Uncertainty-aware forecasting model of HfO_x-memristive devices using mixture density networks

T. Benthien[1], K. Nikiruy[1], B. Spetzler[1], C. Wenger[2], M. Ziegler[1]

[1] Energy Materials and Devices, Department of Materials Science, Kiel University, Kiel, Germany
[2] IHP – Leibniz Institute for High Performance Microelectronics, Frankfurt/Oder, Germany

Stochastic resistive switching is a fundamental challenge shared by all filamentary memristive devices: identical voltage pulses applied to the same device can yield markedly different conductance states due to the stochastic nature of conductive filament formation and dissolution [1]. This cycle-to-cycle variability fundamentally limits deterministic modeling approaches and hinders the reliable design, simulation, and programming of memristive devices for neuromorphic and analog in-memory computing applications. Accurate probabilistic models of switching behavior would enable stochastic device optimization, worst-case circuit design, and efficient automated characterization workflows.

In this work, HfO_x-based memristive devices are used as demonstrators to develop and validate such a framework, combining mixture density networks (MDNs) with an uncertainty-based active learning loop.

The framework introduces two main concepts. The first is a probabilistic MDN model that learns the stochastic switching behavior directly from voltage-pulse measurements: rather than predicting a single conductance value, the MDN outputs the full probability distribution of achievable post-pulse conductance states given pulse polarity, pre-pulse conductance, and applied voltage. This enables the model to capture multimodal switching behavior that conventional deterministic models fundamentally cannot represent. The second concept is an uncertainty-based learning loop that minimizes the number of physical measurements required to build a reliable training dataset. An ensemble of MDNs identifies regions of the pulse-parameter space where model predictions disagree most, indicating gaps in training coverage. These high-uncertainty regions are prioritized for the next round of pulse measurements, leading to a systematic, high-throughput data acquisition protocol.

The MDN accurately predicts the probability distribution of post-pulse conductance states across the relevant parameter range, including multimodal behavior near the switching threshold. This predictive accuracy is already achieved after pretraining on 150 samples of randomized voltage-pulse measurements with a mean error of 7%, demonstrating that a reliable surrogate model can be established with only modest initial data collection. Building on this pretrained model, the uncertainty-based learning loop then selects only the most informative experiments for subsequent measurements. The maximum model disagreement is reduced by approximately 50% through the addition of only 200 targeted samples to the training data. Reference models trained on the same total amount of randomly measured data showed no comparable reduction.

Beyond this specific device, the proposed framework is broadly applicable to memristive technologies where stochastic switching complicates data-driven modeling. By keeping the training dataset small while maintaining high predictive quality, the approach reduces both the computational cost of model training and the physical load on devices - a potential advantage in scenarios where device endurance is limited. The automatable nature of the loop opens a path toward self-characterizing memristive systems, where devices continuously refine their own surrogate models during operation. Probabilistic surrogate models of this kind are a key enabler for stochastic circuit simulations, hardware-aware neural network training, and the programmatic control of memristive arrays in neuromorphic and analog in-memory computing architectures.

References

[1] S. Yarragolla et al., J. Appl. Phys. (131), 134304, 2022