

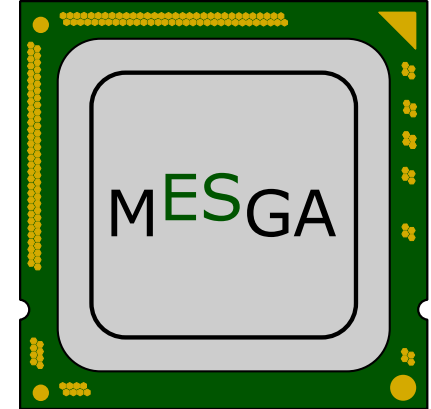
MESGA: An MPSoC Based Embedded System Solution for Short Read Genome Alignment

Vikkitharan Gnanasambandapillai
(Vikki.g@unsw.edu.au)
Arash Bayat
Sri Parameswaran
University of New South Wales, Australia



Agenda

- DNA tests
- Sequencer and current computational platforms
- Embedded system for genomic computation
- MESGA
- Future works & Conclusions



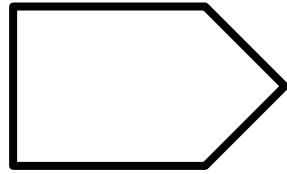
DNA Tests

- Four types bases A, C, G & T
- 3.2 billion base pair
- 99.9% similar
- DNA tests identify genetic markers



Read Genome Sequencing

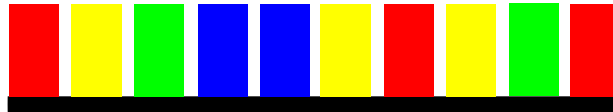
1. Sample preparation



2. Sequencing



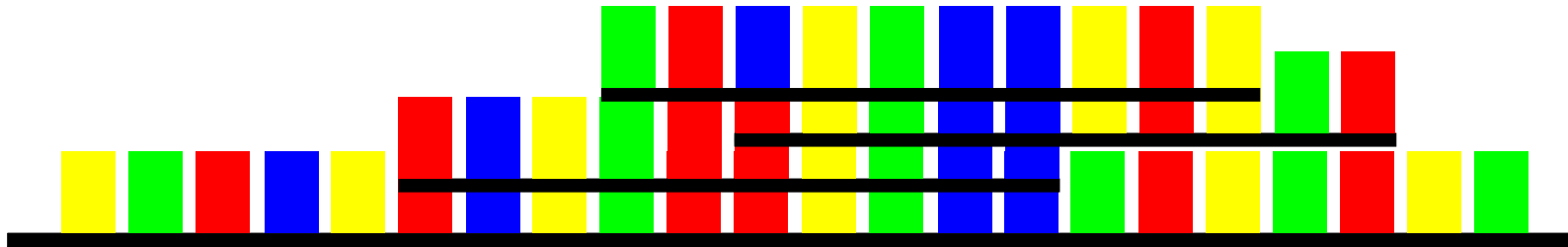
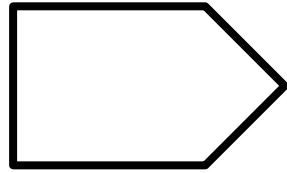
3. Computations



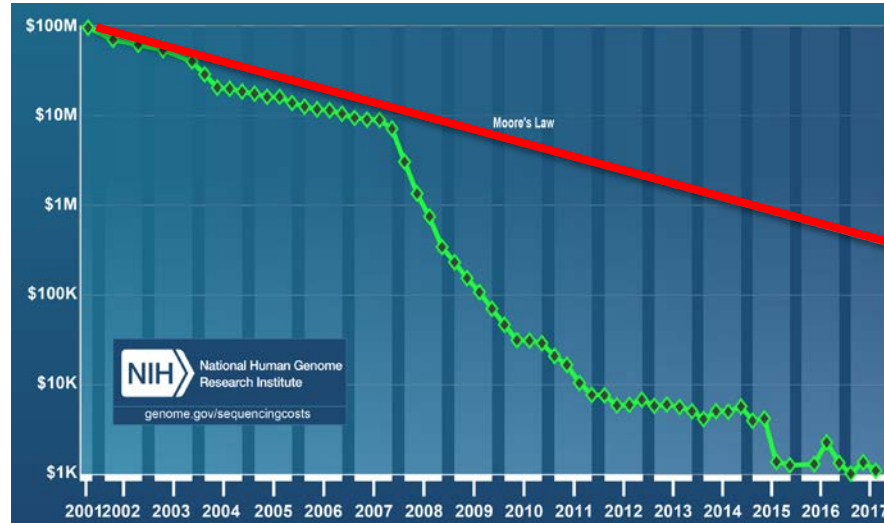
- “Short reads” or “long reads”

Short Read Genome Alignment

- Reference & target
- 30x coverage
- Huge raw data (50 TB)



