COMS10015 lecture: week #1

- Agenda: a non-technical introduction to
 - 1. unit objectives,
 - 2. unit organisation, and
 - 3. some motivation (i.e., why the unit exists).

Part 1: unit objectives, i.e., the "what" (1)

▶ The term **computer architecture** can be explained via analogy, e.g., if

building a house = architect + civil engineer

then

building a computer = *computer* architect + electrical engineer.

Part 1: unit objectives, i.e., the "what" (1)

Objectives

Put simply, after completing this unit you should be able to understand and apply concepts relating to

- 1. how computers are designed and manufactured, e.g., how logic gates are organised to perform computation
- how computers work,
 e.g., how instructions and so programs are executed
- 3. how computers can be used more effectively, e.g., behaviour of high-level programs wrt. low-level resources reading

computer = computer processor + supporting infrastructure (i.e., the wider computer system).

Part 1: unit objectives, i.e., the "what" (3)

AN X64 PROCESSOR IS SCREAMING ALONG AT BILLIONS OF CYCLES PER SECOND TO RUN THE XNU KERNEL, WHICH IS FRANTICALLY WORKINSTHROUGH A/L THE POSIX-SPECIFED ABSTRACTION TO CREATE THE DRAWN SYSTEM UNDERSING OS X, WHICH IN TURN IS STRAINING ITSELF TO RUN FIREBOX AND ITS GECKO RENDERER, WHICH CREATES A FLASH OBSTEMICH SHOPERS, WHICH CREATES A FLASH OBSTEMICH SHOPERS TOZENS OF WIDEN FRAMES EVERY SECOND

BECAUSE I WANTED TO SEE A CAT JUMP INTO A BOX AND FALL OVER.



I AM A GOD.

Theory	
(data structures and algorithms)	
Application	
Compiler	software
Operating System	
Instruction Set Architecture (ISA) •	- interface
Micro-architecture	hardware
Digital (micro-)electronics	nardware
Theory	
(Mathematics and Physics)	

► Important:

1. The unit is delivered by the following members of (academic) staff

Tom Deakin ⇒ Lecturer and Unit Director

Daniel Page ⇒ Lecturer

supplemented by, e.g., $\,$ a wider team who act in/as Teaching Support Roles (TSRs).

► Important:

2. At a high(er) level, the unit is delivered as a set of themes:

```
Theme #1 ⇒ "from Mathematics and Physics to digital logic"
Theme #2 ⇒ "from digital logic to computer processors"
Theme #3 ⇒ "from computer processors to software applications"
```

3. At a low(er) level, the unit involves the following activities:

► Important:

3. The summative assessment for this unit includes

```
summative coursework assignment \rightarrow TB1, week 10 \mapsto 30% weight = 6CP summative exam \rightarrow TB2, assessment period \mapsto 70% weight = 14CP
```

4. The *formative* assessment for this unit includes

```
formative exam #1 \sim TB1, week 6
formative exam #2 \sim TB1, week 12
formative exam #3 \sim TB2, week 18
formative exam #4 \sim TB2, week 24
```

none of which is credit bearing, i.e., it has 0% weight = 0CP.

Important:

- 5. Everything related to the unit is accessible via either
 - the internal-facing Blackboard-based unit web-site

```
https://www.ole.bris.ac.uk
```

or

the external-facing GitHub-based unit web-site

```
https://cs-uob.github.io/COMS10015
```

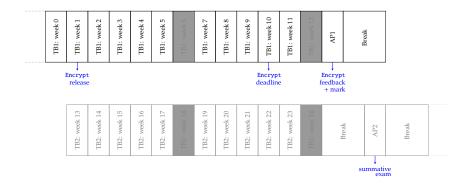
or, more specifically,

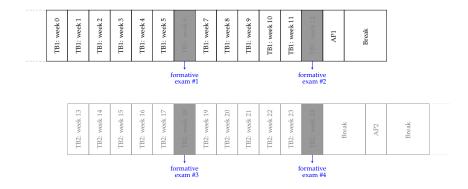
```
unit-wide communication, e.g., announcements ⇒ Blackboard assessment submission, marks, and feedback discussion forum ⇒ Blackboard Blackboard Teams } internal-facing teaching material ⇒ GitHub } external-facing
```

TBI: week 0 TBI: week 2 TBI: week 4 TBI: week 4 TBI: week 5 TBI: week 7 TBI: week 7 TBI: week 10 TBI: week 11
--

TB2: week 13	TB2: week 14	TB2: week 15	TB2: week 16	TB2: week 17	TB2: week 19	TB2: week 20	TB2: week 21	TB2: week 22	TB2: week 23		Break	AP2	Break	
										F				

				ight con per wee						ught cor per wee					
TB1: week 0	TB1: week1	TB1: week 2	TB1: week 3	TB1: week 4	TB1: week 5	TB1: week 6	TB1: week 7	TB1: week 8	TB1: week 9	TB1: week 10	TB1: week 11	TB1: week 12	AP1	Break	
	fo	rmative	e lab. w	orkshee	ets		c	ptional	lab. wo	orksheet	ts	,			1
	week 13	week 14	week 15	week 16	week 17	week 18	week 19	week 20	week 21	week 22	week 23		Dunal	reak AP2	Break





Part 3: unit motivation, i.e., the "why" (1) Computer architecture as hardware design

Question: which processor designs have a larger deployment (i.e., sell more units)?





Question: which processor designs have a larger deployment (i.e., sell more units)?





