The role of NVM, emerging memories and in-memory compute for edge AI
Panel Chair: Boris Murmann, Professor of Electrical Engineering, Stanford University

The design of energy-efficient hardware for edge AI revolves primarily around data movement and memory access. Especially for small models, it is not only feasible to have all memory on chip, but also reduce data movement further using in-memory compute cores. The community is currently investigating a variety of options in this space using compute-SRAM, Flash, RRAM, MRAM, phase change, and other emerging memory technologies. Will these ideas lead to a breakthrough in hardware efficiency? Will they deliver the programmability needed for practical deployment and real-world applications? A panel of experts from industry, academia and research institutes will debate these questions and spell out their future predictions for our audience.