On DRAM Rowhammer and the Physics of Insecurity

Andrew J. Walker, *Member, IEEE*, Sungkwon Lee, and Dafna Beery

***Abstract*—The dynamic random access memory (DRAM) disturb known as rowhammer (RH) has come to dominate the insecurity of computing systems worldwide. Several studies have concentrated on electron injection from a switching cell select transistor and capture by nearby storage node junctions as being the main mechanism for the effect. This article for the first time looks in-depth at RH from the point of view of both electron injection and capture, and capacitive crosstalk. The absence of such comprehensive studies at the silicon level in the literature can be attributed to its sensitive nature within the industry and highlights the difficulty of DRAM scaling. This review article, therefore, forms a broad foundation to extend understanding of this dangerous disturb mechanism and in so doing provides an informed view on the ability of future DRAM technologies to solve RH once and for all.**

***Index Terms*—Crosstalk, dynamic random access memory (DRAM), electron injection, rowhammer (RH).**

I. Introduction

THE story of dynamic random access memory (DRAM) is that of the semiconductor industry itself [1], [2]. Fifty years of scaling a field-effect transistor and a capacitor has resulted in DRAM as the “main memory” in all compute systems from cloud servers through laptops to mobile phones. This success makes such systems prone to any security weakness that may lurk within DRAM itself. The DRAM disturb known as rowhammer (RH) has been causing increasing concern since its public unmasking in [3]. It is characterized by corrupted data in cells close to a row that is turned on and off many times between data refresh. Although the actual problem seems to have been known since 2010, and most probably before that within DRAM Research and Development, with some patent applications in 2012, RH was propelled to the forefront of hardware security issues when it was used to gain computer kernel privileges [4]. After ten years from the first inkling of a problem, the RH disturb varies widely between products from different manufacturers [3], [5] and is only getting worse with DRAM scaling [6].

The confluence of DRAM ubiquity, increasing RH susceptibility, and the possibility to use this unique disturb to gain unauthorized access has created a major hardware security concern across all modern computing systems [7] from cloud servers [8] through personal computers [4], [9] to mobile phones [10].

The response of the industry to the appearance of RH has been one of design mitigation to defend the DRAM chips from an attack [56]. Such approaches are regarded as competitive advantages by the DRAM manufacturers and so remain opaque to the users. The problem is that every mitigation up until now has been hacked [7], [11], [12] leaving all compute systems in jeopardy.

Two connected but distinct groups of RH literature concern us here. The first contains extensive and fascinating studies at the DRAM chip level usually carried out by the large DRAM users in combination with academic institutes [3], [6], [8]. These contain experimental data on RH behavior such as 1) the minimum number of aggressor row activations to induce the first flipped bit (the “hammer count”) and the related “Maximum Activation Count” (MAC) being the maximum number of row activations before any flipped bits are encountered; 2) the locality of flipped bits with respect to the hammered row; 3) the effectiveness of bit flipping when the initial states are ones or zeros; 4) the observation that newer technology DRAMs have lower MACs; and finally 5) the differences between DRAM chips from various (unnamed) vendors. The absence of DRAM vendors in these studies is probably an indication of the sensitive nature of this disturb mechanism but there is no doubt that such an extensive characterization is also being done within these corporations. The second group is concerned with what is happening in the silicon. While possible mechanisms for RH have been put forth within the articles in the first group [3], the second group delves more deeply into the origins of data corruption.

Experimental evidence points to two mechanisms for the RH disturb, namely cell transistor subthreshold leakage and electron injection into the p-well of the DRAM array from the hammered cell transistors and their subsequent capture by the storage node (SN) junctions [13]. Regarding the subthreshold leakage, lower cell transistor threshold voltages have been shown to correlate with higher susceptibility to RH. This is consistent with crosstalk between the switching aggres-

Manuscript received December 29, 2020; revised February 15, 2021; accepted February 15, 2021. Date of publication March 2, 2021; date of current version March 24, 2021. The review of this article was arranged by Editor C. Monzio Compagnoni. *(Corresponding author: Andrew J. Walker.)*

Andrew J. Walker and Dafna Beery are with Spin Memory Inc., Fremont, CA 94538 USA (e-mail: andy.walker@spinmemory.com; dafna.beery@spinmemory.com).

Sungkwon Lee was with Cypress Semiconductor, San Jose, CA 95134 USA. He is now with Infineon Technologies, San Jose, CA 95035 USA (e-mail: sungkwon.lee@infineon.com).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TED.2021.3060362.

Digital Object Identifier 10.1109/TED.2021.3060362

sor wordline and the victim wordlines pulling up the latter sufficiently in the potential to drain away some of the victim cell's stored charge [14], [15].

Regarding the injected electrons from the hammered cell transistors, the blame for these has been placed on two different origins. The first describes a collapsing inversion layer associated with the hammered cell transistor where a population of electrons is injected into the p-well as the transistor's gate turns off [16]. The second describes electron injection from charge traps near the silicon/gate dielectric interface of the cell select transistor [13], [17]. Several studies look into techniques for hampering the migration of these injected electrons. In addition to experimental work showing a degree of mitigation [18], [19], some present simulation results with claims being made for elimination of the RH failure [20].

There are some issues with the current state of affairs. First, for a memory disturb known about for ten years and causing consternation in the hardware security community, there are surprisingly few peer-reviewed studies that investigate the effect within the silicon itself. Second, several of the silicon studies that have been published so far rely on sophisticated device simulation with the result that a physics-based intuitive grasp of what is happening is perhaps difficult to achieve. This then makes it formidable for those not intimately involved in DRAM technology to innovate around the problem. Third, recent literature showing through simulation that certain process changes can reduce RH have led to claims for the eventual complete elimination of the disturb mechanism. Fourth, the literature is sparse that attempts to bridge the data measured at the full chip and described in the first group mentioned above with what is happening in the silicon. Tantalizing questions arise such as why is the effect so local to the hammered row leading to a "blast radius"? Are there multiple interactions taking place to affect neighboring cells? Are existing claims of complete elimination justified? Why are certain cells more susceptible to the effect? Why are victim cells in a certain data state more susceptible to disturb? Why are newer forms of DRAM worse? And finally, is there a path to a root cause fix for a complete elimination of DRAM RH?

The motivation for this work is to address these issues and do so in a way that makes the RH disturb mechanism accessible to a wider audience. The article is divided up as follows: Section II shows the memory cell and array structure for a generic 6F^2 approach to set the scene for subsequent analysis; Section III summarizes the main published experimental results of RH as measured at full chip; Section IV pulls together both experimental and simulation results at the silicon level and highlights some deficiencies; Section V discusses the physics of electron injection, diffusion, and drift in the DRAM array while making analogies with other relevant areas of silicon technologies; Section VI explores capacitive crosstalk as a cause of RH; Discussions follow in Section VII with this article concluding in Section VIII.

II. DRAM Array Structure

Fig. 1 shows a small section of a generic 6F^2 DRAM cell array up to and including the source and drain junction regions of the cell transistor [21]. Although a DRAM is inherently straightforward, fifty years of development have resulted in wonderful complexity. Therefore, a 3-D drawing approach has been taken to help understand the structure for the purposes of studying the RH disturb. Key features of the structure are: the DRAM cells are located within a retrograde p-well that is isolated from the substrate by a deep n-well (not shown); the SN connecting to one plate of the capacitor is in turn connected to an n-type junction in the p-well; the cell transistor is an nMOS device that over the years has evolved through several structural approaches resulting in the present recessed saddle fin form that enables high drive and low leakage all within a small footprint [21]–[23]. Fig. 1(c) shows the cross section that is sometimes presented in the literature to study RH or to show the benefit of certain process changes that are claimed to moderate the disturb, namely the one where the aggressor cell shares the same active region as the victim cell [17]–[20]. Fig. 1(c) shows other cross sections that will come in handy in later analyses in Section V.

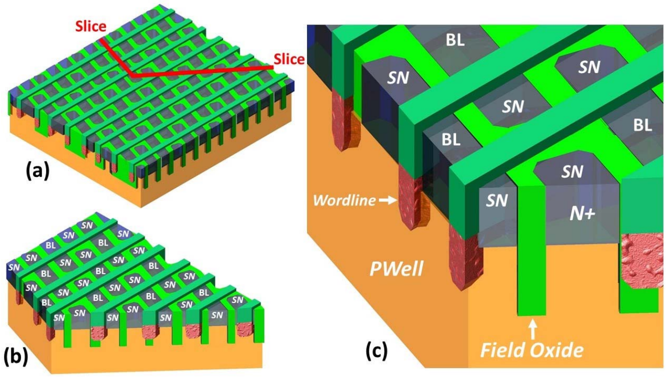


Fig. 1. (a) Schematic perspective illustration of part of a generic 6F^2 DRAM cell array up to the n-type junction level. The red lines show where the array has been schematically sliced through to show relevant cross sections. (b) Shows the result of the slice with the BL and SN diffusions marked. (c) Shows the cross sections in higher magnification. Note the left vertical face shows the shared active aggressor plus victim cells while the other vertical face shows other relevant aggressor and victim pairs.

