

Optimization and Evaluation of Variability in the Programming Window of a Flash Cell With Molecular Metal–Oxide Storage

Vihar P. Georgiev, Stanislav Markov, *Member, IEEE*, Laia Vilà-Nadal, Christoph Busche, Leroy Cronin, and Asen Asenov, *Fellow, IEEE*

Abstract—We report a modeling study of a conceptual nonvolatile memory cell based on inorganic molecular metal–oxide clusters as a storage media embedded in the gate dielectric of a MOSFET. For the purpose of this paper, we developed a multiscale simulation framework that enables the evaluation of variability in the programming window of a flash cell with sub-20-nm gate length. Furthermore, we studied the threshold voltage variability due to random dopant fluctuations and fluctuations in the distribution of the molecular clusters in the cell. The simulation framework and the general conclusions of our work are transferrable to flash cells based on alternative molecules used for a storage media.

Index Terms—Device variability, molecular electronics, multiscale modeling, nonvolatile memory (NVM), polyoxometalates (POMs).

I. INTRODUCTION

THE field of molecular electronics continues to spur interest along the quest for miniaturization and reduction of operational power of electron devices. Recent experimental studies demonstrated the incorporation of redox-active molecules in the dielectric stack of a metal–oxide–semiconductor (MOS) structures and the reversible charging–discharging of the molecules by applying appropriate bias pulses across the device, retaining the redox state between the bias

pulses [1]–[3]. The charge stored in the molecular layer controls the threshold voltage of the structure and the conductivity of the inversion channel in the Si substrate. These studies showed the possibility of capturing the properties of redox-active molecules in the context of nonvolatile memory (NVM) devices, by replacing the charge-storing poly-Si floating gate (FG) in flash-memory cells. A major advantage of doing so is the high localization of the stored charge, which minimizes the cross-cell capacitive coupling—one of the most critical issues with flash memories that arises from charge redistribution on the sides of a poly-Si FG [4]. Although this challenge can be alternatively overcome by the use of charge-trapping dielectrics or metallic nanocluster arrays in place of a FG [5], [6], both these approaches exhibit a relatively large variability—charge-traps exhibit variation in trap-density and trap-energy [7], and the size and density of nanoclusters are difficult to control, which precludes their ultimate miniaturization [8]. On the contrary, chemical synthesis combined with molecular self-assembly of redox-active molecules can yield a regular spatial and energy distribution of charge-storage centers, and thus may allow scaling of the flash cell in the nm range.

Exploration in this direction has been mainly limited to organic molecules, e.g., fullerenes, ferrocene, porphyrins, and Alq₃ [2], [3], [9]–[12]. Although successful integration with contemporary Si-technology has been demonstrated [2], [10], charge retention and thermal budget remain a challenge.

In this paper, the novelty comes from the focus on an alternative class of molecules with unrivaled structural and chemical diversity, and a very attractive electronic and redox properties—these are inorganic metal–oxide clusters formed by early transition metal ions and oxo ligands, and are termed polyoxometalates (POMs), extensively reviewed in [13], [20], and [21]. The Wells–Dawson-type of POM structure selected for this paper is shown in Fig. 1. Such POMs are synthesized in a one-pot reaction from elementary metal–oxide (MO_x) molecules (where the metal M can be, e.g., W, V, Mo, and x is an integer between 4 and 7) that self-assemble to form an elliptical shell about 1.2 nm at length. The cavity of the shell is large enough to include two anions in an intermediate oxidation state, e.g., [S^{IV}O₃]²⁻ that make the molecular cluster highly redox-active. Such POMs possess

Manuscript received October 16, 2013; revised February 26, 2014; accepted March 21, 2014. Date of publication April 16, 2014; date of current version May 16, 2014. This work was supported in part by the Engineering and Physical Sciences Research Council Programme under Grant EP/H024107/1 and in part by Molecular-Metal-Oxide-Nanoelectronics: Achieving the molecular limit. The work of S. Markov was supported in part by the Area of Excellence on Nanoelectronics, University Grant Council of Hong Kong, Hong Kong, under Project AoE/P-04/08. The review of this paper was arranged by Editor S. Bandyopadhyay.

V. P. Georgiev is with the Device Modelling Group, School of Engineering, University of Glasgow, Glasgow G12 8LT, U.K. (e-mail: vihar.georgiev@glasgow.ac.uk).

S. Markov is with the Department of Chemistry, University of Hong Kong, Hong Kong (e-mail: figaro@hku.hk).

L. Vilà-Nadal, C. Busche, and L. Cronin are with Westchem, School of Chemistry, University of Glasgow, Glasgow G12 8QQ, U.K. (e-mail: laiavn@chem.gla.ac.uk; christoph.busche@glasgow.ac.uk; lee.cronin@glasgow.ac.uk).

A. Asenov is with the Device Modelling Group, School of Engineering, University of Glasgow, Glasgow G12 8LT, U.K., and also with Gold Standard Simulations Ltd., Glasgow G12 8LT, U.K. (e-mail: asen.asenov@glasgow.ac.uk).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TED.2014.2315520

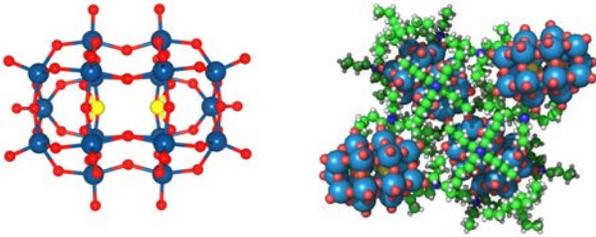


Fig. 1. Ball-and-stick representation of the atomic structure of the Wells-Dawson [20] POM $[W_{18}O_{54}(SO_3)_2]^{4-}$ (left) without and (right) with surrounding counter-cations. W: blue, O: red, and S: yellow. $(CH_3H_7)_4N^+$ (tetrapropylammonium: TPA) cation. C: green, H: white, and N: dark blue.

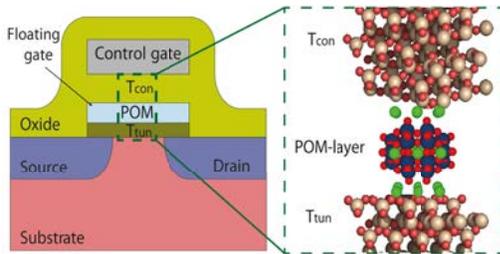


Fig. 2. Schematic diagram of a transistor NVM cell, suggesting substitution of the poly-Si FG with an array of POMs.

multiple, stable, but reversible reduction/oxidation states, and is attractive for multibit charge storage.

