



SCIMA: Experimental Results

How to use On-Chip Memory

SCIMA provides various data placement and utilization scheme according to the characteristics of data access

consecutiveness \ reusability	not-reusable	reusable
	consecutive	use On-Chip Mem. as a stream buffer
stride	use On-Chip Mem. as a stream buffer	reserve On-Chip Mem. for reused data
irregular	use cache	reserve On-Chip Mem. for reused data

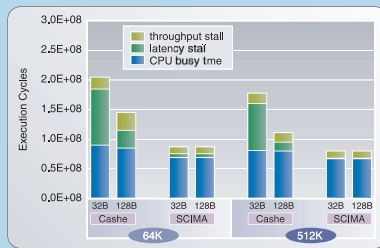
- latency-stall reduction by burst transfer
- latency & throughput-stall reduction by stride transfer
- throughput-stall reduction by software controllability

Latency/Throughput stall is reduced for wide variety of data access

Evaluation Results

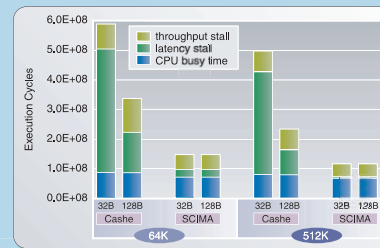
Throughput Ratio = Ratio between on-chip and off-chip memory throughput
 Latency = Memory access latency for off-chip memory (latency for the first data)

FT



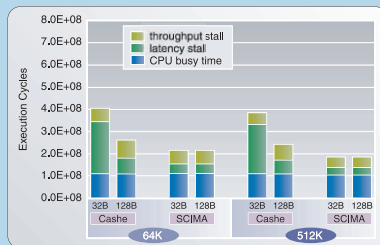
Throughput ratio=2:1
Latency=40

future technology trend



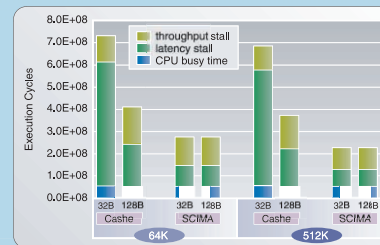
Throughput ratio=8:1
Latency=160

QCD



Throughput ratio=2:1
Latency=40

future technology trend



Throughput ratio=8:1
Latency=160

SCIMA is robust to large throughput ratio and long memory access latency caused by current technology trend of CPU-memory speed gap