It will be obvious that having SN capacitors connected to n-type junctions in a p-well makes them susceptible to any stray electrons in the p-well. Two key factors exacerbate this. First, the ever-decreasing cell capacitance with advancing technologies [24] amplifies the effect of every electron that may drift into the SN junction's depletion region. For a cell capacitance of 20 fF in the 3× DRAM generation, a 100 mV shift in SN voltage would need between 12 000 and 13 000 stray electrons to be picked up. For advanced DRAM processes where the SN capacitance has shrunk below 10 fF [24], this would require less than 5000 electrons. In the latter case, each electron would result in an SN potential shift of about 20 μV making it one of the most sensitive electrometers in existence except for the case of a NAND Flash cell transistor [57]. Second, the retrograde p-well presents a reflecting electric field to any stray electrons within the p-well. Such reflecting fields placed below the silicon surface are deliberately used in other areas of silicon technology to limit the lateral extent of any electron migration.

The first example of this lateral shepherding of electrons using buried reflecting electric fields is in the area of latch-up

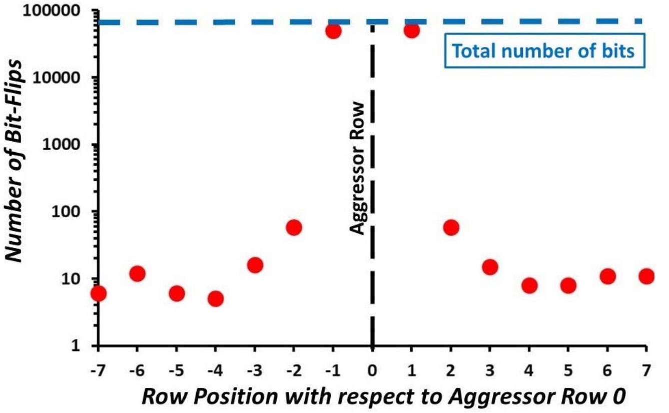


Fig. 2. Number of RH bit-flips accumulated over multiple chips of a DDR4 DRAM as a function of the row position with the assumption that there is a linear mapping between physical and logical addresses (data as in [8]). The concept of a "blast radius" becomes obvious.

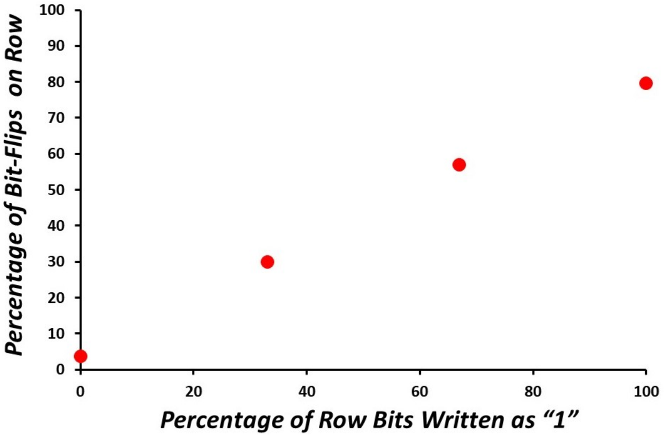


Fig. 3. Percentage of bit-flips in a row as a function of the percentage of bits in the row written as "1" (data as in [8]).

prevention where a minority carrier n-well guard ring placed around an electron injector has an efficiency of electron collection that is orders of magnitude greater when used in an epitaxial layer placed on a heavily p-type doped substrate than in the case of a uniformly doped substrate. The doping gradient from epi to substrate results in a reflecting electric field that directs the electrons toward the surface where they are then efficiently collected by the first minority carrier guard ring [25]. The second example is CMOS image sensors where reflecting fields below the silicon surface are used to reduce photoelectron crosstalk between pixels and increase image sharpness [26]. An increasing p-type doping concentration with depth results in an anisotropic diffusion which directs the photoelectrons vertically toward the surface to be collected and electronically measured within the pixel they were created in and minimizes their lateral transport to neighboring pixels.

III. Experimental Data at Chip Level

It must be said that all the valiant efforts by DRAM users and academic institutions to discern the behavior at the full chip of the RH disturb are like peering through a glass darkly while the DRAM manufacturers probably have a crystal clear view. Despite this, lots of fascinating details have been published at the chip level that we will summarize here. The absence of the DRAM manufacturers in this public discussion presents a challenge in that RH bit-flips are monitored from the perspective of the memory controller without being privy to the logical-to-physical translation that may have been implemented in the DRAM design. Nevertheless, a fairly safe albeit circular appearing assumption is usually made that a hammered row will cause more bit-flips in physically adjacent rows than in rows more physically separated. With this caveat and the assumption of a linear mapping of logical to physical addresses, Fig. 2 shows the number of RH bit-flips in each row in a DDR4 DRAM as a function of the adjacency of the row to the hammered row while the system-level RH mitigation has been turned off as much as possible [8].

The highly local nature of the bit-flipping capability of RH seen in Fig. 2 has been reported on before [3] and shows how the concept of the "blast radius" came about. Any relevant device physics-based model in the DRAM array needs to explain this effect. Fig. 3 shows the importance of the data pattern on RH. The percentage of bits that flip in a victim row is a function of the percentage of bits in that row that have a data value of 1 [8]. Similar behavior again was seen before [3], [16], [27]. The behavior that only charged cells can flip is slightly obscured by the fact that data can be stored as "true-cells" or "anticells" where, for a data 1, the former has the SN junction at a positive potential and the latter is close to ground [3], [8]. Nevertheless, it appears that the RH disturb discharges only those cells that were initially positively charged.

Other relevant full-chip experimental data include the effect of temperature [3], [27], [16] where RH susceptibility has been seen to be the worst within the range of 13 °C to 30 °C [27] although the effect was not as pronounced in [16]. Also, more advanced DRAM technologies suffer more from this disturb in having greater numbers of bit-flips per aggressor wordline activation [3], [6], [8] and a lower hammer count needed to flip the first bit [6].

IV. Experimental and Simulation Data at Silicon Level

Although several comments above have lamented the absence of the main DRAM manufacturers in the technical RH literature, possibly the most important experimental work at the silicon level comes from this group. Ryu *et al.* [13] from SK Hynix showed experimentally that RH susceptibility was correlated with both cell transistor threshold voltage (higher threshold giving lower susceptibility) and cell transistor charge pumping current (higher current giving higher susceptibility). The effect of the threshold voltage is consistent with RH being partially caused by capacitive coupling between the switching aggressor wordline and the adjacent victim wordlines causing subthreshold leakage to drain the charged capacitor. We shall look more closely into this effect in Section VI. Since cell transistor charge pumping current is a measure for MOS interface state density [28]–[30], RH susceptibility correlates with interface state density in the cell transistor. Hydrogen gas annealing was shown to reduce both the interface state density and the susceptibility to RH.

The importance of Ryu's study is that it provides experimental evidence for two causes of RH namely electron injection from interface states in the cell transistor and wordline-to-wordline crosstalk. The hydrogen anneal had an effect on both by reducing the number of electron trapping states and tightening up the cell transistor threshold voltage population. The former results in fewer injected electrons while the latter means that there are fewer victim cells in the lower portion of the threshold voltage distribution that would have otherwise resulted in large subthreshold leakage after their wordlines get pulled up in potential during crosstalk. The straight line with a negative slope of the RH fail bit count versus the threshold voltage when plotted on a log-linear plot [13] will be looked at in Section VI when we deal with crosstalk.

A different source of injected electrons from Ryu was proposed by Park *et al.* [16]. Electrons were said to emanate from the collapsing inversion layer of the aggressor cell's transistor as it turned off. These then found their way to the SN junction of the victim cell that shares the same active region with the aggressor cell. Their migration was helped by an energy barrier reduction in the silicon bulk around the victim cell's gate by a sort of drain-induced barrier lowering effect caused by the aggressor cell's voltages. The curious reader is encouraged to study [16] where the requisite figures explain the effect. Nothing was said about crosstalk-induced subthreshold leakage but this may have been due to the fact that the DRAM was from the $3\times-$nm technology. In this type of study, the understandable lack of experimental splits at the silicon level usually results in a proposal turning into the root cause based on correlations between simulations and full chip behavior.

Hampering the migration of injected electrons to reduce RH sensitivity was experimentally studied by Yang *et al.* [18], [19]. Despite citing references to crosstalk as being the cause of RH, the experiments carried out assumed that the interaction between cells was due to frequently induced electrons in the switching aggressor cell affecting the victim. The precise nature of this was not discussed. It was seen that a deeper n-type junction for the common source connected to the bitline (BL) for the aggressor-victim pair sharing the same active region reduced the number of RH bit-flips. In addition, a deeper p-well implant resulted in a large increase in bit-flips. Nothing was said about crosstalk-induced failures and victim cells that did not share the same active region as the aggressor cell. Yang and Lin [17] also concentrated on electron injection and migration as the single RH cause. A simulation-based study on electron capture and emission from traps based on the experimental work of Ryu *et al.* [13] showed agreement with published experimental results [16], [18]. Nothing was said about a possible crosstalk cause of RH that had been shown by Ryu *et al.* [13].

