

Generalized Pattern Matching Micro-Engine

Yuanwei Fang*, Raihan Rasool[‡], Dilip Vasudevan*, Andrew A. Chien*[†]

University of Chicago *

Argonne National Laboratory[†]

King Faisal University[‡]



THE UNIVERSITY OF
CHICAGO

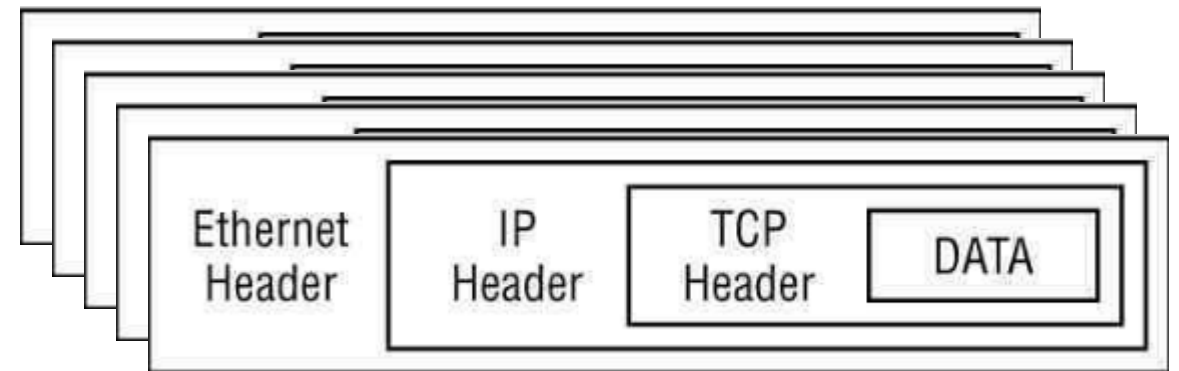


Big Data Applications

- **Deep Packet Inspection**
- **Bioinformatics (DNA Alignment)**
- **JSON/XML Parsing**
- **Signal Triggering**

Deep Packet Inspection

```
^USER\s+w0rm  
^GateCrasher\s+v\d+\x2E\d+\x2C\s+Server  
^[0-9]{1,5}\x00  
^[0-9]{1,5}\x00  
^CWD\s+~root  
^CEL(?!\n)\s[^\n]{100}  
^SITE\s+EXEC  
^PASS\s*\n  
^530\s+(Login|User)
```



High speed network : 100Gb/s

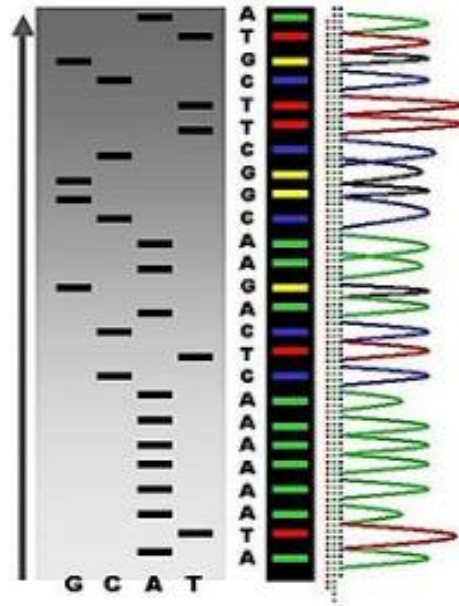
Growing number of patterns: 6000 Snort Rules

Speed requirement: > 75 Tera DFAops/s

Power budget : 200 W

Energy efficiency requirement: > 375Gops/J

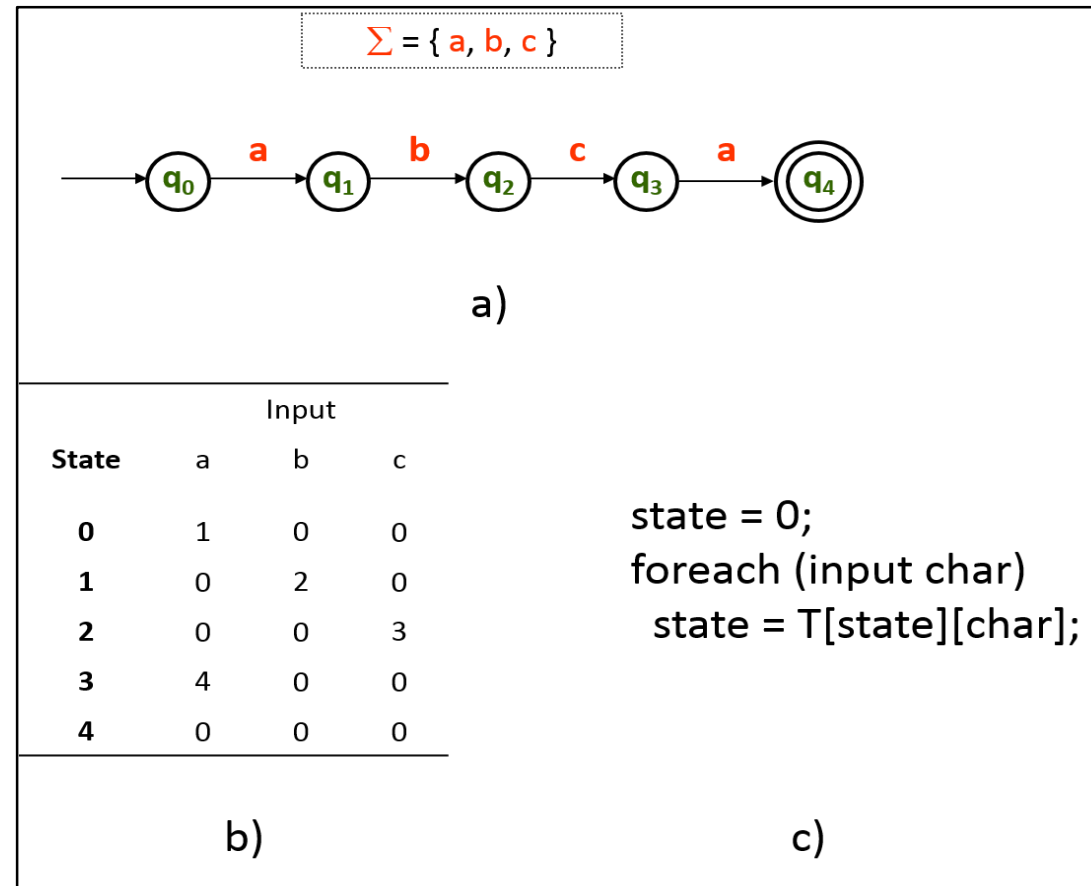
Bioinformatics (DNA Alignment)



