10-5

Hot carrier induced degradation due to multi-phonon mechanism analyzed by lattice and device Monte Carlo coupled simulation

Shirun Ho, Yasuyuki Ohkura*, Maruizumi Takuya,***
Joshi Prasad**, Naoki Nakamura*** and Shoichi Kubo***
Advanced Research Laboratory, Hitachi Ltd., 1-280 Higashi-Koigakubo, Kokubunji, Tokyo, Japan
E-mail: ho@crl.hitachi.co.jp
*Selete Inc., **TATA, ELXSI, Ltd., ***Hitachi ULSI Systems Co., Ltd.

Abstract **– A new multi-phonon model for hydrogen desorption at Si/SiO_2 interface due to hot carriers is proposed for a multi-scale simulation, in which Lattice Monte Calro method is coupled with Device Monte Calro method by using a mediator-based common software platform. The power law between interface trap density and time ($N_{it} \propto t^n$) is demonstrated which shows good agreement with experimental results. It is shown that the multi-phonon mechanism has significant effect on reliability of MOS devices with ultra-short gate under low applied voltage.**

I. INTRODUCTION

Hot carrier damage on gate dielectrics resulting in V_{th} shift and current degradation is a barrier to achieving high performance in ultra-small MOS devices[1]. Experimental results of the performance degradation show an isotope effect[2],[3]. This indicates that mechanisms of defect generation are associated with phonon energy of hydrogen atoms at Si/SiO_2 interface[4],[5]. To study the defect generation due to hot carriers, a device simulation used to analyze non-equilibrium carrier transport under a 100nm gate electrode needs to be incorporated with atomic-scale phenomena on gate dielectrics. Time scale of the defect generation varies from seconds to million-seconds. In contrast, time required for carriers propagating through the gate electrode requires less than pico-seconds. Moreover, the defects at the interface will affect the carrier distribution that determines number of electrons scattered by H phonon, which in turn determines desorption rate of H atoms at the interface. In this paper, a new H desorption model due to multi-phonon mechanism is proposed to study the stochastic desorption rate of H atoms determined by hot electron distribution. Lattice Monte Carlo simulation (LMC) is coupled with device Monte Carlo simulation (DMC) on the Hitachi parallel machines and workstation clusters. A mediator-based common software platform, described in an earlier paper[6], is employed to efficiently combine different kinds of simulation methods to a single simulation. The coupled simulation can be utilized to analyze the intertwined phenomena with varying scales in space and time.

II. MULTI-SCALE SIMULATION FLOW

Fig. 1 shows an analytical region near Si/SiO_2 interface under a gate electrode, in which LMC is coupled with window-DMC hybridized with continuum simulation of the whole device region. For instance, a device with gate length L_g = 240nm and oxide thickness T_{ox} = 5nm is used for DMC of which analytical region is 190nm × 10μm × 250nm under effective gate length in the Si substrate. 5×10^4 particles are used in DMC. Boundary conditions of potential profile and carrier distribution for DMC are obtained by the continuum simulation. Analytical region for LMC is the surrounding region near Si/SiO_2 interface of 250nm × 300nm × 3nm in which periodic boundary condition is applied to width direction. The analytical region is discretized by uniform meshes of 10Å. H atoms and defects exist at the interface according to initial density.

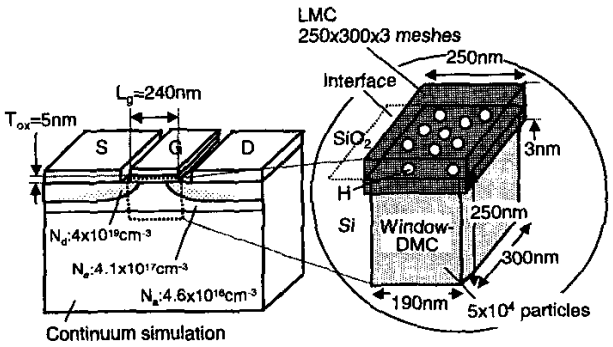

Fig. 1 Analytical region associated with L_g, in which lattice Monte Carlo (LMC) is coupled with device Monte Carlo (DMC) simulation

