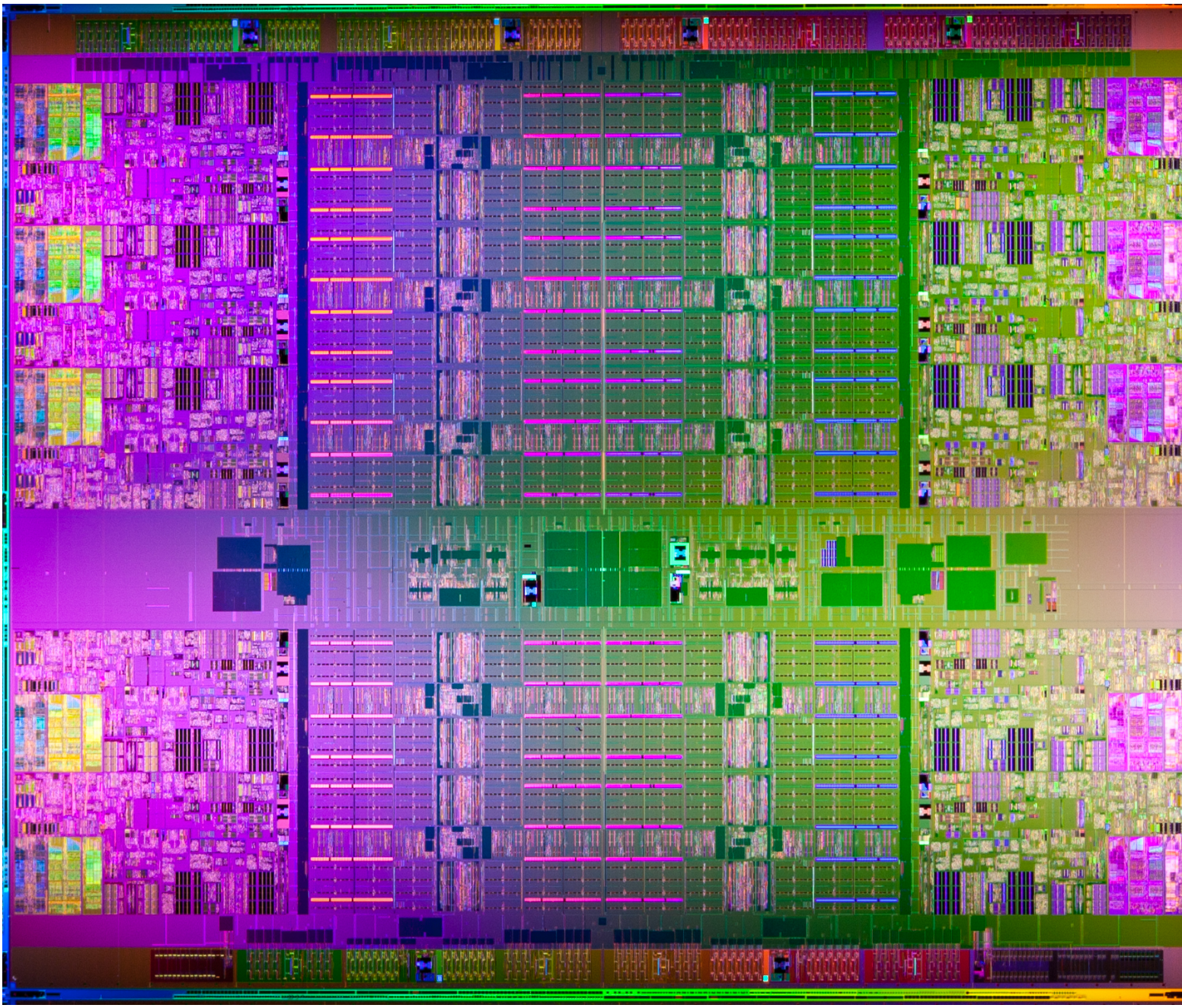


60007

Theory and Practice of Concurrent Programming

Imperial College London

Contents

Chapter 1

Introduction

1.1 Course Structure & Logistics

1.1.1 Structure

Dr Azalea Raad

Prof Alastair Donaldson

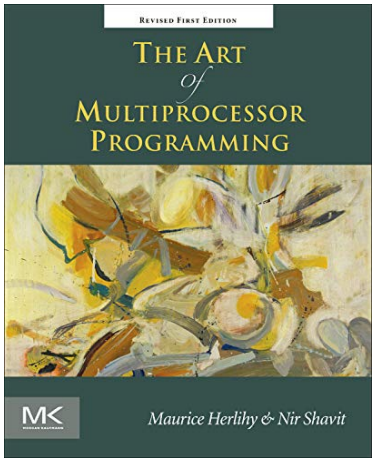
Theory For weeks 2 → 5:

- Intro to synchronisation paradigms (mutual exclusion, readers-writers, producer-consumer)
- Low-level concurrent semantics (sequential consistency, Intel-x86)
- High-level concurrent semantics (concurrent objects, linearisability)
- Transactional memory (serialisability)

Practical For weeks 5 → 8:

- Threads and locks in C++
- Implementing locks
- Concurrency in Haskell
- Race-free concurrency in Rust
- Dynamic data-race detection

1.1.2 Extra Materials



The Art of Multiprocessor Programming
About 65% of the theory course.

1.2 Preface for Concurrency

1.2.1 Moore's Law

Moore's Law — **Definition 1.2.1**

An empirical (supported by observation) law that states the density of transistors in an integrated circuit will double approximately every two years.

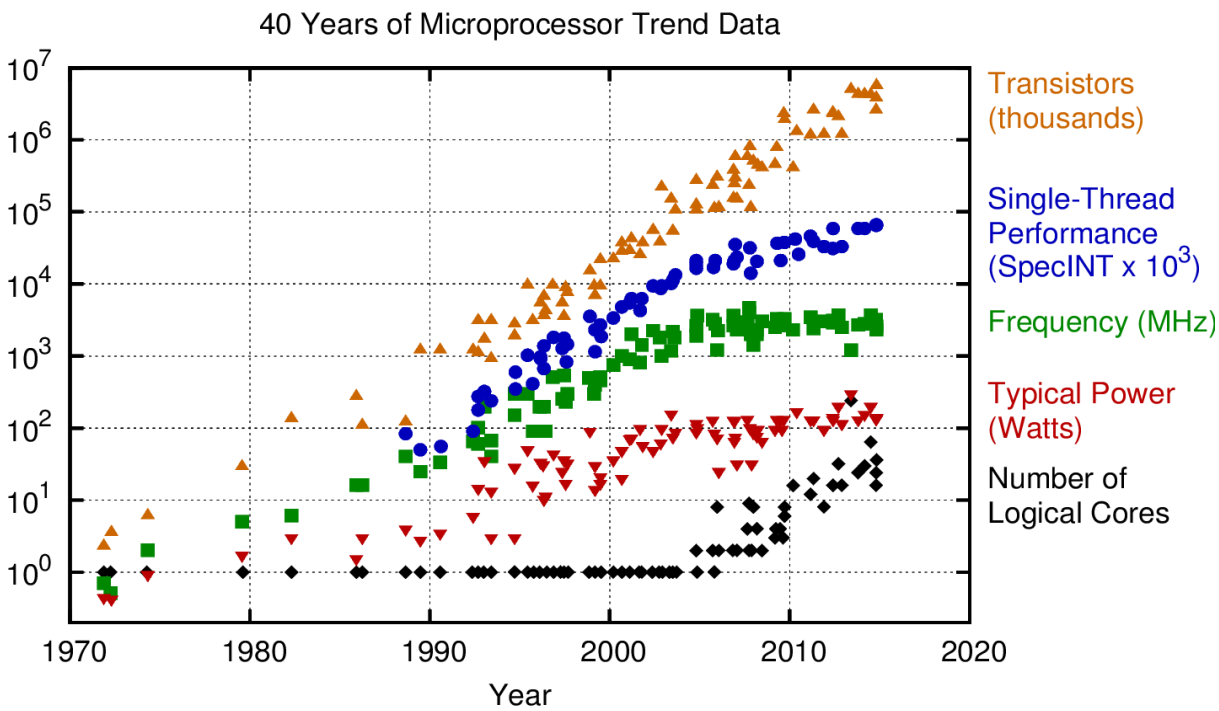
- The observation is named after Gordon Moore (co-founder and later CEO of Intel).
- This law no longer holds, and sequential performance improvements have declined.

Dennard/MOSFET Scaling — **Definition 1.2.2**

$$\text{Power} \propto \text{Transistor Size}$$

A scaling law stating that as transistor density increases, the power requirements stay constant.

- Increasing transistor density results in power staying constant (less power per transistor) and lower circuit delay.
- This allows for higher switching frequency ⇒ higher clocks frequencies ⇒ better sequential performance).



Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010-2015 by K. Rupp

The performance improvements typically expected yearly (moore's law and Dennard scaling) no longer apply.