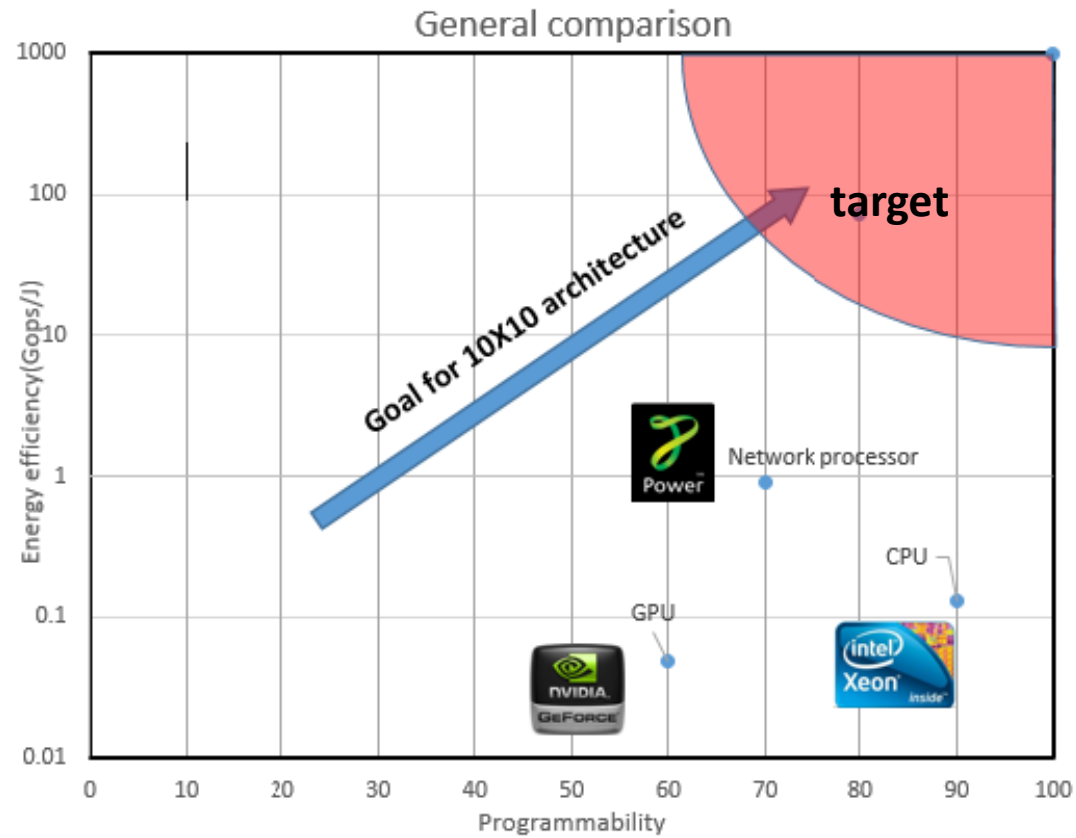
Bioinformatics database: millions of species
Speed requirement: > 1 Tera DFAops/s
Power budget : 200 W
Energy efficiency requirement: > 5 Gops/J

Genome size:
130G base pairs

Deterministic Finite Automata (DFA)



Programmable Approaches



Intel Xeon E5-2600: 17G DFAops/second with 130W, 0.13Gops/J ;

Approach

- **Workload**

M input characters(M DFA transitions)

