Demonstrating Thermal Stability and Retention Reliability of 5-bit Multi-Level ReRAM for In-Memory Computing

Samarth Jain[1,2], Naveen Kumar N[2], Anubhav Mukarjee[2], Xuanyao Fong[1]*

[1][Department of Electrical and Computer Engineering, National University of Singapore, SINGAPORE]

[2][Beyond Moore Labs, USA]

*E-mail: [elefongx@nus.edu.sg]

Resistive Random-Access Memory (ReRAM) is a promising candidate for next-generation non-volatile memory and neuromorphic hardware. However, the practical deployment of ReRAM-based multi-level cells (MLCs) hinges on the stability and reliability of multiple resistance states. In this study, we demonstrate the thermal stability of 32 distinct resistance states in a 5-bit ReRAM device. We conducted a series of electrical tests, including forming resistance evaluation[1], multi-level state programming, and long-term retention analysis at room temperature and 125°C for 20 distinct commercial 130nm CMOS ASIC with ReRAMs. The Hardware setup is shown in **Fig 1**.

After forming, the ReRAM cells were successfully programmed into 32 well-separated resistance levels (S0 to S31). Each state was verified over multiple program/erase cycles, showing minimal overlap and high readout accuracy as shown in **Fig 2(a)**. Retention tests at room temperature for S0 and S31 demonstrated negligible drift over 100,000 seconds. Retention at elevated temperature (125°C) was further examined for S0 and S31. Both exhibited stable behavior with resistance drift remaining within safe margins[2] as shown in **Fig 2(b)**. This indicates the robustness of the memory states under thermal stress, essential for high-density, harsh-environment applications.

Compared to state-of-the-art multi-bit ReRAM implementations, which primarily emphasize increasing bit capacity per cell (e.g., up to 7 bits using gradual SET/RESET schemes), most prior studies have not rigorously investigated the retention characteristics of individual states under thermal stress. While these works showcase excellent memory windows and endurance at ambient conditions, they often lack thorough validation of each state's stability at elevated temperatures[3]. The relaxation behavior which effects just after programming cells is measured and two representative resistance states, S0 (lowest resistance) and S9 (intermediate resistance), measured over time at elevated temperature (125°C) is shown in **Fig 3**. Both states exhibit a characteristic two-phase retention profile: an initial rapid decrease in resistance within the first few minutes, attributed to the relaxation effect, followed by a long-term stabilization phase. The opposite relaxation trends observed in S0 and S9 are consistent with earlier reports. Prior studies show that low-resistance states (LRS) often relax upward in resistance due to partial filament rupture or vacancy back-diffusion, whereas intermediate states are more unstable and may drift downward as additional vacancies migrate into the filament, reinforcing conduction paths[4,5]. **Fig. 4** shows that nine ReRAM devices exhibit a significant spread in their programming latency, which affects both SET and RESET operations. This spread in pulse counts translates to a relatively low programming speed, making ReRAM more suitable for inference tasks, where the focus is on processing pre-programmed data, rather than for applications requiring fast, real-time programming of multiple states.

These results confirm that 5-bit ReRAM can reliably support 32 distinct resistance states with strong endurance and thermal stability. The study underscores its potential in future neuromorphic and in memory computing systems.