A fascinating approach to hindering the migration of injected electrons was presented by Gautam in a simulation study [20]. Low workfunction nanowires were introduced into the gate conductors leading to energy valleys and barriers in the silicon between an aggressor cell and victim cell sharing the same active region. Despite citing references where wordline to wordline interference was based on crosstalk [15] and experimental evidence suggesting crosstalk as an RH cause [13], the study concentrated on electron migration in the silicon between aggressor and victim cells sharing the same active region as the single cause for RH. Indeed, the claim was made that such an approach has the potential to eliminate RH as a failure mechanism.

In summary for this section, there have been surprisingly few silicon level articles on RH. To the best of our knowledge, only two show the effect of experimental silicon splits on RH sensitivity [13], [18]. Only one of these contains a crosstalk effect by showing RH sensitivity as a function of cell threshold voltage [13]. The rest of the studies, all simulation-based, while referencing crosstalk, concentrate on electron injection from the aggressor cell and migration to the victim cell as being the one and only cause for RH. This is a vulnerability given the fact that the crosstalk behavior was reported several years ago and there is no reason to assume that it has been solved. A further weakness with these aforementioned studies is their tendency to deal only with the case where aggressor and victim cells share the same active region. We shall explore these deficiencies in the next two sections.

V. Electron Injection, Diffusion, and Drift

Irrespective of their origin, it is envisaged that a fixed population of electrons is injected from the vicinity of the switching aggressor cell transistor. This population may vary with temperature as has been proposed to explain RH temperature effects [17]. In our analysis, we look at the fundamental migration of these electrons once injected into the p-well. Subsequently, we carry out device simulation to test out various boundary conditions. Initially, we look at the effects of electron injection into a p-well where there is no preexisting electric field.

When a delta function of electrons of total dose ΔN_0 per unit area is injected at time zero with the source subsequently removed, the electron packet spreads out with time while losing some of the initial population to recombination. The governing equation in one dimension is given as [31]

$$\Delta n(x,t) = \frac{\Delta N_0}{2\sqrt{(\pi D_n t)}} \exp\left(\frac{-t}{\tau_n}\right) \exp\left(\frac{-x^2}{4D_n t}\right) \qquad (1)$$

where $\Delta n(x,t)$ is the number of electrons per unit volume at a distance x from the injection point at a time t, D_n is the minority carrier electron diffusion constant, and τ_n is the electron minority carrier lifetime [32], [33]. Fig. 4 shows the electron density as a function of distance from the injection point with time as a parameter according to the above equation. We have set ΔN_0 to 1000/cm^2, D_n to 20 cm^2/s and τ_n to 0.2 μs for a doping of the p-well at 10^{17}/cm^3 and a temperature of 300 K [32]–[34]. Fig. 5 shows how the subsequent electron "pulse" moves laterally within the p-well after injection from the switching aggressor cell. We can extract an effective speed for the motion of the peak electron density which works out to about 5000 nm/ns in this case. Fig. 6 shows the peak electron density as a function of distance from the injection point. Several observations can be made from this. First, the electron lifetime plays almost no role within the "blast radius" of the RH disturb. We can see that the diffusion length is about

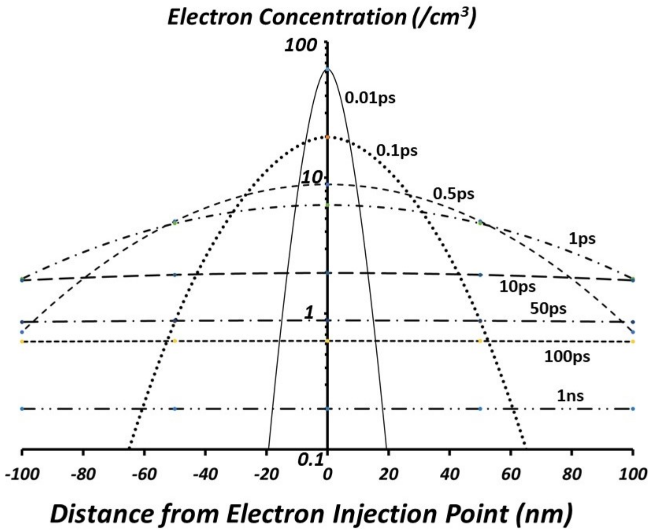


Fig. 4. Evolution of the electron density as a function of distance from the injection point with time as a parameter according to (1). Parameter values used in the equation are described in the text.

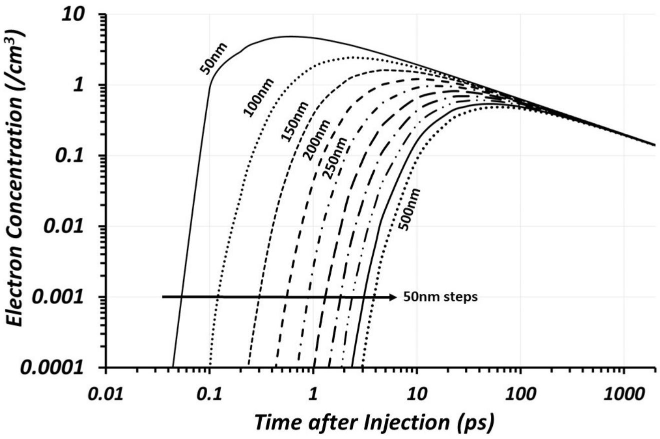


Fig. 5. Electron density as a function of time after injection with the distance from the injecting point as parameter using data generated by (1) and parameter values given in the text.

20 000 nm which is equivalent to almost 700 DRAM cells with a wordline pitch of 30 nm. Second, the peak electron density falloff with distance in Fig. 6 does not match the extreme falloff seen in the DRAM data in Fig. 2. Clearly, something else must be constraining the injected electrons from lateral migration.

The introduction of retrograde wells where the doping density increases with depth was pivotal in the development of high-density CMOS [54], [55]. Such wells have fascinating effects on charge motion within the silicon due to inherent electric fields. With this knowledge, we now return to the concept of the reflecting electric field in the p-well that was introduced in Section II. An estimate for the magnitude of such a reflecting field pointing in a direction toward the depth can be made using the same approximation as in [26]. With z as the variable increasing with depth from the silicon surface, the field can be expressed as

$$E = \frac{kT}{qN_A}\frac{dN_A}{dz} \tag{2}$$

where N_A is the p-type doping concentration as a function of the depth z and with all other symbols having their usual definitions. As an example and after some calculation, for a p-well increasing in concentration by an order of magnitude over a depth of 200 nm with the increase being linear with depth on a logarithmic scale for concentration, the electric field at 300 K comes out at about 7 kV/cm.

This then results in a force on each injected electron that directs it toward the surface. We can mathematically model the effect of this reflecting field on the injected electron population by generating an envelope function that can modify the electron density $\Delta n(x, t)$ in (1). The reflecting field will accelerate the electrons toward the surface. We shall make the following assumptions to be able to derive a modified function for the electron migration in the p-well: the retrograde p-well concentration increases by a factor m such that the boron concentration on a logarithmic scale is linear with depth z and is given by

$$\log_{10}(N_A(z)) = mz + C \tag{3}$$

where C is a constant and the other parameters have already been defined. We use the following assumptions: m is defined over a depth of z_0; the electron motion follows elementary physics of charged particles in an electric field; the electric field is given by (2); the injected electrons at time zero are at the farthest point from the surface being coincident with the peak of the retrograde p-well (this is of course not strictly correct but this simplification allows the problem to be simplified so that we can see the main effects of this field); the vertical distance traveled by the electrons due to the reflecting field is a measure of their density at that lateral distance from injection; the time t after which no more lateral electron motion takes place is given by the time it takes for the electrons to travel from rest vertically over a distance z_0. After some manipulation, we can express the 1-D governing equation for electron motion after injection in the presence of a reflecting electric field as follows:

$$\Delta n(x,t) = \frac{\Delta N_0}{2\sqrt{(\pi D_n t)}} \exp\left(\frac{-t}{\tau_n}\right) \exp\left(\frac{-x^2}{4D_n t}\right)\left(1 - \frac{kT.m.\mathrm{LN}(10)}{2m_n z_0}t^2\right) \tag{4}$$

where in addition to symbols that are universally defined and those that have been defined above, we have m_n as the electron effective mass (taken as 0.26 times the free electron rest mass) and time t measured from the moment of electron injection.

Fig. 7 shows the peak electron density as a function of distance from the injection point in the presence of the reflecting field with m from (3) and (4) as a varying parameter. The values of the other parameters used in generating the figure are included.