Integration of various POMs with SiO_2 has been experimentally demonstrated with different deposition and self-assembly techniques, achieving regular arrays of controllable POM-density of up to $5\text{--}7 \times 10^{13} \text{ cm}^{-2}$ [14]–[16], which is relevant for flash-type memory cells. Conduction through POM layers has also been studied and low mobility reported [17], which is favorable for NVM applications relying on charge localization.

However, the use of POMs as the charge-storage element of a flash cell has not been researched and is addressed theoretically for the first time in this paper. Specifically, we establish a novel simulation framework, linking atomic-level molecular simulations to 3-D numerical simulations of a flash cell device (Section II). Furthermore, we discuss the design of an 18-nm flash cell, schematically shown in Fig. 2, in terms of programming window and variability as design constraints (Section III). Noting the crucial impact that variability has on contemporary semiconductor technology, evaluation of functionality, and performance of novel devices must be done with fluctuations in mind. We report statistical simulations of ensembles of flash cells, subjected to random fluctuations of impurities in the substrate, and to irregularities in the distribution of POMs within the charge-storage layer in Section IV. Finally, conclusions are drawn in Section V.

II. SIMULATION METHODOLOGY

Our simulation flow links results from microscopic molecular simulations within the density functional theory (DFT), as implemented in the commercial simulator automatic direction finder 2008 [18], to mesoscopic transistor simulations with the commercial 3-D numerical device simulator GARAND [19].

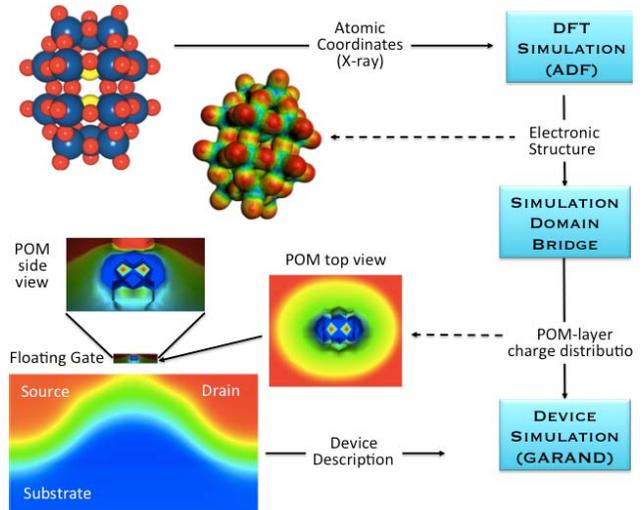


Fig. 3. Simplified block diagram of the simulation methodology.

The simplified simulation flow diagram is shown in Fig. 3. Central to this flow is the custom-built simulation domain bridge, connecting the two distinct simulation domains. The motivation for using this hierarchy of modeling approaches is the complexity of the problem. Accurate description of the POM clusters requires first principles calculations on atomic level, involving around 100 atoms, while the descriptions of the current flow through the flash cell demands mesoscopic modeling, applied to a system of millions of atoms, yet subjected to important quantum mechanical effects in the conducting channel of the device, and complex electrostatic effects arising from device nonuniformities and physical boundaries.

The DFT is the most comprehensive theoretical approach for understanding the structural, electronic, and magnetic properties of POMs [21]. In this paper, DFT calculations provide the atomic and electronic structure of a POM in a given redox state, that is, the equilibrium atomic coordinates, validated by spectroscopic measurements [20], and the partial charges on each atom of the molecule for a given redox state. Note that such calculations are feasible for an individual POM only, while a flash cell of contemporary dimensions needs around 10 POMs to realize efficient storage. The role of the simulation domain bridge is to compose the spatial charge distribution corresponding to a given spatial and redox configuration of a set of POMs and their balancing cations that together form the storage layer of the flash cell. This charge distribution is imported by GARAND as a set of fixed fractional charges in the gate-oxide of the flash cell transistor, as schematically shown in Figs. 2 and 3. In this way, we are able to calculate current–voltage characteristics of the flash memory cells for different spatial and redox configurations of POMs in the oxide, i.e., for different amounts of stored charge representing different logical states of the cell.

Spin-unrestricted formalism is used within the generalized gradient approximation for the exchange and correlation energies, employing the Becke and Perdew functionals [22]. Valence electrons are described with Slater-type basis of triple- ζ functions plus polarization quality, inner electrons being frozen. Scalar relativistic corrections are included using

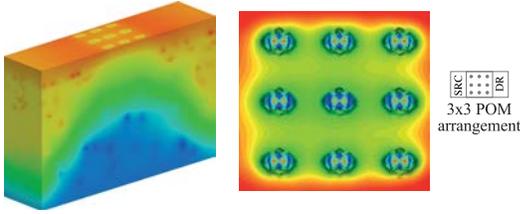


Fig. 4. 3-D electrostatic potential in the lower part of the oxide and the substrate, and 2-D map of the potential across of the plane through the center of the POMs, arranged in a 3×3 regular grid 4.5 nm from the Si-SiO₂ interface, as schematically illustrated.

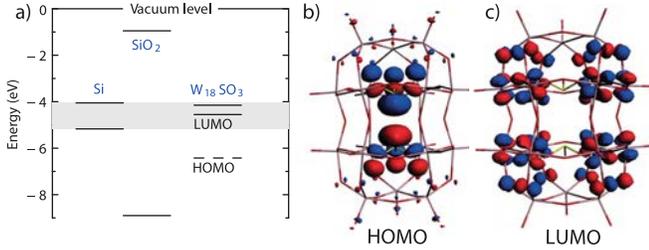


Fig. 5. (a) Energy diagram comparing the conduction and valence band edges of Si and SiO₂ to the HOMO and LUMO levels of the [W₁₈O₅₄(SO₃)₂]⁴⁻ POM. Wireframe model and (b) HOMO and (c) LUMO of the POM in its parent and 2 \times reduced state. Upon reduction the LUMO acquires extra electrons, which leads to a substantial redistribution of the charge density.

the zero-order regular approximation [23]. Simulated structures are fully optimized considering the solvent effects by means of the conductor-like screening model [24].