A flow chart of spatial and temporal multi-scale simulation is shown in Fig. 2. LMC provides H atoms and defect configuration f^L to DMC using the mediator-based platform according to correlative specification between particles and meshes. The in-sphere correlation is employed to transform the charge of defects in LMC to charge distribution associated with correlated mesh points inside of the sphere centered at the defects in DMC. The physical values are transformed and transferred automatically by mediators associated with each simulation. In DMC, real scattering of H atoms and defects as well as the conventional phonon and impurity scattering are taken into account. The time step Δt^D in DMC is set to 1fs. Number of electrons scattered by H atoms at the interface and their respective

energy distributions f^D are accumulated in DMC for 5 ps and sent to LMC. The correlative specification of the nearest point is used to transform the number of electrons and the energy distribution to a lattice point which is the nearest point from the electrons. These physical values are used to analyze the desorption rate of H atoms in LMC based on the new multi-phonon model. The temporal distribution of H atoms and defects at the interface can be obtained by solving rate equations for the defect generation using Monte Carlo method. The time step Δt^L of LMC is set to more than 1ms. After forming new configuration of H atoms and defects, the physical values are transformed and transferred to DMC. The procedure is executed iteratively to communicate new configurations of H atoms, defects and electron distribution, in which LMC and DMC are executed simultaneously on the parallel machines and workstation clusters.

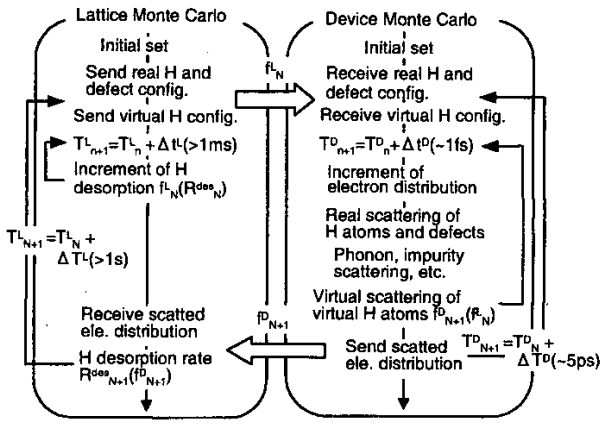

Fig. 2 Flow chart of multi-scale simulation

III. ACCELERATION ALGORITHMS

Since the coupled simulation has different spatial and time scales, new algorithms are proposed to accelerate execution time of the multi-scale simulation. The following acceleration algorithms have been implemented.

1) Virtual scattering of electrons at H atoms in DMC

Besides the real scattering due to H atoms which exist at the interface randomly, virtual H atoms with density $10^{13}cm^{-2}$ are configured uniformly. Electrons inside of a virtual scattering radius r^v_s centered at virtual H atoms with high scattering rate coefficient R^v_s are assembled and allowed to undergo free flight. These virtually scattered electrons are sent to LMC and scaled to calculate the desorption rate according to a real scattering radius r^r_s and a real scattering rate R^r_s. For instance, r^r_s=2.5Å, $R^r_s=3x10^{11}s^{-1}$ and r^v_s=10Å, $R^v_s=10^{14}s^{-1}$ are used. The execution time is accelerated by 10^4 times faster, as the number of scattered electrons is proportional to be $r_s^3*R_s$

2) Replication of an analytical region in LMC over DMC

The periodic boundary condition is applied to the analytical region in LMC over the entire Si/SiO_2 interface in width direction. This reduces total number of real H atoms to solve rate equations for the defect generation in LMC. Furthermore, using replication of the analytical region, the execution time to assemble the virtual scattered electrons in the whole region in DMC can be reduced according to total number of the virtual H atoms.