It is clear that the reflecting field has the necessary impact on electron migration that would explain the blast radius at full DRAM chip level as shown in Fig. 2. It is also interesting to note from (4) that taking z_0 to be the depth of the p-well, we can see that deeper p-wells will result in less lateral localization and potentially more disturbed cells as was seen in [18] and [19]. The relevance of this analysis can be seen by

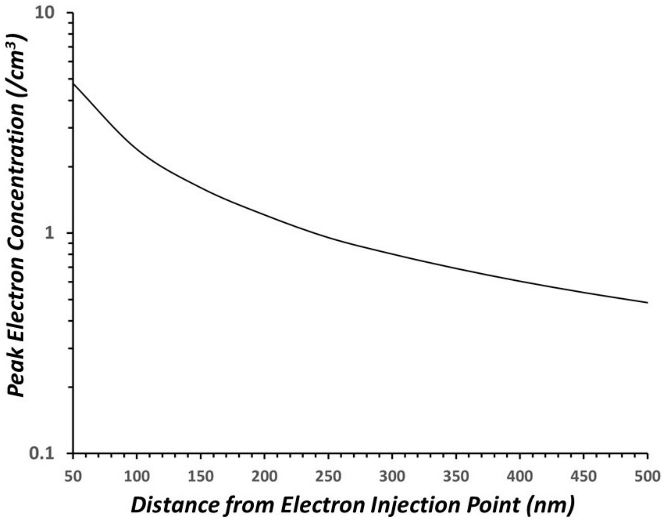


Fig. 6. Peak electron density as a function of distance from the injection point using data generated by (1) and parameter values given in the text.

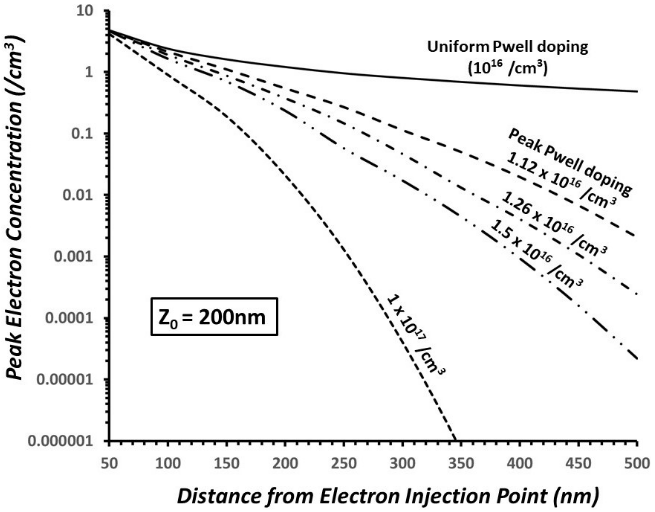


Fig. 7. Peak electron density as a function of distance from the injection point with the structure of the p-well as a parameter and using data generated by (3) and (4) and parameter values given in the text. The link between m in (3) and z_0 is given by $m = (1/z_0)\log_{10}(N_{A2}/N_{A1})$ where N_{A1} is the lower p-type concentration at $z = 0$ at which point the retrograde well concentration increases to a peak level of N_{A2} at $z = z_0$. Here, N_{A1} is taken as $10^{16}/\text{cm}^3$.

reference to Fig. 1. Each hammered cell transistor along the switching wordline is a source for electron injection. All N+ junctions within the memory array have the ability to collect these electrons if they find their way into the vicinity of the junctions' depletion regions. Clearly, only those junctions that are within a few cells' distances from the hammered wordline will have any chance of picking up these electrons. In addition, junctions that are a field oxide away from the injection point will also pick up charge. Therefore, any RH solutions that only deal with the SN junction that shares the same active region as the hammered transistor can be regarded as incomplete.

This electron migration analysis cannot explain the effect of temperature on RH where a minimum in wordline activations needed for the first flipped bit with increasing temperature has been seen [27]. Increasing the temperature from 250 to 350 K causes the minority electron diffusion length to increase by about 20% in a p-type region with a doping concentration closely equivalent to our p-well [35]. Since the product of D_n and τ_n is the square of the diffusion length, the impact to $\Delta n(x,t)$ is minimal apart from a further localization of the electrons due to an increase in the reflecting field as given by (2).

TABLE I
Structural and Electrical Parameters for the DRAM Simulations Involving Cell Program and RH Electron Injection Mimicking Using the Forward Biasing of the BL N+ to p-Well Junction

DRAM Cell Achitecture			
GOX(at bottom)	55A/60A		
WL width	30nm		
Metal Saddle-Fin width	21nm		
Gate Trench(WL) depth	160nm		
Storage Node(SN) depth	130nm		
STI Isolation Depth	210nm		
Bias Condition (DRAM Cell Program)		Bias Condition (RH Simulation)	
Vth (with Vbb=-0.8V)	0.91V	Vth (with Vbb=-0.8V)	0.91V
Metal gate WF	4.54eV	Metal gate WF	4.54eV
SNA, SNB Capacitance	10fF	SNA, SNB Capacitance	10fF
Victim Note Capacitance	NA	Victim Note Capacitance	10fF
WL Waveform (Program cycle)		BL aggressor waveform (injection)	
Von/Voff	2.0V/0V	Von/Voff	-1.4V/0V
Rise time/Fall time	10ns/10ns	Rise time/Fall time	10ns
On/Off duration	100ns/100ns	On/Off duration	20ns/100ns
V(SNA) & V(SNB)	0.5V	V(SNA) & V(SNB)	0V
Vpwell	-0.8V	Vpwell	-0.8V

To look more deeply into the effect of SN distance from the hammered cell and the localization effects of the p-well, we carried out 2-D simulations using the Sentaurus TCAD suite [36]. Since our suite version did not have the ability to incorporate trap-assisted capture and emission of electrons as was simulated by Yang and Lin [17] and our investigation was not intended to address the subtleties of electron injection from a collapsing inversion layer as in Park *et al.* [16], we mimicked the injection of electrons by forward biasing the BL N+ to p-well junction in a pulsed fashion. While the injection location may differ from reality by about a minimum feature size, we feel the general electron transport properties can still be ascertained. A 2-D DRAM structure was assembled in simulation with Table I giving the structural and electrical parameters used for both 1) writing to the cell for TCAD calibration and 2) injecting electrons from the BL junction.

The structural parameters are consistent with a 3× nm memory generation. The p-well and source/drain junction profiles were optimized to achieve typical DRAM performance. All SN's were connected to 10 fF capacitors and biased at 0.5 V. The p-well body was biased at −0.8 V and isolated with deep n-well (not shown). The following models are used in the simulation: impact ionization mode (impact.i); band-to-band tunneling (btbt); dopingdependent Shockley–Read–Hall recombination; field-dependent mobility (fldmob); Philips unified mobility (Phumob); Slotboom bandgap narrowing model (bgn); surface mobility model (srfmob).

Fig. 8 shows the 2-D cross section used in the simulations and similar to TEM cross sections of DRAM cell arrays [37].

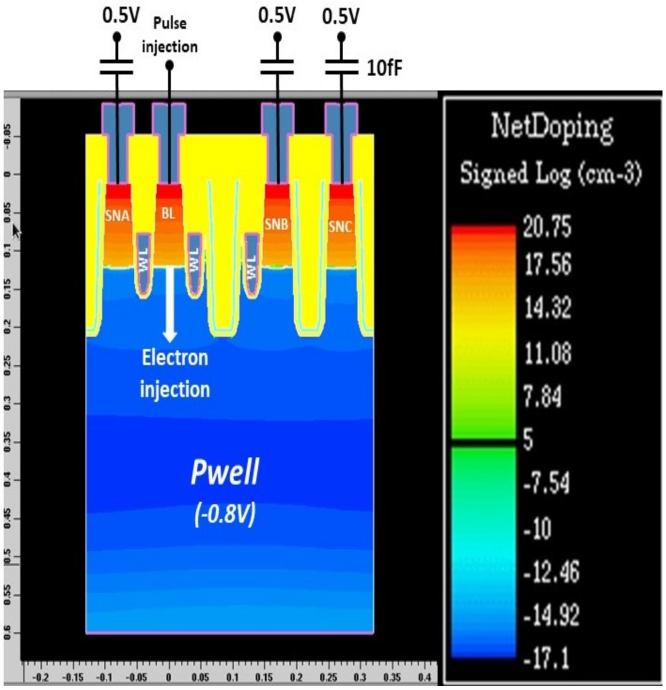
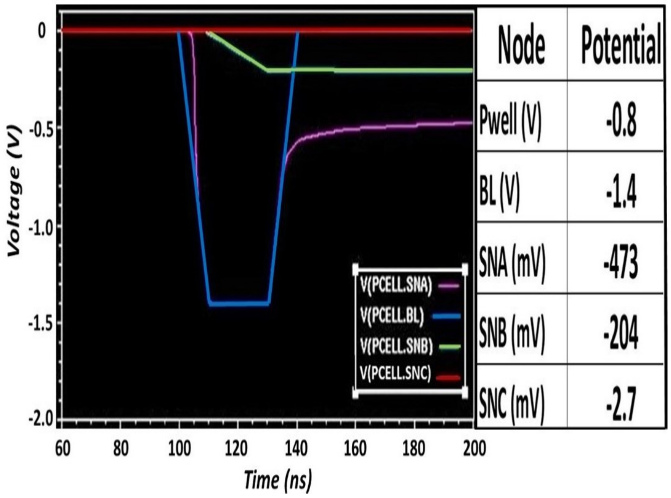


Fig. 8. 2-D simulation structure to mimic RH electron injection. The BL junction is forward biased and the potentials on the three victim nodes are monitored. Three types of collector are simulated: SNA shares the same active as the aggressor cell; SNB is a nearest victim not sharing the same active; SNC is a victim node separated from the aggressor by at least one other collecting node (SNB).