Fig. 5 shows an important result of the DFT calculations for the [W₁₈O₅₄(SO₃)₂]⁴⁻ POM. The energy levels of its highest occupied and lowest unoccupied molecular orbitals (HOMO and LUMO, respectively) are favorably aligned below the CB of Si and they could be effectively insulated by a relatively high potential barrier of SiO₂. Fig. 5 shows also the spatial localization of the HOMO and LUMO of the POM in its parent and 2 \times -reduced states. These representations indicate a subtle atomic rearrangement but a substantial shift of electron density between different redox states of the POM since the LUMO (due mostly to the 5d atomic orbitals of W) acquires extra electrons upon reduction.

The spatial charge distributions of a POM molecule for different redox states are used by the simulation domain bridge to construct a POM layer with a controlled spatial and redox configuration of POMs. The fractional charges associated with each atomic coordinate are obtained from the self-consistent charge-density calculation in DFT, based on the approach of multipole derived charges [25], up to quadruple moment. This approach exactly reproduces the dipole and quadruple moments of the molecule, and enables the accurate simulation of the electrostatic effects of the POMs in the flash cell. The POMs are negatively charged and in their parent oxidation state are balanced by positively charged cations—the green structure in Fig. 1. The presence of the cations in the POM layer is modeled as a set of fractional point charges distributed around each POM. The total positive charge balances out the negative charge of the parent POMs, so that any reduction of the POMs would lead to the presence of extra electron

charges in the gate-stack. This provides localized balancing of each POM, essential for modeling flash cells with broad dispersion of the position and number of POMs in the gate dielectric. The distribution of fractional charges constructed in this way by the simulation domain bridge is imported by the device simulator and modeled as a distribution of fixed oxide charges.

The 3-D numerical simulations of the flash cell, performed with GARAND, deploy the drift-diffusion transport formalism and include density-gradient quantum-corrections, essential for the accurate modeling of decananometer devices [26]. Results reported next are based on the [W₁₈O₅₄(SO₃)₂]⁴⁻ POM.

III. FLASH CELL DESIGN

A. Analytical Considerations

We consider the design and optimization of a flash cell with 18-nm gate length, focusing on the programming window ΔV_T , while minimizing short-channel effects (SCEs) and variability. It is based on a previously studied template of an 18-nm transistor [27], but without pocket implants near source and drain, and a lower doping of $\sim 5.5 \times 10^{19} \text{ cm}^{-3}$ in these contacts, similar to the contemporary flash cells studied elsewhere [28]. The problem is greatly simplified by not considering explicitly the programming, erasure, retention, and reliability characteristics of the cell. In this way, the particularities of the dielectric stack become unimportant, and optimization can proceed assuming all dielectric layers are of SiO₂, focusing on the evaluation of SCE and variability, both of which depend directly on the effective oxide thickness of the entire gate-stack.

The programming window ΔV_T is defined as the difference between the V_T of the programmed cell—all POMs in the storage layer being one-time reduced, storing 1 net charge each, and the V_T of the erased cell—no net charge being stored in the POM layer. Assuming that the redox-active molecules form a regular planar grid parallel to the substrate as in Fig. 4, one can adopt the sheet-charge approximation (SCA), to express ΔV_T as a function of the number of POMs and the number of net-charges per POM. The SCA implies that the entire gate-stack is split in two parts—the control oxide T_{CON} , and the tunneling oxide T_{TUN} , as schematically shown in Fig. 6 (inset), the total oxide thickness being $T_{\text{OX}} = T_{\text{CON}} + T_{\text{TUN}}$. If N_S is the sheet-density of POM clusters and each of them is reduced n -times, the corresponding sheet charge is (with q being the unit charge)

$$Q_S = -qnN_S \quad (1)$$

and the impact of the n -times reduced POMs on the threshold voltage of the cell is [29] (with ϵ_{OX} the oxide permittivity)

$$\Delta V_T = -Q_S T_{\text{CON}} / \epsilon_{\text{OX}}. \quad (2)$$

This relation is typically written in terms of the total oxide capacitance and the distance from the charged plane to the oxide/substrate interface. However, in the form of (2), it shows that the programming window ΔV_T is proportional to the thickness of the control oxide T_{CON} , and independent of the

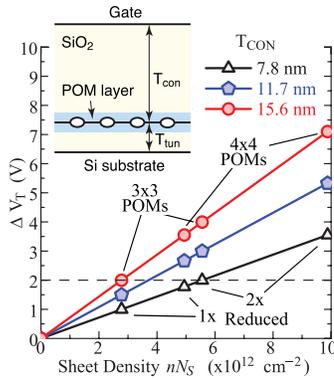


Fig. 6. Dependence of the threshold shift ΔV_T on the sheet-density of POMs and N_S for three values of T_{CON} , based on the SCA (inset) of (1) and (2). Symbols indicate N_S corresponding to the 3×3 and 4×4 arrangements for 1 \times and 2 \times reduced POMs.

thickness of the tunneling oxide; the value of V_T does, of course, depend on both.

Given the gate area of the template flash cell is $18 \times 18 \text{ nm}^2$, we consider two sheet densities N_S of POM clusters, approximately 3×10^{12} and $5 \times 10^{12} \text{ cm}^{-2}$, corresponding to 3×3 and 4×4 rectangular planar arrangements of the POMs. Fig. 6 shows the dependence of ΔV_T on Q_S obtained from (1), for three values of T_{CON} . Symbols mark the sheet density corresponding to the first and second reductions of each POM in the storage layer for the two configurations, as indicated. A 2 V programming window can be obtained even for the lower density of POMs (the 3×3 arrangement), if T_{CON} is 15.6-nm thick, which is used for T_{CON} throughout the rest of this paper.

The choice of tunnel oxide, in terms of physical thickness, conduction, and valence band offset from Si, intrinsic defect density, and so on, is of fundamental importance for the retention, and program/erase characteristics. As tunneling through the oxide is not modeled in this paper, we select the tunnel oxide based on the following considerations. In terms of retention, the limit of high-quality thermal oxide has been projected to 7–8 nm for conventional flash RAM with poly-Si FG [30]. For charge-trap memories, it is as low as 3–5 nm [31]–[35], and for nanocrystal memories down to 2 nm [36]–[38]. Sub-5-nm tunnel oxide has also been used in molecular memories based on organic redox-active molecules [3]. The thinning of the tunnel oxide is possible due to the relatively deep levels of traps in Si_3N_4 and nanocrystals, relative to the oxide conduction band [39]. The same is also true for the $[\text{W}_{18}\text{O}_{54}(\text{SO}_3)_2]^{4-}$ POM as shown in Fig. 5, which compares the position of its HOMO and LUMO with the position of the conduction and valence bands of Si and SiO_2 . Upon consecutive reductions, the LUMO adopts the first and second extra electrons. As the LUMO levels are well below (–4.05 eV), the SiO_2 conduction band, good retention is expected.

