

A DEEP DIVE INTO TAGE-SC-L

Alberto Ros

University of Murcia, Spain

Jun 21, 2025

MY MOTIVATION

- Understand TAGE-SC-L code in depth
- Focus on improving MPKI
- Long distance from key concepts to code
- The dive was indeed deep



Isadora Moon in a deep dive

THE END RESULT

- Very hard to improve MPKI
- So not such a deep dive in performance
 - (like trying to dive in the Dead Sea)
- But I hope some ideas are interesting



Me at the Dead Sea, ISCA'13

- PROLOG: A 192KB TAGE-SC-L
- ACT 1: History is not how it was told
- ACT 2: How to be more confident
- ACT 3: Be careful when feeding the beast
- CLOSING: Other optimizations and fine tuning

PROLOG: A 192KB TAGE-SC-L

- TAGE-SC-L is state of the art
 - So a 192KB TAGE-SC-L should be the baseline
- A new baseline: $\approx 192\text{KB}$ TAGE-SC-L
 - Bimodal table: $\times 8$ entries
 - TAGE tables:
 - $\times 2$ entries
 - From 10+20 (low+high history lengths) to 14+30
 - Extra bit for tags in the low tables (9 bits)
 - Loop predictor: $\times 2$ entries
 - Statistical corrector: $\times 2$ entries (except $\times 4$ for path history)
 - Other: NHIST=42, BORN=9, NNN=2, BORNTICK=512*3

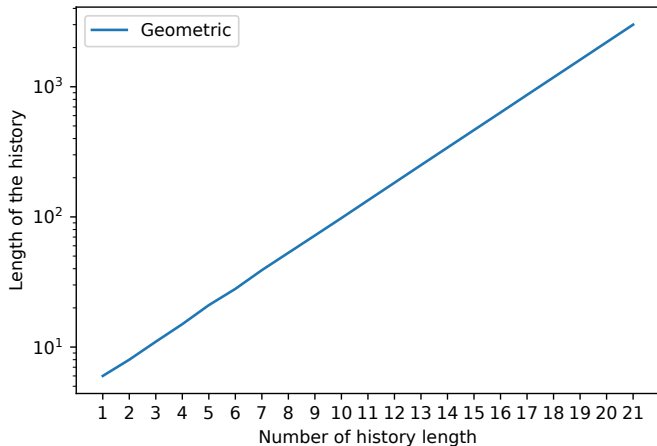
PROLOG: A 192KB TAGE-SC-L

- TAGE-SC-L is state of the art
 - So a 192KB TAGE-SC-L should be the baseline
- A new baseline: $\approx$ 192KB TAGE-SC-L
 - Bimodal table: $\times 8$ entries
 - TAGE tables:
 - $\times 2$ entries
 - From 10+20 (low+high history lengths) to 14+30
 - Extra bit for tags in the low tables (9 bits)
 - Loop predictor: $\times 2$ entries
 - Statistical corrector: $\times 2$ entries (except $\times 4$ for path history)
 - Other: NHIST=42, BORN=9, NNN=2, BORNTICK=512*3

