As transistor sizes continue to scale, we are about to witness stunning levels of chip integration, with 1,000 (simple) core on a single die, and increasing levels of die stacking. Transistors may not be much faster, but there will be many more of them. In these architectures, energy and power will be the main constraint, efficient communication and synchronization a major challenge, and programmability an unknown.

In this context, this talk presents some of the technologies that we will need to deploy to exploit these architectures. Cores need to flexibly operate at a range of voltages, and techniques for efficient energy use such as power gating and voltage speculation need to be widespread. To enable data sharing, we need to rethink synchronization and fence hardware for scalability. Hardware extensions to ease programming will provide a competitive edge. A combination of all of these techniques—and additional disruptive technologies—are needed.

Josep Torrellas is the Saburo Muroga Professor of Computer Science at the University of Illinois at Urbana-Champaign. He leads the Center for Programmable Extreme-Scale Computing, a center focused on architectures for extreme energy and power efficiency. He has been the director of the Intel-Illinois Parallelism Center (I2PC), a center created by Intel to advance parallel computing. He has made contributions to parallel computer architecture in the areas of shared memory multiprocessor organizations, cache hierarchies and coherence protocols, thread-level speculation, and hardware and software reliability. He is a Fellow of IEEE and ACM. He received the 2015 IEEE CS Technical Achievement Award.