Amdahl's Law in Multi-core Era

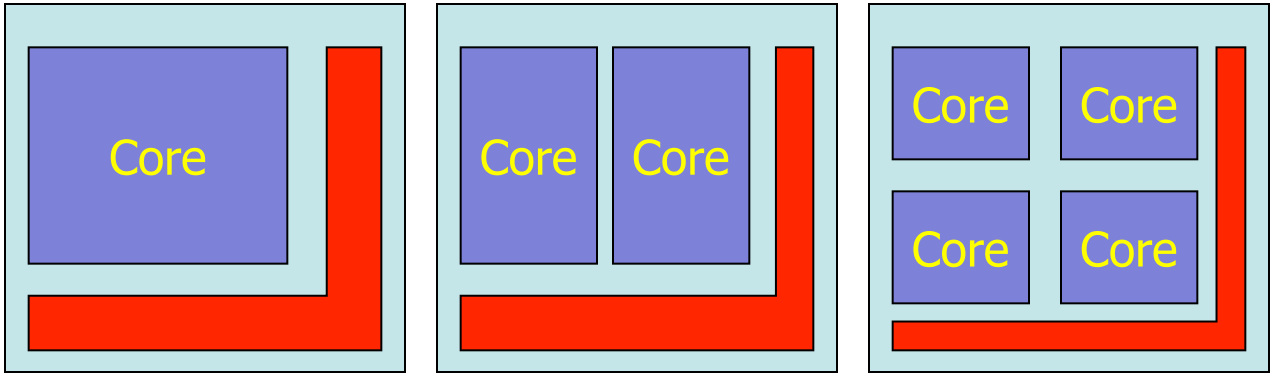
Virendra Singh
Associate Professor
Computer Architecture and Dependable Systems Lab
Department of Electrical Engineering
Indian Institute of Technology Bombay
http://www.ee.iitb.ac.in/~viren/
E-mail: viren@ee.iitb.ac.in

EE-739: Processor Design

Lecture 17 (19 March 2015)

CADSL

Why Multicore ?



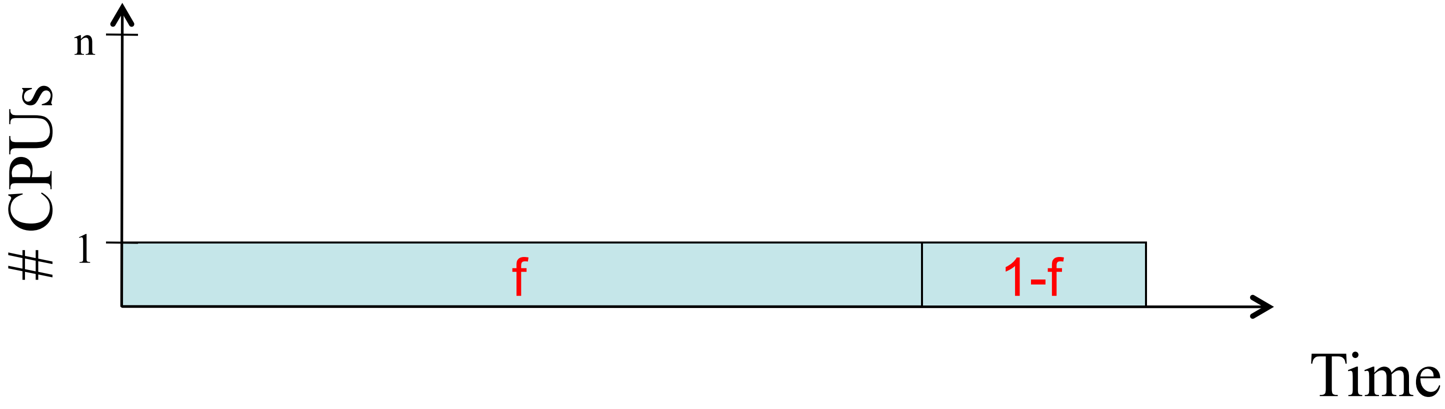
	Single Core	Dual Core	Quad Core
Core area	A	~A/2	~A/4
Core power	W	~W/2	~W/4
Chip power	W + O	W + O’	W + O’ ’
Core performance	P	0.9P	0.8P
Chip performance	P	1.8P	3.2P

Multi-Core Era

- How many Cores?
- Should Cores use simple pipelines or powerful multi-issue out-of-order dynamic pipeline designs?
- Should Cores use same or different micro-architectures?
- How to manage the power?