N DFA rules perform on the M input characters

- **Goal**

Compute $N \times M$ transitions efficiently

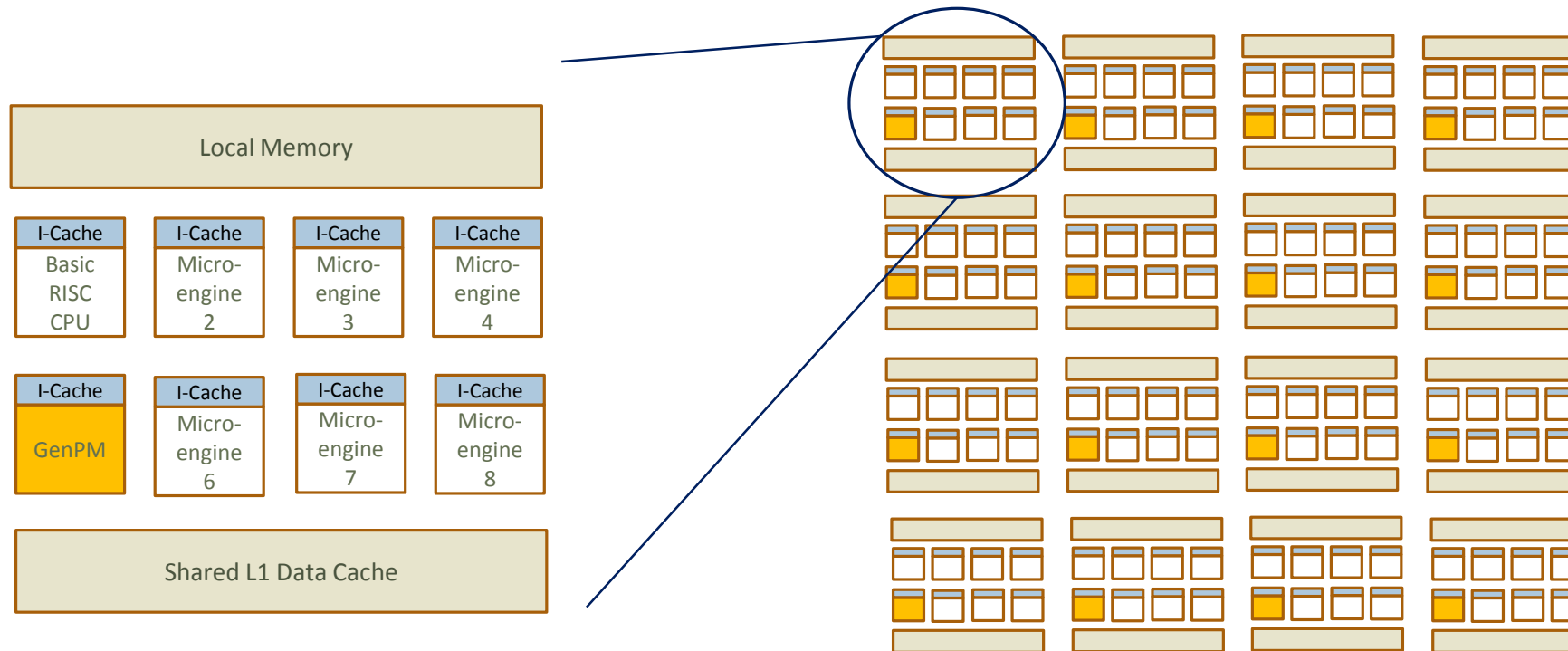
- **Approach**

Parallelize DFA execution

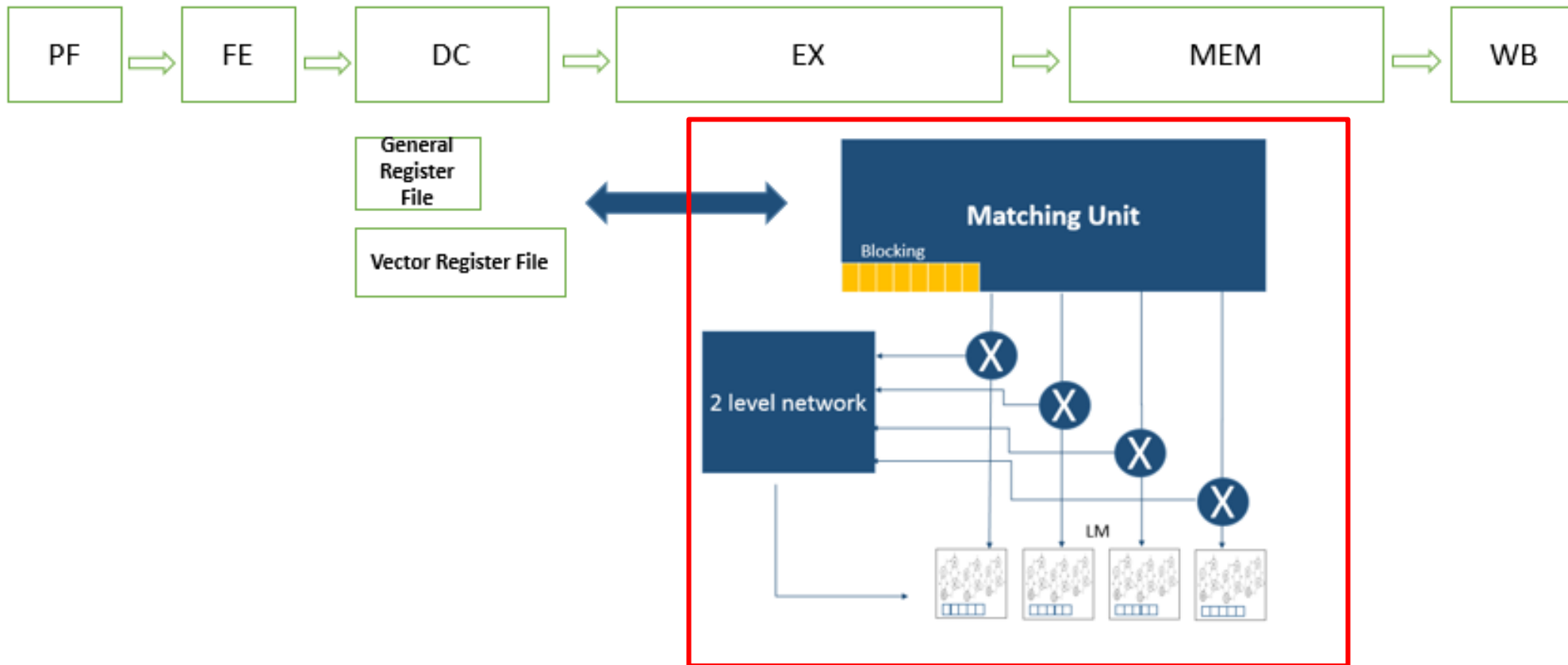
Fused Instruction

What Is Micro-Engine

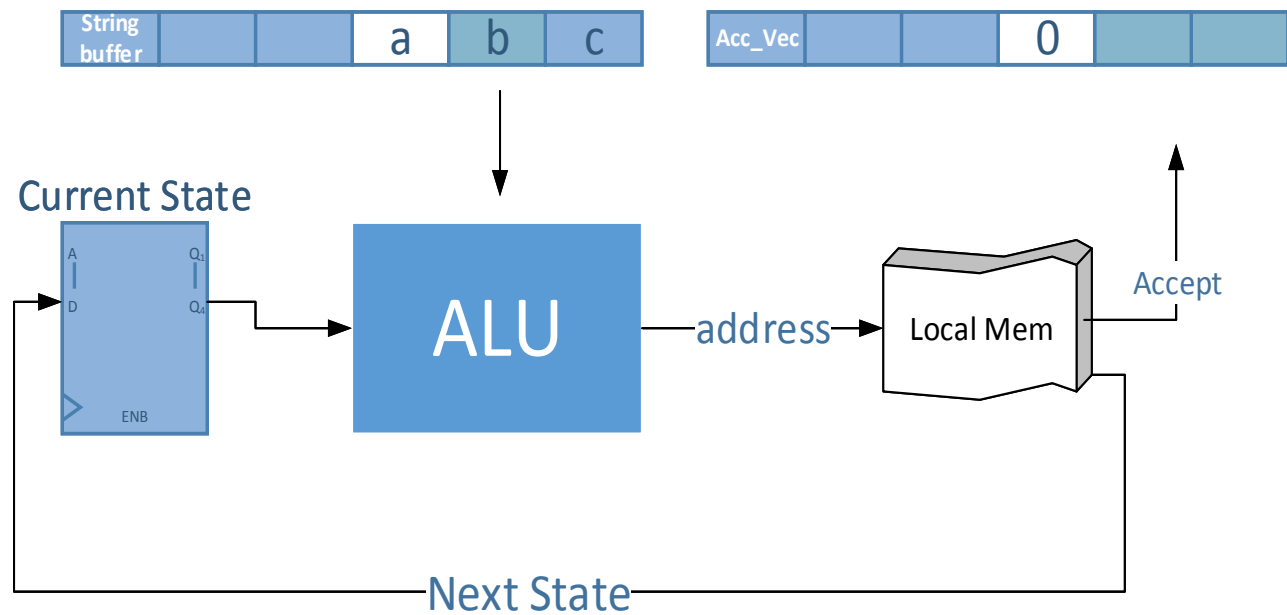
Generalized Pattern Matching Micro-Engine (GenPM) is one micro-engine of 10x10 approach



