Diffusion is implemented as random walk: At each time point, two random numbers are generated for each diffusible particle that define the direction and the magnitude of its movement. These random numbers are distributed so that they generate Brownian motion, thus simulating diffusion. Reactions are implemented in a similar way: For each reaction, a reaction probability is calculated on the basis of the binding constants and the local reactant concentrations (see the figure). At each time point a reaction event occurs if the random number generated for that particular reaction is greater than the reaction probability. Thus, one MCell run generates a realistic sequence of diffusion steps and reactions for each particle. Average concentration values are obtained by averaging results from many MCell runs.

Obtaining a realistic model is essential for this type of computer simulation. Coggan et al. used electron-tomographic reconstructions of ciliary ganglia containing clusters of somatic spines to extract and convert to digital form realistic surfaces of pre- and postsynaptic membranes and to identify synaptic and extrasynaptic regions. Two major classes of nicotinic acetylcholine receptors (nAChRs) are known to be present in postsynaptic neurons: α7-nAChRs, which are mostly extrasynaptic, and α3*-nAChRs, which are concentrated at the postsynaptic density. The authors populated the digitalized surfaces with both nAChR types, specified detailed kinetic models for neurotransmitter-induced opening, and included ACh hydrolysis by acetylcholinesterase. As is generally the case in modeling biochemical processes, not all parameters needed to make a realistic model of synaptic transmission are available in the literature. The authors used available experimental data, combined with their simulations when appropriate, to estimate the parameters needed for their model: the number of ACh molecules per vesicle, the distribution densities of both nAChR types and of acetylcholinesterase, the ACh binding constants, and the transition rates between the states of nAChR in each kinetic model.

Coggan et al. used their model to simulate peak currents at the postsynaptic neuron after release of a single synaptic vesicle from different sites for a range of parameters. They show convincingly that their simulations are consistent with experimental data only if, or even all, neurotransmitter release is ectopic. Ectopic release can explain some previously unresolved experimental observations, such as that the postsynaptic current arising from presynaptic stimulation is dominated by the contribution from α7-nAChRs, although these receptors are almost excluded from the postsynaptic density (6). Among the parameters used in the model, the uncertainty in the mean number of ACh molecules per vesicle had the strongest influence on the postsynaptic peak currents.

Models are rarely perfect, and here too there is of course room for improvement. For example, some parameters used by Coggan et al. were obtained from the neuromuscular junction, rather than the ciliary ganglion, and the binding of choline to α7-nAChRs was not taken into account. Further computational studies of synapses or other systems might benefit from the better structural models that can be obtained by using more recent methods for preserving samples for electron microscopy, such as rapid freezing followed by freeze substitution, or cryo preparation (7, 8). Although the simulation data presented by Coggan et al. fit experimental data well, there are still quantitative differences, probably due to uncertainties in the parameter values—the results of a modeling study are valid only for the conditions specified by the model. Nevertheless, Coggan et al. provide an example of how computer simulations based on carefully built, realistic models provide important insights that cannot be obtained by current experimental methods. We hope to see more such studies in the future.

References

APPPLIED PHYSICS

Where Do the Dopants Go?
Scott Roy and Asen Asenov

The rapid growth of the semiconductor industry in the past 40 years has largely been a result of the ever-decreasing size of CMOS (complementary metal-oxide semiconductor) switching elements, which form the underlying logic circuits in practically every modern digital system. As the size of CMOS switches—and of the field-effect transistors from which they are made—is reduced, integrated circuits constructed from these devices improve in speed, device density, and cost per function. The result is an intense industrial drive toward miniaturization, as predicted by Moore (1). Here, we consider key physical limitations to the miniaturization process and the evolution of modeling tools used to understand these limitations.

Each CMOS transistor consists of a source, a drain, a channel, and a gate (see the figure). The International Roadmap for Semiconductors (2) predicts that CMOS transistors with gate lengths of 7 nm will be mass-produced by 2018. Devices with channel lengths of 45 nm are already in production, and individual research devices with channel lengths of 4 nm have been demonstrated (3). CMOS devices will continue to shrink over the next two decades, but as they approach the scale of the silicon lattice, the precise atomic configuration of their structure will become critically important to their macroscopic properties.

