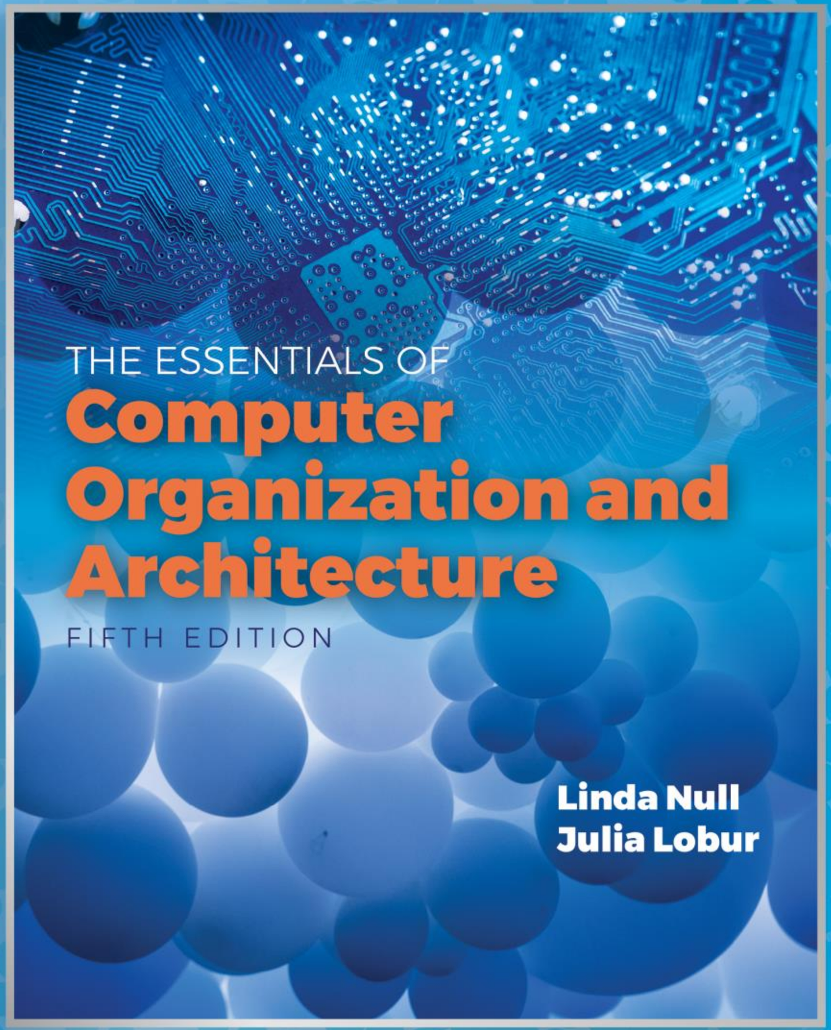
This is the fourth lecture of Chapter 9

# Chapter 9

## Alternative Architectures (D)



# Quick review of last lecture (1)

- Parallel and Multiprocessor Architectures
  - Shared Memory Multiprocessors
    - Tightly-coupled multiprocessor
    - Distributed shared memory multiprocessor
    - Uniform memory access (UMA)
    - Nonuniform memory access (NUMA)
    - Cache coherence problems
    - Write-through with update
    - Write-through with invalidate
    - Write-back (exclusive rights to the data)
  - Distributed computing
    - Very loosely-coupled processing units
    - Cloud computing

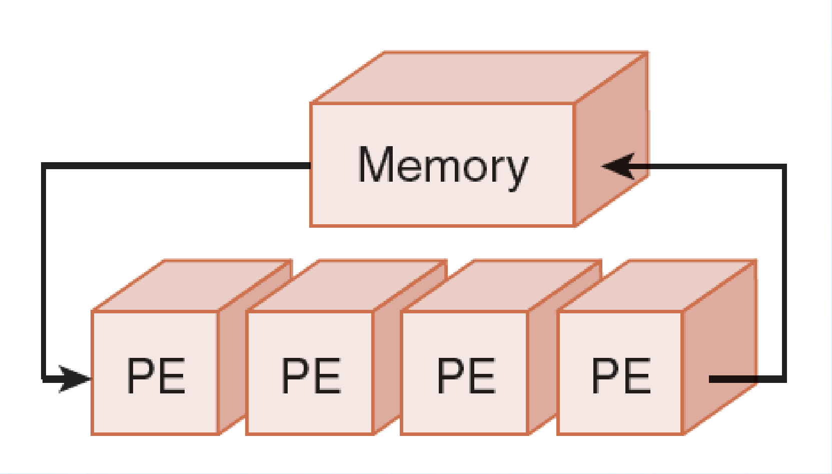
# Quick review of last lecture (2)