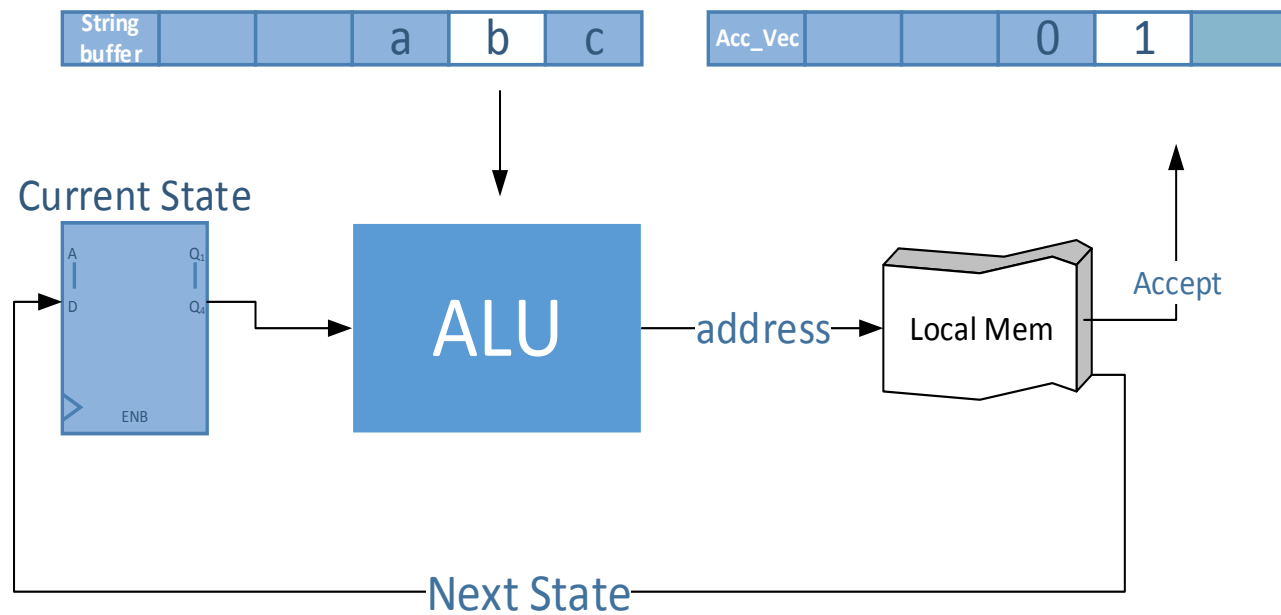
GenPM Micro Architecture



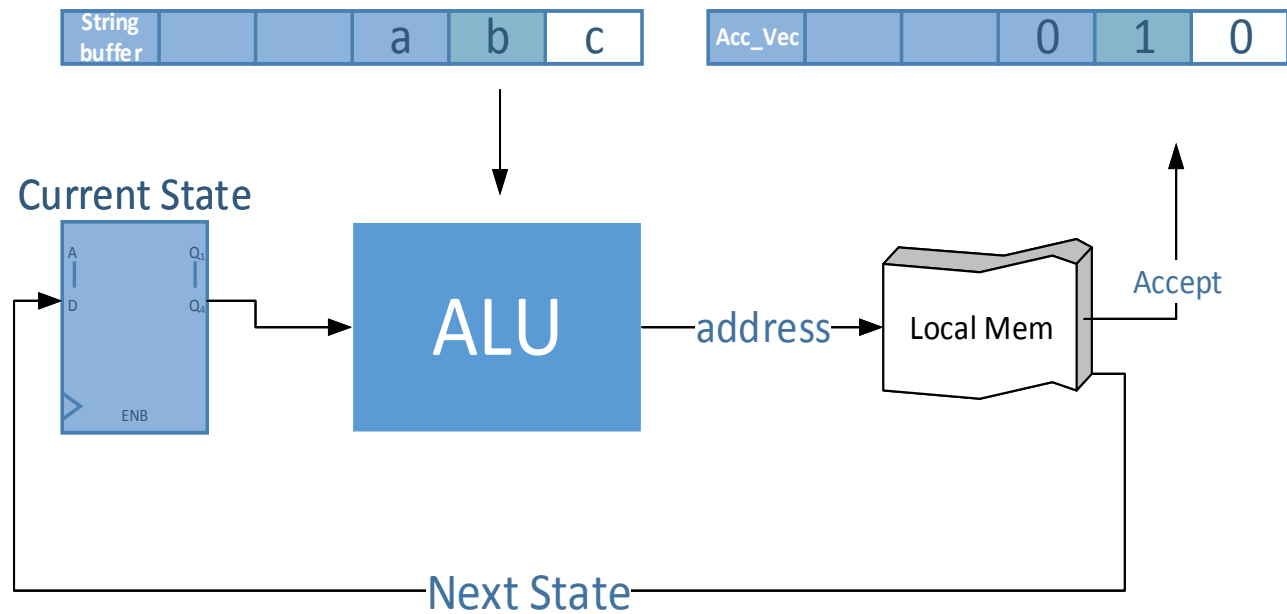
Fused Instructions: Multi-Step



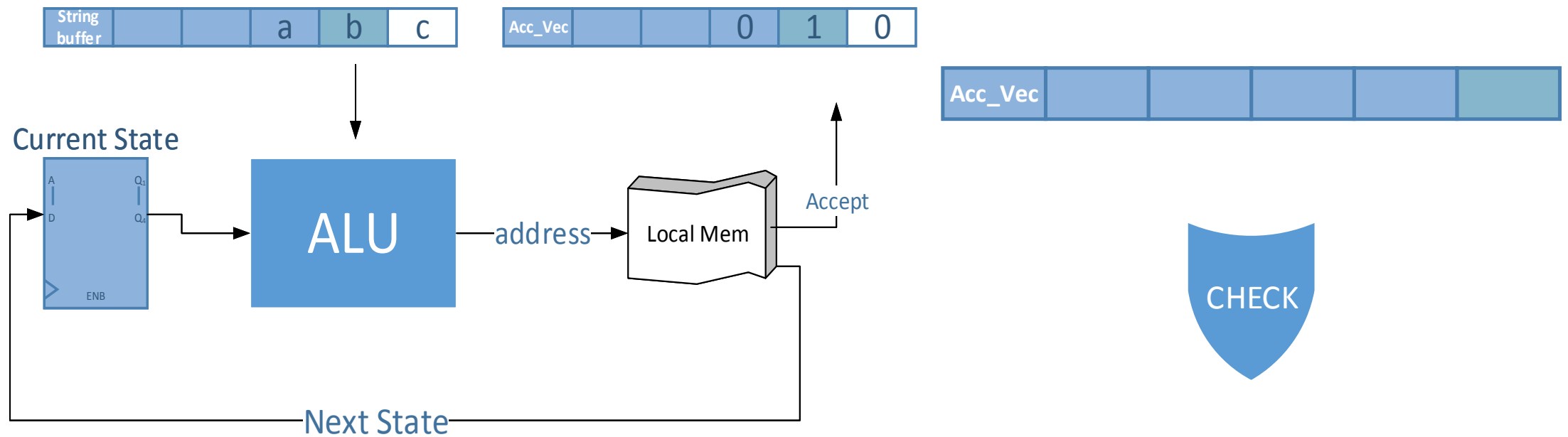
Fused Instructions: Multi-Step



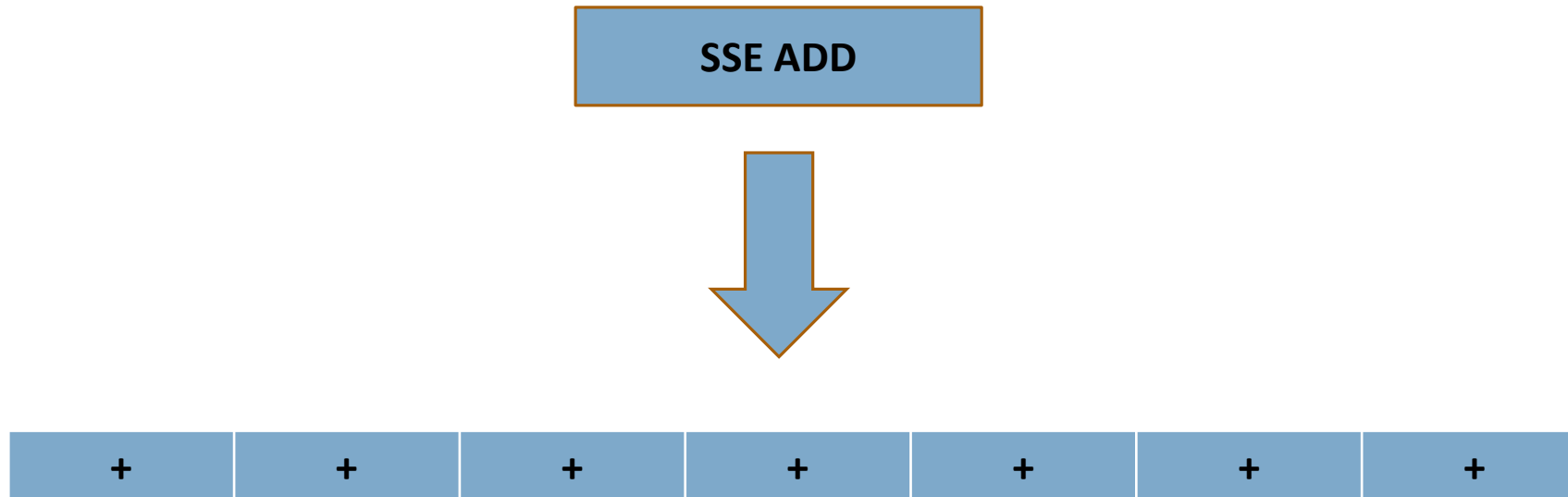
Fused Instructions: Multi-Step



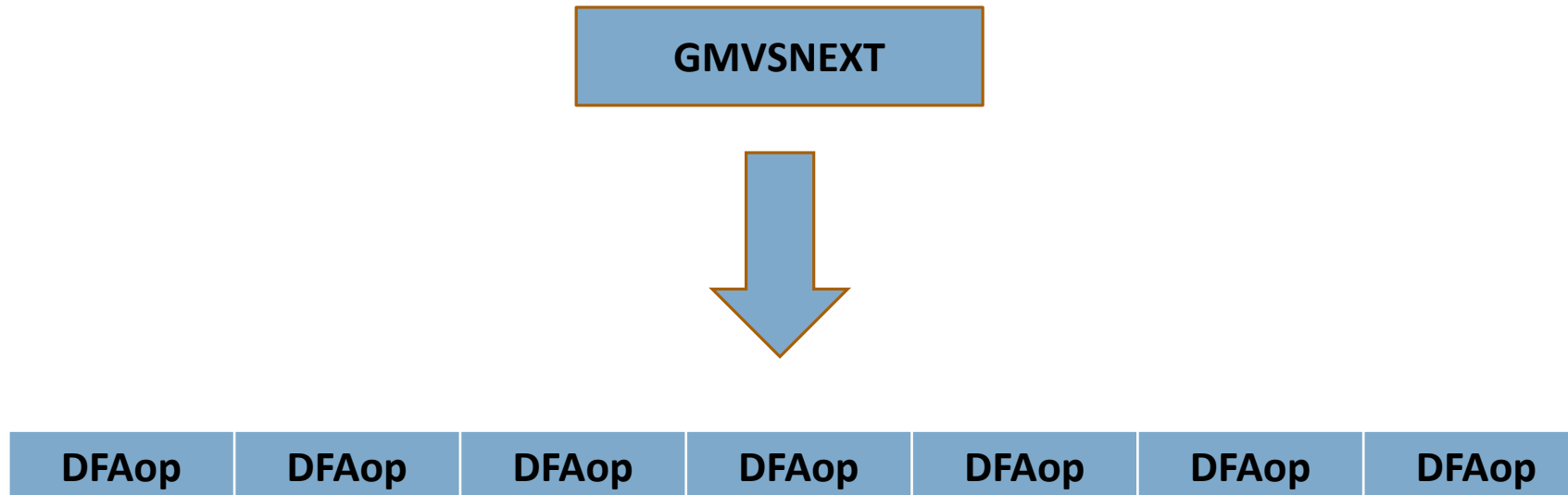
Fused Instructions: Multi-Step



Parallel DFA: Vector Instruction



Parallel DFA: Vector Instruction



GenPM Code Example

```
string_ptr += 128;  
BR_ptr = 0;  
current_len += 128;  
flag = 1;  
gmdbufld( string_ptr , vecbuf_len , od );  
od = !od;
```

Data movement

```
while( flag ){  
    gmvsnext( BR_ptr , od , step );  
    BR_ptr += 8;  
    if( BR_ptr < vecbuf_len ){ flag = 1; }else{ flag = 0; }  
    if( gmvsacc(0) ){
```