Sequencers Vs Computers

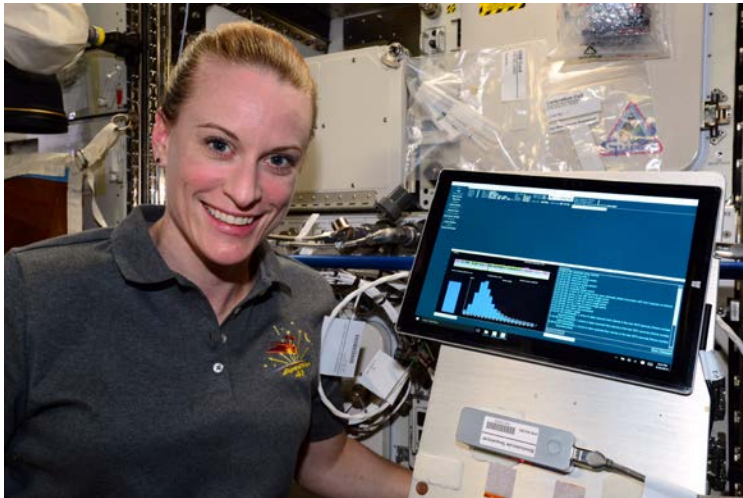


Enormous improvement		Moderate improvement
Small		Big
Portable		Remote cloud computers
Cheap		?
Fast		?

Remote Sequencing

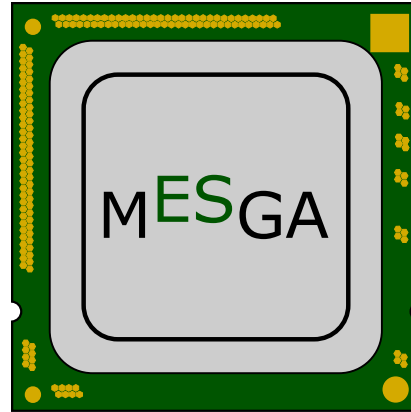


- Real-time, portable genome sequencing for Ebola surveillance
 - Identify infections agent & monitor spreading rate
 - Sequenced in west Africa and processed in UK
 - High speed satellite internet would cost 2000\$



- DNA Sequencing in Space Monitor
 - Astronauts' health & identify life beyond earth
 - Sequence in ISS and process on earth
 - Would take 15 months to diagnose

Sequencers vs Computers vs Embedded Processor



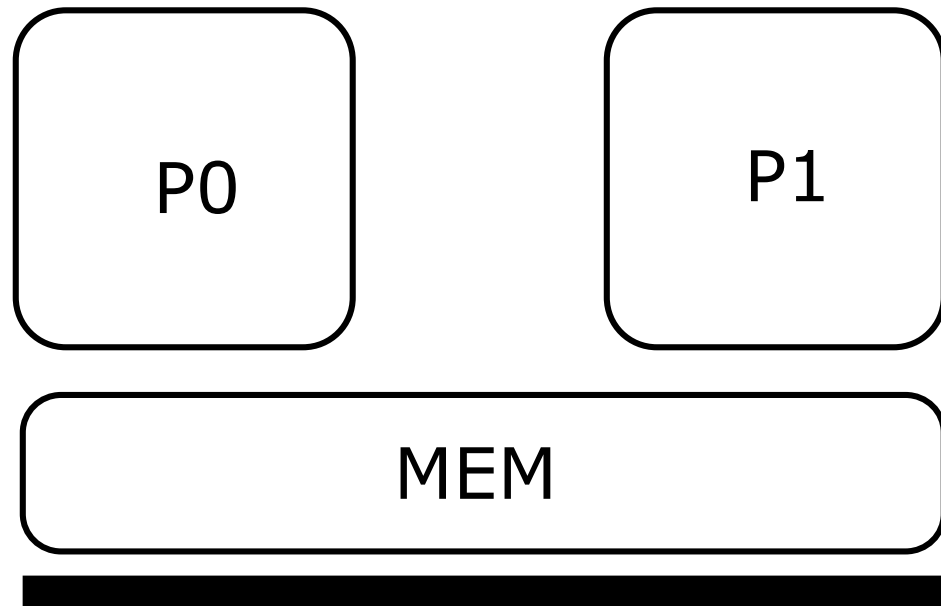
Exponential improvement		Linear improvement
Small	Small	Big
Portable	Portable	Remote
Cheap	Cheap	Expensive
Fast	Fast	Slow
	Secure	Unsecure
	Slower CPU	Faster CPU
	Smaller memory	Bigger Memory

MESGA

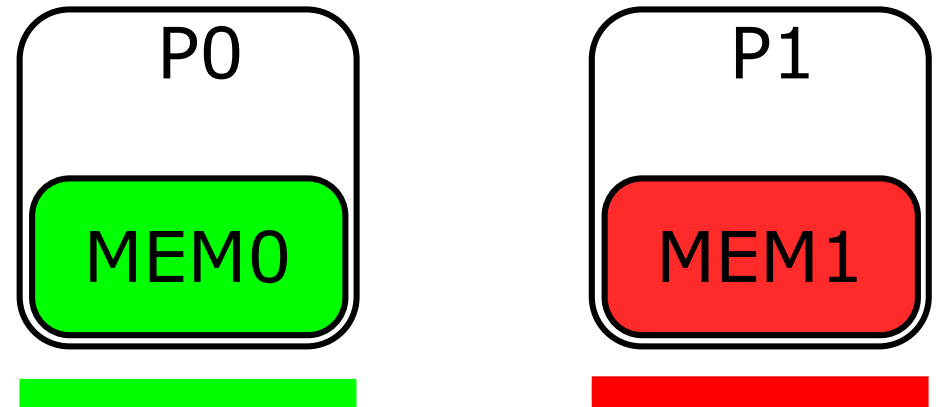
- Reference partition
- MPSoC pipelined architectures
- Processor
- Experiment setup & Results