- Alternative Parallel Processing Approaches
  - Dataflow Computing
    - A data flow graph
  - Neural networks
    - Perceptrons
    - Supervised learning
    - Unsupervised learning

# 9.5 Alternative Parallel Processing Approaches (14 of 15)

## 9.5.3 Systolic Array

- Where neural nets are a model of biological neurons, *systolic array* computers are a model of how blood flows through a biological heart.
- Systolic arrays, a variation of SIMD computers, have simple processors that process data by circulating it through vector pipelines.



# 9.5 Alternative Parallel Processing Approaches (15 of 15)

- Systolic arrays can sustain great throughout because they employ a high degree of parallelism.
- Connections are short, and the design is simple and scalable. They are robust, efficient, and cheap to produce. They are, however, highly specialized.
- They are useful for solving repetitive problems that lend themselves to parallel solutions using a large number of simple processing elements.
  - Examples include sorting, image processing, and Fourier transformations.

- Systolic-Array Implementation of Matrix Multiplication

- $c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + a_{i3}b_{3j} + a_{i4}b_{4j}$

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \times \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & b_{24} \\ b_{31} & b_{32} & b_{33} & b_{34} \\ b_{41} & b_{42} & b_{43} & b_{44} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \\ c_{31} & c_{32} & c_{33} & c_{34} \\ c_{41} & c_{42} & c_{43} & c_{44} \end{bmatrix}$$

$$c_{11} = a_{11} \bullet b_{11} + a_{12} \bullet b_{21} + a_{13} \bullet b_{31} + a_{14} \bullet b_{41}$$
$$c_{12} = a_{11} \bullet b_{12} + a_{12} \bullet b_{22} + a_{13} \bullet b_{32} + a_{14} \bullet b_{42}$$
$$c_{13} = a_{11} \bullet b_{13} + a_{12} \bullet b_{23} + a_{13} \bullet b_{33} + a_{14} \bullet b_{43}$$
$$c_{14} = a_{11} \bullet b_{14} + a_{12} \bullet b_{24} + a_{13} \bullet b_{34} + a_{14} \bullet b_{44}$$
$$c_{21} = a_{21} \bullet b_{11} + a_{22} \bullet b_{21} + a_{23} \bullet b_{31} + a_{24} \bullet b_{41}$$
$$c_{22} = a_{21} \bullet b_{12} + a_{22} \bullet b_{22} + a_{23} \bullet b_{32} + a_{24} \bullet b_{42}$$
$$c_{23} = a_{21} \bullet b_{13} + a_{22} \bullet b_{23} + a_{23} \bullet b_{33} + a_{24} \bullet b_{43}$$
$$c_{24} = a_{21} \bullet b_{14} + a_{22} \bullet b_{24} + a_{23} \bullet b_{34} + a_{24} \bullet b_{44}$$
$$c_{31} = a_{31} \bullet b_{11} + a_{32} \bullet b_{21} + a_{33} \bullet b_{31} + a_{34} \bullet b_{41}$$
$$c_{32} = a_{31} \bullet b_{12} + a_{32} \bullet b_{22} + a_{33} \bullet b_{32} + a_{34} \bullet b_{42}$$
$$c_{33} = a_{31} \bullet b_{13} + a_{32} \bullet b_{23} + a_{33} \bullet b_{33} + a_{34} \bullet b_{43}$$
$$c_{34} = a_{31} \bullet b_{14} + a_{32} \bullet b_{24} + a_{33} \bullet b_{34} + a_{34} \bullet b_{44}$$
$$c_{41} = a_{41} \bullet b_{11} + a_{42} \bullet b_{21} + a_{43} \bullet b_{31} + a_{44} \bullet b_{41}$$
$$c_{42} = a_{41} \bullet b_{12} + a_{42} \bullet b_{22} + a_{43} \bullet b_{32} + a_{44} \bullet b_{42}$$
$$c_{43} = a_{41} \bullet b_{13} + a_{42} \bullet b_{23} + a_{43} \bullet b_{33} + a_{44} \bullet b_{43}$$
$$c_{44} = a_{41} \bullet b_{14} + a_{42} \bullet b_{24} + a_{43} \bullet b_{34} + a_{44} \bullet b_{44}$$

