2023 IEEE Nanotechnology Materials and Devices Conference (NMDC) | 979-8-3503-3546-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/NMDC57951.2023.10343611

Materials-to-applications evaluation framework: assessing memristor technologies for neural network implementations

G. Bersuker, J. Farmer, D. Veksler, A.-M. El-Sayed, T. Durrant, D. Z. Gao, A. Shluger

***Abstract*— Practical needs in technology capability assessment for extremely low-energy neuromorphic computing is addressed via a novel development/analysis concept integrating atomic-level material modeling, statistical simulations of charge transport in a device material stack and verification of the modeling scheme against measurements emulating circuitry operation conditions for applications in specific neural networks (NN). This multi-scale concept - from materials to applications - directly links materials to their electrical properties, and the latter to NN algorithms. Such link enables identifying structural features controlling device characteristics and the range of operation conditions delivering performance targets for a given technology implementation. In comparison to widely employed memristor analyses primarily based on TCAD-type methodology with adjustable phenomenological parameters, the proposed approach allows to deliver feedback on favorable material compositions and cell architecture/dimensions to modify memristor fabrication process. Implementation of this technology evaluation approach to carbon nanotube (CNT) memristors enables identifying structural and operation conditions delivering optimal performance ahead of actual circuitry fabrication.**

.

I. Introduction

Most demanding applications for future electronics hardware such as autonomous sensing/analysis (IoT), mobile and environmentally stable computing impose strict limitations on energy consumption. To simultaneously meet speed, power and density targets prospective devices are required to satisfy a wide range of conditions that need to be assessed at the initial development stages, prior to circuit design and fabrication. Considered options (avoiding “von Neumann bottleneck”) include novel computing architectures, in particular neuromorphic computing, involving non-volatile-memory (NVM) technologies capable of 3D integration, among other structural/integration features. Furthermore, to achieve on-chip training, analog devices - memristors (in crossbars structure) - should be utilized. Memristor technology enabling low power-high performance neuromorphic computing must deliver well-controlled multi-level memory updates to satisfy operation conditions in the considered NN type (DNN, SNN, etc.). Hence, hardware evaluation should be done in coordination with intended software use - co-evaluation of hardware and algorithms for specific applications.

General concerns to be addressed in technology evaluation include Classification accuracies; Reliability: for each type of neuromorphic computing implementations cells should be evaluated within the entire range of considered NC circuitry operation conditions; Variability: Drift and fluctuations of device parameters detrimentally impact NC operations. Such variability is a new reliability issue for neuromorphic classification. In particular, switching variability for each individual device occurring when the conductance value fluctuates or drifts during cycling between memory states.

The objective of the proposed framework is to expand and verify resistive random-access memories (RRAM) evaluation methodology, which can be applied to a wide range of materials, beyond carbon nanotubes. In its present form, it combines atomic-level material modeling (DFT and mesoscopic force fields) and statistical simulations of currents via conductive paths through the intrinsically stochastic structure of material fabrics. Initial verification of the proposed physical model for memory update processes included matching electrical measurements of cell operations performed under circuitry-relevant sub-ns pulse durations to simulations employing material modeling parameters. Assessment of the impact of hardware non-idealities, such as memory-update variability, on NN learning characteristics has been conducted. This approach was applied to evaluation of carbon nanotubes (CNT NRAM), Fig. 1, metal oxides (HfO2 OxRAM) [2] and 2D material-based memristors (MoS2 and hBN [3]).