Reference Partition

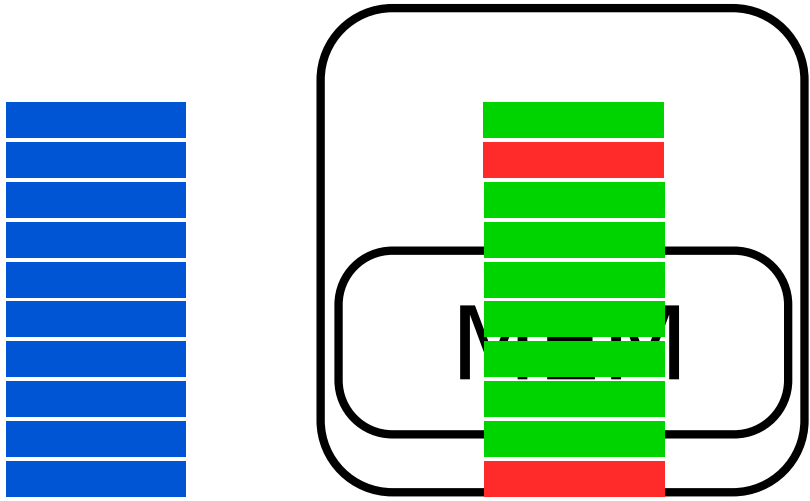
- Whole ref in shared memory
 - Memory bottleneck
 - Limited memory size



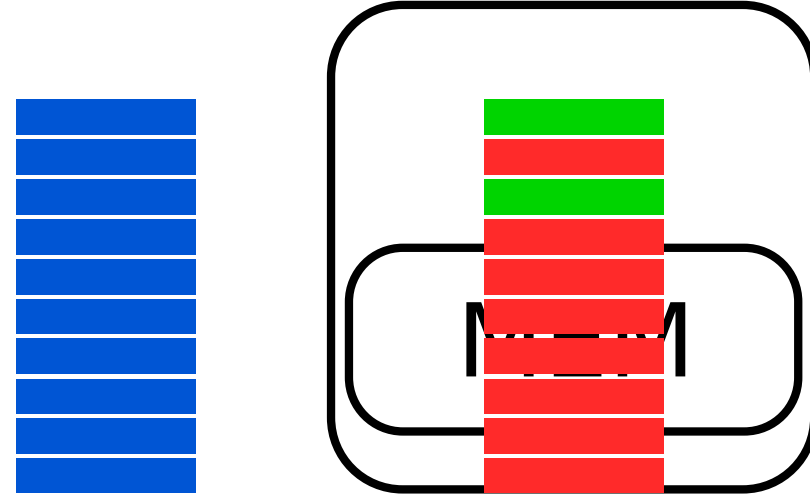
- Partial ref in local memory



Reference Partition Cont.