Mead and Keyes recognized in the 1970s (4, 5) that below a critical size, devices can no longer be described, designed, modeled, or understood as continuous semiconductors with smooth boundaries and interfaces. At nanometer-scale dimensions, the number and position of the dopant atoms, introduced to alter the electrical properties of regions of a field effect transistor, will vary between devices (see the figure); as a result, each transistor will be microscopically different. The variation in dopant position between devices leads to measurable differences in macroscopic parameters such as drive current, threshold voltage, and leakage. Further size reduction reduces the number of dopants, exacerbating the variations and hence the differences in device performance.

Furthermore, with decreasing device size, the interface roughness of typical gate oxides (one or two atomic layers) becomes comparable to the gate thickness itself. Thus, each device will have a unique gate thickness and a unique pattern of interface roughness. The use of high-permittivity (high-k) gate insulators as a replacement...
for present gate oxides will allow thicker gates, which may ease this source of variation for one or two technology generations (3 to 6 years). However, atomic-scale variation in the positions of impurity atoms, local variations in the silicon/silicon dioxide interface above the channel, and local variations of the thickness of the silicon dioxide introduce, between each transistor and its neighbor, fluctuations in device electrostatics, electron transport, and gate leakage, respectively. The granularity of the photoresist used to pattern the gate will introduce further local variations in the shape of the gate itself.

With existing technology, it is impractical to image the detailed atomic structure of individual nanometer-scale CMOS transistors and to link their structure to the corresponding device characteristics. Over the past decade, researchers aiming for an understanding of intrinsic parameter fluctuations in nanometer-scale CMOS transistors have therefore resorted to numerical simulation, using increasingly detailed mathematical models.

It was not until the mid-1990s that a body of experimental evidence confirmed fluctuations in device current-voltage characteristics due to atomic-scale structure variation, beyond that expected from variations in fabrication equipment, for devices with channel length below 1 μm (6, 7). This gave the first serious impetus to model the problem, although initial theoretical analyses were qualitative in nature or used two-dimensional (2D) simulation techniques. It was realized early on that these 2D models could not capture all the relevant physics and that the results were inconsistent with experiment. A realistic dopant distribution is inherently 3D, defining a potential landscape under the gate. When the device is turned on, local variations in the potential landscape form chains of potential “valleys,” allowing the formation of percolation paths between source and drain. However, a realistic potential landscape will rarely produce percolation paths that run parallel to the average electron flow in idealized devices, as assumed in the 2D model.

Wong and Taur were the first to report a full 3D simulation of field-effect transistors under the influence of random discrete doping (8). They used a drift-diffusion simulator, which models electron transport as incompressible fluid flow, considering the area under the gate as a checkerboard of smaller, interconnected devices, each with a different density of dopant atoms. The results showed the two classic fingerprints of randomly distributed, discrete dopants: a spread in the device threshold voltages, and a lowering of the mean threshold voltage relative to that of a continuously doped system (the formation of percolation paths will always allow current to flow at a lower gate potential than for an idealized device).

These 3D simulations foreshadowed today’s techniques and used a computationally efficient 3D device simulator, but they were not immediately adopted into commercial simulators. A 3D simulation requires substantial computational resources; the simulations are made even more demanding by the need to accurately model over atomic length scales. In addition, statistical information is required for each field effect transistor design, necessitating many 3D simulations for each design. Because of these formidable computational obstacles, 2D and analytical approximations remained popular, with researchers accepting substantial limitations in accuracy as a price for acceptable simulation times (9, 10).

Yet in the past few years, the same device scaling that proves so problematic to simulation has led to radical improvements in computational power. Combined with modern, highly parallel simulation codes, this means that statistical simulations that ran for weeks in the late 1990s can now be completed in a weekend. Present simulation tools are still mainly drift diffusion–based and now include random distributions of dopants in the channel, source, drain, and gate of a field effect transistor (see the figure). In addition, variations in gate thickness and atomic-scale roughness in the pattern of the gate edges are accounted for, and the quantum nature of channel electrons is modeled through a density gradient adjustment to the device’s potential profiles.

There is even a move to replace the drift-diffusion core of modern commercial and research device simulation codes with the computationally more expensive Monte Carlo approach (11) to correctly account for electron transport at the high lateral fields present in modern devices in the on-state. However, threshold voltage variations and the subthreshold operation of devices (which govern device leakage) remain the most important parameters for a circuit designer, and these can be adequately modeled with drift-diffusion simulators.