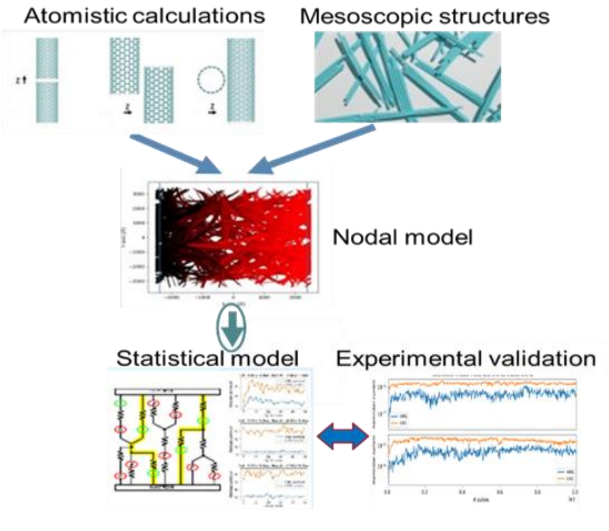


Fig. 1. RRAM evaluation methodology in application to CNT technology. Material modeling results are used to construct/parametrized a nodal current model, which, in turn, is incorporated in statistical simulations of the cell operations via a network of conductive paths formed by contacting CNTs and verified by electrical measurements.

G. Bersuker, J. Farmer, D. Veksler are with M2D Solutions, Austin TX, USA

A.-M. El-Sayed, T. Durrant, D. Z. Gao are with Nanolayers Research Computing LTD, London, UK

A. Shluger is with Department of Physics and Astronomy, UCL, London, UK

979-8-3503-3546-0/23/$31.00 ©2023 IEEE

To demonstrate the scope and employed methodology, below we focus on the results obtained for CNT material [4, 5, 6], Fig.2. CNT technology has demonstrated excellent switching characteristics and reliability, including latency, cycling endurance, data retention, and radiation hardness.

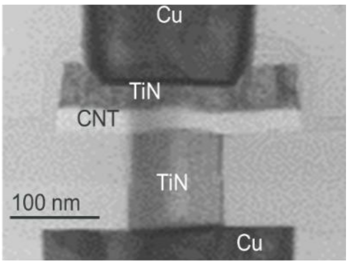


Figure 2. TEM of cross-section CNT cell bit with Cu BEOL

II. APPROACH

The switching mechanism considered here is an energy-driven tubes rearrangement via locally generated energy determined by the duration/magnitude of a current flow I(t) through the inter-tubes junction: $E = \int V(t)I(t)dt$, Fig. 3. An additional factor affecting CNTs switching is an electric field across the film inducing tubes torque, local charging and details of the film fabric.

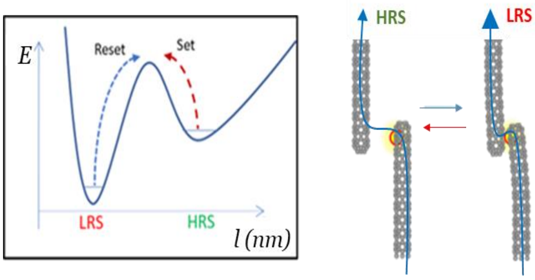


Figure 3. Double-well potential that defines switching processes, Set and Reset, between high- and low-resistance states (HRS and LRS). The minima and barriers are calculated based on mutual position/orientation of contacting CNTs along the conductive path.

Using Density Functional Tight Binding (DFTB) combined with non-equilibrium Green's functions (NEGF) modeling techniques we have carried out preliminary atomistic simulations on CNT junctions while systematically varying their chirality and geometry (length and mutual orientation of nanotubes) in order to understand their electrical conductivity [7, 8]. The electric current was found to strongly depend on the distances between the interacting nanotubes and was fitted to an equation depending on them. The charge transport parameters extracted from the atomistic modeling were used to simulate electrical conduction through realistic cell-size CNT films [9].

Structure - Material modeling

To model the complex behaviour of CNT films, we have taken a multi-scale approach to describe various elements involved. Properties of individual tubes are affected by their anhydride and hydrogen terminations, Fig. 4

On a larger scale, the film itself is treated using a mesoscopic potential, where the nanotubes are coarse-grained into connected cylindrical segments, Fig.4. This allows us to make representative models of CNT films with dimensions relevant to experimental devices.