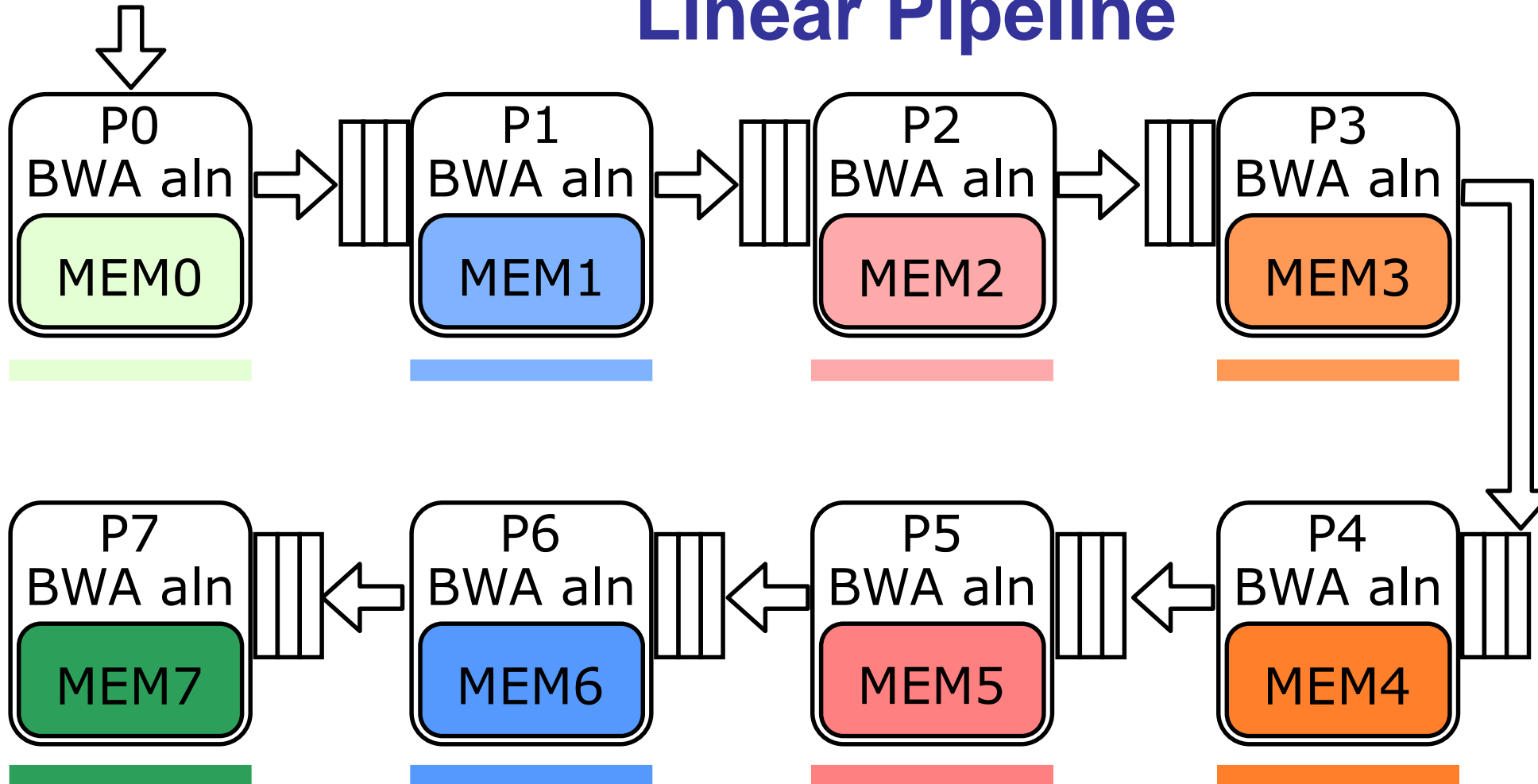
Whole ref



Partial ref

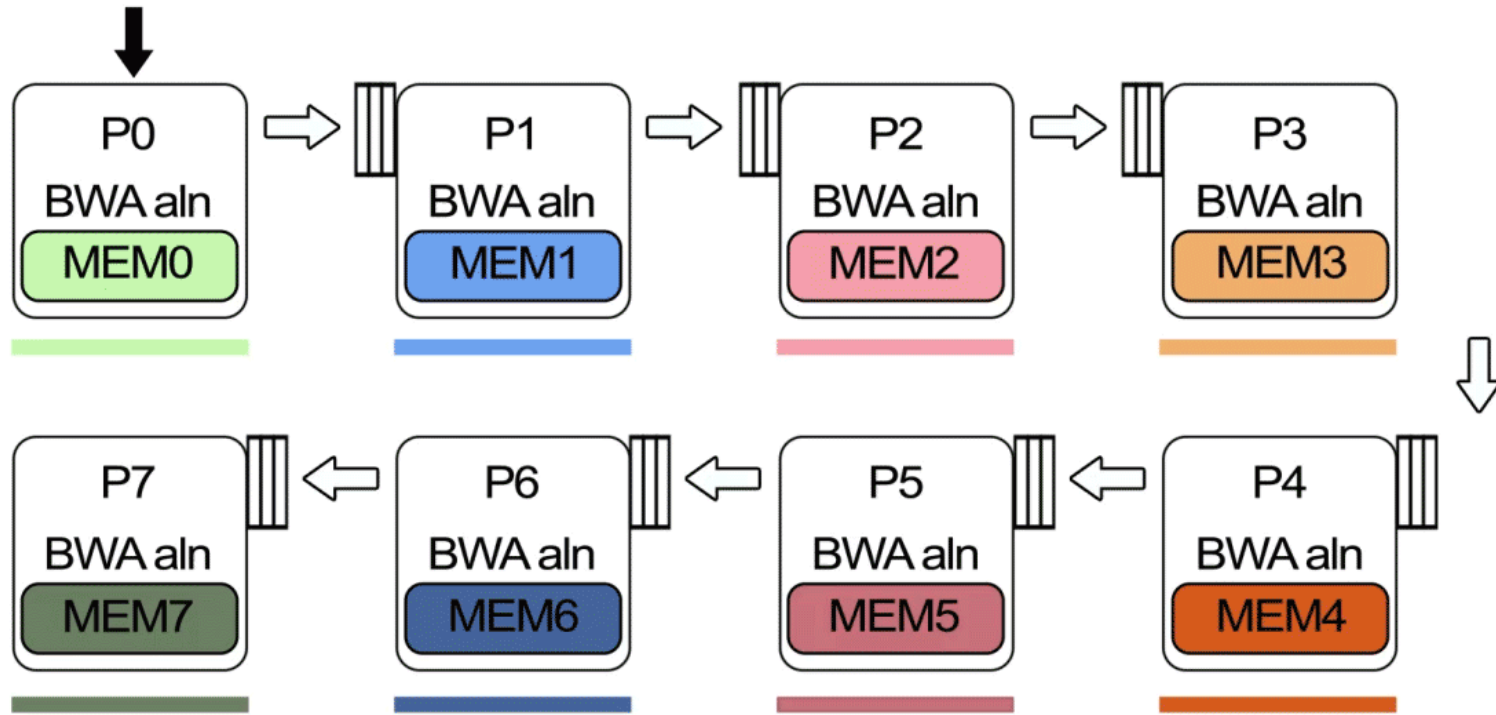
- Faster alignment
- Higher failure rate