Accordingly, throughout this paper, the tunnel oxide is kept at 4.5 nm. Similar to [2], the tunnel oxide comprises 3-nm high-quality SiO_2 , and a half of the redox-active layer, i.e., 1.5 nm (see the blue-shaded region in the inset of Fig. 6) of which the POM-balancing cations form an insulating barrier of permittivity very close to that of SiO_2 [40].

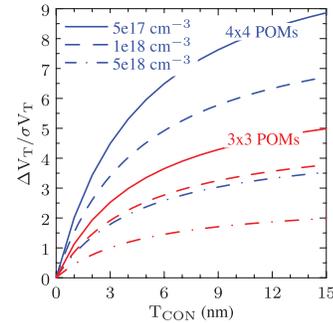


Fig. 7. Dependence of the ratio between ΔV_T and σV_T on T_{CON} , for two POM arrangements (blue and red) and three levels of substrate doping (line, dash, dot-dash). The ratio improves with enlarging T_{CON} , but the trend slows down, as T_{CON} starts to dominate T_{OX} .

Having selected T_{CON} and T_{TUN} , we can estimate the amount of random dopant fluctuation (RDF)-induced statistical variability (SV), which is proportional to T_{OX} . Fig. 7 shows the ratio between the programming window ΔV_T and the standard deviation of V_T , σV_T , versus T_{CON} . ΔV_T is calculated from (2), while σV_T is obtained from (11) in [41]

$$\sigma V_T = 3.19 \times 10^{-8} (T_{CON} + T_{TUN}) N_A^{0.4} (WL)^{-0.5} \quad (3)$$

where N_A is the acceptor concentration in the substrate, WL is the product of gate width and length, respectively. The doping concentrations are representative of typical FG-flash cells ($5 \times 10^{17} \text{ cm}^{-3}$), sub-4-nm charge trapping flash cells ($1 \times 10^{18} \text{ cm}^{-3}$), and high-performance logic bulk MOSFET ($5 \times 10^{18} \text{ cm}^{-3}$). Note in Fig. 7 that the ratio $\Delta V_T/\sigma V_T$ shows saturation as the T_{CON} becomes significantly larger than T_{TUN} , and becomes the dominant component of T_{OX} . Therefore, enlarging T_{CON} beyond 15 nm does not offer any advantage. Another important observation is that at $N_A = 5 \times 10^{18} \text{ cm}^{-3}$ and the 3×3 POM configuration, the ratio $\Delta V_T/\sigma V_T$ becomes as low as 2, which may not be tolerable even if read-verify-write algorithm is deployed at circuit level. Therefore, we chose $N_A = 1 \times 10^{18} \text{ cm}^{-3}$ in this paper, anticipating tolerable variability and acceptable SCE.

Fig. 8 shows a color-map of the simulated drain-current density at $V_{DS} = 50 \text{ mV}$, $V_{GS} = 0.3 \text{ V}$ and neutral oxide. There is approximately two times enhancement of the current magnitude near the shallow-trench-insulation (STI) compared with the center of the device. This is due to the penetration of the electric field through the STI, inducing additional carriers. The phenomenon is important for both SCEs and the programming window, as discussed below.

Fig. 9 compares the V_T roll-off and subthreshold slope characteristics of the flash cell with and without considering the effect of STI. Although the absolute value of the threshold voltage of the erased cell is reduced by approximately 1 V (not shown for clarity), the SCEs seem to be suppressed appreciably by the STI. Recall that V_T reduction is due to the enhancement of the gate control through the STI. With respect to SCE, the addition of STI means that an increasing fraction of inversion charge is controlled by the gate contact, at a given level of inversion density, and therefore a reduced

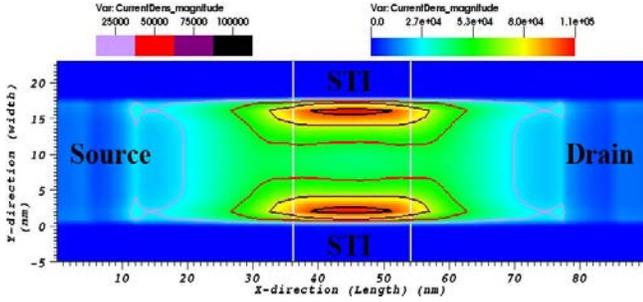


Fig. 8. 2-D color map of the magnitude of the drain current flowing 1 nm below the Si/oxide interface, showing two times larger current nearer the STI. The white lines delineate the extent of the control-gate contact. Doping density in the channel and source/drain is 10^{18} cm^{-3} and $5.5 \times 10^{19} \text{ cm}^{-3}$.

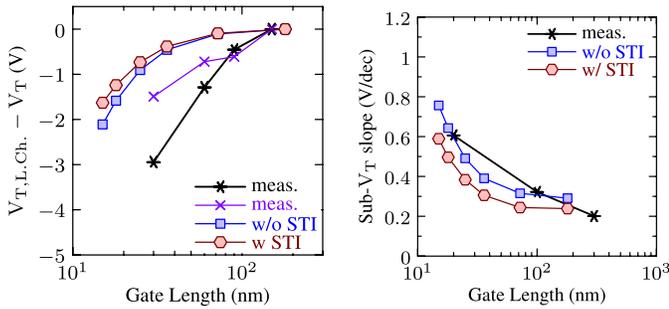


Fig. 9. Dependence of the (left) threshold voltage roll-off and (right) subthreshold slope on the gate length of the intrinsic transistor cell, and comparison against experimental data from CT-flash cells of similar dimensions [35], [44]. There is appreciable improvement of the short-channel characteristics of the cell in the simulations including STI.

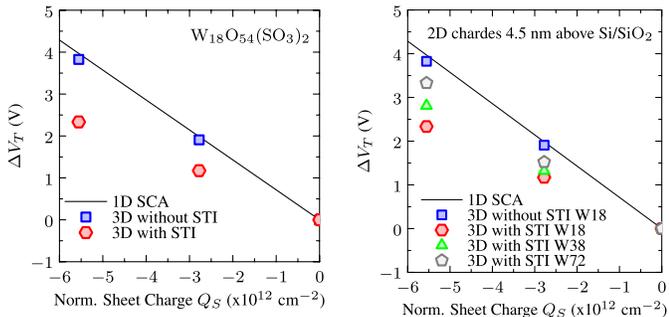


Fig. 10. Comparison between the SCA and the numerical result of the simulated flash cell, showing substantial reduction of the programming window owing to the finite MOS-capacitor area (with STI).

influence of the drain contact should be expected. In turn, both V_T roll-off and sub- V_T slope should improve.