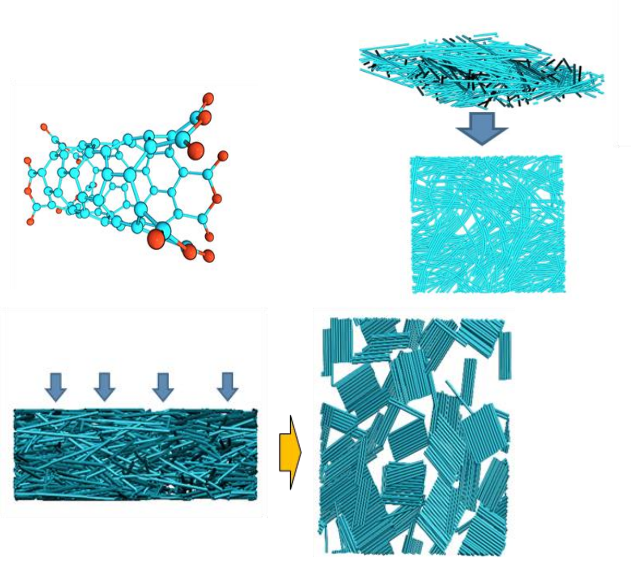


Figure 4. CNTs can have different terminations due to processing. Starting from random configuration in solution (used in fabrication), a force is applied downward to reduce the layer thickness (increase density). Nanotubes evolve in time in a new reduced volume until experimental density is reached. Stability of films at a given density in the range of 0.3 – 0.6 g/cm3 was verified. A fully densified structure exhibits enhanced bundle formation.

The currents through these mesoscopic structures can be evaluated using our previously parameterized current model and compared to electrical measurements of these devices, Fig.5.

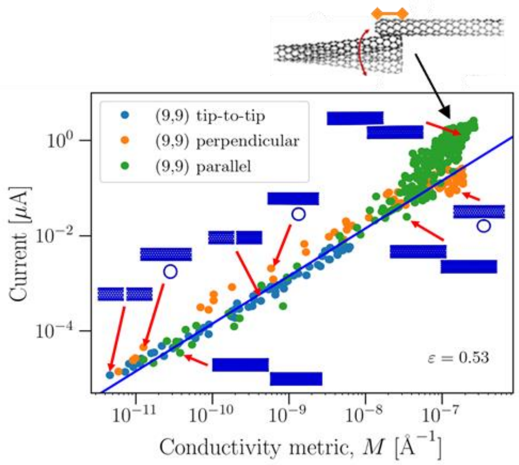


Figure 5. Current through the junction is dominated by conduction pathways between individual C atoms in contacting CNTs. Total contact conduction is a sum all individual tunnelling pathways accounting for all mutual C-C distances in possible mutual orientations and positions (see schematics of mutual CNT-CNT shifts) of neighbouring CNTs.

Electronic structure simulations were used to calculate the current across a wide range of representative junctions between nanotubes, Fig.5. This data allowed us to develop simple models capable of predicting current based on the geometric properties of a nanotube pair, e.g., the minimum distance between them.

Material modelling calculations extracted two main features of the CNT fabrics: (a) bundled CNTs are strongly coupled and require significant energy in order to separate, Fig.4. Additionally, a large surface area of contacting CNTs enables effective (low resistance) electron tunnelling between CNTs. (b) Contacts between stretching CNTs in different bundles have a smaller area. Hence, less energy is required to modify such inter-tubes contacts, which exhibit greater resistance due to smaller electron tunnelling. Hence, it is reasonable to expect that the electrical switching in such CNT fabrics is dominated by the type (b) contact.

The presently developed model is a proof of concept, demonstrating that the current can be directly calculated in physics-based models of a CNT fabric structure. Dynamical effects in the fabric formation can also be included, as the films' force field allows for time-dependent evolution of the structure. We demonstrate how structural information and physical parameters can be extracted to be employed in statistical device simulations.

Simulations

The developed Python simulation package (NRAMPY) connects local transport properties and CNT fabric structure obtained by modeling to the observable electrical characteristics of NRAM cells.