Linear Pipeline

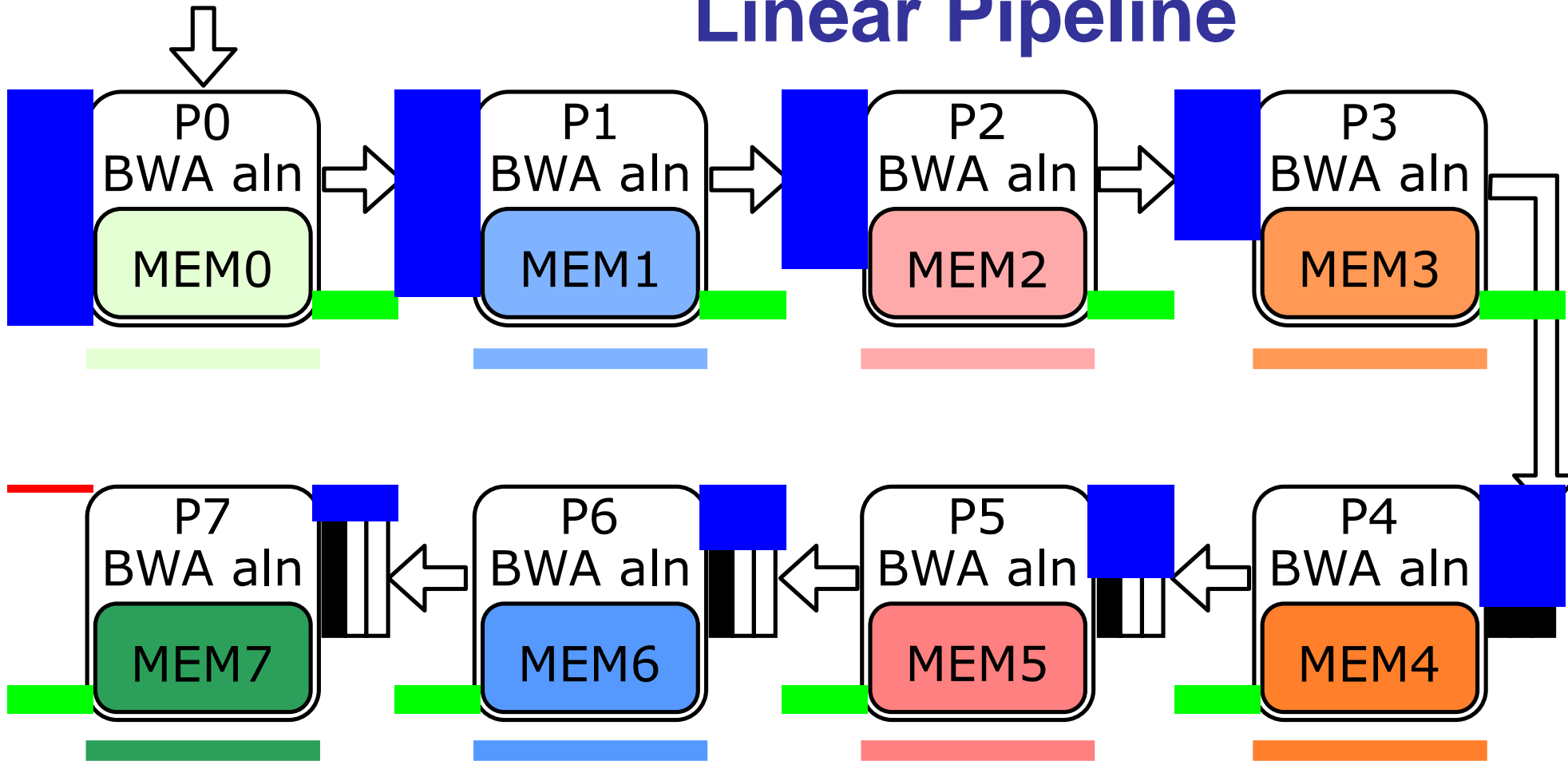


- Same processor
 - Except ref
- An external interface
- FIFO interfaces

Linear Pipeline

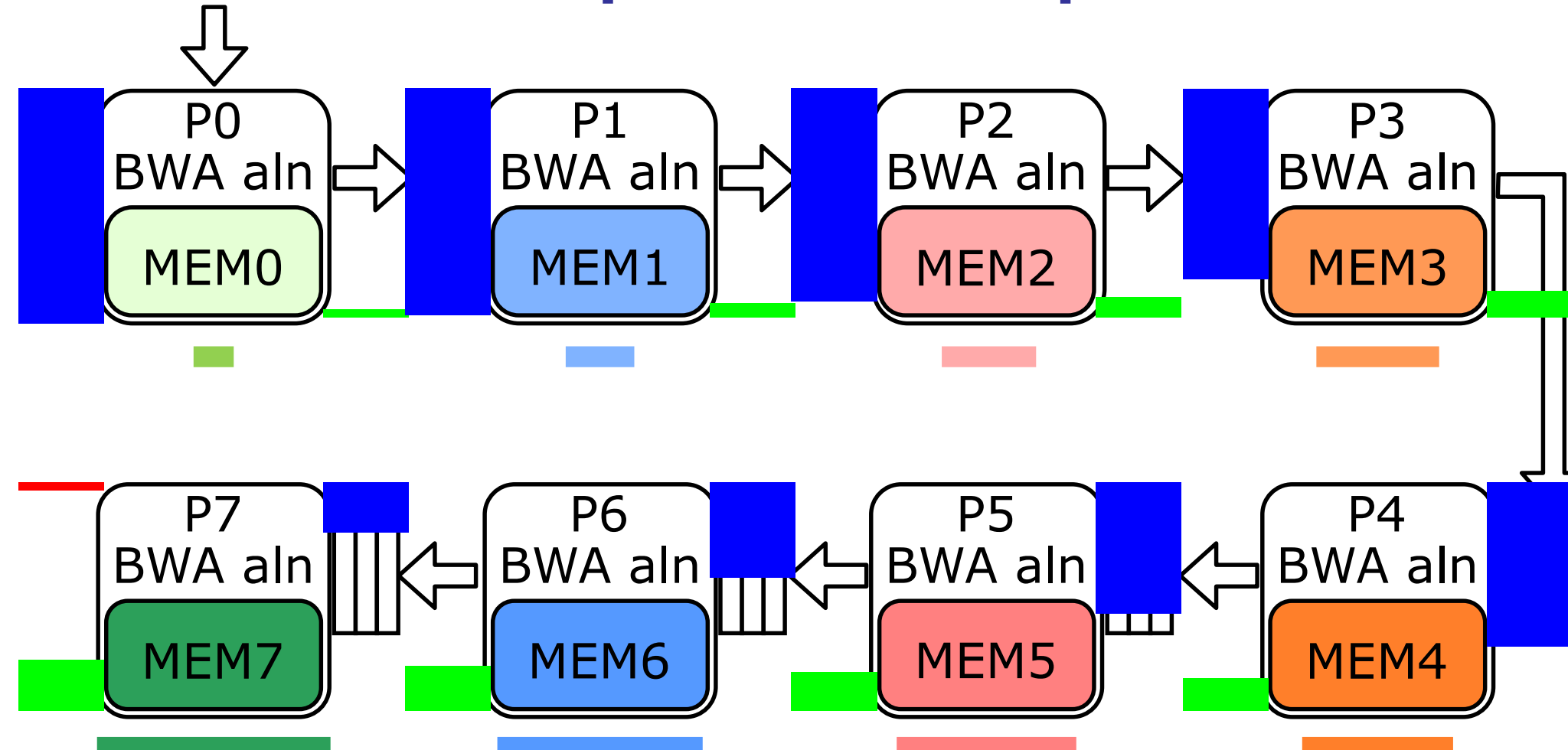