3.4405 MPKI (146.1 CycWpPKI)

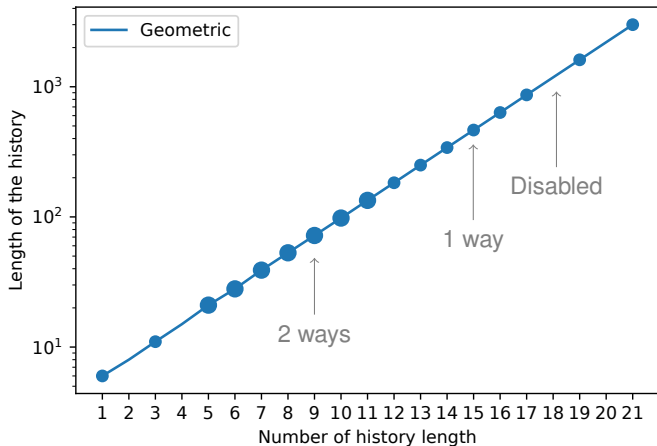
ACT 1: HISTORY IS NOT HOW IT WAS TOLD

- TAGE (and other predictors) use GEometric history lengths



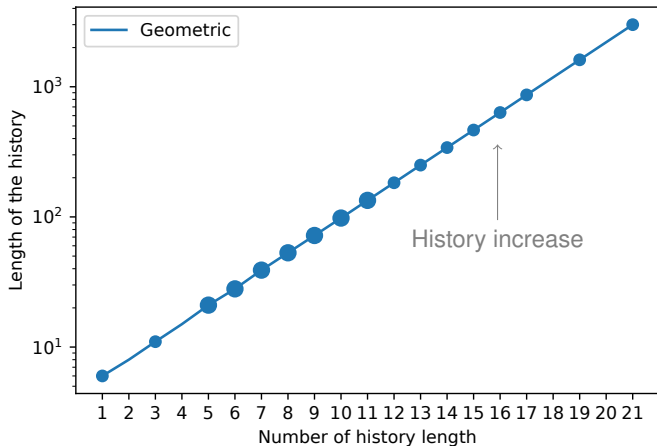
ACT 1: HISTORY IS NOT HOW IT WAS TOLD

- TAGE (and other predictors) use GEometric history lengths



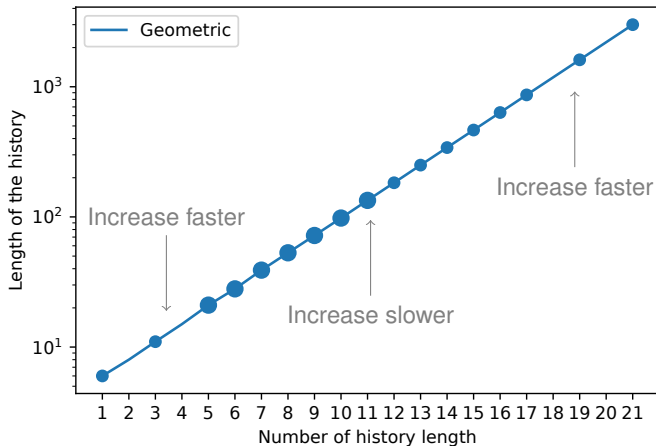
ACT 1: HISTORY IS NOT HOW IT WAS TOLD

- TAGE (and other predictors) use GEometric history lengths



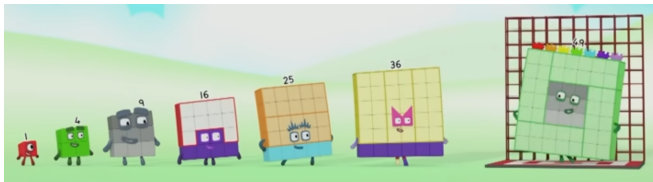
ACT 1: HISTORY IS NOT HOW IT WAS TOLD

- TAGE (and other predictors) use GEometric history lengths



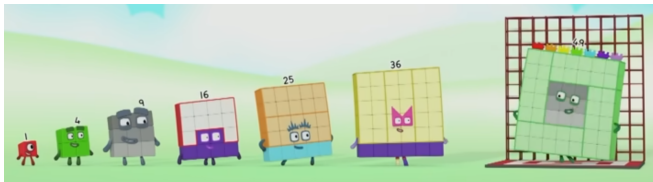
ACT 1: HISTORY IS NOT HOW IT WAS TOLD

- First approach: **Square** party!

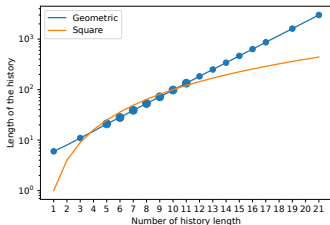


ACT 1: HISTORY IS NOT HOW IT WAS TOLD

- First approach: **Square** party!