- Sequential (Single-Thread) performance improvements have declined.
- Parallelism is being exploited to improve performance (uniprocessors are virtually extinct).
- Shared-memory multiprocessor systems have lost out to multicore processors.

Amdahl's Law — **Definition 1.2.3**

$$S_{\text{latency}}(s) = \frac{1}{(1-p) + \frac{p}{s}} \text{ where } p = \text{parallel portion and } s = \text{threads}$$

Amdahl's law describes the speedup of a program, associated with the number of threads.

- Can be applied to other resources.
- Versions of the equation exist for different proportions using different numbers of threads.
- As the number of threads increases the sequential part of the program becomes a bottleneck.

1.2.2 Concurrency Difficulties

Writing correct, concurrent code is difficult.

race Condition — **Definition 1.2.4**

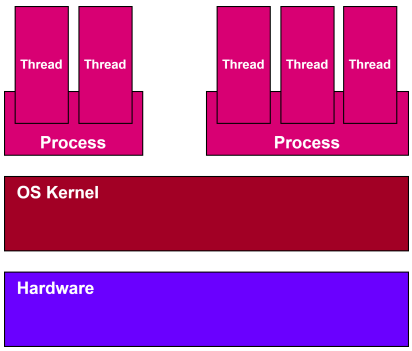
A potential for a situation where the result of a program depends on the non-deterministic timing or interleaving of threads.

- Where multiple threads access data (non-atomically) and at least one writes.
- Where a lack of enforced ordering on some events causes differing results (e.g output to user)

- A process can have multiple threads executing in parallel.
- Cannot determine at compile time the relative speed of execution of threads (many delays are unpredictable; cache misses, page faults, interrupts).
- Cannot predict how long threads will be blocked (e.g I/O) or when threads will be scheduled (or use up their time quantum).

Hence we must use synchronisation mechanisms to regulate accesses to shared data that can result in a race condition.

1.2.3 OS Concepts



- Operating system provides process and thread abstractions.
- A process contains one or more threads (streams of instructions being executed).
- A process has its own address space, all threads in the process share this address space.
- The OS kernel contains a scheduler which schedules processes & their threads.

Chapter 2

Concurrency In C++

2.1 Threads

To interact with threads the `thread` header must be included.

- It provides a standard, implementation independent, interface for handling threads.
- Provides the `std::thread` class

```
#include <thread>

namespace std {
  class thread {
  public:
    // types
    class id;
    using native_handle_type = /* implementation-defined */;

    // construct/copy/destroy
    thread() noexcept;

    // Constructor takes a function to start from, and its arguments (all type checked)
    template<class F, class... Args> explicit thread(F&& f, Args&&... args);

    // Destructor (terminates current thread if the thread has not been joined)
    ~thread();

    // Attempting to copy a thread is not allowed. Hence delete ensures no compile.
    thread(const thread&) = delete;

    // Can create thread from a thread r-value (copy)
    thread(thread&&) noexcept;

    // Attempting to copy a thread (via an immutable reference).
    // This is not allowed, so if this operator is used it will not compile.
    thread& operator=(const thread&) = delete;

    // Assign a thread from an (r value - e.g expression, literal) reference
    thread& operator=(thread&&) noexcept;

    // members
    void swap(thread&) noexcept;
    bool joinable() const noexcept;

    // Wait for this thread to terminate.
    void join();
```

```
    // Allow the thread to continue executing after the thread handler (this)
    // is destroyed
    void detach();

    // Get the unique id of the thread
    id get_id() const noexcept;

    native_handle_type native_handle();

    // static members
    static unsigned int hardware_concurrency() noexcept;
  };
}
```

Lambda **Definition 2.1.1**

A lambda is a small function that can be defined in an expression, capture values in its scope (and above), and be passed as a value.

```
// [captures] (arguments) {body}

// a basic lambda with no captures
auto my_lambda = [] (int a, int b) -> int {return a + b;}

// using the lambda
int c = my_lambda(1, 2);

auto another_lambda = [c&] (int d) {return c + d;}
```

We can construct using `std::thread`'s constructors.

```
// idiomatic constructor
std::thread my_thread(StartFunction, arg1, arg2, ...)

// call constructor and assign
std::thread my_thread = std::thread(StartFunction, arg1, arg2, ...)
auto my_thread = std::thread(StartFunction, arg1, arg2, ...)

// pass lambda as function
std::thread my_thread(StartFunction, arg1, arg2, ...)
```