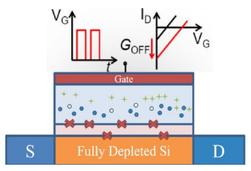
Long-term Retention and Analysis of CTT for In-memory (IM) Analog Compute

Siyun Qiao, Steven Moran, Zhe Wan, Sudhakar Pamarti, Subramanian S. Iyer, *UCLA*

Introduction

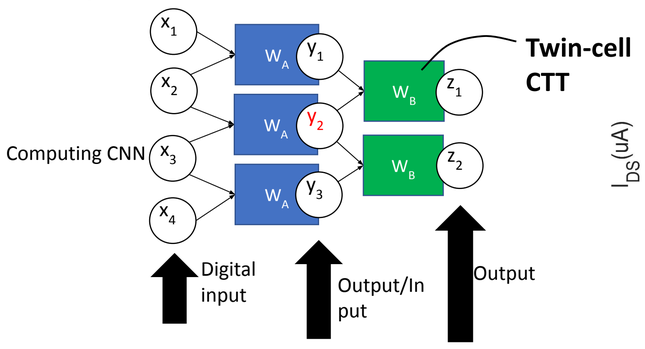
Background

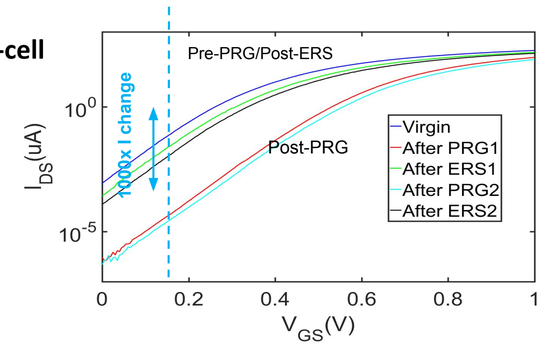
- CTT for in-memory analog compute engine
- Input data stored in CTT for arithmetic/logic operation



Evaluations of CTT as Analog NVM Device

- Range of current for data representation
- Noise-induced current fluctuations
- Data encoding precision
- Long-term retention characteristics

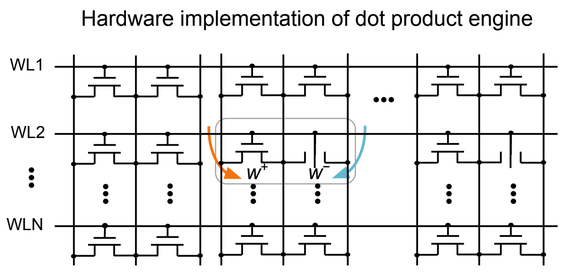




Target Current for Data Encoding

Application: NeuroCTT0p3

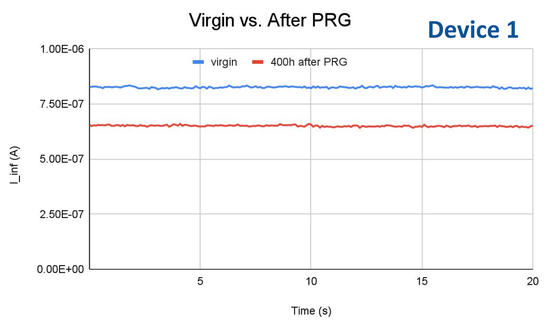
- Multiply-accumulate (MAC) operation for neural network inference
- Twin-cell architecture
  - Single device: (100, 700) nA
  - Twin-cell range: (-600, 600) nA



Noise-induced Current Fluctuations

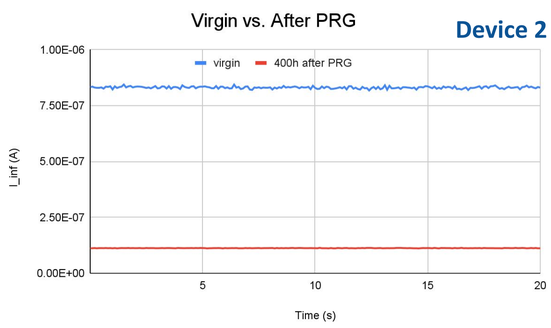
Root Causes

- 1/f noise and random telegraph noise (RTN)
- May exist other unknown sources of noise



Device 1

- 830 —program→ 600 nA
- Virgin state
  - ±9 nA fluctuation, or 1.1%
- 400 hours after program
  - ±9 nA fluctuation, or 1.4%



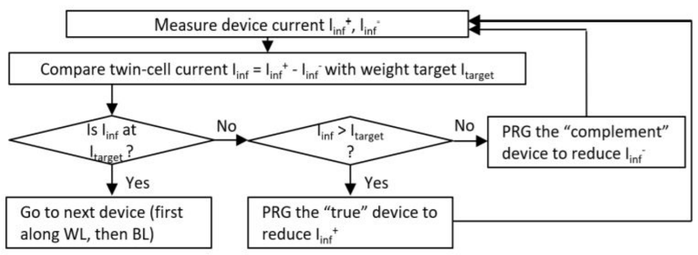
Device 2

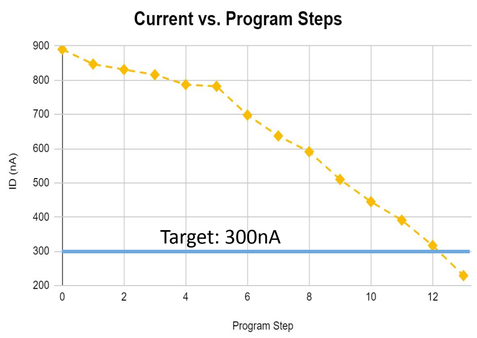
- 830 —program→ 100 nA
- Virgin state
  - ±13 nA fluctuation, or 1.6%
- 400 hours after program
  - ±1.5 nA fluctuation, or 1.3%

