深圳市大数据研究院
Shenzhen Research Institute of Big Data

深圳国际工业与应用数学中心
SHENZHEN INTERNATIONAL CENTER FOR INDUSTRIAL AND APPLIED MATHEMATICS

NEURAL INFORMATION PROCESSING SYSTEMS

Constant Bit-size Transformers Are Turing Complete

Qian Li

SICIAM,

Shenzhen Research Institute of Big Data

Yuyi Wang

CRRC Zhuzhou Institute &

Tengen Intelligence Institute

Background

- Transformers (with CoT steps) can simulate Turing machines.
- But existing proofs required larger model to handle longer inputs.

Source	Precision	Dimension	Window length	COT / TM step
Perez et al. 2021	$O(\log t)$	$O(1)$	$O(t)$	1
Bhattamishra et al. 2020	unbounded	$O(1)$	$O(t)$	1
Merrill & Sabharwal, 2024	$O(\log t)$	$O(1)$	$O(t)$	1
Li et al. 2024	$O(1)$	$O(\log t)$	$O(t \log t)$	$O(\log t)$
Qiu et al. 2024	$O(\log t)$	$O(1)$	$O(t \log t)$	$O(\log t)$

- **Question**: Is it necessary to continue scaling up the bit-size (=precision*#parameters) of transformers to handle longer inputs?

Our Results

- **Question**: Is it necessary to continue scaling up the bit size of transformers to handle longer inputs?
- **Our answer**: No

Source	Precision	Dimension	Window length	COT / TM step
Perez et al. 2021	$O(\log t)$	$O(1)$	$O(t)$	1
Bhattamishra et al. 2020	unbounded	$O(1)$	$O(t)$	1
Merrill & Sabharwal, 2024	$O(\log t)$	$O(1)$	$O(t)$	1
Li et al. 2024	$O(1)$	$O(\log t)$	$O(t \log t)$	$O(\log t)$
Qiu et al. 2024	$O(\log t)$	$O(1)$	$O(t \log t)$	$O(\log t)$
This work	$O(1)$	$O(1)$	$O(s)$	$O(s)$

Our Result

- **WINDOW[$s(n)$]**: Decision problems solvable by a constant bit-size Transformer using $O(s(n))$-long window
- **SPACE[$s(n)$]**: Decision problems solvable by a Turing Machines using $O(s(n))$ space.

Theorem: **WINDOW[$s(n)$]**=**SPACE[$s(n)$]**.

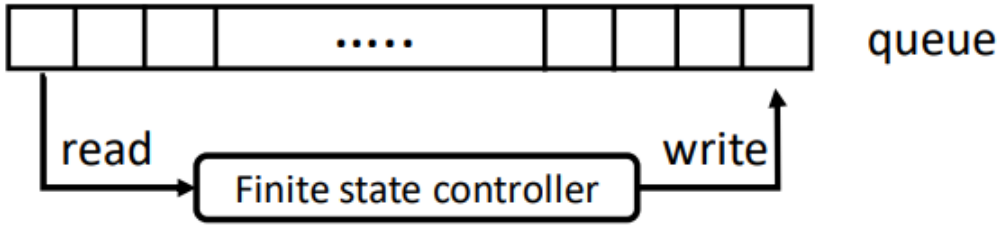
Messages:

- **Constant bit-size:** To handle longer inputs, it is unnecessary to continue scaling up the bit-size of Transformers.
- **WINDOW[$\mathbf{poly}(n)$]=PSAPCE**: Poly(n)-long windows suffice to solve SAT, Sokoban, etc.
- **General reasoning ability**: A single, constant bit-size transformer can compute any computable function, as long as the description of a relevant TM is loaded in the prompt.