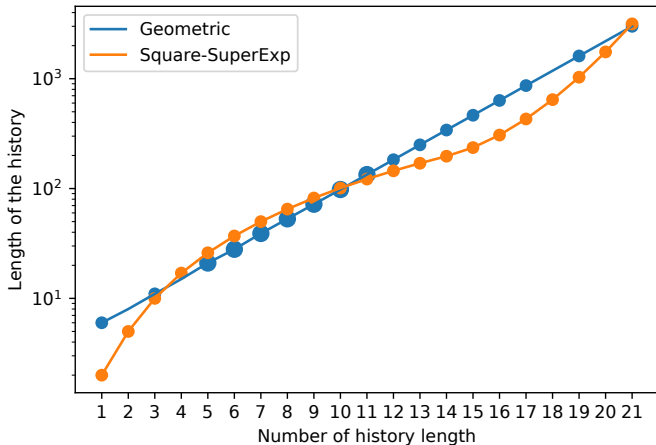


- Starts increasing faster than the geometric series but after some point it goes slower



ACT 1: HISTORY IS NOT HOW IT WAS TOLD

- Solution: At some point in the square sequence switch to a **generalized geometric** sequence



ACT 1: HISTORY IS NOT HOW IT WAS TOLD

- General formula for the proposed series

$$h_n = \begin{cases} h_1, & n = 1 \\ h_{n-1} + d \cdot (n - 1) + k, & 1 < n < t \\ \text{round}(h_{n-1} \cdot (f \cdot (n - t + 1) + m)), & n \geq t \end{cases}$$

ACT 1: HISTORY IS NOT HOW IT WAS TOLD

- General formula for the proposed series

$$h_n = \begin{cases} h_1, & n = 1 \\ h_{n-1} + d \cdot (n - 1) + k, & 1 < n < t \\ \text{round}(h_{n-1} \cdot (f \cdot (n - t + 1) + m)), & n \geq t \end{cases}$$

- Quadratic** sequences grow by a factor that increases by 2 (d) at each step

ACT 1: HISTORY IS NOT HOW IT WAS TOLD

- General formula for the proposed series

$$h_n = \begin{cases} h_1, & n = 1 \\ h_{n-1} + d \cdot (n - 1) + k, & 1 < n < t \\ \text{round}(h_{n-1} \cdot (f \cdot (n - t + 1) + m)), & n \geq t \end{cases}$$

- Quadratic** sequences grow by a factor that increases by 2 (d) at each step

+1

ACT 1: HISTORY IS NOT HOW IT WAS TOLD

- General formula for the proposed series

$$h_n = \begin{cases} h_1, & n = 1 \\ h_{n-1} + d \cdot (n - 1) + k, & 1 < n < t \\ \text{round}(h_{n-1} \cdot (f \cdot (n - t + 1) + m)), & n \geq t \end{cases}$$

- Quadratic** sequences grow by a factor that increases by 2 (d) at each step

	+3
	+1

ACT 1: HISTORY IS NOT HOW IT WAS TOLD

- General formula for the proposed series

$$h_n = \begin{cases} h_1, & n = 1 \\ h_{n-1} + d \cdot (n - 1) + k, & 1 < n < t \\ \text{round}(h_{n-1} \cdot (f \cdot (n - t + 1) + m)), & n \geq t \end{cases}$$

- Quadratic** sequences grow by a factor that increases by 2 (d) at each step

		+5
		+3
		+1

ACT 1: HISTORY IS NOT HOW IT WAS TOLD

- General formula for the proposed series

$$h_n = \begin{cases} h_1, & n = 1 \\ h_{n-1} + d \cdot (n - 1) + k, & 1 < n < t \\ \text{round}(h_{n-1} \cdot (f \cdot (n - t + 1) + m)), & n \geq t \end{cases}$$

- **Quadratic** sequences grow by a factor that increases by 2 (d) at each step
- **Generalized geometric** sequences have multipliers increasing by a factor (f)

ACT 1: HISTORY IS NOT HOW IT WAS TOLD

$$h_n = \begin{cases} h_1, & n = 1 \\ h_{n-1} + d \cdot (n - 1) + k, & 1 < n < t \\ \text{round}(h_{n-1} \cdot (f \cdot (n - t + 1) + m)), & n \geq t \end{cases}$$