Figure 3.1: Multiplication of matrices of size 4 4.

- Elements $a_{ir}$ and $b_{rj}$ arrive at processor $P_{ij}$ simultaneously for the operation $a_{ir} \bullet b_{rj}$ to be performed.
- $c_{ij}$ is initialized to 0 in $P_{ij}$ , for all $i, j = 1, 2, 3, 4$. At the end, processor $P_{ij}$ will contain $c_{ij}$ , for $1 \leq i, j \leq 4$.
- Whenever a processor $P_{ij}$ receives two inputs $b$ and $a$ from the north and the west, respectively, it performs the following set of operations, in this order:
  1. it calculates $a \bullet b$;
  2. it adds the result to the previous value $c_{ij}$ , and stores the result in $c_{ij}$ ;
  3. it sends **a** to $P_{i,j+1}$, unless $j = 4$; and
  4. it sends **b** to $P_{i+1,j}$, unless $i = 4$.
- This algorithm takes time **O(n)**, for **n×n** matrices.

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + a_{i3}b_{3j} + a_{i4}b_{4j}$$

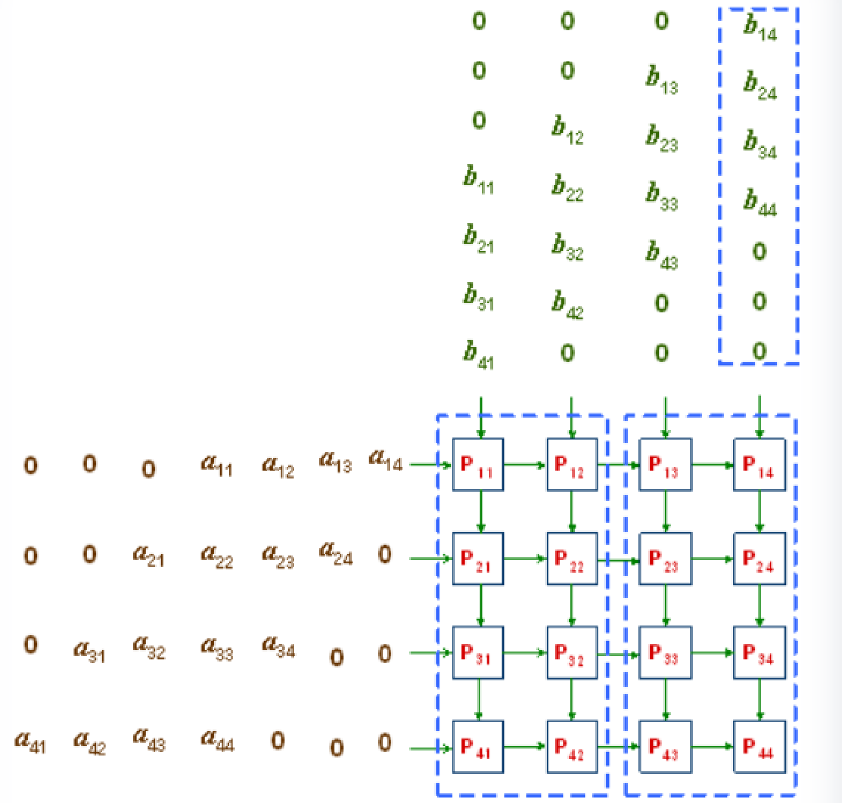


Figure 3.2: A 4 x 4 mesh (systolic array) of processors for matrix multiplication.