Fig. 10 (left-hand side) compares the analytical result of ΔV_T versus Q_S , obtained from (1) and (2), with the result from 3-D simulations. The charge distribution of $[\text{W}_{18}\text{O}_{54}(\text{SO}_3)_2]^{4-}$ POM is used in the 3-D simulations. The figure clarifies that while the SCA may be acceptable for the case without STI (i.e., large, capacitor-like structures), the presence of the STI (realistic memory cell) significantly reduces the programming window. We verified that this is due to the finite area of the MOS capacitor, by simulating cells of larger widths, the result of which is shown in Fig. 10

(right-hand side). Note that a four times increase in the gate width recovers only about 50% of the difference from the ideal SCA. This result is very important as it extends beyond the particular choice of POMs as redox-active storage layer. Specifically, when an experimental programming window is reported from measurements of a large-area capacitor-like structure with a given density of redox-active molecules in the oxide, twice that density is actually needed to achieve the same programming window in a sub-20-nm flash cell. The issue is augmented due to variability in the substrate, and in the charge storage centers, as is discussed in the following.

IV. STATISTICAL VARIABILITY

Although feasibility of molecular devices is typically evaluated on either large or ultimately scaled structures featuring a single molecule, for a quantitative understanding of the potential of novel devices, in light of continuous scaling of their size, it is imperative to consider more realistic structures and the effects of variability at the nanoscale. Consequently, to obtain realistic results for the programming window of the bulk flash cell discussed in the previous section, we introduce two principal sources of SV. The first source of SV, RDF, has been extensively studied and is known to have a crucial impact on device and circuit performance in current CMOS technology [41]. The second source of SV is the random distribution of the POMs in the FG, termed POM fluctuations (POMF). Both position and number of the POMs vary. We choose a somewhat low density of POMs— $\sim 3 \times 10^{12} \text{ cm}^{-2}$, leading to nine POMs in a cell on average, which emphasizes the influence of their positional fluctuations and which can be regarded as a limiting case of POM-induced variability. It also leads to an intermolecular distance of about 3 nm, which is in agreement with experiments [17] even when as many as 21 POMs per cell are used due to the assumed Poisson distribution of their number.

Our numerical device simulations comprise three sets of 2000 flash cells each, subjected to the influence of RDF and POMF individually, and in combination (RDF+POMF). We assume the $[\text{W}_{18}\text{O}_{54}(\text{SO}_3)_2]^{4-}$ POM cluster as a charge storage center, possessing three easily accessible redox states. These are parent ($n = 4$), 1x reduced ($n = 5$), 2x reduced ($n = 6$). The corresponding amount of net charge per POM in the oxide is 0, $-q$ and $-2q$, q being the unit charge. By embedding the charge distributions obtained from DFT for each POM in the corresponding redox state, three different bits are encoded by virtue of three distinct V_T values of the flash cell. These are reported in Table I as Nominal V_T . Table I reports also the average (μ) and the standard deviation (σ) of the three V_T 's for each ensemble. More information about the distribution of V_T is shown in Fig. 11, showing the probability density function (pdf) for each ensemble and bit, and a histogram of the number of POMs in the ensembles with POMF.

The results show that the variability induced by POMF is in the same order as that induced by RDF. In fact, the standard deviation σV_T of the POMF ensemble with two-times reduced POMs is larger than that of the RDF ensemble with two-times reduced POMs. The RDF ensemble exhibits a decreasing

TABLE I
NOMINAL THRESHOLD VOLTAGE OF THE CELL ENCODING THREE BITS,
AND AVERAGE AND STANDARD DEVIATION VALUES FOR THE
ENSEMBLES

Bit	RDF			POMF		RDF+POMF	
	Nomi- nal V_T	V_T	σV_T	V_T	σV_T	V_T	σV_T
(Hex) redox state	V_T	V_T	σV_T	V_T	σV_T	V_T	σV_T
	(V)	(V)	(mV)	(V)	(mV)	(V)	(mV)
(1) parent	1.778	1.821	448	1.778	0	1.803	458
(2) 1x red.	2.948	2.976	443	2.941	376	2.958	583
(3) 2x red.	4.107	4.122	437	4.063	742	4.080	857

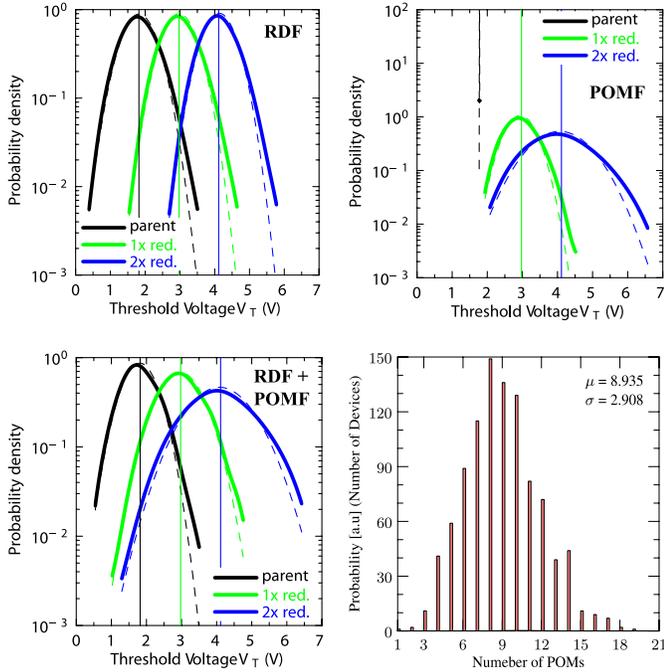


Fig. 11. PDF of the V_T distribution for each bit in the three ensembles as indicated. Histogram of the number of POMs is also shown for the ensembles with POMF.

trend (by nearly 10%) with the increase of net negative charge in the oxide, which is ascribed to the increasing control of the stored charge over channel conductance, while the fluctuations in depletion charge remain invariant per device. This effect will be stronger for higher POM-densities in the oxide. The trend in σV_T versus redox state of the POMs for the POMF ensemble is opposite— σV_T increases with the increase of net negative charge in the oxide, but reflects a known dependence of the variance on the number of oxide charges [42]. This dependence is much stronger and it is reflected in the RDF+POMF ensemble too, as observed in Fig. 11 and Table I. RDF and POMF are statistically almost independent with only a small negative covariance [less than 5% of $\sigma^2(\text{RDF})$].