3) Varying time step for H desorption in LMC

To solve rate equations for H desorption in LMC, the time step can vary according to desorption rate of H atoms. Based on the H desorption rate of multi-phonon mechanism, time required for 5% desorption of total number of H atoms for each iteration cycle ΔT^L is evaluated and divided equally into 10^3 time steps of Δt^L. Time required for H desorption varies from seconds to more than million-seconds. Therefore, time step in LMC is set to few milliseconds in the first few cycles and increase to few thousands of seconds in later cycles according to the desorption rate. This technique is very significant in view of appropriate load balancing in multi-scale coupled simulations.

IV. Multi-Phonon Model

Under given bias condition, electrons propagating through the gate can be scattered by H atoms at the interface and the H atoms will absorb and emit the phonon leading to excited and de-excited states. When the excited phonon state reaches to the threshold energy, the H atoms undergo desorption and dangling bond gets created which becomes interface traps.

Excitation and de-excitation of the phonon states of H atoms depend upon phonon energy of H atoms and kinetic energy of scattering electrons. To mask complexities involved in each scattering event, the typical H-Si bending phonon energy is taken into account. Total number of the scattered electrons is separated into two categories. Emission scattering electrons, N_u, are number of electrons capable of emitting phonon after scattering with H atoms and average energy of emission scattering electrons is denoted by E_u. Similarly, absorption scattering electrons, N_d, are number of electrons capable of absorbing energy after scattering with H atoms and average energy of absorption scattering electrons is denoted by E_d. The electrons possessing kinetic energy greater than the typical phonon energy are capable to excite H atoms. Whereas, all scattering electrons are capable to de-excite H atoms, provided H atoms are not in the ground state. Fig.3 and Fig.4 show dependence of N_u, N_d and E_u, E_d on the position under the gate electrode, respectively. N_d has a sharp peak structure near source side. Number of electrons at the source side is high and all scattering electrons can de-excite the phonon state. N_u has a small peak at the middle region of the gate. The electron energy near source side is too low to excite

H atoms. As electrons are accelerated along the gate, both E_u and E_d have peak near the drain side. The lowest energy of E_u that can excite H atoms at the source side is almost same as typical phonon energy.

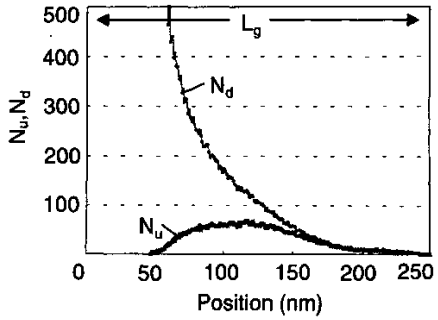

Fig.3 Dependence of N_u, N_d on position

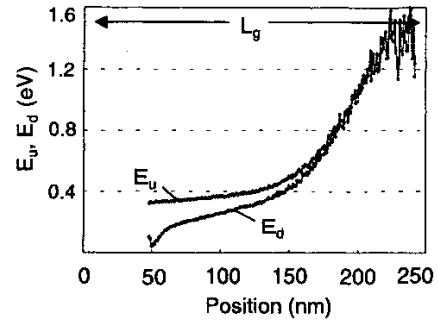

Fig.4 Dependence of E_u, E_d on position

Effective probability of excitation from lower to higher phonon level, $P_u(N_u, E_u)$, and probability of de-excitation from higher to lower level, P_d (N_d, E_d), are obtained by averaging probability of excitation and de-excitation over all phonon level of H atoms and number of scattering events m, respectively.

$$P_u(N_u, E_u) = \frac{\Sigma_n f(m,n)\, \pi_{u,n}(N_u, E_u)}{\Sigma_n f(m,n)} \quad (1)$$

$$P_d(N_d, E_d) = \frac{\Sigma_n f(m,n)\, \pi_{d,n}(N_d, E_d)}{\Sigma_n f(m,n)} \quad (2)$$