- Parameters of the selected history
 - $h_1 = 2, d = 2, k = 1, f = 0.1, m = 1.1, t = 15$

ACT 1: HISTORY IS NOT HOW IT WAS TOLD

$$h_n = \begin{cases} h_1, & n = 1 \\ h_{n-1} + d \cdot (n - 1) + k, & 1 < n < t \\ \text{round}(h_{n-1} \cdot (f \cdot (n - t + 1) + m)), & n \geq t \end{cases}$$

- Parameters of the selected history

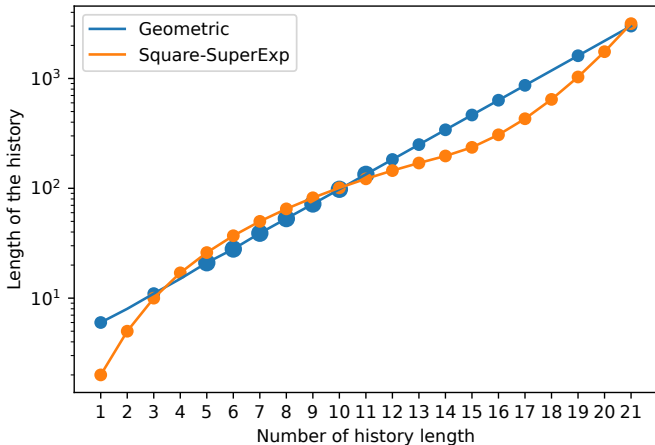
- $h_1 = 2, d = 2, k = 1, f = 0.1, m = 1.1, t = 15$
 - $d = 1?$



- Not that bad, but squares seem better than stairs :-)

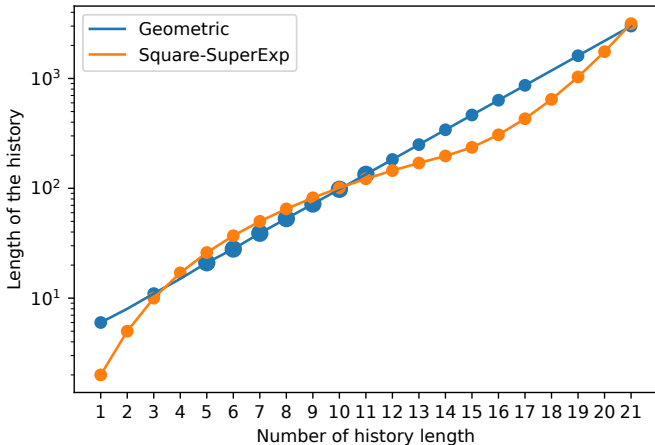
ACT 1: HISTORY IS NOT HOW IT WAS TOLD

- Just 1 way per history length! $\Rightarrow$ Simpler design



ACT 1: HISTORY IS NOT HOW IT WAS TOLD

- Just 1 way per history length! $\Rightarrow$ Simpler design



3.4276 MPKI (145.9 CycWpPKI)

ACT 2: HOW TO BE MORE CONFIDENT

- TAGE-SC-L has *choosers* to decide the prediction provider
- The more information the choosers have about the confidence of each component, the more accurate predictions are taken
- **There main choosers** in TAGE-SC-L
 - Alternate vs. hit bank prediction
 - Loop vs. TAGE (TAGE-L)
 - SC vs. TAGE-L



ACT 2: HOW TO BE MORE CONFIDENT

LOOP CHOOSER

- TAGE-SC-L discard loop predictions if SC is confident
 - But... the loop predictor is highly accurate
 - ⇒ The chooser should be **Loop vs. TAGE-SC**
- Loop confidence is binary
 - The threshold is $\text{loop iterations} \times \text{repetitions} > 128$
 - ⇒ Let's add **fine grain 0..7** (number of significant bits)
- Single chooser
 - ⇒ Let's add **more choosers**