It is important to note that the large σV_T in the POMF ensemble is mostly due to the Poissonian number of POMs per cell, rather than to their positional configuration. This is proved by the fact that an ensemble with an identical number of nine POMs per device, randomly distributed in the POM layer, exhibits σV_T of 42 and 91 mV for the 1 \times and 2 \times reduced

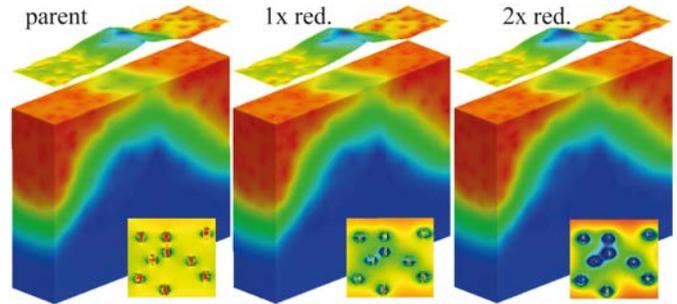


Fig. 12. Potential barrier profile along the channel (2-D elevated plot) and distribution of the electron density (3-D) in the cell with highest V_T when erased (parent state of POMs) from the ensemble with RDF and POMF; other cells have higher V_T when programmed, however. Insets: color maps of the potential under the gate and in a plane through the POM layer.

POMs, respectively (not shown for brevity). This is almost an order of magnitude lower than that of the POMF ensemble.

Fig. 12 shows electron density (3-D) and potential along the channel (2-D elevated plot) for the cell with the highest V_T when erased, i.e., in the parent state of POMs, for RDF+POMF ensemble. The effect of the POMs on the potential through the POM layer is shown as inset color maps. The specific configuration of dopants and POMs results in the highest potential barrier in the ensemble for this cell. However, when the cells in the ensemble are programmed and all POMs are 2 \times reduced, several cells have substantially larger V_T . This is due to the percolative conduction in sub- V_T regime (corresponding to a readout of the flash cell) and illustrates the importance of detailed numerical simulations on a statistical scale. Note also that the transistor featured in Fig. 12 does not exhibit extreme, but close to average programming window ΔV_T .

The analysis of V_T and its distribution help to understand the variability in the programming window, ΔV_T . The average values of ΔV_T , for the transitions from parent to 1 \times reduced (1 \times to 2 \times reduced) state, are approximately 1.16 (1.12) V for the RDF, POMF, and RDF+POMF ensembles, which compares well with the 1.17 (1.16) V nominal window separating two bits, as deduced from Table I. This reflects an apparently well-behaved dispersion of V_T , which compares well with a Gaussian, except for the upper tails, in Fig. 11.

However, when we look at the cumulative distribution function (cdf) of ΔV_T in Fig. 13 (1-cdf), we see rather large variability imparted by POMF. Specifically, the distributions of ΔV_T for the RDF ensemble are with narrow tails, feature a uniform slope to at least 3σ ($\sigma \Delta V_T$ are 20 and 38 mV), and the average transitions are well separated. For the RDF+POMF ensemble the distributions are very broad. $\sigma \Delta V_T$ values in this case are over 10 \times larger (380 and 745 mV correspondingly) and the tail crosses the average of the next bit before 3σ is reached. We have verified that once again the large variability is due mainly to the number of fluctuations, rather than to the positional configuration differences. The latter contribute only about twice of the variability associated with RDF.

We should recall that flash variability is also evaluated in conjunction with the incremental step-pulse programming algorithm [11]. This requires time-domain simulations, in

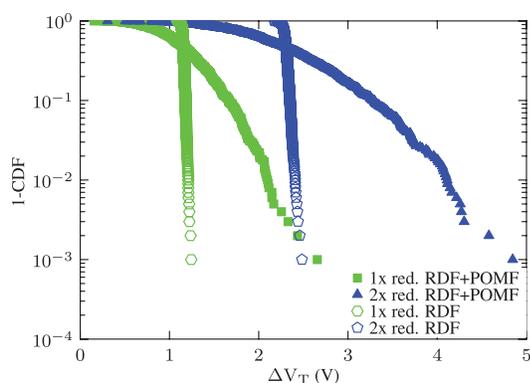


Fig. 13. Cumulative probability distribution of ΔV_T for two consecutive reductions of all POMs with respect to their parent state for two ensembles (open symbols) RDF and (closed symbols) RDF+POMF.

conjunction with gate current modeling for the simulation of program/erase, and is the current focus of our efforts.

V. CONCLUSION

We reported a novel multiscale modeling framework that enables the simulation of MOSFET-based flash cell memories with redox-active molecules as storage centers. It is unique in its ability to capture the electrostatic interaction between the molecular charge distribution obtained *ab initio* and the distribution of dopant atoms in the substrate of the transistor, as well as in its ability to account for the impact of physical boundaries in realistic devices. We showed that the approach provides both qualitative and quantitative aid toward the design and optimization of such molecular flash cells, particularly in the context of SV, which is of great relevance. The accurate description of the complex electrostatic potential arising from the molecular charge distribution in the oxide is also a fundamental prerequisite for modeling the tunneling mechanisms associated with the program-erase and retention characteristics of molecular flash cells, which is our current focus.

We reported also a careful design of a hybrid flash cell based on a bulk-MOSFET with 18-nm gate length and molecular storage media, by replacing the traditional poly-Si FG with a layer of redox-active POMs, $[W_{18}O_{54}(SO_3)_2]^{4-}$. These molecules are interesting since they can serve as monoenergetic trap centers, with deep levels of about 4.1 eV below the conduction band of SiO_2 . They show potential improvement of the device variability if self-assembly of the POM layer could minimize their number-density fluctuation. Variation in the number of POMs per device dominates the variability in the programming window, by a factor of over 10, when compared with RDF. Further work will look to testing physical devices and closing the loop on molecular design and device characteristics.