Gene Amdahl

Amdahl' s Law

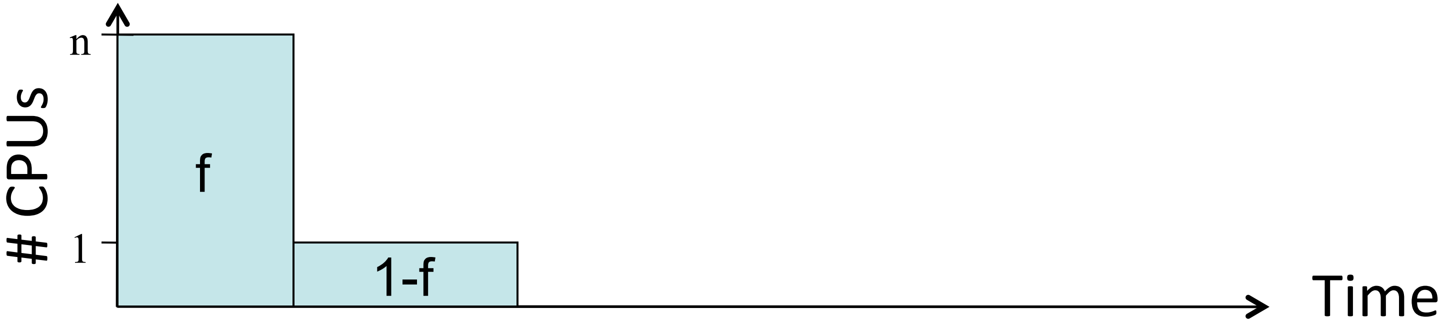


f – fraction that can run in parallel

1-f – fraction that must run serially

Every one know's Amdahl's law, but quickly forget's it.
Thomas Puzak, IBM, 2007

Amdahl's Law



f – fraction that can run in parallel

1-f – fraction that must run serially

$$Speedup = \frac{1}{(1-f) + \frac{f}{n}}$$

$$\lim_{n \to \infty} \frac{1}{1-f+\frac{f}{n}} = \frac{1}{1-f}$$

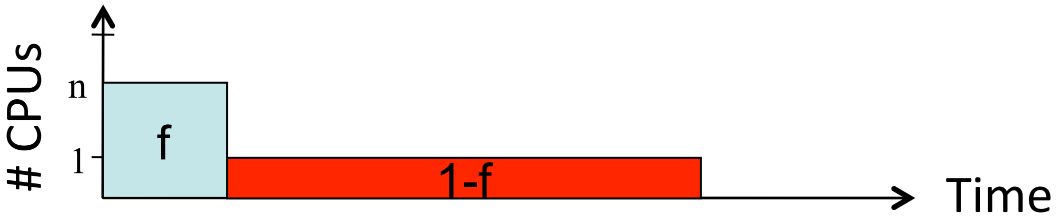
Amdahl's Law

Corrolaries:

- Attack the common case
  - – When f is small, optimization will have little effect
- The aspects you ignore also limits the speedup
  - As n approaches infinity, speed up is bounded by

$$\boxed{\lim_{n\to\infty} \frac{1}{1-f+\frac{f}{n}} = \frac{1}{1-f}}$$

Fixed Chip Power Budget



- Amdahl’s Law
  - Ignores (power) cost of n cores
- Revised Amdahl’s Law
  - More cores → each core is slower
  - Parallel speedup < n
  - Serial portion (1-f) takes longer
  - Also, interconnect and scaling overhead

Pollack's Law

Given the same process technology, the state-of-the-art processor

- provides 1.5 to 1.7 times higher performance and
- Consumes 2 to 3 times the die area compared with its previous-generation counterpart