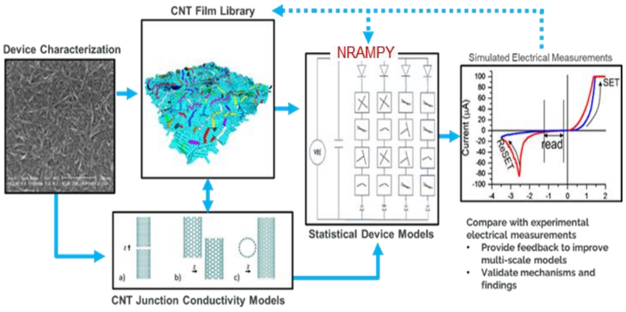

CNT Film Library
Device Characterization
NRAMPY
Simulated Electrical Measurements
Current (µA)
100
80
60
40
20
0
-20
-40
-60
-80
-100
-4
-3
-2
-1
0
1
2
SET
RESET
read
Statistical Device Models
a)
b)
c)
CNT Junction Conductivity Models
Compare with experimental electrical measurements
• Provide feedback to improve multi-scale models
• Validate mechanisms and findings

Figure 6. Within the developed simulation scheme, statistical 2D device model (NRAMPY) is constructed from randomly packed 3D CNT fabric (CNT Film library) accounting for possible inter-CNT contacts (CNT Junction Conductivity model). The cell structure presents a set of individual conductive paths with randomly oriented contacting CNTs. Simulations allow to predict electrical characteristics based on the cell architecture/dimensions/material fabrication process. It enables validation of the switching mechanism and modeling findings via comparison to electrical measurements, as well as provides feedback to improve modeling.

Using material modeling data, Figs. 3-5, NRAMPY constructs a unique one-bit cell, which is represented by numerous conductive paths formed by chains of contacting CNTs between the top and bottom electrodes, Fig. 9. Each path consists of several single CNTs and CNT bundles, both fixed and switchable inter-CNTs junctions and Schottky contacts between CNT chain, thin interfacial sub-oxide layer (3nm TiO_x) and metal electrodes. Simulations employ statistical distribution of the physical parameters defining the cell conductivity, such as area density of conductive paths, number of bundles in each path, CNT compositions in bundles, % of types of inter-bundles CNT contacts, etc. Conductance of each path is updated according to the changes of resistivity in individual junctions, Fig. 3, and the energy release occurring during the switching process that induces local temperature variations.

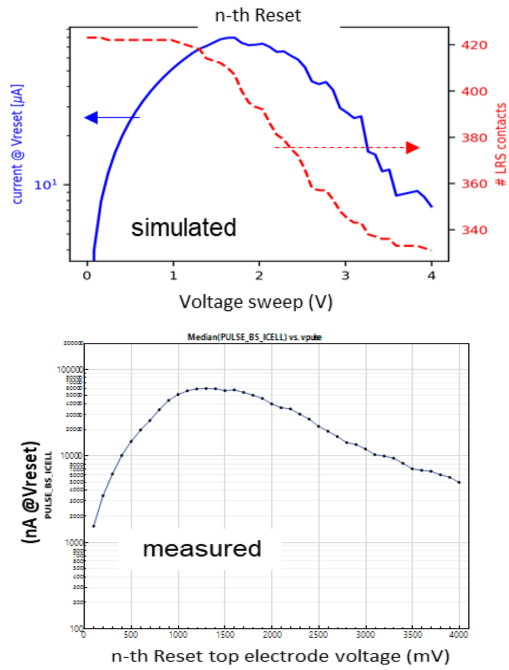


Figure 7. Example of the measurement and simulation of a DC I-V Reset process during one of the multiple Set/Reset switching cycles. The match is reached for the entire I-V range by accounting for about 70 junctions (out of 420 total – see Y-axis on the right in top graph) distributed throughout ~360 conductive paths, which switched from Low (LRS) to High (HRS) resistance state.