Where, f (m, n) is the number of pathways from 0 to n phonon levels at m_{th} scattering event and $\pi_{u,n}$ $\pi_{d,n}$ are normalized probabilities for excitation and de-excitation at n_{th} phonon excitation level, respectively. While calculating $\pi_{d,n}$, relaxation time for H phonon, τ_{relax}, is taken into account. Typical relaxation time for τ_{relax} is 10 ns[5]. The desorption probability of H atoms due to the multi-phonon mechanism can be derived by the following expression.

$$P_{des} = \Sigma_m f(m,n) P_u^{(m+n)/2}(E_u, N_u) P_d^{(m-n)/2}(E_d, N_d) \quad (3)$$

The threshold energy, E_{th}=3.7eV, for H bond breaking and the H-Si bending phonon energy, ω_0=0.29eV, are used. Hence, the number of phonon excited level, n, becomes 13 for H desorption and 18 for deuterium desorption resulting in the isotope effect. As shown in Fig. 5, partial desorption probability shows finite stochastic probability at scattering events at m = 13, 15, 17, -----. For instance, m=17 consists of 15 times of excitation and 2 times of de-excitation resulting in H desorption. Total desorption probability can be obtained by summing up all partial desorption probability and saturates to a constant value P_{des} after time τ_{des}. As a consequence, the desorption rate of H atoms, R_{des}, can be obtained by dividing P_{des} by τ_{des}.

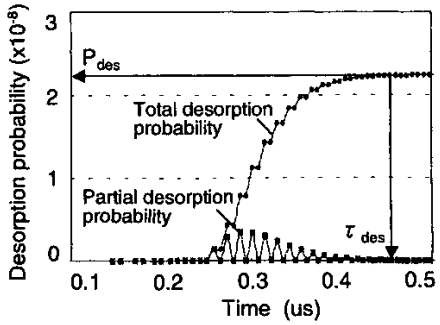

Fig. 5 Total and partial desorption probability at position of 150nm

V. RESULTS AND DISCUSSION

In Fig. 6, R_{des} as well as P_{des} and τ_{des} are shown to be strongly dependent on the position under the gate electrode.

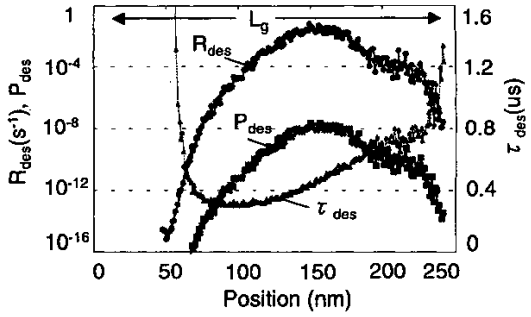

Fig. 6 Desorption rate R_{des} and probability P_{des} and time τ_{des}

Maximum R_{des} appears at the middle region. This is because N_u becomes large and E_u is higher than E_d at the middle of the gate leading to high P_{des}. Whereas, the R_{des} decreases near source side, as N_d becomes high leading to low P_{des} and τ_{des} becomes long in the source region. In the drain side, R_{des} and P_{des} have a slight flat structure and decrease

steeply, where the phonon relaxation time becomes shorter than the typical time for excitation.

Based on the multi-phonon desorption rate, the coupled simulation is applied to the device with L_g=240nm and T_{ox}=5nm. In Fig. 7, the universal power law $N_{it} \propto t^{\alpha}$ is shown to be in good agreement with experimental results of L_g=350nm and T_{ox}=5.5nm, where the stress bias of $(V_g - V_{th})/T_{ox}$ is same in order to set both of the electron concentration to be equal. Power of defect generation is obtained by straight line in fitting the region 2 s to 10^3 s, of which power α is 0.51. Please note that the value of power is a direct consequence of positional dependence of the desorption rate shown in Fig.6. Initial portion of the power law is associated with H desorption at the middle region of the gate, whereas saturation region observed after 10^3 s is associated with H desorption near the source and drain region.