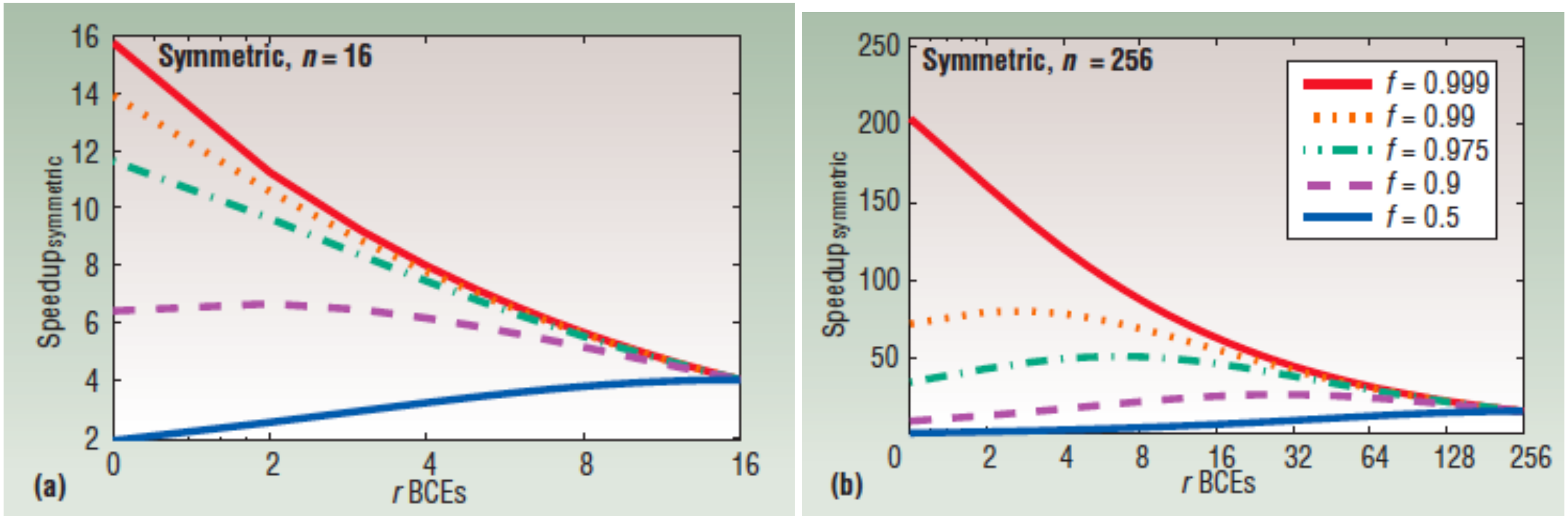
This means that a processor that consumes T times more transistors can provide only $\sqrt{T}$ times higher performance

Symmetric Multi-cores

- Need cost model
- Multi-core chip of a given technology can contains at most n base core equivalents (BCE)

$$\text{Speedup}_{\text{symmetric}}(f,n,r)=\frac{1}{\frac{1-f}{perf(r)}+\frac{f\cdot r}{perf(r)\cdot n}}$$