Multi-step parallel
DFA execution

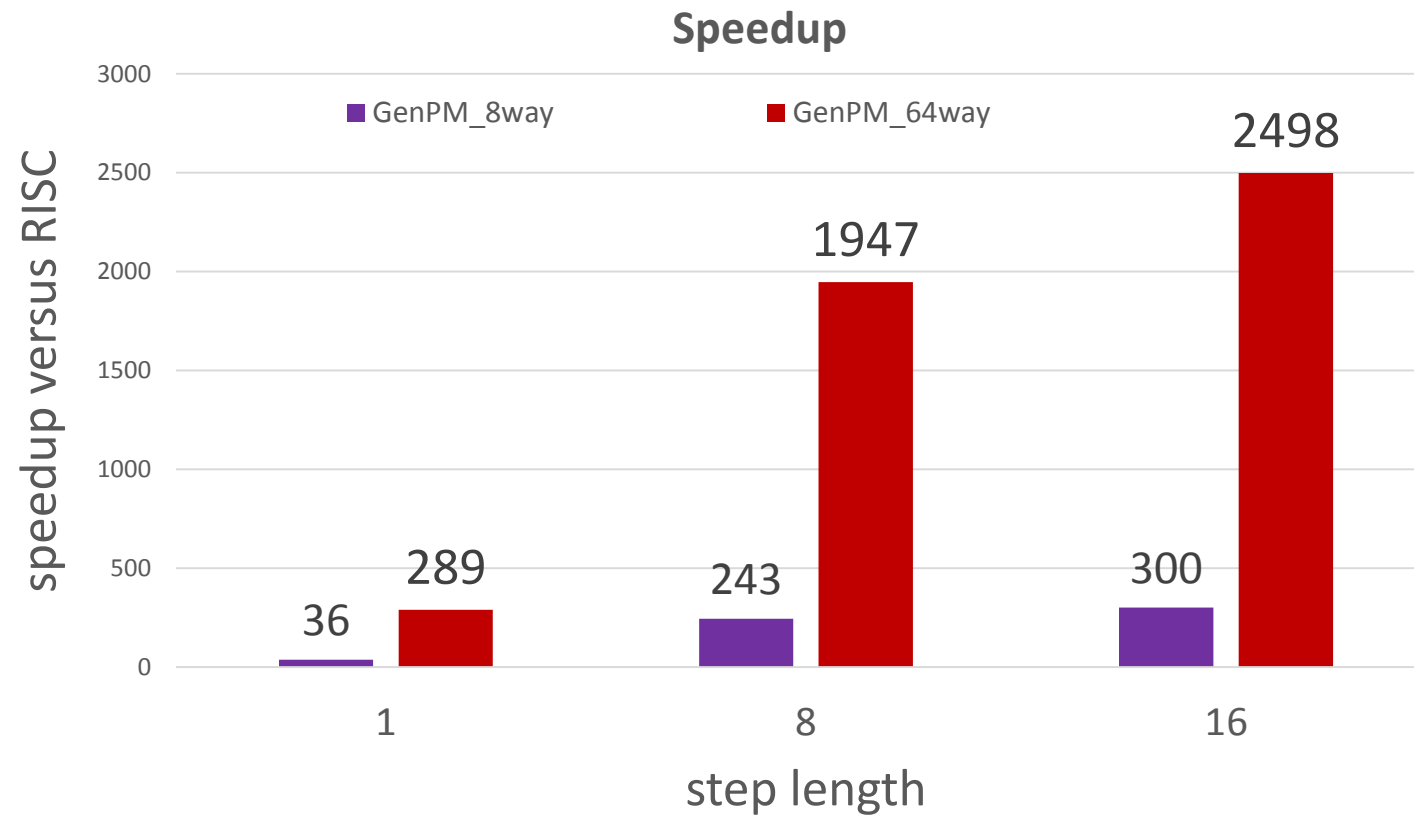
```
        gmstrideback();  
        BR_ptr = BR_ptr - 8;  
        step = 1;  
        for( k = 1; k <= 8 ; k++){  
            gmvsnext( BR_ptr , od , step );  
            BR_ptr++;  
            if( gmvsacc( 0 ) )  
                for( ck = 1; ck <= 8; ck++)  
                    if( gmvsacc( ck ) )  
                        printf("\nDFA[%d] at %d get %d accepts, longest_length=%d\n",ck,BR_ptr+current_len-128,gmvsacc(ck),gmvsent(ck-1)+1);  
        }  
        step = 8;  
    }  
}
```