REFERENCES

- [1] X. Duan, Y. Huang, and C. M. Lieber, "Nonvolatile memory and programmable logic from molecule-gated nanowires," *Nano Lett.*, vol. 2, no. 5, pp. 487–490, 2002.
- [2] T. Shaw *et al.*, "Integration of self-assembled redox molecules in flash memory devices," *IEEE Trans. Electron Devices*, vol. 58, no. 3, pp. 826–834, Mar. 2011.

- [3] S. Paydavosi *et al.*, "High-density charge storage on molecular thin films—Candidate materials for high storage capacity memory cells," in *Proc. IEEE IEDM*, Dec. 2011, pp. 24.4.1–24.4.4.
- [4] J.-D. Lee, J.-D. Choi, and S.-H. Hur, "Effects of floating-gate interference on NAND flash memory cell operation," *IEEE Electron Device Lett.*, vol. 23, no. 5, pp. 264–266, May 2002.
- [5] M. White, Y. Yang, A. Purwar, and M. French, "A low voltage SONOS nonvolatile semiconductor memory technology," *IEEE Trans. Compon., Packag., Manuf. Technol. A*, vol. 20, no. 2, pp. 190–195, Jun. 1997.
- [6] S. Tiwari, F. Rana, K. Chan, H. Hanafi, W. Chan, and D. Buchanan, "Volatile and non-volatile memories in silicon with nano-crystal storage," in *Proc. IEEE IEDM*, Dec. 1995, pp. 521–524.
- [7] B. Kumar, P. Nair, R. Sharma, S. Kamohara, and S. Mahapatra, "Lateral profiling of trapped charge in SONOS flash EEPROMs programmed using CHE injection," *IEEE Trans. Electron Devices*, vol. 53, no. 4, pp. 698–705, Apr. 2006.
- [8] J. Shaw, T.-H. Hou, H. Raza, and E. Kan, "Statistical metrology of metal nanocrystal memories with 3-D finite-element analysis," *IEEE Trans. Electron Devices*, vol. 56, no. 8, pp. 1729–1735, Aug. 2009.
- [9] T. Pro *et al.*, "Investigation of hybrid molecular/silicon memories with redox-active molecules acting as storage media," *IEEE Trans. Nanotechnol.*, vol. 8, no. 2, pp. 204–213, Mar. 2009.
- [10] Z. Liu, A. A. Yasseri, J. S. Lindsey, and D. F. Bocian, "Molecular memories that survive silicon device processing and real-world operation," *Science*, vol. 302, no. 5650, pp. 1543–1545, Nov. 2003.
- [11] T.-H. Hou, U. Ganguly, and E. C. Kan, "Programmable molecular orbital states of C_{60} from integrated circuits," *Appl. Phys. Lett.*, vol. 89, no. 25, pp. 253113-1–253113-3, Dec. 2006.
- [12] X. Duan, Y. Huang, and C. M. Lieber, "Nonvolatile memory and programmable logic from molecule-gated nanowires," *Nano Lett.*, vol. 2, no. 5, pp. 487–490, 2002.
- [13] D.-L. Long, E. Burkholder, and L. Cronin, "Polyoxometalate clusters, nanostructures and materials: From self assembly to designer materials and devices," *Chem. Soc. Rev.*, vol. 36, no. 1, pp. 105–121, 2007.
- [14] A. M. Douvas, E. Makarona, N. Glezos, P. Argitis, J. A. Mielczarski, and E. Mielczarski, "Polyoxometalate-based layered structures for charge transport control in molecular devices," *ACS Nano*, vol. 2, no. 4, pp. 733–742, 2008.
- [15] B. Fleury, M. Billon, F. Duclairoir, L. Dubois, A. Fanton, and G. Bidan, "Electrostatic immobilization of polyoxometalates on silicon: X-ray photoelectron spectroscopy and electrochemical studies," *Thin Solid Film*, vol. 519, no. 11, pp. 3732–3738, 2011.
- [16] C. Musumeci, M. Rosnes, F. Giannazzo, M. Symes, L. Cronin, and B. Pignataro, "Smart high-k nanodielectrics using solid supported polyoxometalate-rich nanostructures," *ACS Nano*, vol. 5, no. 12, pp. 9992–9999, 2011.
- [17] N. Glezos, P. Argitis, D. Velessiotis, and C. D. Diakoumakos, "Tunneling transport in polyoxometalate based composite materials," *Appl. Phys. Lett.*, vol. 83, no. 3, pp. 488–490, 2003.
- [18] Vrije Univ., Amsterdam, The Netherlands. (2008). *SCM, Theoretical Chemistry* [Online]. Available: <http://www.scm.com>
- [19] (2011, Jan. 10). *GARAND (Computer Program)* [Online]. Available: <http://www.goldstandardsimulations.com/>
- [20] N. Fay *et al.*, "Structural, electrochemical, and spectroscopic characterization of a redox pair of sulfite-based polyoxotungstates: α - $[W_{18}O_{54}(SO_3)_2]^{4-}$ and α - $[W_{18}O_{54}(SO_3)_2]^{5-}$," *Inorg. Chem.*, vol. 46, no. 9, pp. 3502–3510, 2007.
- [21] X. López, J. J. Carbó, C. Bo, and J. M. Poblet, "Structure, properties and reactivity of polyoxometalates: A theoretical perspective," *Chem. Soc. Rev.*, vol. 41, no. 22, pp. 7537–7571, 2012.
- [22] C. Lee, W. Yang, and R. Parr, "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density," *Phys. Rev. B*, vol. 37, no. 2, pp. 785–789, 1988.
- [23] S. Faas, J. G. Snijders, J. H. V. Lenthe, E. V. Lenthe, and E. J. Baerends, "The ZORA formalism applied to the Dirac-Fock equation," *Chem. Phys. Lett.*, vol. 246, no. 6, pp. 632–640, 1995.
- [24] C. C. Pye and T. Ziegler, "An implementation of the conductor-like screening model (COSMO) of solvation within the ADF package," *Theoretical Chem. Accounts*, vol. 101, no. 6, pp. 396–408, 1999.
- [25] M. Swart, P. T. van Duijnen, and J. G. Snijders, "A charge analysis derived from an atomic multipole expansion," *J. Comput. Chem.*, vol. 20, no. 1, pp. 79–88, 2001.
- [26] A. Asenov, G. Slavcheva, A. R. Brown, J. H. Davies, and S. Saini, "Increase in the random dopant induced threshold fluctuations and lowering in sub-100 nm MOSFETs due to quantum effects: A 3-D density-gradient simulation study," *IEEE Trans. Electron Devices*, vol. 48, no. 4, pp. 722–729, Apr. 2001.