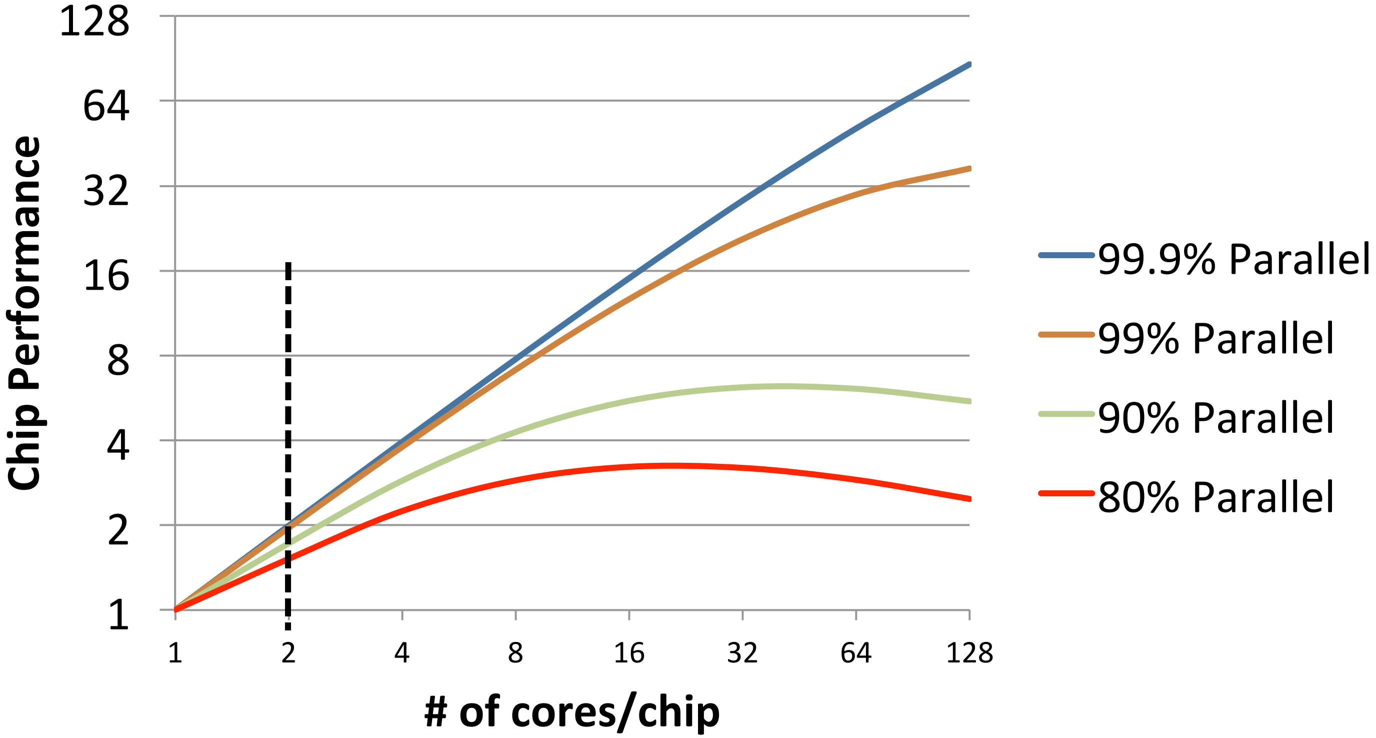
Symmetric Multi-cores



Symmetric Multi-cores

- Should target increasing f through architectural support, compiler techniques, programming model improvement
- Using more BCE per core, r>1, can be optimal, even when performance grows only r.
  - Seek methods of increasing core performance even at high cost

