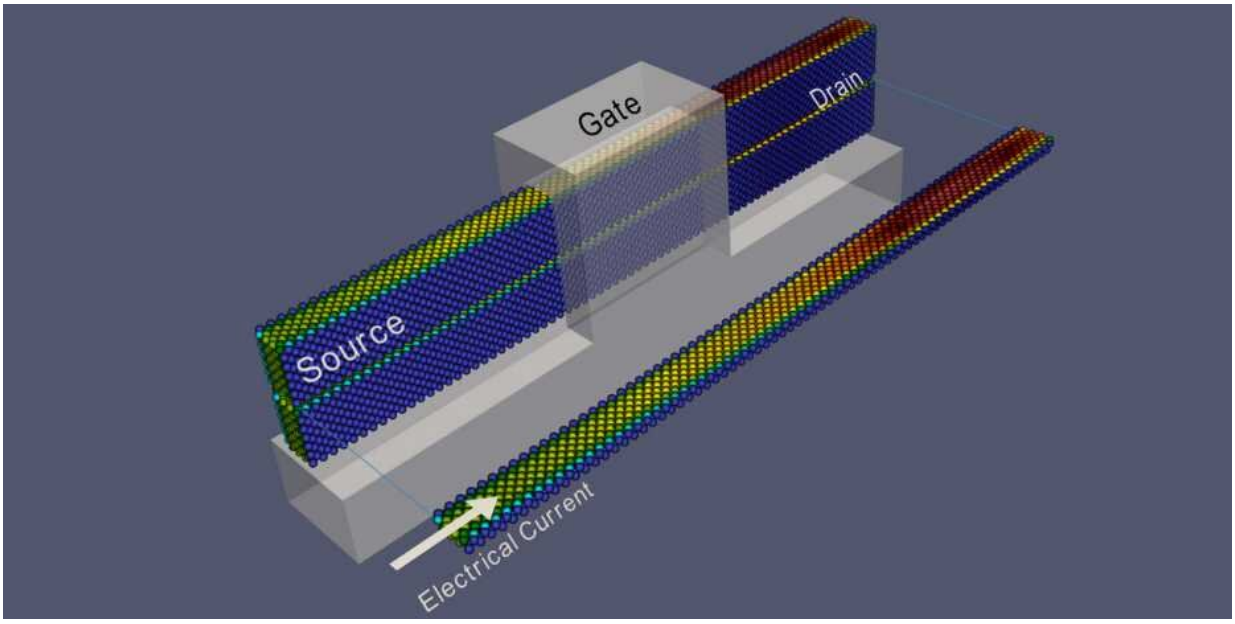
# A 'simulation booster' for nanoelectronics

November 22 2019, by Simone Ulmer




Self-heating in a so-called Fin field-effect transistor (FinFET) at high current densities. Each constituting Silicon atom is coloured according to its temperature. Credit: Jean Favre, CSCS

Two research groups from ETH Zurich have developed a method that can simulate nanoelectronics devices and their properties realistically, quickly and efficiently. This offers a ray of hope for the industry and data center operators alike, both of which are struggling with the (over)heating that comes with increasingly small and powerful transistors.

Chip manufacturers are already assembling transistors that measure just a few nanometers across. They are much smaller than a human hair, whose diameter is approximately 20,000 nanometers in the case of finer strands. Now, demand for increasingly powerful supercomputers is driving the industry to develop components that are even smaller and yet more powerful at the same time.

However, in addition to physical laws that make it harder to build ultra-scaled transistors, the problem of the ever increasing heat dissipation is putting manufacturers in a tricky situation—partly due to steep rises in cooling requirements and the resulting demand for energy. Cooling the computers already accounts for up to 40 percent of power consumption in some data centers, as the research groups led by ETH professors Torsten Hoefler and Mathieu Luisier report in their latest study, which they hope will allow a better approach to be developed. With their study, the researchers have been awarded the ACM Gordon Bell Prize, the most prestigious prize in the area of supercomputers, which is awarded annually at the SC supercomputing conference in the United States.

