A new approach to failure analysis and yield enhancement of very large-scale integrated systems

Shuhei Amakawa
Microelectronics Research Centre
Cavendish Laboratory
Cambridge CB3 0HE, United Kingdom
sa264@cam.ac.uk

Kazuo Nakazato and Hiroshi Mizuta
Hitachi Cambridge Laboratory
Cavendish Laboratory
Cambridge CB3 0HE, United Kingdom
nakazato@phy.cam.ac.uk

Abstract

A method of failure analysis is developed based on probability theory. Unlike Monte Carlo methods, it produces accurate results even when the probabilities of interest differ from one another by many orders of magnitude. The method is applied to the analysis of the leakage-current distribution of double-gate MOSFETs and the microscopic failure mechanism is identified that limits the final yield. It explains experimental data very well. The insight into the failure mechanism gives clear guidelines for yield enhancement and facilitates device design together with the quantitative yield prediction.

1. Introduction

Development of very large-scale integrated systems (VLSIs) has been spurred continuously in the past decades. Since it is usually not possible to repair faulty component devices in a VLSI, each device in a chip can become a single point of failure unless some redundancy is introduced. Therefore, VLSIs have to be designed based on the characteristics of worst devices rather than those of average devices. Even if a chip is equipped with some redundant devices, today’s scale of integration is becoming so high that the yield requirement is still very severe. For instance, a 1 Gbit DRAM with \(10^3\) redundant cells would require the device yield of at least 99.9999%. The final chip yield is governed by the device yield.

Once the major cause of failure is somehow identified or assumed, one could use a Monte Carlo method to study yield problems. This is a very common approach. One serious practical limitation of Monte Carlo methods is that events of interest (failures) must occur very many times for the result to make any quantitative sense. However, if failure probabilities are very low, most of the computation time is consumed by uninteresting events, without producing much useful result. To cope with this, we develop using probability theory a framework for calculating the probability distribution of stochastic events whose probabilities may be extremely low. Although our approach is fairly general and may have various applications, we shall illustrate it by means of a concrete example: probability distribution of leakage current in double-gate (DG) MOSFETs (Fig. 1).

2. The problem defined

DG MOSFETs are expected to scale well below 0.1 \(\mu\m\) [1, 2]. Our DG MOSFETs have poly-Si channels and were fabricated with a 0.2 \(\mu\m\) process technology. The distribution of measured leakage currents in the off state is plotted in Fig. 2 as dots. The device yield is given by the cumulative probability at a given tolerable leakage current. Most of the DG MOSFETs had ideal characteristics, but some of them exhibited very large leakage currents as shown in the graph. The measured leakage current increases rapidly above the cumulative probability of around 80%. This part of the distribution is known as the tail distribution and has actively been studied in the context of DRAM retention time [3–7].

Characteristics of the average devices, those in the main distribution, were adequately explained by conventional two-dimensional device simulation. In an attempt to explain the characteristics of the faulty devices in the tail distribution, we modified a two-dimensional device simulator to include a generic model of field-enhanced thermally-assisted trap-to-band tunnelling proposed by Lui and Migliorato [8]. This model unifies major mechanisms of carrier generation and recombination mediated by traps: the SRH generation-recombination [9], Poole-Frenkel barrier lowering, and thermally-assisted tunnelling. It turned out that the trap model provides a very good account of the leakage-current characteristics. Thus we concluded that the field-enhanced carrier generation by traps is the main cause of the failure. We now want to reproduce the experimental leakage-current distribution.
by modelling and predict how the yield changes depending on various factors, of which some are controllable.

3. Calculation procedure

First, the device is meshed and a two-dimensional device simulation is run without incorporating the new trap model. This is to calculate the leakage current due to other mechanisms (mainly diffusion) and also to obtain the electric field and electron and hole densities, which are to be used to calculate the trap-generated current later. The implicit assumption here is that the presence of traps does not affect the electric field and carrier densities very much.

Next, the computation of the distribution of trap-generated currents follows. We assume a spatially uniform trap distribution in the channel. Since the channel is poly-Si, traps take continuous energy levels within the energy gap [8]. A trap is either donor type or acceptor type. We use the trap level distribution shown in Fig. 3. The electric field and the carrier densities are assumed to be constant within a finite element. Given the electric field, carrier densities, the trap type and the trap level, the current generated by a trap in the finite element can be calculated using Lui and Migliorato’s formula [8].

The leakage current differs from device to device because each device has different number of traps of either type at different sites at different energy levels. The probability to find a certain trap configuration can be calculated if the total trap density is specified. The leakage current for that trap configuration can also be calculated as explained above. Therefore, we can compute the probability density function \( p(I_{\text{trap}}) \) of the trap-generated current by taking into account all possible trap configurations.