Linear Pipeline

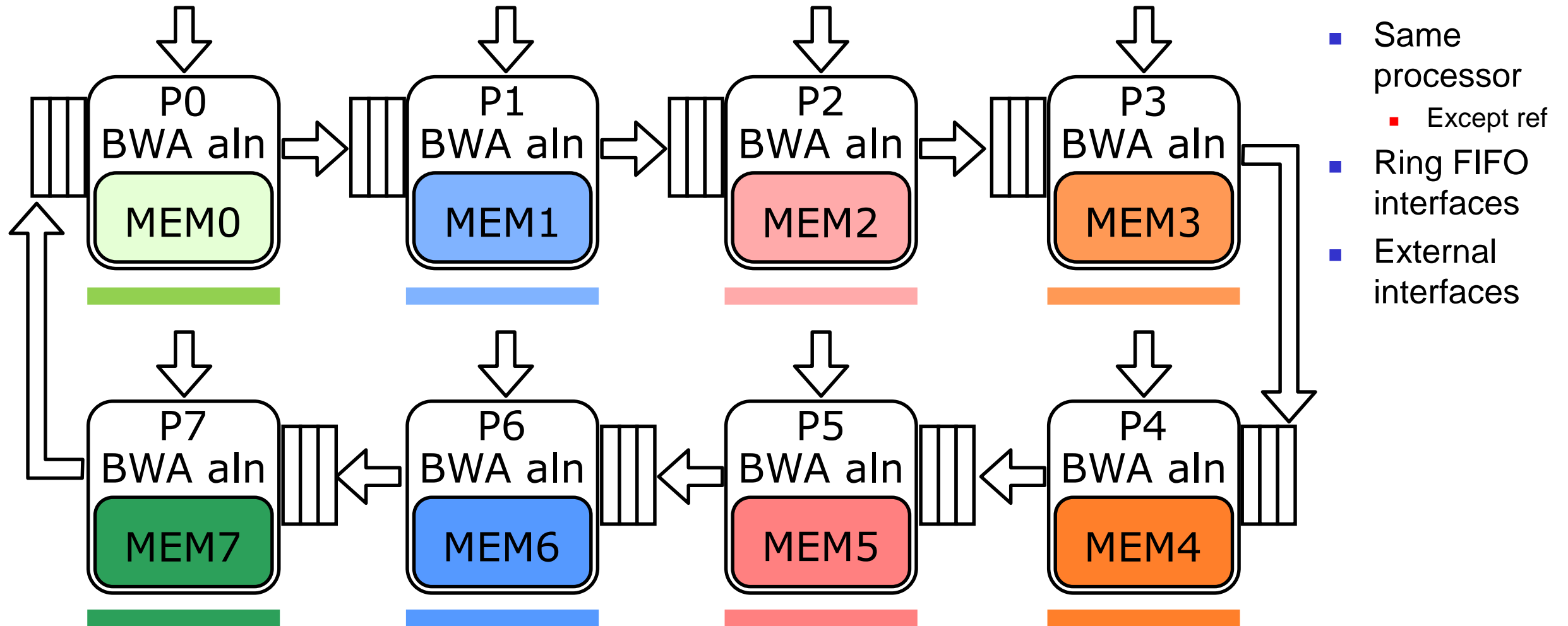


- Same processor
 - Except ref
- An external interface
- FIFO interfaces
- Pipeline freeze
- Unbalanced load