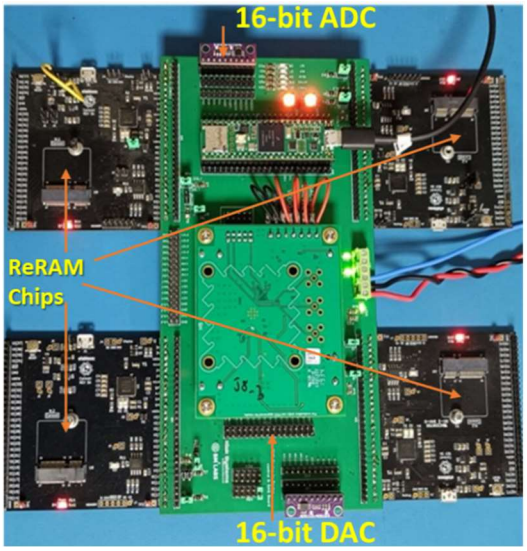
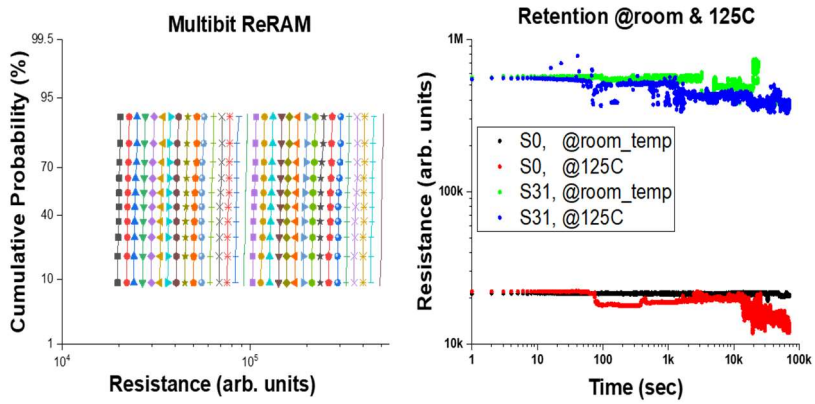


Figure 1. Hardware setup used for resistance state programming and retention testing.



(a) (b)

Figure 2(a). Cumulative Probability Distribution of Resistance for All 32 States in 5-bit ReRAM Cell. **Figure 2(b).** Retention of S0 and S31 state at both room temperature and 125°C.

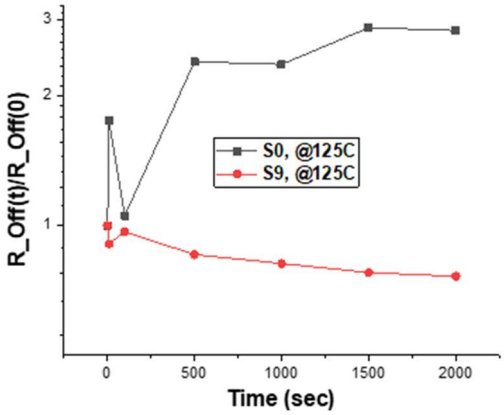


Figure 3. Median resistance relaxation of S0 and S9 states from four ReRAM chips at 125°C. Both states show initial rapid drift followed by stabilization.

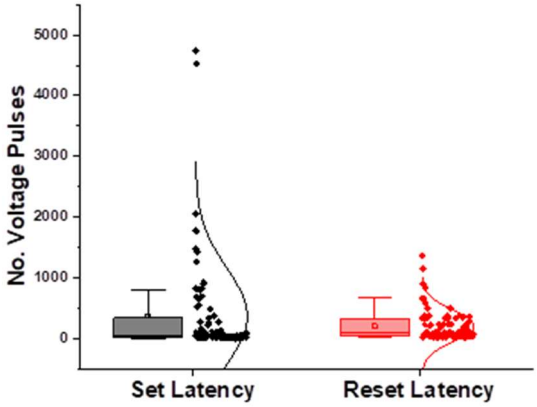


Figure 4. Latency distribution for SET and RESET operations in resistive switching devices. Box plots with overlaid scatter points show variability in the number of voltage pulses required, ranging from ~500 to 4000 pulse

References

[1] Hsieh et al., IEDM, 35.6, 2019.
[2] Bianchi, S. et al., *Nat. Commun.*, **14**, 1565, 2023.
[3] Mingyi et al., *Nature*, 615, 823–829, 2023.
[4] Wang, C. et al., IEEE Electron Device Lett., 37(2), 18
[5] He, W. et al., IEEE IRPS, 1–5, 2021.