Simulations successfully match I-V characteristics during forming and switching operations in NRAM cells of different thicknesses and dimensions under a wide range of operation conditions, Figs.7,8 [10]. Perfect match between simulations and measurements on a large number of devices and switching/read DC/pulse conditions confirms modelling results (Figs.4, 5).

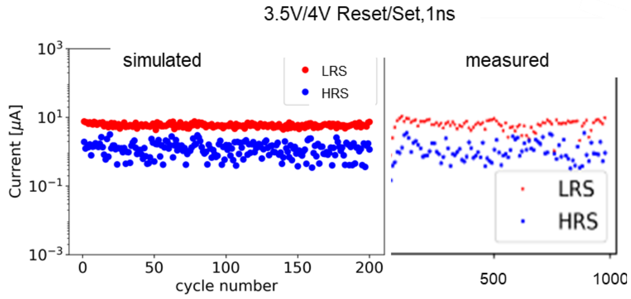


Figure 8. Example of a Read current simulation (at V= 0.5V) of cell switching cycling under 1ns pulse at (Reset/Set) = (3.5V/4.0) V. Simulations match mean values of conductance distribution and conductance variability by accounting for randomness of the initial distribution of junctions' resistance values.

Operation processes

Atomic-level structural changes determining electron transport in resistive NVM are driven by local heating induced by dissipated energy E, which is controlled by the magnitude of the current through a given conductive path at a given moment in time, I(t). Larger amount of energy released under longer programming time (t + Δt) may induce additional structural changes that can further modify current (either to higher > or lower < magnitude) I(t + Δt) ≠ I(t): longer time → higher released energy → more structural modifications affecting electron transport. Thus, the NVM cell characteristics are strongly affected by operation time duration. For this reason, device assessment should be performed under the test conditions which are close to actual circuitry operations (~ GHz).

We implemented ultra-short pulse measurements with the pulse duration going down to 100 ps.

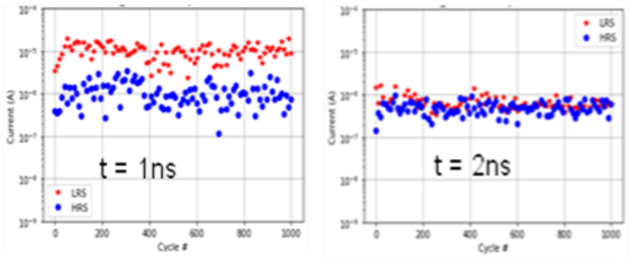


Figure 9. Cycling with 1 ns and 2 ns pulse durations and associated Set and Reset I-V characteristics. Symbols are the read current values taken at 0.5V after each pulse switching event.

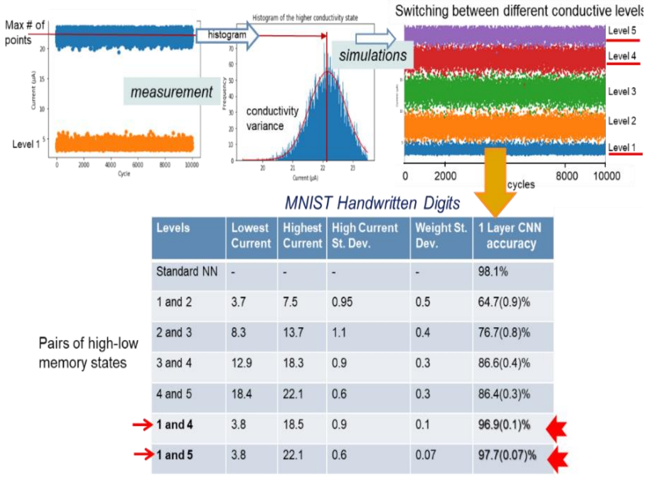
Critical importance of the test time duration for assessing technology capability under the circuitry-relevant operation conditions is demonstrated in Fig. 9. Higher released energy caused by the increase of switching pulse duration from 1 ns to 2 ns induces significant structural changes. These additional structural changes happen to result in more resistive HRS due to larger CNTs separations: consequently, the current through such contacts is low and cannot generate sufficiently high energy enabling the switch to LRS, Fig.3, - the system got stuck in HRS as seen in a collapse of the memory window.