Loop confidence	TAGE has max conf.	TAGE vs. SC
3 bits	1 bit	1 bit

ACT 2: HOW TO BE MORE CONFIDENT

ALTERNATE PREDICTION CHOOSER

- Alternate (alt) prediction
 - It is the second longest matching history length
 - Checked when the confidence of the longest (hit) prediction is weak (0)
 ⇒ Let's check it also for **medium confidence (1)**
- Alternate prediction uses 16 choosers
 - It uses the hit bank and the confidence of the alt prediction
 ⇒ Let's add **more choosers and confidence** information

Hit bank chunk	Alternate confidence	Alt bank hit	Hit conf.
3 bits	2 bits	1 bit	2 bit

ACT 2: HOW TO BE MORE CONFIDENT

STATISTICAL CORRECTOR (SC) PREDICTION CHOOSER

- SC prediction
 - Perceptron-like predictor
 - High priority over TAGE-L (but it also learns from TAGE)
 - Two choosers: medium vs. high
 - Otherwise TAGE only chosen if high confidence and SC low
- ⇒ Let's add **more choosers and confidence** information

TAGE confidence and alt chooser	SC sum bin in threshold
2 bits	2 bits

ACT 2: HOW TO BE MORE CONFIDENT

STATISTICAL CORRECTOR (SC) PREDICTION CHOOSER

- SC prediction
 - Perceptron-like predictor
 - High priority over TAGE-L (but it also learns from TAGE)
 - Two choosers: medium vs. high
 - Otherwise TAGE only chosen if high confidence and SC low
- ⇒ Let's add **more choosers and confidence** information

TAGE confidence and alt chooser	SC sum bin in threshold
2 bits	2 bits

3.4216 MPKI (145.7 CycWpPKI)

ACT 3: BE CAREFUL WHEN FEEDING THE BEAST

- TAGE is able to predict very difficult patterns
 - But it requires a lot of storage
 - ⇒ Entry should be allocated wisely

- The loop prediction has very high accuracy
 - TAGE-SC-L can allocate TAGE entries on loop hit-predictions
 - ⇒ Do **not allocate** in that case
 - ⇒ Simpler logic... and without using *random*

- Replacement policy (more work needed)
 - ⇒ Increase **useful bit** for the alt bank (on hit over miss)

ACT 3: BE CAREFUL WHEN FEEDING THE BEAST

- TAGE is able to predict very difficult patterns
 - But it requires a lot of storage
 - ⇒ Entry should be allocated wisely

- The loop prediction has very high accuracy
 - TAGE-SC-L can allocate TAGE entries on loop hit-predictions
 - ⇒ Do **not allocate** in that case
 - ⇒ Simpler logic... and without using *random*

- Replacement policy (more work needed)
 - ⇒ Increase **useful bit** for the alt bank (on hit over miss)

3.4156 MPKI (145.5 CycWpPKI)

CLOSING: OTHER OPTIMIZATIONS AND FINE TUNING

- Other optimizations
 - Fast loop direction change
 - Reset loop age on mismatch
- Fine tuning
 - Small IMLI SC table removed
 - SC bias hashes use TAGE (not TAGE-L) prediction
 - Increase SC second and third history-tracking tables

CLOSING: OTHER OPTIMIZATIONS AND FINE TUNING

- Other optimizations
 - Fast loop direction change
 - Reset loop age on mismatch

- Fine tuning
 - Small IMLI SC table removed
 - SC bias hashes use TAGE (not TAGE-L) prediction
 - Increase SC second and third history-tracking tables

3.4120 MPKI (145.4 CycWpPKI)

CONCLUDING REMARKS

- A number of simple optimizations to TAGE-SC-L
 - Do not add complexity to the branch predictor design
- New history length series: Quadratic-SuperExponential
 - Can be integrated in most modern branch predictors
- Let's give it a name
 - TASQ-SC-L (from Sequence Quadratic)

A DEEP DIVE INTO TAGE-SC-L

Alberto Ros

University of Murcia, Spain

Jun 21, 2025



ECHO, ERC Consolidator Grant (No 819134)