Node	Potential
Pwell (V)	-0.8
BL (V)	-1.4
SNA (mV)	-473
SNB (mV)	-204
SNC (mV)	-2.7

Fig. 9. Voltage pulse to forward bias the BL junction to mimic electron injection during RH. The potential on the other nodes are tracked with the final potential changes on SNA, SNB, and SNC shown in the table.

Each hammered transistor is an injection source while all other N+ diffusions are potential collectors. Our cross section shows three collectors: the storage node A (SNA) that shares the same active region with the hammered transistor; the storage node B (SNB) that is closest to the hammered transistor but does not share the same active region; the storage node C (SNC) that represents the next closest SNs that are separated from the hammered transistor by at least one other collecting node.

Fig. 9 shows the simulated pulse to forward bias the BL junction to p-well and tracks the voltages on the other SNs SNA, SNB, and SNC. Also shown are the changes in the node potentials as a result of the injection. The storage node A junction (SNA) closest to the injector and sharing the same active collects almost 70% of the injected electrons while the one (SNB) a field oxide region separated from the injector picks up about 30%. The storage node C (SNC) further away collects less than 1% of the injected electrons. This reinforces the concept of the blast radius as shown in Fig. 7 with the p-well doping rising by an order of magnitude into the depth of the silicon.

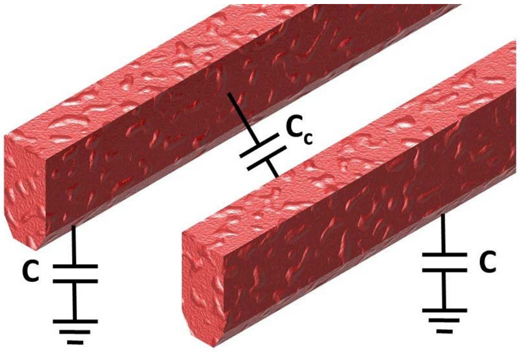


Fig. 10. Schematic of two parallel wordlines with the associated capacitances used in the crosstalk calculations [14].

As regards the effect of temperature, it must be said that our simulated injection method does not lend itself to be a credible approach since forward-biased junction behavior itself is strongly temperature-dependent. Therefore, for electron injection and collection as functions of temperature, we defer to the results in [17]. For capacitive crosstalk, however, a temperature model can be derived.

VI. Capacitive Crosstalk

It is quite surprising that the clear experimental evidence for a capacitive crosstalk effect in RH published by Ryu *et al.* [13] in 2017 has since not been dealt with in subsequent RH analyses. We shall do so here.

The space between wordlines is the minimum feature size in a 6F^2 DRAM array. The buried conductor nature of the wordline affords some natural electrical shielding between adjacent wordlines by way of inversion layers when the transistors are on and high capacitance heavily doped regions tied to either BLs or SN capacitors. However, there are regions of dielectric where electric field lines from one wordline can impinge on an adjacent wordline. Such coupling is exacerbated by the fact that the voltage (V_{PP}) at around 3 V applied to the wordline to turn on the cell transistors is the highest voltage in the chip being charge pumped internally higher than the external power supply voltage [21]. Quite sophisticated simulation can be let loose on this coupling issue [15] but we shall turn to Sakurai [14] to build an insightful model that links the crosstalk peak voltage to the subthreshold leakage of the cell transistors and include the effect of temperature.

Fig. 10 shows our setup with parallel wordlines and associated capacitances. When a pulse of peak voltage V_{PP} is applied to a particular wordline, a crosstalk noise voltage is induced in the adjacent wordlines on both sides of the driven wordline whose peak value V_{WL1} is given by

$$V_{\mathrm{WL1}} = \frac{1}{2}\left(\frac{\eta}{1+\eta}\right)V_{\mathrm{PP}} \tag{5}$$

where $\eta = C_c/C$ with C_c as the total coupling capacitance between two adjacent wordlines and C as the total capacitance of each wordline to ground.

This induced voltage causes the subsequent wordline to also be boosted and so on to the next much like a damped wave emanating from the driven wordline. For the Nth wordline from the driven wordline, it will be obvious that its peak voltage V_{WLN} will be as follows:

$$V_{\mathrm{WLN}} = \left[\frac{1}{2}\left(\frac{\eta}{1+\eta}\right)\right]^N V_{\mathrm{PP}}. \tag{6}$$

The simplifying assumption used in (5) and (6) as was used by Sakurai [14] is that the source resistance of the circuit driving the wordlines is negligible compared to the line resistance. This seems reasonable in the case of a DRAM wordline where the tungsten gate will probably result in tens of ohms per cell. Such induced wordline voltages will increase the subthreshold leakage of the associated cell transistors and partially reduce the positive potential of their charged SNs. Since this leakage is highly temperature-dependent, we need to build this into the model. We can choose either of the two paths here. The first involves using the temperature coefficient of the threshold voltage [38] in an expression for the subthreshold current of the DRAM cell select transistor. The second, which we use here since we have access to actual data, involves using the measured gate modulated energy barrier at the source that governs conduction from source to channel [39]. In this way, the temperature behavior of both the threshold voltage and the inverse subthreshold slope is taken care of. The subthreshold conduction can then be given as [39]

$$I_{\mathrm{ds}} = I_{\mathrm{ds0}}\exp\left(\frac{-qE_{\mathrm{bar}}}{kT}\right) \tag{7}$$

where E_{bar} is the energy barrier at the source measured in volts and I_{ds0} is the source-drain current close to threshold (strictly speaking at the point where the current is no longer temperature-activated). Using the barrier measurement for a crystalline silicon nMOS transistor from [39] as an example, we can approximate E_{bar} as a function of the gate to source voltage (V_{gs}) as follows:

$$E_{\mathrm{bar}} = 0.7\left(1 - V_{\mathrm{gs}}\right) \tag{8}$$

with the domain of V_{gs} between zero and one volt for an enhancement mode nMOS transistor. The actual relationship for the DRAM cell transistor will be similar to first order but would have to be measured for a more accurate model. Equating V_{gs} with the peak voltage induced through capacitive crosstalk on the Nth wordline due to a pulse of V_{PP} volts on the "zeroth" wordline and combining (6)–(8), we can express the subthreshold leakage current in a cell transistor with the Nth wordline as gate as follows:

$$I_{\mathrm{dsN}} = I_{\mathrm{ds0}}\exp\left\{\frac{-0.7q}{kT}\left[1 - \left(\frac{1}{2}\left(\frac{\eta}{1+\eta}\right)\right)^N V_{\mathrm{PP}}\right]\right\} \tag{9}$$

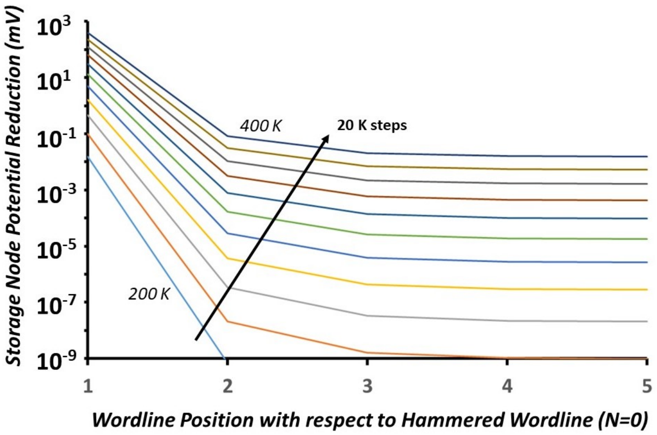


Fig. 11. Reduction in SN potential as a function of wordline position with respect to the hammered wordline with temperature as a parameter. Details of the use of (9) for this calculation are given in the text.

where all symbols have been defined already. Such boosted threshold leakage drains the charge away from a cell in the charged state. To estimate how much, we make the following assumptions: a boosted wordline remains at its peak value given by (6) for 10 ns; η is 0.5; I_{ds0} is 0.2 μA at 1 V V_{gs} [23]; the SN capacitance is 10 fF. Fig. 11 shows the shift down in SN potential as a function of N and with temperature as a parameter. In this case, the injecting wordline is hammered 50 000 times. It is clear that the cells on the first ($N = 1$) wordline take the brunt of the crosstalk effect while cells further away from the hammered wordline are shielded.