Now that rapid, robust, and accurate simulation tools have been developed, a wealth of applications present themselves in the fields of electronic devices, circuits, and systems. For instance, such simulations may help to develop devices that are resistant to fluctuation effects. Double-gate transistors (in which a 10-nm silicon channel is gated above and below) require no channel doping and are therefore immune to dopant fluctuation effects, but are subject to body thickness variations due to local roughness above and below the channel. The new simulation tools are being used to predict when industry should move to double-gate transistors, and when double-gate transistors themselves will become unviable.

In addition, methodologies have been developed to make use of the extracted data on fluctuations as an input to industrial circuit design tools (12, 13). For the first time, circuit designers can quantitatively analyze how atomic-scale variations in field-effect
transistors will affect the yield and functionality of commercial digital logic and memory circuits.

References

ASTRONOMY

The First Generations of Stars

Timothy C. Beers

The very first stars that formed after the Big Bang, some 13 to 14 billion years ago, are likely to have been quite massive and extremely short-lived; no examples are expected to remain in the universe today. However, they may have left behind their “calling cards” by producing a distinctive distribution of elements recorded in the atmospheres of long-lived stars that formed just after these massive progenitors. Stars that are extremely iron-poor (hyper metal-poor stars) are believed to be very old, and are thus possible candidates for second-generation stars.

On page 451 of this issue, Iwamoto et al. (1) describe a model that attempts to account for the elemental abundances in two hyper metal-poor stars. The stars, HE 0107-5240 (2) and HE 1327-2326 (3), contain less than 1/100,000 of the iron observed in the Sun. Furthermore, they are both greatly enhanced, relative to the Sun, in the light elements carbon, nitrogen, and oxygen (for HE 0107-5240; studies of the oxygen abundance in HE 1327-2326 are under way); these are the most important elements for the formation of life, at least of the form with which we are familiar.

Star formation in the Milky Way and throughout the present universe is poorly understood. This is because it takes place in a complex environment, where one has to account for the effects of the elements produced by previous generations of stars, the influence of magnetic fields, and star formation—triggering events such as shocks from nearby supernovae. In the very early universe, the physics of star formation is thought to have been much simpler, because only hydrogen, helium, and a small amount of lithium were present; stars most likely formed via radiative cooling by molecules involving these elements.

Modern computational models of early star formation predict that most stars that formed in the early universe were probably quite massive, on the order of several hundred times the mass of the Sun. Such stars burn their fuel extremely rapidly (within a few million years after their birth) and then explode. Astronomers are uncertain which elements might form in these very massive stars during their explosive death throes, but current calculations indicate that they should eject large amounts of iron and only small amounts of carbon (4, 5).

This prediction is incompatible with the elements observed in the hyper metal-poor stars modeled by Iwamoto et al. (1). However, contemporary observations show that when stars form, they do so with a distribution of masses. The distribution of stellar masses in the early universe may have included first-generation stars with only 25 times the mass of the Sun. Iwamoto et al. suggest that these lower-mass first-generation stars are responsible for the elemental abundance patterns now observed in the second-generation stars HE 0107-5240 and HE 1327-2326.

These two stars—the most iron-poor stars known today—are inferred to have masses that are ~80% that of the Sun. Stars in this mass range have very long lifetimes, because they burn their fuel slowly—so slowly that, if they were born in the early universe, they could still be detected today. According to one popular model, the formation of such low-mass stars was triggered by shocks from the explosions of the massive first-generation stars (see the figure). These second-generation low-mass stars provide our only means to quantify the distribution of elements that were formed by their long-gone progenitors. Hence, stars like HE 0107-5240 and HE 1327-2326 are the “scribes” of the early universe. Their atmospheres retain the memory of the composition of the gas from which they formed.

The search for stars with extremely low metallicity (astronomers refer to all elements heavier than hydrogen and helium as “metals”) began 50 years ago, when it was recognized that stars with lower metallicity than the Sun exist in the Milky Way. In the past decade, several thousand stars with iron abundances less than 1% of the solar abundance have been identified. Many of the most extreme examples have been studied at high spectral resolution with the world’s largest telescopes. These studies have shown that, of the 12 stars with the lowest iron abundance known to date, five exhibit highly enhanced light elements such as carbon, nitrogen, and, in some cases, oxygen (6). It seems inescapable that early element production favored the light elements. The early production of light elements is a crucial ingredient in models for early star formation, because these species pro-

The author is at Michigan State University and the Joint Institute for Nuclear Astrophysics, East Lansing, MI 48824, USA. E-mail: beers@pa.msu.edu