Linear Pipeline with Optimized Partition



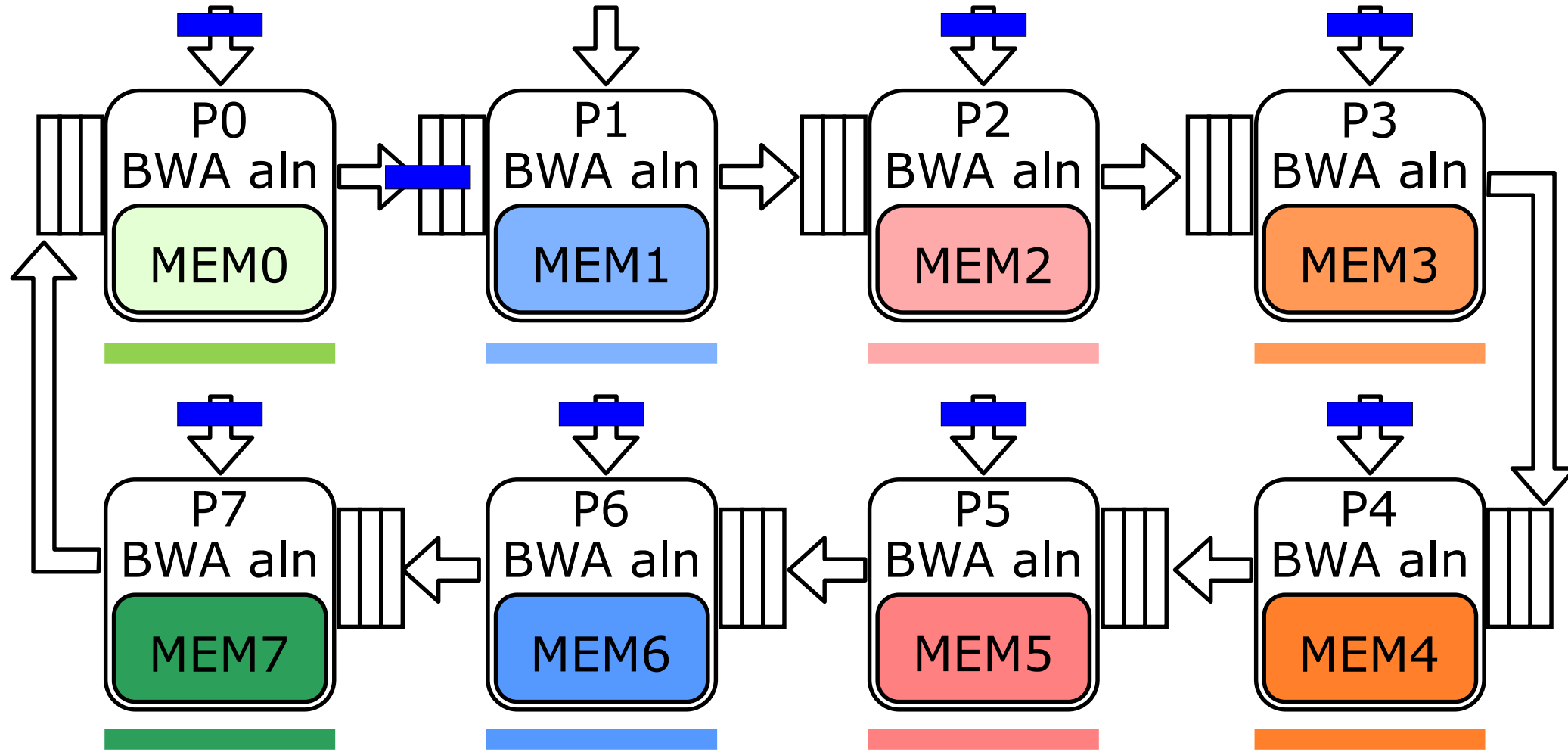
Ring Pipeline



Ring Pipeline

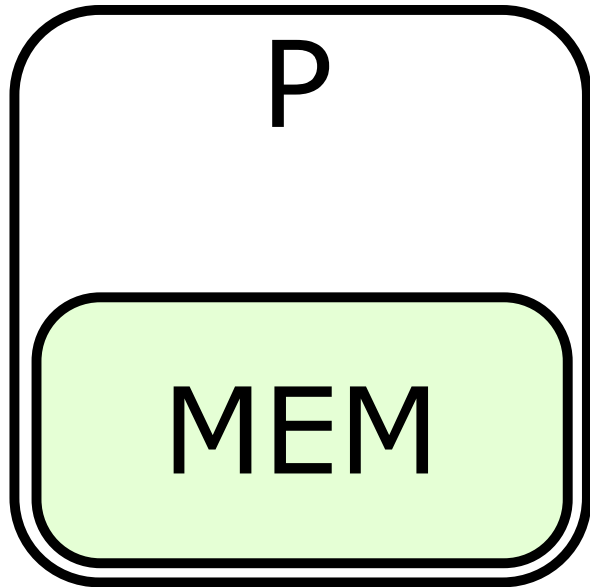
- An animation is removed to reduce the file size to 10MB.
- Please contact the presenter for the original slide

Ring Pipelined



- Same processor
 - Except ref
- Ring FIFO interfaces
- External interfaces
- Balanced load
- Less pipeline freezing

MESGA Processor

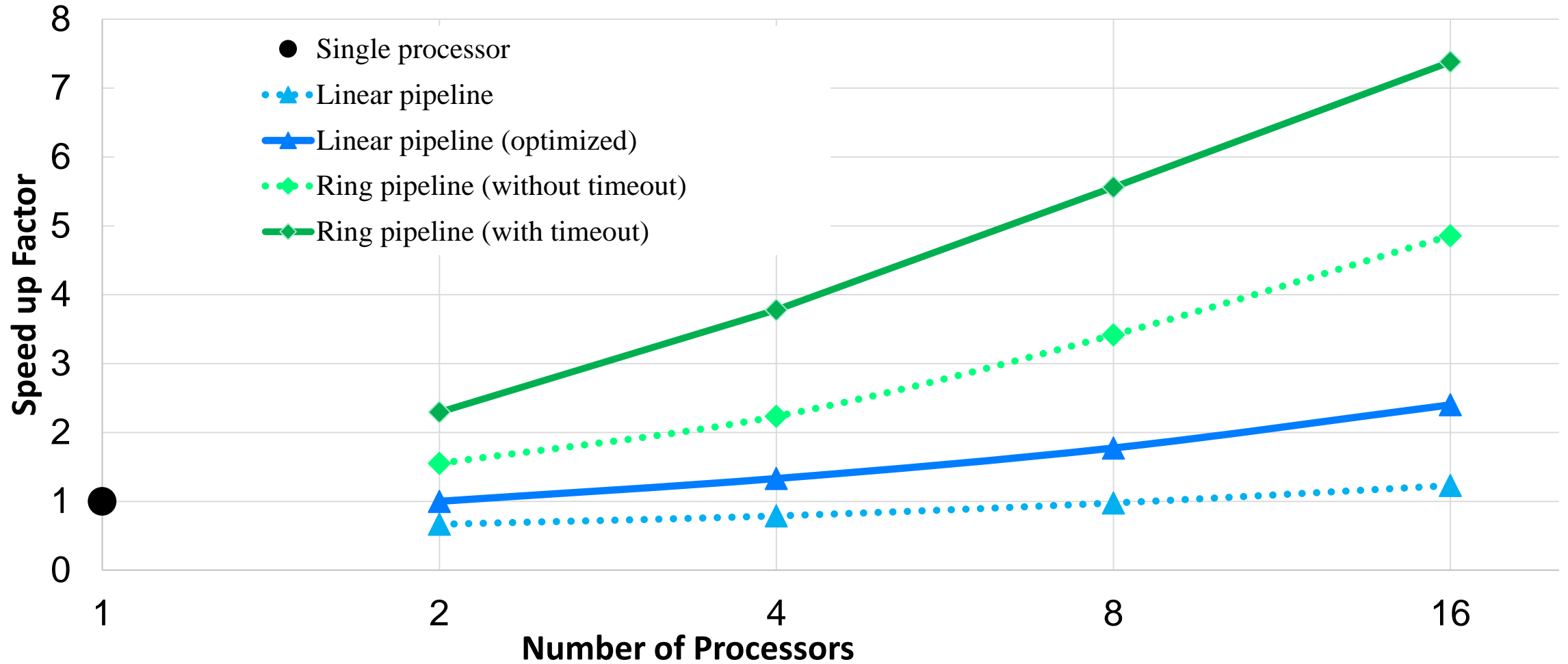


- 32-bit Xtensa LX6.0 processors
 - Clock speed 1092 MHz
 - Size 0.08 mm²
- 8kB data cache
- Up to 2GB local memory

Experiments Setup

- Full human genome reference
- One million reads each 125 base pair length
 - 0.09 % SNP mutation error
 - 0.01 % indel mutation error
 - 2 % uniform sequencing error

Results



MESGA vs Other Solutions

	GPU	FPGA	MESGA
Purpose	Hardware accelerator	Hardware accelerator	Stand alone & Complete
Feature support	Limited	Limited	Full
Flexibility	Low	Low	High

Future works & Conclusions

- Customised Instructions
- Inter-processor communication optimization
- Long read alignment
- Whole human genome can be processed in embedded system
 - Cheap
 - Small
 - Secure
- Performance improvement

Thank You