The behavior of the RH failure count as seen in [13] can now be understood. At any fixed temperature, the "1" in (8) can be replaced by the threshold voltage plus a small constant which signifies where the current is no longer temperature-activated and the source to channel barrier is effectively zero. If we assume that the RH failure count is proportional to I_{dsN} as given by (9) but with our introduction of threshold voltage, we can then take the logarithm of the resultant RH failure count to show a straight line as a function of threshold voltage with a slope of minus 0.7 q/kT. This works out to about 27 dec/V at 300 K. For the two decades seen in RH failure count in [13], a threshold voltage distribution of about 75 mV would be needed which is of the right order usually seen in transistor manufacturing. It would be ideal to compare this to the data in [13] but with arbitrary units replaced by actual ones.

Fig. 12 shows the SN shift on the cells with $N = 1$ as a function of temperature. As the temperature increases, the effect becomes worse. This is consistent with the experimental behavior for the hammering threshold reduction reported in [27] but would be unable to explain the U-shaped threshold curve with temperature [27]. To do so would involve an effect that reduces the potential rise due to crosstalk as the temperature increases. The unhammered wordlines depend on nMOS transistors in driver circuitry to hold them close to the ground. As temperature increases, their inversion layer resistance increases and so cannot explain a sluggish capacitive boosting at a higher temperature. Although the nMOS transistor junction leakage will increase, it will not be

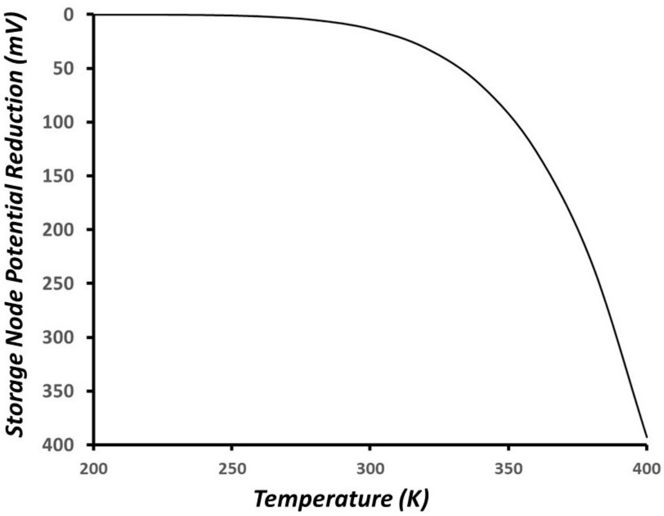


Fig. 12. Reduction in SN potential as a function of temperature for the victim cells on the adjacent wordline ($N = 1$) to the aggressor wordline. The parameters used are the same as in Fig. 11 and are given in the text.

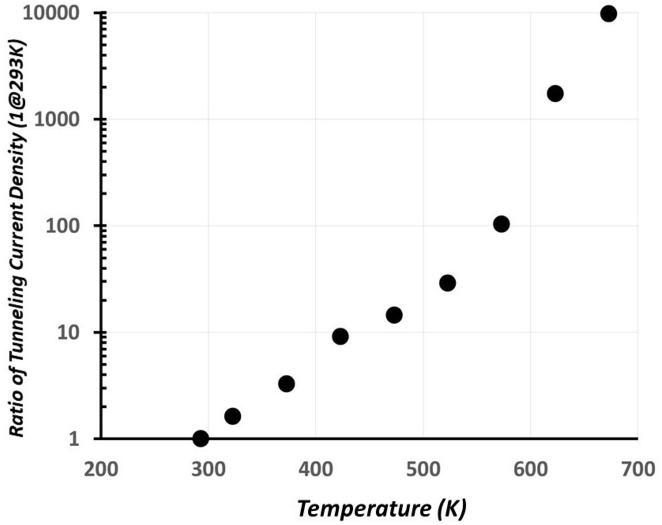


Fig. 13. Ratio of tunneling current density as a function of temperature to the value at 293 K at low field equivalent to 1.5 V over 4 nm of oxide [41].

enough to deflate the boosted wordline potential at a higher temperature [39].

One leakage mechanism that perhaps could explain an increase in hammering threshold with increasing temperature [27] is Fowler–Nordheim (FN) tunneling through thin dielectrics. FN currents increase with temperature [40], [41] and are magnified by asperities [42].

A DRAM wordline is a conductor with thin gate dielectric between it and other nodes such as the p-well and the source and drain of the select transistors. The unusual shape of the saddle fin transistor introduces nonplanar surfaces that could act as asperities for tunneling along the length of the wordline. Fig. 13 shows the tunneling current density normalized to the value at 20 °C as a function of temperature for 1.5 V over 4 nm of dielectric. The graph has been reconstructed from the data for the preexponential and exponential FN terms in [41]. Interesting current increases are seen above 500 K probably due to tunneling from excited levels in the conduction band. The temperatures relevant to DRAM operation are below 400 K where we see slightly less than a decade current increase going from 300 to 400 K. The highly nonplanar nature of the wordline – silicon interface could be the source of electric field enhancement leading to further increases in current from that seen in Fig. 13. Without experimental data from a DRAM wordline we shall leave this at the hypothetical level and just mention that a model can be built using this effect that reduces the boosted crosstalk voltage as given in (6) with larger reductions at higher temperatures. This then counteracts the rise in subthreshold leakage as given in (9). The result is a U-shape in the equivalent of Fig. 12 that matches the experimental RH threshold seen in [27].

VII. Discussion

Not only does the RH disturb bear all the hallmarks of being a brick in the wall of DRAM scalability but it has become a serious security concern for computing systems worldwide. The relative scarcity of published literature linking chip-level behavior to what is happening in silicon has been our motivation for this study. It appears that two main mechanisms can fully describe the disturb. First, electron injection and capture by positively charged SN junctions. Second, capacitive crosstalk that lifts the potential of adjacent wordlines allowing cell transistors to leak in the subthreshold region that in turn deflate the SN potential of a charged capacitor. The first is a localized effect that depends on the reflecting electric field in the p-well. Similar electron lateral localization has been deliberately used elsewhere. The second is an even more localized effect in that only the immediately adjacent wordline to the hammered one need be considered. In this way, the blast radius can be understood. This study also implies that both mechanisms should be considered if RH is to be fully mitigated or defeated. Furthermore, studies that probe only the victim cell associated with the same active region as the aggressor cell are not complete in their analysis of even the first mechanism.

This study highlights the main parameters that affect the severity of RH. In the electron diffusion and drift mechanism, the steepness of the retrograde p-well affects the electron localization. The distance from the injector affects how many electrons are captured. In the crosstalk mechanism, coupling capacitance between wordlines is fundamental with the advanced DRAM processes by definition suffering the worst. Any cell transistor threshold voltages in the lower part of the distribution will be the most susceptible. In all cases, the ever-shrinking SN capacitance amplifies the disturb whatever mechanism is looked at since each electron reaching a charged node has an effect on the node's potential that is inversely proportional to the node's capacitance. Indeed, the quite dramatic reduction in hammer count seen in newer DRAM generations [6] can mostly be attributed to the SN capacitance reducing at a rate of about 3 fF per node [24]. We can also predict the existence of susceptible cells within a technology node as those with one or more of the following characteristics: a cell that contains an SN capacitance whose value is in the lower end of the distribution; a cell with a transistor whose threshold voltage is in the lower end of

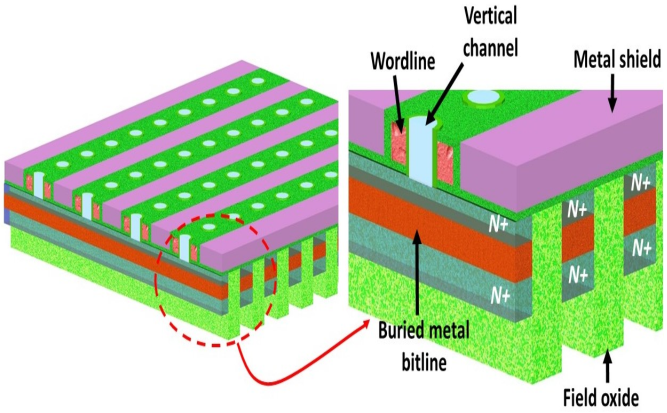


Fig. 14. Schematic perspective illustration of part of a DRAM cell array (capacitors not shown) that combats the two mechanisms of RH. The SNs are connected to the top end of the three terminal vertical channel transistor where floating body effects have been extinguished [45]. This shields the SNs from any migrating electrons emanating from other cells. Crosstalk is quenched through a metal shield placed between wordlines.

distribution; a cell that is physically far away from where the victim wordline is held at ground; a cell close to a particularly egregious electron emitter.

Eliminating RH means dealing with both highlighted mechanisms. Deep trench isolation [43] would perhaps minimize mechanism 1 when the aggressor and victim cells do not share an active region. This could be combined with an existing proposal to minimize the same-active interaction [18], [20]. Shielding the SN from any migrating electrons in the p-well would deal with the first mechanism completely. A three-terminal vertical channel cell transistor would do so as long as any floating body effects have been dealt with [21], [44], [45]. To deal with the second mechanism, we note that a conductor tied to a particular potential and placed between a hammered wordline and a victim wordline would act as a shield [14]. Indeed, we see this in our study for N equal to 2. This form of shielding would not be possible with the existing saddle fin cell transistor and its associated buried wordline.

