completely eliminates any static reading window.

IV. Conclusion

We have investigated some options to enable scaling of the 1T-DRAM concept down to a gate length of 10 nm, considering both the effects of TAT and of random doping distribution. The most promising solution is represented by an underlap of $≈16$ nm between the drain/source junctions and the gate, which reduces the electric field during HOLD and separates the channel from regions where the random dopant distribution can cause significant variations of the potential and of the electric field. This solution leads to an increase of the total device area but can yield retention time in excess of 100 ms. We have also shown that the possibility of using field plates does not improve retention and that the operation of 1T-DRAMs based on JL transistors is completely undermined by variability due to random dopants.

REFERENCES