NSF NANOSCALE SCIENCE AND ENGINEERING GRANTEES CONFERENCE:

NANO AND AI CONVERGENCE

DECEMBER 9-10, 2024

"AI at the Edge – from materials to algorithms"

AVIK GHOSH

Professor, Department of Electrical and Computer Engineering and Department of Physics,
University of Virginia

Bio: Avik Ghosh is Professor in the Charles Brown Dept of Electrical and Computing Engineering and the Dept of Physics at the University of Virginia. Ghosh has authored 150+ refereed papers and two books with World Scientific in the area of computational nano-materials and devices. He is a Fellow of the Institute of Physics (IOP), a senior member of the IEEE, and site director of the NSF MIST Center on multifunctional integrated systems. He has received the IBM Faculty Award, the NSF CAREER Award, and UVA's All University Teaching Award.

Abstract: With the unprecedented growth in mobile computing and AIoT (Artificial IoT) in the fourth industrial revolution (information), there is a clear need to push processing to the edge, at the location where data is being generated. This will require processors that operate at low Size, Weight and Power (SWaP), and AI has an important role to play. A candidate example is magnetic devices which are now integrable with CMOS for various applications from non-volatile memory to artificial neurons to sources of true random noise. At the materials level, there is a need to expand the suite of magnetic materials ("Materials Genome") for optimal properties, where exploratory ML can play a critical role. At a circuit and architecture level, coupled magnetic neurons can perform specialized tasks such as NP-hard optimization problems and real-time prediction of time-series data. At an algorithmic level, one challenge with on-chip computing and limited memory is how to encode new information without continuous data overwrite ('catastrophic forgetfulness'). Brain-inspired processes such as a supervised Hebbian algorithm associated with adaptive synaptogenesis could help optimize the reuse of neurons and synapses to process information with very high sparsity of interconnections.