Fixed Power Scaling

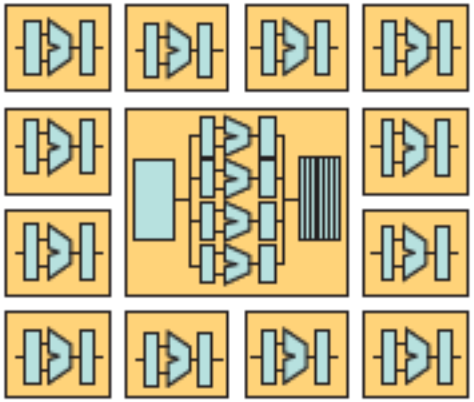


- Fixed power budget forces slow cores
- Serial code quickly dominates

Symmetric Multi-cores

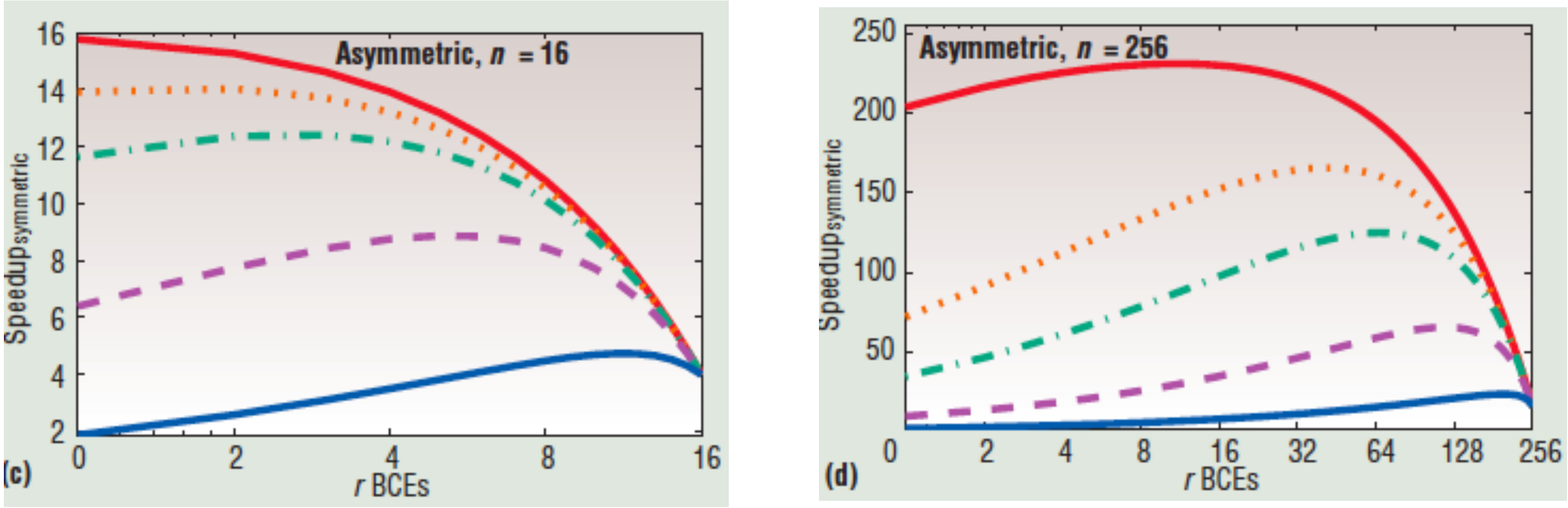
- Should target increasing f through architectural support, compiler techniques, programming model improvement
- Using more BCE per core, r>1, can be optimal, even when performance grows only r.
  - Seek methods of increasing core performance even at high cost