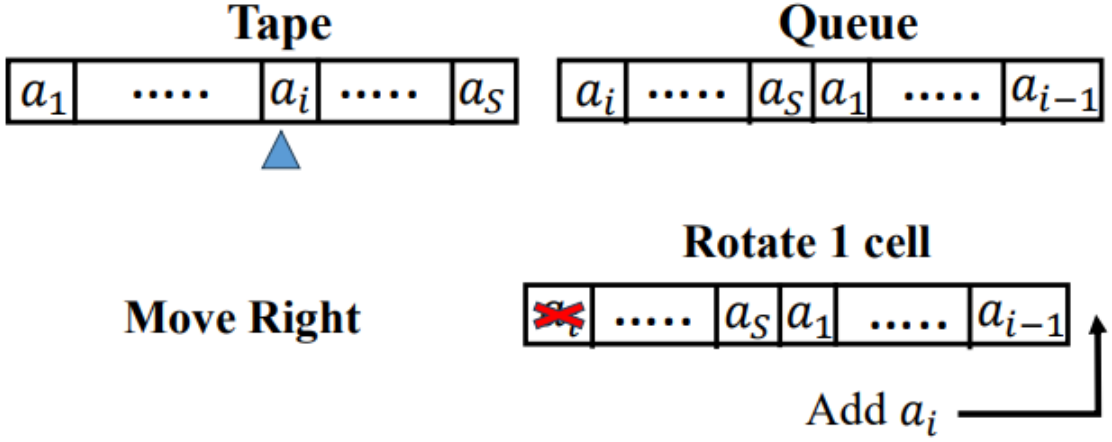
Proof Idea: TM ⇒ Post Machine ⇒ Transformer

What is Post Machine (PM)?

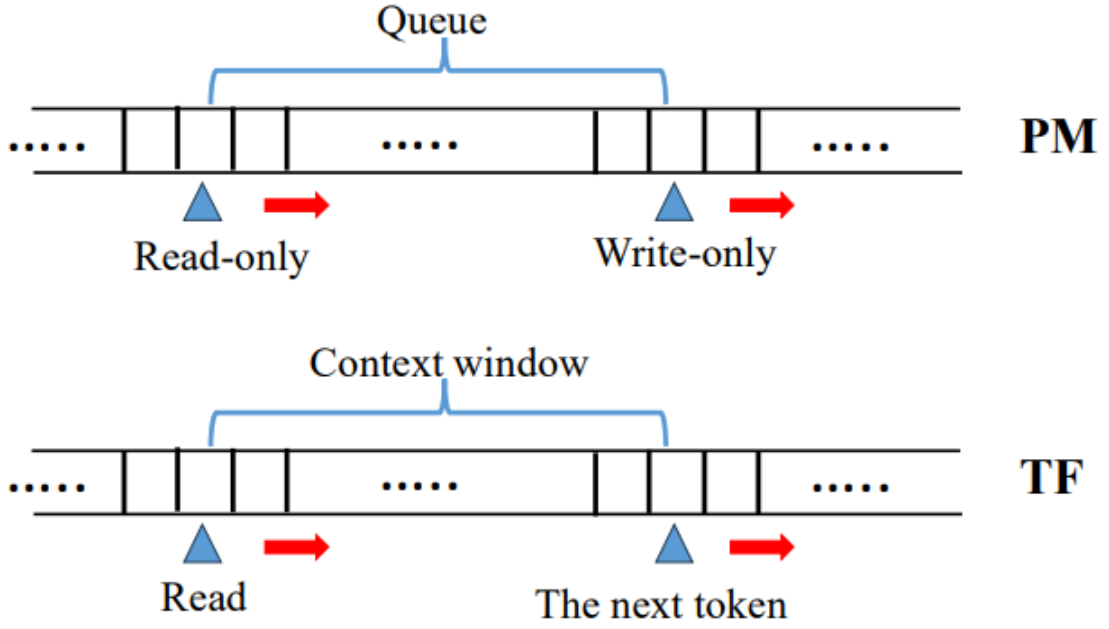
An automaton equipped with a queue.



Step I: (S,T) space-time TM ⇒ (S,T*S) space-time PM



Step II: (S,T*S) space-time PM ⇒ (S,T*S) window-cot TF



Discussion and Future Directions

- **Simulation Efficiency**
  - **Weakness 1**: our constructions requires $\Omega(s(n))$ CoT steps to simulate one TM step.
  - **Future direction 1**: whether the slowdown can be avoided.

- **Positional encodings**
  - **Weakness 2**: our construction employs a nonstandard relative PE. Moreover, it explicitly depends on the assumed space upper bound.
  - **Future direction 2**: whether it can be replaced with standard PEs. and it is unclear how our PE could be inferred automatically.