To make today's nanotransistors more efficient, the research group led by Luisier from the Integrated Systems Laboratory (IIS) at ETH Zurich simulates transistors using software named OMEN, which is a so-called quantum transport simulator.

OMEN runs its calculations based on what is known as density functional theory, allowing a realistic simulation of transistors in atomic resolution and at the quantum mechanical level. This simulation visualizes how electrical current flows through the nanotransistor and how the electrons interact with crystal vibrations, thus enabling researchers to precisely identify locations where heat is produced. In turn, OMEN also provides useful clues as to where there is room for improvement.

# Improving transistors using optimized simulations

Until now, conventional programming methods and supercomputers only permitted researchers to simulate heat dissipation in transistors consisting of around 1,000 atoms, as data communication between the processors and memory requirements made it impossible to produce a realistic simulation of larger objects.

Most computer programs do not spend most of their time performing computing operations, but rather moving data between processors, main memory and external interfaces. According to the scientists, OMEN also suffered from a pronounced bottleneck in communication, which curtailed performance. "The software is already used in the semiconductor industry, but there is considerable room for improvement in terms of its numerical algorithms and parallelization," says Luisier.

Until now, the parallelization of OMEN was designed according to the physics of the electro-thermal problem, as Luisier explains. Now, Ph.D. student Alexandros Ziogas and the postdoc Tal Ben-Nun—working under Hoefler, head of the Scalable Parallel Computing Laboratory at ETH Zurich—have not looked at the physics but rather at the dependencies between the data. They reorganized the computing operations according to these dependencies, effectively without considering the underlying physics. In optimizing the code, they had the help of two of the most powerful supercomputers in the world—"Piz Daint" at the Swiss National Supercomputing Centre (CSCS) and "Summit" at Oak Ridge National Laboratory in the US, the latter being the fastest supercomputer in the world. According to the researchers, the resulting code—dubbed DaCe OMEN—produced simulation results that were just as precise as those from the original OMEN software.

For the first time, DaCe OMEN has reportedly made it possible for researchers to produce a realistic simulation of transistors ten times the

size, made up of 10,000 atoms, on the same number of processors—and up to 14 times faster than the original method took for 1,000 atoms. Overall, DaCe OMEN is more efficient than OMEN by two orders of magnitude: on Summit, it was possible to simulate, among other things, a realistic transistor up to 140 times faster with a sustained performance of 85.45 petaflops per second—and indeed to do so in double precision on 4,560 computer nodes. This extreme boost in computing speed has earned the researchers the Gordon Bell Prize.

## Data-centric programming

The scientists achieved this optimization by applying the principles of data-centric parallel programming (DAPP), which was developed by Hoefler's research group. Here, the aim is to minimize data transport and therefore communication between the processors. "This type of programming allows us to very accurately determine not only where this communication can be improved on various levels of the program, but also how we can tune specific computing-intensive sections, known as computational kernels, within the calculation for a single state," says Ben-Nun. This multilevel approach makes it possible to optimize an application without having to rewrite it every time.

Data movements are also optimized without modifying the original calculation—and for any desired computer architecture. "When we optimize the code for the target architecture, we're now only changing it from the perspective of the performance engineer, and not that of the programmer—that is, the researcher who translates the scientific problem into code," says Hoefler. This, he says, leads to the establishment of a very simple interface between computer scientists and interdisciplinary programmers.

The application of DaCe OMEN has shown that the most heat is generated near the end of the nanotransistor channel and revealed how it

spreads from there and affects the whole system. The scientists are convinced that the new process for simulating electronic components of this kind has a variety of potential applications. One example is in the production of lithium batteries, which can lead to some unpleasant surprises when they overheat.

**More information:** Alexandros Nikolaos Ziogas et al. A data-centric approach to extreme-scale ab initio dissipative quantum transport simulations, *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis on - SC '19* (2019). DOI: 10.1145/3295500.3357156

Provided by ETH Zurich