Asymmetric Cores

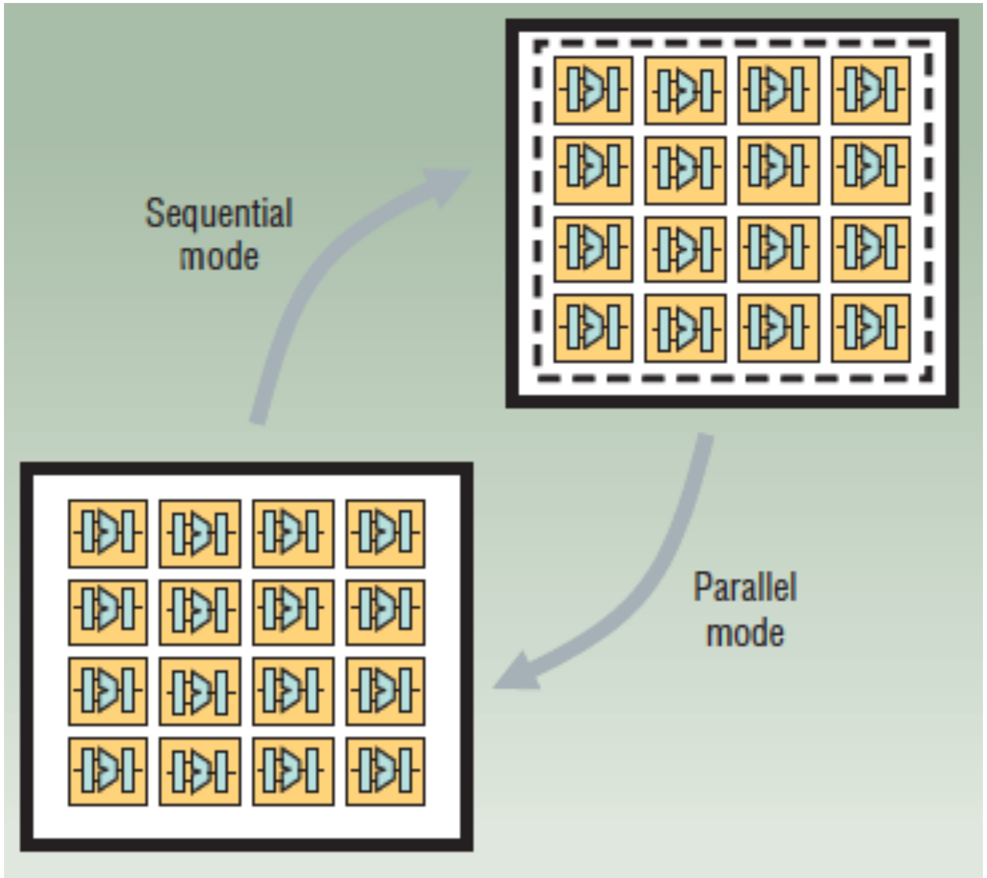


$$\text{Speedup}_{\text{asymmetric}}(f,n,r) = \frac{1}{\frac{1-f}{perf(r)} + \frac{f}{perf(r)+n-r}}$$

- Should investigate method of increasing performance of sequential code

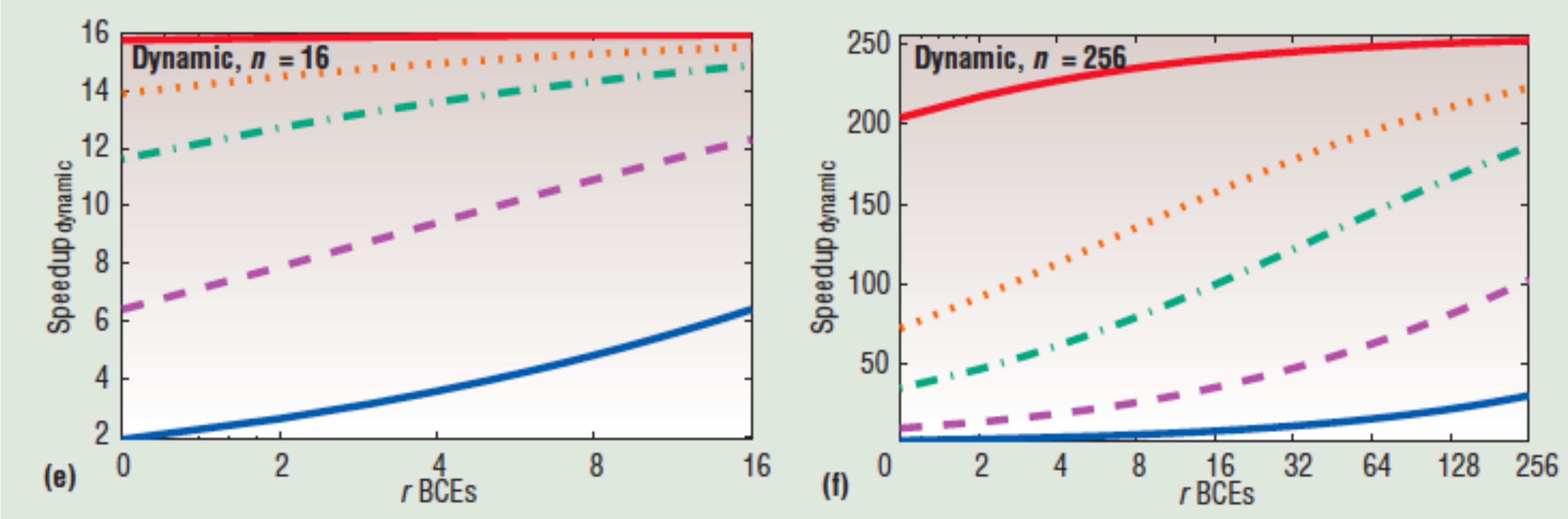


Dynamic Cores



Dynamic Cores

$$\text{Speedup}_{\text{dynamic}}(f,n,r) = \frac{1}{\frac{1-f}{perf(r)} + \frac{f}{n}}$$