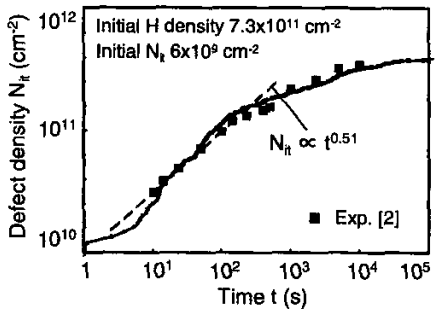

Fig. 7 Power law between defect density and time at stress condition of $(V_g - V_{th})/T_{ox}$=0.15(V/nm), V_d=3V

The coupled simulation is also applied to an ultra-small device with L_g=13nm, T_{ox}=1nm and halo p-doping profile obtained from MIT web-site[7]. The power law of 13nm gate device is compared to that of 240nm in Fig.8.

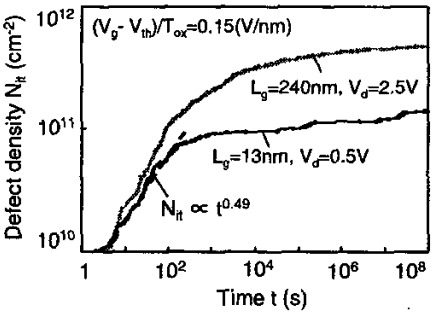

Fig. 8 Dependence of power law between defect density and time on L_g

$(V_g - V_{th})/T_{ox}$ is set to be same. The operation bias of V_d for L_g=13nm and L_g=240nm is set to be 0.5V and 2.5V predicted in ITRS'01, respectively. It is found that the interface traps are generated under low V_d of 0.5V in L_g=13nm device, of which power α is 0.49 for a short period up to 10^2 s. The power decreases after long current stress because N_u becomes too small to lead to H desorption at the source side, of which electron energy E_u needs to be higher than the typical H phonon energy. This shows that hot carrier damage due to the multi-phonon mechanism will occur slowly in the future device when low operation voltage is higher than H phonon energy.

VI. CONCLUSION

Multi-scale simulation, in which Lattice Monte Carlo is coupled with device Monte Carlo method, has been employed to analyze defect generation due to hot electrons. The new multi-phonon model and new acceleration algorithms have been proposed to realize space and time scalability. The power law between interface traps and time obtained from the coupled simulation shows good quantitative agreement with experimental results. It is found that 10nm gate device with low operation voltage still has damage on dielectrics but occurs slowly. The mediator-based coupled simulation allows cooperative designs for high-performance and high-reliability. It also allows to efficiently enhancing complex interaction mechanisms such as H diffusion and tunneling current by communicating H atoms, defects and electron distribution in dielectrics.

ACKNOWLEDGEMENT

The authors would like to thank the national projects of Japan Science and Technology Cooperation and Real Word Computing Partnership for supporting this work.

REFERENCES

[1] M. Alam, B. Weir and P. Silverman, Extended abstracts of international workshop on gate insulator, pp.30-34, 2001
[2] K Cheng, J. Lee and J. W. Lyding, Appl. Phys. Lett., ol.77, 15, pp.2358-2360, 2000
[3] K Cheng, K Hess and J. W. Lyding, IEEE Electron Device Lett., Vol.22, 5, pp.203-205, 2001
[4] K. Hess, A. Haggag, W. McMahon, B. Fischer, K. Cheng, J. Lee and J. Lyding, IEDM, pp.93-96, 2000
[5] T.-C. Shen, C. Wang, G. C. Abeln, J. W. Lyding, R. E. Walkup et al., Science, Vol.268, 16, pp.1590-1592, 1995
[6] S. Ho, Y. Ohkura, J. Prasad, Y. Kawashima, S. Kubo et al., SISPAD 2001, Springer WienNewYork, pp.372-375
[7] A. Dimitri, et al., http://www-mtl.mit.edu/Well/device13/topology13.html