Data Encoding Precision

Uni-directional Programming

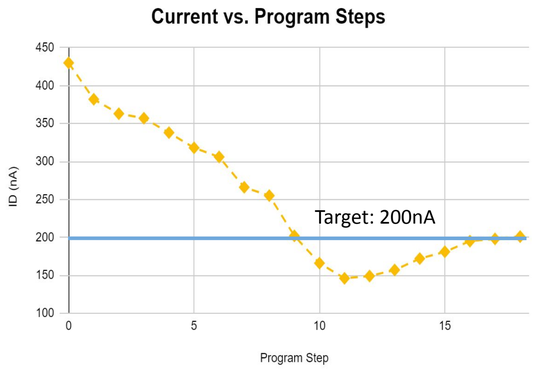
- Set target level
- Stop when reading exceeds target
- Lack fine control capability

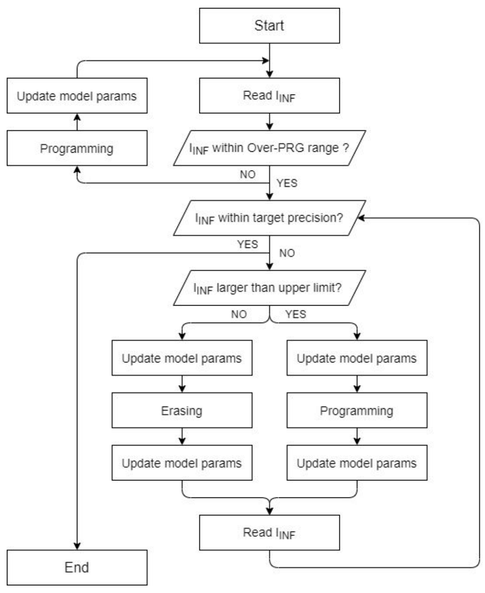




Bi-directional Programming

- Set target level
- Fine-tune current level by PRG/ERS
- Achieve high encoding precision
  - Up to 0.5%

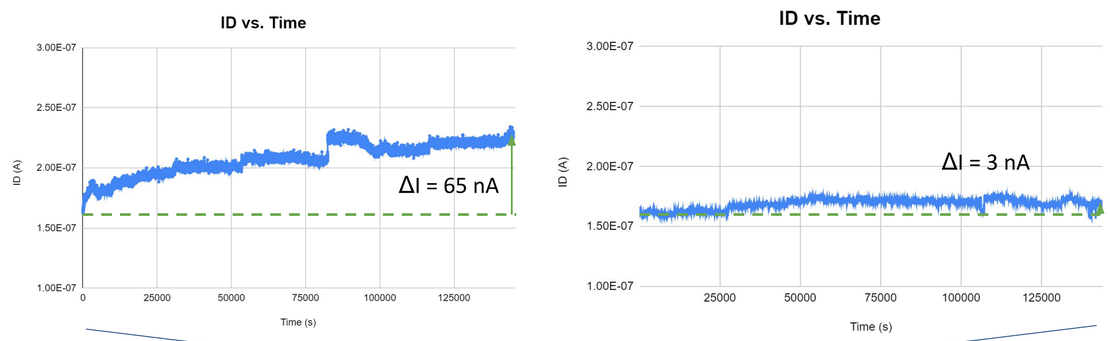




Long-term Retention Characteristics

Characteristics of Data Retention

- Large current drift in uni-directional programmed device
- Better long-term stability offered by bi-directional programming



40-hour measurement

A comparison between analog IM compute device candidates

Flash *vs.* **CTT** *vs.* **ReRAM** *vs.* **PCM**

	Flash	CTT	ReRAM	PCM
Material Property/ Technology Node	Floating Gate Transistor	Logic nFET (FDSOI)	Compatible with different tech nodes	Compatible with different tech nodes
Cell Configuration	2T (in two arrays)	2T	2T2R	1T1R
Target Resolution Per Cell	4-bit	> 3-bit	Binary (Analog: 3-bit)	Binary (Analog: 3-bit)
Dynamic Range	1 ~ 1000nA (<1% STD/mean)	-600 ~ 600nA (<1% STD/mean)	R_{on}/R_{off} <= 20	R_{on}/R_{off} <= 100
Operating Voltage	2 - 6V	0.2V	---	---
Data Retention	---	Up to 200h reported	Up to 10^4s (< 3h) reported	Up to 10^4s (< 3h) reported
Durability	Digital: > 10^4 P/E cycles	Digital: > 10^4 P/E cycles	---	Digital: > 10^4 P/E cycles

Future Work

- Design experiment for large-sample CTT long-term retention characterization
- Modeling retention characteristics with physics-based model

Conclusions

- Twin-cell configuration with (-600, 600) nA current range
- Minimum noise-induced current fluctuations
- High precision bi-directional data encoding scheme
- Encouraging data retention characteristics with bi-directional programming

This work is supported in part by the UCLA CHIPS Consortium, the Defense Advanced Research Projects Agency (DARPA), and UC Multicampus Research Programs and Initiatives (UC-MRPI). We would like to thank GlobalFoundries and ON-Semiconductor for their support.

Reference

1. X. Gu, *et. al*, "Charge-Trap Transistors for CMOS-Only Analog Memory," in IEEE Transactions on Electron Devices, Oct. 2019