However, both approaches to shielding could be combined as shown in Fig. 14 where an approach is shown with a buried BL conductor directly underneath the three-terminal select transistor in which floating body effects have been extinguished. The transistor channel is grown through selective epitaxy [45] from largely monocrystalline silicon on top of a buried metal silicide acting as seed [46]–[50]. The use of a three-terminal vertical select transistor means that the space between the wordlines is available for a metal shield that would be tied to a fixed potential. In these ways, both RH mechanisms would be addressed at the expense, to some extent, of wordline switching speed and process complexity.

3-D approaches to DRAM have also been proposed [51]–[53]. The benefit will be the ability to place active circuitry underneath the memory array. The use of thin-film transistors in the vertical or horizontal form will counteract the first mechanism of RH. However, the second mechanism will eventually become a challenge as larger 3-D dimensions get gobbled up in the race for a lower cost per bit.

VIII. Conclusion

The RH disturb is consistent with being caused by two fundamental mechanisms namely electron injection and capture, and capacitive crosstalk. From a transistor perspective, the first is bipolar and the latter is field effect. Elimination of this disturb means that both mechanisms have to be dealt with. The basic physics of each mechanism has been explained. The actual contribution of each to the final disturb at each DRAM node needs to be addressed with experimental data.

This review study adds a foundation to the publicly available literature on this topic in the hope that many more minds can focus on solutions to this dangerous computer insecurity.

Acknowledgment

The authors would like to thank Lucian Cojocar of Microsoft Corporation for sharing the experimental data that formed the basis of Fig. 2 and also like to say that this article was written during the time of Covid-19 and is dedicated to all those who have suffered and to those who have passed away during this pandemic. Andrew J. Walker would like to thank his son Lucas for checking the mathematics.

References

[1] R. H. Dennard, “Technical literature [reprint of ‘field-effect transistor memory’ (US Patent No. 3,387,286)],” *IEEE Solid-State Circuits Soc. Newslett.*, vol. 13, no. 1, pp. 17–25, Winter 2008, doi: 10.1109/N-SSC.2008.4785686.

[2] S. Deleonibus, “Marvels of microelectronic technology: The 1T-1C dynamic random access memory, from a groundbreaking idea to a business benchmark,” *IEEE EDS Newslett.*, vol. 26, no. 4, pp. 1–9, Oct. 2019.

[3] Y. Kim *et al.*, “Flipping bits in memory without accessing them: An experimental study of DRAM disturbance errors,” in *Proc. ACM/IEEE 41st Int. Symp. Comput. Archit. (ISCA)*, Jun. 2014, pp. 361–372, doi: 10.1109/ISCA.2014.6853210.

[4] M. Seaborn and T. Dullien, “Exploiting the DRAM rowhammer bug to gain kernel privileges,” *Black Hat*, vol. 15, p. 71, Mar. 2015. [Online]. Available: https://www.youtube.com/watch?v=0U7511Fb4to

[5] K. Park, D. Yun, and S. Baeg, “Statistical distributions of row-hammering induced failures in DDR3 components,” *Microelectron. Rel.*, vol. 67, pp. 143–149, Dec. 2016, doi: 10.1016/j.microrel.2016.10.014.

[6] J. S. Kim *et al.*, “Revisiting rowhammer: An experimental analysis of modern DRAM devices and mitigation techniques,” in *Proc. ACM/IEEE 47th Annu. Int. Symp. Comput. Archit. (ISCA)*, May 2020, pp. 638–651, doi: 10.1109/ISCA45697.2020.00059.

[7] P. Frigo *et al.*, “TRRespass: Exploiting the many sides of target row refresh,” in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2020, pp. 747–762, doi: 10.1109/SP40000.2020.00090.

[8] L. Cojocar *et al.*, “Are we susceptible to rowhammer? An end-to-end methodology for cloud providers,” in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2020, pp. 712–728, doi: 10.1109/SP40000.2020.00085.

[9] Apple Inc. (Jun. 2015). *About the Security Content of Mac EFI Security Update 2015-001*. [Online]. Available: https://support.apple.com/en-us/HT204934

[10] V. van der Veen *et al.*, “Drammer: Deterministic rowhammer attacks on mobile platforms,” in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Oct. 2016, pp. 1675–1689, doi: 10.1145/2976749.2978406.

[11] L. Cojocar, K. Razavi, C. Giuffrida, and H. Bos, “Exploiting correcting codes: On the effectiveness of ECC memory against rowhammer attacks,” in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2019, pp. 55–71, doi: 10.1109/SP.2019.00089.

[12] O. Mutlu and J. S. Kim, “Rowhammer: A retrospective,” *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 39, no. 8, pp. 1555–1571, Aug. 2020, doi: 10.1109/TCAD.2019.2915318.

[13] S.-W. Ryu *et al.*, “Overcoming the reliability limitation in the ultimately scaled DRAM using silicon migration technique by hydrogen annealing,” in *IEDM Tech. Dig.*, San Francisco, CA, USA, Dec. 2017, pp. 21.6.1–21.6.4, doi: 10.1109/IEDM.2017.8268437.

[14] T. Sakurai, “Closed-form expressions for interconnection delay, coupling, and crosstalk in VLSIs,” *IEEE Trans. Electron Devices*, vol. 40, no. 1, pp. 118–124, Jan. 1993, doi: 10.1109/16.249433.

[15] M. Redeker, B. F. Cockburn, and D. G. Elliott, “An investigation into crosstalk noise in DRAM structures,” in *Proc. IEEE Int. Workshop Memory Technol., Design Test. (MTDT)*, Bendor, France, Jul. 2002, pp. 123–129, doi: 10.1109/MTDT.2002.1029773.

[16] K. Park, C. Lim, D. Yun, and S. Baeg, “Experiments and root cause analysis for active-precharge hammering fault in DDR3 SDRAM under 3 × nm technology,” *Microelectron. Rel.*, vol. 57, pp. 39–46, Feb. 2016, doi: 10.1016/j.microrel.2015.12.027.

[17] T. Yang and X.-W. Lin, “Trap-assisted DRAM row hammer effect,” *IEEE Electron Device Lett.*, vol. 40, no. 3, pp. 391–394, Mar. 2019, doi: 10.1109/LED.2019.2891260.

[18] C.-M. Yang, C.-K. Wei, Y. J. Chang, T.-C. Wu, H.-P. Chen, and C.-S. Lai, “Suppression of row hammer effect by doping profile modification in saddle-fin array devices for sub-30-nm DRAM technology,” *IEEE Trans. Device Mater. Rel.*, vol. 16, no. 4, pp. 685–687, Dec. 2016, doi: 10.1109/TDMR.2016.2607174.

[19] C.-M. Yang *et al.*, “Scanning spreading resistance microscopy for doping profile in saddle-fin devices,” *IEEE Trans. Nanotechnol.*, vol. 16, no. 6, pp. 999–1003, Nov. 2017, doi: 10.1109/TNANO.2017.2738667.

[20] S. K. Gautam, S. K. Manhas, A. Kumar, M. Pakala, and E. Yieh, “Row hammering mitigation using metal nanowire in saddle fin DRAM,” *IEEE Trans. Electron Devices*, vol. 66, no. 10, pp. 4170–4175, Oct. 2019, doi: 10.1109/TED.2019.2931347.

[21] A. Spessot and H. Oh, “1T-1C dynamic random access memory status, challenges, and prospects,” *IEEE Trans. Electron Devices*, vol. 67, no. 4, pp. 1382–1393, Apr. 2020, doi: 10.1109/TED.2020.2963911.

[22] K.-H. Park, K.-R. Han, and J.-H. Lee, “Highly scalable saddle MOSFET for high-density and high-performance DRAM,” *IEEE Electron Device Lett.*, vol. 26, no. 9, pp. 690–692, Sep. 2005, doi: 10.1109/LED.2005.854381.

[23] S.-W. Chung *et al.*, “Highly scalable saddle-fin (S-fin) transistor for sub-50 nm DRAM technology,” in *Symp. VLSI Technol., Dig. Tech. Papers*, Honolulu, HI, USA, Jun. 2006, pp. 32–33, doi: 10.1109/VLSIT.2006.1705202.

[24] S.-K. Park, “Technology scaling challenge and future prospects of DRAM and NAND flash memory,” in *Proc. IEEE Int. Memory Workshop (IMW)*, Monterey, CA, USA, May 2015, pp. 1–4, doi: 10.1109/IMW.2015.7150307.

[25] R. R. Troutman, “Epitaxial layer enhancement of n-well guard rings for CMOS circuits,” *IEEE Electron Device Lett.*, vol. EDL-4, no. 12, pp. 438–440, Dec. 1983, doi: 10.1109/EDL.1983.25794.