Find precise matching
position

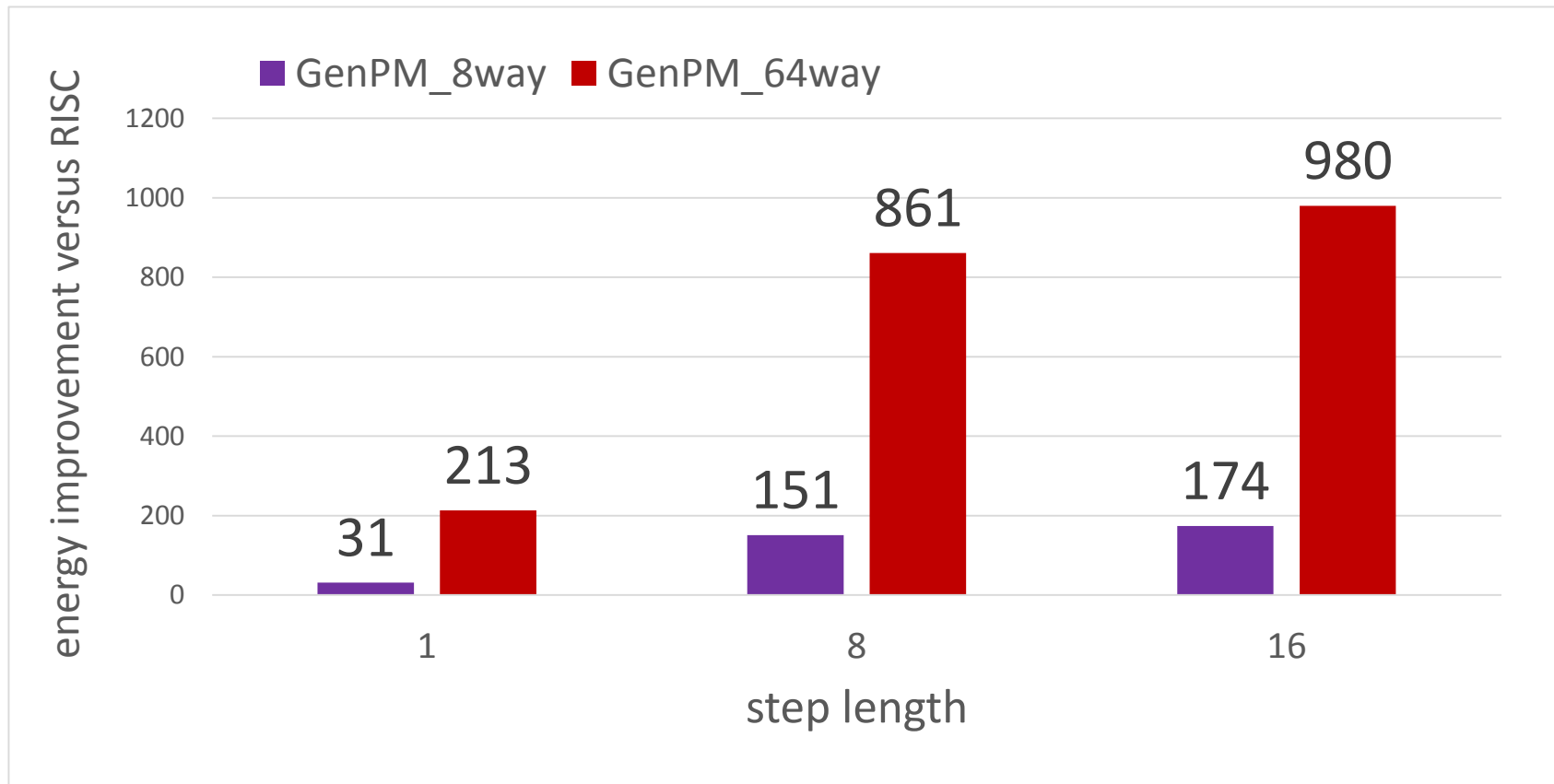
Methodology

- *Design space: Parallelism and step length*
- *Baseline*
 - *32-bit 6-stage in-order RISC*
 - *4GB DDR3 DRAM*
 - *32KB L1 I-cache, 24KB L1 D-cache, 512KB L2 (modeled on Intel Silverthorne)*
- *GenPM*
 - *1MB Local memory (up to 64 banks)*
 - *Vector and Fused Instructions*
- *Performance/Power Model*
 - *Core : 32nm synthesis by Synopsys Processor Designer*
 - *Memories : MARSSX86/CACTI 6 + DRAMSim2*
- *Workload*
 - *64 Snort rules from 2.9.5.6 snapshot, 10KB random network dump*