- [27] S. Markov *et al.*, "Drain current collapse in nanoscaled bulk MOSTETs due to random dopant compensation in the source/drain extensions," *IEEE Trans. Electron Devices*, vol. 58, no. 8, pp. 2385–2393, Aug. 2011.
- [28] S. M. Amoroso, A. Maconi, C. M. Compagnoni, A. Mauri, A. S. Spinelli, and A. L. Lacaita, "Three-dimensional simulation of charge-trap memory programming—Part II: Variability," *IEEE Trans. Electron Devices*, vol. 58, no. 7, pp. 1872–1878, Jul. 2011.
- [29] A. S. Grove, *Physics and Technology of Semiconductor Devices*. New York, NY, USA: Wiley, 1967.
- [30] K. Naruke, S. Taguchi, and M. Wada, "Stress induced leakage current limiting to scale down EEPROM tunnel oxide thickness," in *IEDM Tech. Dig.*, 1988, pp. 424–427.
- [31] S. D. Chae *et al.*, "Ultrashort SONOS memories," *IEEE Trans. Nanotechnol.*, vol. 3, no. 4, pp. 417–425, Dec. 2004.
- [32] C. H. Lee, K. I. Choi, M. K. Cho, Y. H. Song, K. C. Park, and K. Kim, "A novel SONOS structure of SiO₂/SiN/Al₂O₃ with TaN metal gate for multi-giga bit flash memories," in *IEEE IEDM Tech. Dig.*, Dec. 2003, pp. 26.5.1–26.5.4.
- [33] A. Chung, J. Deen, J.-S. Lee, and M. Meyyappan, "Nanoscale memory devices," *Nanotechnology*, vol. 21, no. 41, pp. 412201–412223, 2010.
- [34] H. T. Lue *et al.*, "Study of electron and hole injection statistics of BE-SONOS NAND flash," in *Proc. IEEE Int. Memory Workshop*, May 2010, pp. 978–982.
- [35] H. T. Lue *et al.*, "Scaling feasibility study of planar thin floating gate (FG) NAND flash devices and size effect challenges beyond 20 nm," in *IEEE IEDM Tech. Dig.*, Dec. 2011, pp. 9.2.1–9.2.4.
- [36] A. Nainani *et al.*, "Development of a 3D simulator for metal nanocrystal flash memories under NAND operation," in *IEEE IEDM Tech. Dig.*, 2007, pp. 947–950.
- [37] T.-H. Hou, C. Lee, V. Narayanan, U. Ganguly, and E. C. Kan, "Design optimization of metal nanocrystal memory—Part II: Gate-stack engineering," *IEEE Trans. Electron Devices*, vol. 53, no. 12, pp. 3103–3109, Dec. 2012.
- [38] G. Molas *et al.*, "Thorough investigation of Si-nanocrystal memories with high-k interpoly dielectrics for sub-45 nm node flash NAND applications," in *IEEE IEDM Tech. Dig.*, Dec. 2007, pp. 453–456.
- [39] J. Robertson and M. J. Powell, "Gap states in silicon nitride," *Appl. Phys. Lett.*, vol. 44, no. 4, pp. 415–417, 1984.
- [40] N. Glezos, A. M. Douvas, P. Argitis, F. Saurenbach, J. Chrost, and C. Livitsanos, "Electrical characterization of molecular monolayers containing tungsten polyoxometalates," *Microelectron. Eng.*, vol. 83, nos. 4–9, pp. 1757–1760, 2006.
- [41] A. Asenov, A. R. Brown, J. H. Davies, S. Kaya, and G. Slavcheva, "Simulation of intrinsic parameter fluctuations in decanometer and nanometer-scale MOSFETs," *IEEE Trans. Electron Devices*, vol. 50, no. 9, pp. 1837–1852, Sep. 2003.
- [42] S. Markov, S. M. Amoroso, L. Gerrer, F. A. Lema, and A. Asenov, "Statistical interactions of multiple oxide traps under BTI stress of nanoscale MOSFETs," *IEEE Trans. Electron Devices*, vol. 34, no. 5, pp. 686–688, May 2013.
- [43] A. Mauri *et al.*, "Comprehensive investigation of statistical effects in nitride memories," *IEEE Trans. Electron Devices*, vol. 57, no. 9, pp. 2116–2123, Sep. 2010.
- [44] C. C. Hsieh *et al.*, "A novel BE-SONOS NAND flash using non-cut trapping layer with superb reliability," in *IEEE IEDM Tech. Dig.*, 2010, pp. 114–117.



Vihar P. Georgiev received the Ph.D. degree from the University of Oxford, Oxford, U.K., in 2011.

He joined the Device Modelling Group, School of Engineering, University of Glasgow, Glasgow, U.K., in 2011, where he is currently a Post-Doctoral Research Assistant.



Stanislav Markov (M'09) received the Ph.D. degree in electrical and electronics engineering from the University of Glasgow, Glasgow, U.K., in 2009.

He has been with the Department of Chemistry, University of Hong Kong, Hong Kong, since 2012.



Laia Vilà-Nadal received the Ph.D. degree from the Rovira i Virgili University, Tarragona, Spain, in 2011.

She joined the Cronin Group, Westchem, School of Chemistry, University of Glasgow, Glasgow, U.K., in 2011, as a Post-Doctoral Researcher.



Christoph Busche received the Ph.D. degree from the Institute of Inorganic Chemistry, University of Heidelberg, Heidelberg, Germany, in 2009.

He has been a member of the Cronin Group, Westchem, School of Chemistry, University of Glasgow, Glasgow, U.K., since 2010, as a Post-Doctoral Research Assistant.



Leroy Cronin received the Ph.D. degree in chemistry from the University of York, York, U.K., in 1997.

He joined the University of Glasgow, Glasgow, U.K., as a Lecturer in 2002, Reader in 2005, Professor in 2006, Gardiner Professor in 2009, and has been the Regius Professor of Chemistry since 2013. He leads the Complex Chemical Systems Research Group.



Asen Asenov (M'96–SM'05–F'11) received the Ph.D. degree in physics from the Bulgarian Academy of Science, Sofia, Bulgaria, in 1989.

He is currently the James Watt Chair in Electrical Engineering with the University of Glasgow, Glasgow, U.K. He is also a Co-Founder, CEO, and Director of Gold Standard Simulations Ltd., Glasgow.