Algorithms

Variability of cell switching between memory states came up as one of the major obstacles for practical implementation of memristor technologies. However, different implementations of neuromorphic algorithms have different tolerance with respect to switching variability. Our study aims to reduce the solution space, which can be extremely large in the machine learning architectures implemented with memristors. This task generally requires accounting for a variety of available memristor technologies in combination with multitude circuit configurations used to carry out machine learning algorithms. Our simulations are intended to identify combinations of devices structures and circuit operations that suit best specific hardware applications of interest. Below we demonstrate effectiveness of the proposed methodology in identifying memory characteristics of CNT memristors delivering optimum performance in a considered type of neural network.

Probabilistic NN (PNN)

TensorFlow PNN where weights and biases are sampled from a distribution (unlike a standard NN was employed to assess its capability to mitigate switching variability [11]. In TensorFlow, training is done using the mean and variance of the distribution for each weight and bias (instead of weights and biases directly). It allows to establish a target for device characteristics that would enable its use in NN, Fig.10.



Levels	Lowest Current	Highest Current	High Current St. Dev.	Weight St. Dev.	1 Layer CNN accuracy
Standard NN	-	-	-	-	98.1%
1 and 2	3.7	7.5	0.95	0.5	64.7(0.9)%
2 and 3	8.3	13.7	1.1	0.4	76.7(0.8)%
3 and 4	12.9	18.3	0.9	0.3	86.6(0.4)%
4 and 5	18.4	22.1	0.6	0.3	86.4(0.3)%
1 and 4	3.8	18.5	0.9	0.1	96.9(0.1)%
1 and 5	3.8	22.1	0.6	0.07	97.7(0.07)%

Figure 10. With the given switching variability (upper left graph), classification accuracy 86% is unacceptably low. Simulations show that in order to increase the classification accuracy above 94%, memory window needs to be kept at the level 1-4 or higher with standard deviation under 0.1.

SUMMARY

Within the proposed evaluation framework, we developed physical models describing operation processes in CNT material discussed here, as well as in metal oxides and other materials of interest. Statistical charge transport simulations based on output of mesoscopic and atomistic material modeling have been verified against device measurements in the wide range of switching conditions down to circuity-relevant sub-ns operations frequencies.

References

1. International Roadmap for Devices and Systems (IRDS™) 2020 Edition
2. G. Bersuker et al, Metal oxide resistive random access memory (RRAM) technology. In: *Advances in Non-volatile Memory and Storage Technology*, Elsevier, 2019
3. Teja Nibhanupudi et al, “Experimental demonstration of ultra-fast switching in 2D hexagonal Boron Nitride based Resistive memory devices”, 80th Device Research Conference (DRC) 2022
4. Iijima, S. Helical microtubules of graphitic carbon. Nature 354, 56, 1991
5. D. Gilmer et al, "NRAM: a disruptive carbon-nanotube resistance-change memory." Nanotechnology 29.13 134003, 2018
6. Rueckes et al, Carbon nanotube-based nonvolatile random access memory for molecular computing. Science 2000, 289, 94
7. T. R. Durrant et al, Electrical Conductivity of Single-Walled Carbon Nanotube Junctions, Phys. Stat. Sol. 2200118, 16, 2022
8. D. Z. Gao et al, Multiscale Materials Modelling of Nanotube-based Devices, INTERSECT 2022
9. G. Bersuker et al, “Evaluation framework assessing memristor technologies for neural network implementations”, 15th International Conference on Modern Materials and Technologies, Perugia, 2022
10. J. Farmer et al, Combining measurements and modeling/simulations analysis to assess carbon nanotube memory cell characteristics, IEEE IRPS 2022; 11. G. Bersuker et al, "Neuromorphic Hardware: Optimizing Carbon Nanotube Cell Characteristics", MRQW-2021