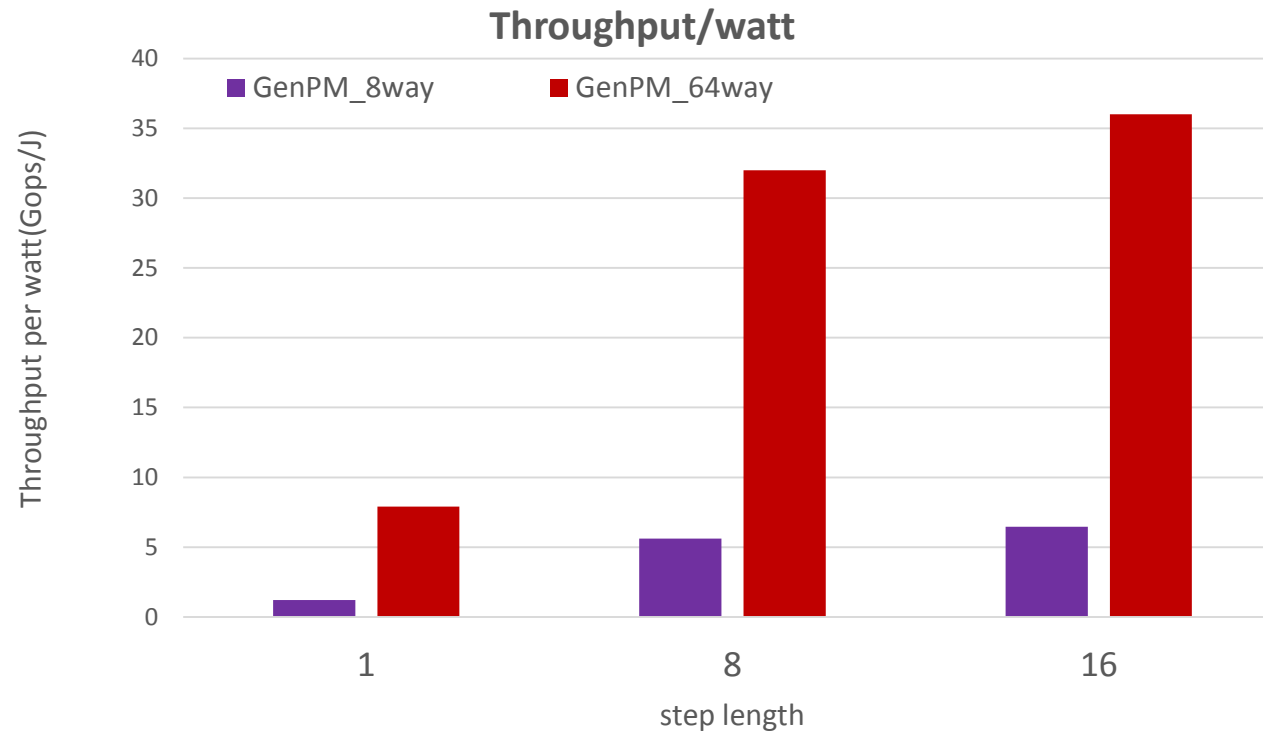
Performance



Energy Efficiency

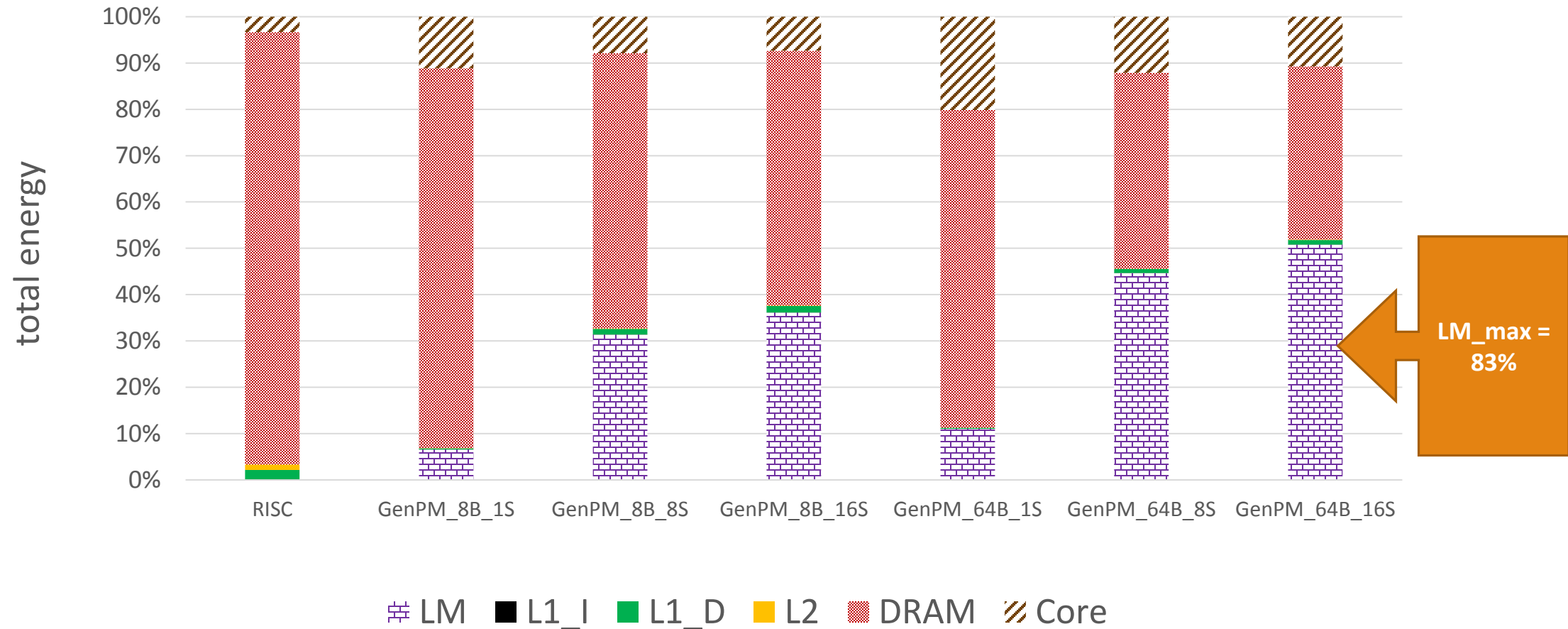


Throughput/watt (absolute)

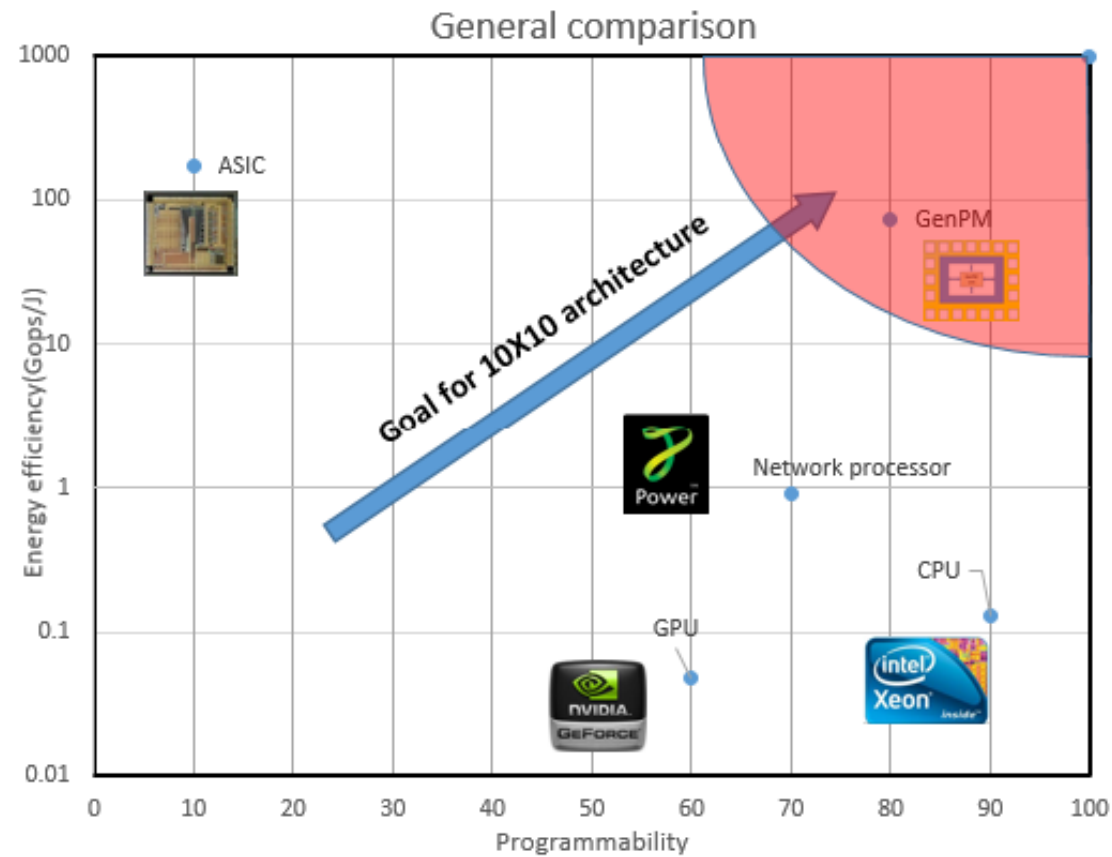


Scale to a 75W chip, GenPM delivers > 2.6 Tera DFAops/second

Energy Breakdown



General Comparison



Related Work

ASIC: [Brodie, et.al. ISCA 2006], [Titanic System RXP], [Cisco SCE]

FPGA: [Yang Xu, et.al. ANCS 2011], [T Song, et.al. INFOCOM 2008], [I Sourdis et.al. VLSI 2008]

CPU: [Mytkowicz et.al. ASPLOS 2014] , [Intel HyperScan]

GPU: [Vasiliadis G, et.al. CCS 2011], [Lin CH, et.al. INFOCOM 2012]

SoC: [C Johnson et.al. ISSCC 2010], [Cavium Octeon], [IBM PowerEN]

Summery

- *GenPM is a high performance and energy efficient accelerator for pattern matching workloads*
- *ISA exploits parallelism and multi-step execution*
- *Scale to a 75W chip, GenPM delivers > 2.6 Tera DFAops/second*
- *GenPM approaches ASIC efficiency and integrates it into a programmable core*

Future Work

- DFA table compression
- Scale up with multiple GenPM micro-engines
- Explore more applications

Acknowledgements

- Defense Advanced Research Projects Agency (DARPA)
- Agilent Technologies (now Keysight Technologies)
- Synopsys Academic program
- Dr. Tung Hoang and members of the Large Scale Systems Group in the Department of Computer Science