Energy Efficient Computing

- Energy efficiency is becoming increasingly important
- Use of simple analytical model at early design stage, we can architect a systems with better energy efficiency

Model for P*

- Power consumed by idle processors is k times power consumed by the active processor
- Power consumed by the active processor is 1 unit
- Sequential phase power consumption
  - 1 + (n-1) k
- Parallel phase power consumptio
- Average power consumption

$$W = \frac{(1-f)\times\{1+(n-1)k\}+\frac{f}{n}\times n}{(1-f)+\frac{f}{n}}$$

$$= \frac{1+(n-1)k(1-f)}{(1-f)+\frac{f}{n}}$$

Energy efficient computing

- Performance Per watt
  - Reciprocal of energy

$$\frac{Perf}{W}=\frac{1}{(1-f)+\frac{f}{n}}\times\frac{(1-f)+\frac{f}{n}}{1+(n-1)k(1-f)}$$

$$=\frac{1}{1+(n-1)k(1-f)}$$

- Performance per joule
  - Reciprocal of energy delay

$$\frac{Perf}{J}=\frac{1}{(1-f)+\frac{f}{n}}\times\frac{1}{1+(n-1)k(1-f)}$$

Performance of P*

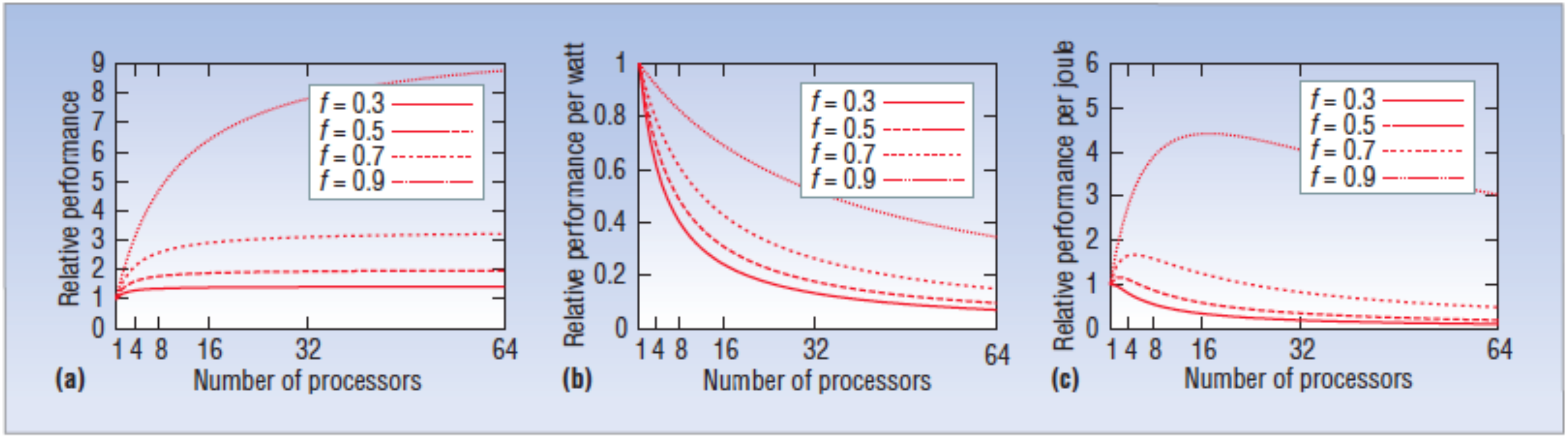


Figure 2. P scalability. P*, a symmetric many-core processor that replicates a state-of-the-art superscalar processor on a die, consumes a high amount of energy to complete the task: (a) performance, (b) performance per watt, and (c) performance per joule, where k = 0.3.*

Thank You