[26] B. Dierickx and J. Bogaerts, “NIR-enhanced image sensor using multiple epitaxial layers,” *Proc. SPIE*, vol. 5301, pp. 205–212, Jun. 2004, doi: 10.1117/12.525726.

[27] K. Park, S. Baeg, S. Wen, and R. Wong, “Active-precharge hammering on a row induced failure in DDR3 SDRAMs under 3× nm technology,” in *Proc. IEEE Int. Integr. Rel. Workshop Final Rep. (IIRW)*, South Lake Tahoe, CA, USA, Oct. 2014, pp. 82–85, doi: 10.1109/IIRW.2014.7049516.

[28] J. S. Brugler and P. G. A. Jespers, “Charge pumping in MOS devices,” *IEEE Trans. Electron Devices*, vol. ED-16, no. 3, pp. 297–302, Mar. 1969, doi: 10.1109/T-ED.1969.16744.

[29] A. B. M. Elliot, “The use of charge pumping currents to measure surface state densities in MOS transistors,” *Solid State Electron*, vol. 19, no. 3, pp. 241–247, Mar. 1976, doi: 10.1016/0038-1101(76)90169-6.

[30] G. Groeseneken, H. E. Maes, N. Beltran, and R. F. De Keersmaecker, “A reliable approach to charge-pumping measurements in MOS transistors,” *IEEE Trans. Electron Devices*, vol. ED-31, no. 1, pp. 42–53, Jan. 1984, doi: 10.1109/T-ED.1984.21472.

[31] K. K. Ng, *Complete Guide to Semiconductor Devices*, 2nd ed. Hoboken, NJ, USA: Wiley, 2002.

[32] J. A. del Alamo and R. M. Swanson, “Modelling of minority-carrier transport in heavily doped silicon emitters,” *Solid-State Electron.*, vol. 30, no. 11, pp. 1127–1136, Nov. 1987, doi: 10.1016/0038-1101(87)90077-3.

[33] M. S. Tyagi and R. Van Overstraeten, “Minority carrier recombination in heavily-doped silicon,” *Solid-State Electron.*, vol. 26, no. 6, pp. 577–597, Jun. 1983, doi: 10.1016/0038-1101(83)90174-0.

[34] S. M. Sze, *Semiconductor Devices: Physics and Technology*. Hoboken, NJ, USA: Wiley, 1985.

[35] D. B. M. Klaassen, “A unified mobility model for device simulation,” in *Int. Tech. Dig. Electron Devices*, San Francisco, CA, USA, Dec. 1990, pp. 357–360, doi: 10.1109/IEDM.1990.237157.

[36] *Sentaurus Sdevice User’s Manual*, Synopsys, Mountain View, CA, USA, 2018.

[37] D. James, “Recent advances in memory technology,” in *Proc. 24th Annu. SEMI Adv. Semiconductor Manuf. Conf. (ASMC)*, Saratoga Springs, NY, USA, May 2013, pp. 386–395, doi: 10.1109/ASMC.2013.6552766.

[38] F. M. Klaassen and W. Hes, “On the temperature coefficient of the MOSFET threshold voltage,” *Solid-State Electron.*, vol. 29, no. 8, pp. 787–789, Aug. 1986, doi: 10.1016/0038-1101(86)90180-2.

[39] A. J. Walker, S. B. Herner, T. Kumar, and E.-H. Chen, “On the conduction mechanism in polycrystalline silicon thin-film transistors,” *IEEE Trans. Electron Devices*, vol. 51, no. 11, pp. 1856–1866, Nov. 2004, doi: 10.1109/TED.2004.837388.

[40] M. Lenzlinger and E. H. Snow, “Fowler-Nordheim tunneling into thermally grown SiO_2,” *J. Appl. Phys.*, vol. 40, no. 1, pp. 278–283, Jan. 1969, doi: 10.1063/1.1657043.

[41] R. Kies, C. Papadas, G. Pananakakis, and G. Ghibaudo, “Temperature dependence of Fowler-Nordheim emission tunneling current in MOS structures,” in *Proc. 24th Eur. Solid State Device Res. Conf. (ESSDERC)*, Edinburgh, U.K., Sep. 1994, pp. 507–510.

[42] N. M. Ravindra and J. Zhao, “Fowler-Nordheim tunneling in thin SiO_2 films,” *Smart Mater. Struct.*, vol. 1, no. 3, pp. 197–201, Sep. 1992, doi: 10.1088/0964-1726/1/3/002.

[43] Y. Kitamura *et al.*, “Suppression of crosstalk by using backside deep trench isolation for 1.12 μm backside illuminated CMOS image sensor,” in *IEDM Tech. Dig.*, San Francisco, CA, USA, Dec. 2012, pp. 24.2.1–24.2.4, doi: 10.1109/IEDM.2012.6479093.

[44] S. Hong, “Memory technology trend and future challenges,” in *IEDM Tech. Dig.*, San Francisco, CA, USA, Dec. 2010, pp. 12.4.1–12.4.4, doi: 10.1109/IEDM.2010.5703348.

[45] J.-W. Han *et al.*, “Surround gate transistor with epitaxially grown Si pillar and simulation study on soft error and rowhammer tolerance for DRAM,” *IEEE Trans. Electron Devices*, vol. 68, no. 2, pp. 529–534, Feb. 2021, doi: 10.1109/TED.2020.3045966.

[46] A. H. Van Ommen, C. W. T. Bulle-Lieuwma, J. J. M. Ottenheim, and A. M. L. Theunissen, “Ion beam synthesis of heteroepitaxial Si/$CoSi_2$/Si structures,” *J. Appl. Phys.*, vol. 67, no. 4, pp. 1767–1778, Feb. 1990, doi: 10.1063/1.345602.

[47] C. W. T. Bulle-Lieuwma, A. H. van Ommen, and L. J. van IJzendoorn, “Microstructure of heteroepitaxial Si/$CoSi_2$/Si formed by Co implantation into (100) and (111) Si,” *Appl. Phys. Lett.*, vol. 54, no. 3, pp. 244–246, Jan. 1989, doi: 10.1063/1.101446.

[48] C. W. T. Bulle-Lieuwma, A. F. de Jong, A. H. van Ommen, J. F. van der Veen, and J. Vrijmoeth, “Determination of the coordination number of Co atoms at the $CoSi_2$(A,B)/Si(111) interface by transmission electron microscopy,” *Appl. Phys. Lett.*, vol. 55, no. 7, pp. 648–650, Aug. 1989, doi: 10.1063/1.102439.

[49] A. H. Reader, A. H. Van Ommen, P. J. W. Weijs, R. A. M. Wolters, and D. J. Oostra, “Transition metal silicides in silicon technology,” *Rep. Prog. Phys.*, vol. 56, no. 11, pp. 1397–1467, Nov. 1993, doi: 10.1088/0034-4885/56/11/002.

[50] S. Mantl, “Molecular beam allotaxy: A new approach to epitaxial heterostructures,” *J. Phys. D, Appl. Phys.*, vol. 31, no. 1, pp. 1–17, Jan. 1998, doi: 10.1088/0022-3727/31/1/002.

[51] S.-Y. Lee and D. K. Schroder, “3D IC architecture for high density memories,” in *Proc. IEEE Int. Memory Workshop*, Seoul, South Korea, May 2010, pp. 1–6, doi: 10.1109/IMW.2010.5488391.

[52] T. Atsumi *et al.*, “DRAM using crystalline oxide semiconductor for access transistors and not requiring refresh for more than ten days,” in *Proc. 4th IEEE Int. Memory Workshop*, Milan, Italy, May 2012, pp. 1–4, doi: 10.1109/IMW.2012.6213660.

[53] A. Belmonte *et al.*, “Capacitor-less, long-retention (>400s) DRAM cell paving the way towards low-power and high-density monolithic 3D DRAM,” in *IEDM Tech. Dig.*, Dec. 2020, pp. 609–612.

[54] A. Stolmeijer, “A twin-well CMOS process employing high-energy ion implantation,” *IEEE Trans. Electron Devices*, vol. ED-33, no. 4, pp. 450–457, Apr. 1986, doi: 10.1109/T-ED.1986.22511.

[55] P. H. Woerlee *et al.*, “Device characterisation of a high density half-micron CMOS process,” in *Proc. 18th Eur. Solid State Device Res. Conf. (ESSDERC)*, Montpellier, France, 1988, pp. C4-33–C4-36, doi: 10.1051/jphyscol:1988405.

[56] E. Lee, S. Lee, G. E. Suh, and J. H. Ahn, “TWiCe: Time window counter based row refresh to prevent row-hammering,” *IEEE Comput. Archit. Lett.*, vol. 17, no. 1, pp. 96–99, Jan. 2018, doi: 10.1109/LCA.2017.2787674.

[57] C. M. Compagnoni and A. S. Spinelli, “Reliability of NAND flash arrays: A review of what the 2-D–to–3-D transition meant,” *IEEE Trans. Electron Devices*, vol. 66, no. 11, pp. 4504–4516, Nov. 2019, doi: 10.1109/TED.2019.2917785.