Let \( U_m \) be the volume of the \( m \)th finite element. The mean trap number in it is \( \langle n \rangle = U_m N_T \), where \( N_T = \sum_j N_{Tj} \) is the total trap density. \( N_{Tj} \) is the density of traps at the \( j \)th energy level \( E_{Tj} \). \( N_T \) is given according to Fig. 3. For notational simplicity, let \( j \) index not only the energy level but also the trap type. The probability to find \( n \) traps in the \( m \)th element can be approximated to the Poisson distribution [10] \( P_m(n) = \langle n \rangle^n e^{-\langle n \rangle}/n! \). If there are \( n \) traps in the \( m \)th finite element, the probability of \( i \) of them being at \( E_{Tj} \) is given by a binomial distribution. Therefore, the probability of finding \( n \) traps in the \( m \)th element and \( i \) of them being at \( E_{Tj} \) is

\[
P_{mj}(n, i) = P_m(n) \binom{n}{i} \left( \frac{N_{Tj}}{N_T} \right)^i \left( 1 - \frac{N_{Tj}}{N_T} \right)^{n-i}.
\]

The probability of having \( i \) traps at \( E_{Tj} \) in the \( m \)th element is

\[
P_{mj}(i) = \sum_{n=i}^{R_m} P_{mj}(n, i),
\]

where \( R_m \) is a number chosen such that \( \langle n \rangle/R_m \ll 1 \). \( P_{mj}(i) \) satisfies \( \sum_{i=0}^{R_m} P_{mj}(i) = 1 \). Let \( \Delta I_{mj} \) denote the current generated by a trap at \( E_{Tj} \) in the \( m \)th finite element. Then \( i \Delta I_{mj} \) will be generated by \( i \) such traps with the probability \( P_{mj}(i) \). \( I_{\text{trap}} \) is the sum of \( i \Delta I_{mj} \) over \( j \), \( i \) and \( m \). Evidently, the functional relationship between \( I_{\text{trap}} \) and the probability density is implicit, but can be computed numerically by iteration. The calculation procedure is summarised in Fig. 4. We would like to emphasise that the traps are treated as discrete, countable entities. This discreteness will turn out very important.

4. Results and discussion

Calculated leakage-current distributions (i.e. the integrals of the probability density functions) are shown in Fig. 2 for several trap densities. One of the results agree
Trap level density
Trap number

field for all trap configurations. But the potential profile
within a device. So far we have used the same
bution should be more uniform.
the case of single-crystal MOSFETs in which trap distri-
aries of poly-Si. We were, however, unable to repro-
be effective, but this would be primarily due to the reduc-
the drain. Improvement of the channel quality would also
concentration exists that minimises the tail current. Other
is truly effective is to reduce
I. 

Probabilities and currents are summed over all possible
in each device may differ slightly from others’ for many
reasons, such as existence of defects and grain bound-
aries. Then, we note that a slight difference in a local
potential profile can result in a significant difference in its
slope (electric field), especially when the slope is steep
there. Because of poly-Si grain boundaries, there should
be much greater potential fluctuation in our devices than
in single-crystal MOSFETs. Thus some ad hoc local field
enhancement is added in some areas, with altering the cal-
culation procedure accordingly. A similar model was pro-
bposed by Hiraiwa et al. [5]. The enhanced electric field
raises \( I_t \), a great deal because of the field-induced bar-
rier lowering [8], and gives rise to the tail distribution as
shown in Fig. 7. The amount of field enhancement has a
far greater effect on the tail current than the area in which
the field is enhanced because only the former affects \( I_t \).

Reduction of the trap density has little effect on the tail
current if the mean current is less than \( I_t \) (Fig. 5). What
is truly effective is to reduce \( I_t \) itself. The important point
here is the sensitivity of \( I_t \) to the electric field. As an
example of yield enhancing design, DG MOSFETs with
additional lightly-doped drain and source regions are sim-
ulated. The highest field near the drain should become
lower if properly designed, thereby leading to a lower \( I_t \)
and a better yield. Figure 8 shows that an optimum donor
concentration exists that minimises the tail current. Other
examples may include the use of thicker gate oxide near
the drain. Improvement of the channel quality would also
be effective, but this would be primarily due to the reduc-
tion of the electric field fluctuation rather than the reduc-
tion in the trap number.

In conclusion, we successfully described the distribu-
tion of leakage currents in DG MOSFETs starting from a
microscopic model. We showed that \( I_t \), the maximum
possible current generated by a single trap, governs the de-
vice yield. Our method proved to be useful for yield pre-
diction and device design. Transistors should be designed
such that \( I_t \) is very much lower than the tolerable leakage
current at the specified cumulative probability. Our
method does not have any convergence problems as in the
conventional Monte Carlo approach.

Figure 4. The calculation procedure of the trap-
generated current distribution. First, the electric field
and the carrier densities are calculated by a device sim-
ulator. In each finite element, the current generated by
one trap is calculated for each trap level and trap type.
Probabilities and currents are summed over all possible
trap configurations.

Figure 5. The mean current, the median current, and
the tail current at the cumulative probability of 99.999 %
versus total trap density. \( I_t = \max_{j,m}(\Delta I_{mj}) \).
Figure 6. The maximum current that can be generated by a single trap in the off state, as a function of the trap position. The current peaks occur near the drain $n^+ - p^-$ junction where the electric field becomes very high. The peak current value, referred to as $I_t$, appears in Fig. 2. The simulated channel length is 0.35 $\mu$m and the channel width is 0.13 $\mu$m. $V_{DS} = 1.8$ V, $V_{GS} = -0.5$ V.

Figure 7. Cumulative probability distributions of leakage currents in DG MOSFETs with local electric field fluctuation. The electric field is enhanced by some factor in some areas. $E_{mean}$ is the field calculated by the device simulator. The dots are the experimental data.

Figure 8. Dependence of the tail current on the donor concentration in the lowly-doped drain and source regions. $I_{trap}$ is the trap-generated leakage current. $I_{dev}$ is the leakage current calculated by the device simulator. $I_{total} = I_{trap} + I_{dev}$. The lowly-doped regions are 50 nm in length each.