# 9.6 Quantum Computing (1 of 8)

- Computers, as we know them are binary, transistor-based systems.
- But transistor-based systems strain to keep up with our computational demands.
- We increase the number of transistors for more power, and each transistor smaller to fit on the die.
  - Transistors are becoming so small that it is hard for them to hold electrons in the way in which we're accustomed to.
- Thus, alternatives to transistor-based systems are an active area or research.

# 9.6 Quantum Computing (2 of 8)

- Computers are now being built based on:
  - Optics (photonic computing)
  - Biological neurons
  - DNA
- One of the most intriguing is quantum computers.
- Quantum computing uses quantum bits (qubits) that can be in multiple states at once.
- The “state” of a qubit is determined by the spin of an electron.
- A thorough discussion of “spin” is under the domain of quantum physics.

# 9.6 Quantum Computing (3 of 8)

- A qubit can be in multiple states at the same time.
  - This is called *superpositioning*.
- A 3-bit register can simultaneously hold the values 0 through 7.
  - 8 operations can be performed at the same time.
- This phenomenon is called *quantum parallelism*.
  - A system with 600 qbits can superposition 2600 states.

# 9.6 Quantum Computing (4 of 8)

- D-Wave Computers is the first quantum computer manufacturer.
- D-Wave computers having 512 qubits were purchased separately by University of Southern California and Google for research purposes.
- Quantum computers may be applied in the areas of cryptography, true random-number generation, and in the solution of other intractable problems.

# 9.6 Quantum Computing (5 of 8)

- Making effective use of quantum computers requires rethinking our approach to problems and the development of new algorithms.
  - To break a cypher, the quantum machine simulates every possible state of the problem set (i.e., every possible key for a cipher) and it “collapses” on the correct solution.
- Examples include Schor’s algorithm for factoring products of prime numbers.
- Many others remain to be discovered.

# 9.6 Quantum Computing (6 of 8)

- These systems are not constrained by a fetch-decode-execute cycle; however, quantum architectures have yet to settle on a definitive paradigm analogous to von Neumann systems.
- Rose’s Law states that the number of qubits that can be assembled to successfully perform computations will double every 12 months; this has been precisely the case for the past 9 years.
  - This “law” is named after Geordie Rose, D-Wave’s founder and chief technology officer.

# 9.6 Quantum Computing (7 of 8)

- One of the largest obstacles to the progress of quantum computation is the tendency for qubits to decay into a state of *decoherence*.
  - Decoherence causes uncorrectable errors.
- Advanced error-correction algorithms have been applied to this problem and show promise.
- Much research remains to be done, however.

# 9.6 Quantum Computing (8 of 8)

- The realization of quantum computing has raised questions about technological singularity.
  - Technological singularity is the theoretical point when human technology has fundamentally and irreversibly altered human development.
  - This is the point when civilization changes to an extent that its technology is incomprehensible to previous generations.
- Are we there, now?

# Conclusion (1 of 4)

- The common distinctions between RISC and CISC systems include RISC's short, fixed-length instructions. RISC ISAs are load-store architectures. These things permit RISC systems to be highly pipelined.
- Flynn's Taxonomy provides a way to classify multiprocessor systems based upon the number of processors and data streams. It falls short of being an accurate depiction of today's systems.

# Conclusion (2 of 4)

- Massively parallel processors have many processors, distributed memory, and computational elements communicate through a network. Symmetric multiprocessors have fewer processors and communicate through shared memory.
- Characteristics of superscalar design include superpipelining, and specialized instruction fetch and decoding units.

# Conclusion (3 of 4)

- Very long instruction word (VLIW) architectures differ from superscalar architectures because the compiler, instead of a decoding unit, creates long instructions.
- Vector computers are highly-pipelined processors that operate on entire vectors or matrices at once.
- MIMD systems communicate through networks that can be blocking or nonblocking. The network topology often determines throughput.

# Conclusion (4 of 4)

- Multiprocessor memory can be distributed or exist in a single unit. Distributed memory brings to rise problems with cache coherency that are addressed using cache coherency protocols.
- New architectures are being devised to solve intractable problems. These new architectures include dataflow computers, neural networks, systolic arrays, and quantum computers.