- Answer: probably ARM, because the embedded processor market is so large;
 - 1. keep in mind that
 - $computer \ = \ \{desktop\ computer, laptop\ computer, embedded\ computer, \ldots\},$
 - i.e., there's a lot more to the topic than traditional models of computing, and
 - 2. knowledge driven industries demand you "understand" not just "do".

Part 3: unit motivation, i.e., the "why" (2) Computer architecture as hardware design

- ▶ Question:
 - 1. what is this device, and
 - 2. who is the designer and/or vendor?



Part 3: unit motivation, i.e., the "why" (2) Computer architecture as hardware design

- ► Ouestion:
 - 1. what is this device, and
 - 2. who is the designer and/or vendor?



- Answer: this is an open-hardware based laptop called Novena [7];
 - designed by Andrew "bunnie" Huang and Sean "xobs" Cross,
 - houses an ARM-based processor plus Xilinx-based FPGA,
 - ▶ raised ~ \$780,000 via CrowdSupply campaign: you can buy one,
 - if equipped with the right set of skills, you can design and manufacture something similar.

Part 3: unit motivation, i.e., the "why" (3)

High-level applications of computer architecture

Quote

People who are really serious about software should make their own hardware.

 $- \textit{Kay} \; (\texttt{https://en.m.wikiquote.org/wiki/Alan_Kay})$





















▶ Question: ranging from the late 1970s to the late 1990s, spot the difference(s).











- ► Answer: advances in at least two fields, namely
 - 1. improved 2D and 3D computer graphics techniques, and
 - 2. improved general-purpose processors and display technologies (including a lineage of special-purpose GPUs)

suggesting computer architecture evolves symbiotically with other fields.

Question: what do these logos relate to?





Question: what do these logos relate to?





- ▶ Answer: the Meltdown [9] and Spectre [8] security vulnerabilities, which
 - 1. can be broadly classified as micro-architectural side-channel attacks, and, as such,
 - 2. demand deep understanding of processor (micro-)architecture to either
 - mount (i.e., use), or
 - prevent.

https://www.meltdownattack.com/meltdown.png https://www.spectreattack.com/spectre.png













Question: ranging from the late 1990s to the early 2000s, spot the difference(s).





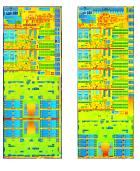


Answer:

- left-hand picture is *the* Google data center circa 1997,
- ▶ right-hand picture is *a* Google data center circa 2007

suggesting efficient software stacks demand care with respect to computer architecture: a holistic approach to CS is unavoidable.

Question: identify these objects, and spot the difference(s).



Question: identify these objects, and spot the difference(s).



Answer:

- the left-hand picture is a dual-core Haswell model Intel processor, whereas
- the right-hand picture is a quad-core Haswell model Intel processor,

which highlights a trend: parallelism and concurrency are inherent, suggesting that understanding, coping with, and exploiting them are all vital.

git # b282dbb9 @ 2025-09-03

https://download.intel.com/newsroom/kits/core/4thgen/gallery/images/

Question: given the list of operations (circa 2010)

L1 cache access Branch misprediction

L2 cache access

Memory access

Read 1 MB sequentially from memory

IP packet round-trip (local area network)

Disk seek

Read 1 MB sequentially from disk

IP packet round-trip (internet)

estimate the associated latencies (i.e., how long they take)?

Question: given the list of operations (circa 2010)

L1 cache access	0.5 ns
Branch misprediction	5.0 ns
L2 cache access	7.0 ns
Memory access	100.0 ns
Read 1 MB sequentially from memory	250000.0 ns
IP packet round-trip (local area network)	500000.0 ns
Disk seek	10000000.0 ns
Read 1 MB sequentially from disk	20000000.0 ns
IP packet round-trip (internet)	150000000.0 ns

estimate the associated latencies (i.e., how long they take)?

git # b282dbb9 @ 2025-09-03

Question: given the list of operations (circa 2010)

```
L1 cache access
                                                             0.5 \, \mathrm{ns}
                                                                              5.0 \cdot 10^{-10} \text{ s}
                                                             5.0 ns
                                                                        = 5.0 \cdot 10^{-9}
Branch misprediction
                                                                        = 7.0 \cdot 10^{-9}
L2 cache access
                                                             7.0 ns
                                                          100.0 ns
                                                                        = 1.0 \cdot 10^{-7}
Memory access
                                                      250000 0 ns
                                                                              2.5 \cdot 10^{-4}
Read 1 MB sequentially from memory
IP packet round-trip (local area network)
                                                      500000 0 ns
                                                                              5.0 \cdot 10^{-4}
Disk seek
                                                   10000000.0 ns
                                                                        = 1.0 \cdot 10^{-2} \text{ s}
                                                                              2.0 \cdot 10^{-2} s
Read 1 MB sequentially from disk
                                                   200000000.0 ns
                                                  150000000.0 ns
                                                                              1.5 \cdot 10^{-1} s
IP packet round-trip (internet)
```

estimate the associated latencies (i.e., how long they take)?

Part 3: unit motivation, i.e., the "why" (8) Low-level applications of computer architecture

Question: given the list of operations (circa 2010)

L1 cache access	0.5 ns	=	$5.0 \cdot 10^{-10} \text{ s}$	~	1 s	econds
Branch misprediction	5.0 ns	=	$5.0 \cdot 10^{-9}$ s	\simeq	10 s	econds
L2 cache access	7.0 ns	=	$7.0 \cdot 10^{-9} \text{ s}$	\simeq	10 s	econds
Memory access	100.0 ns	=	$1.0 \cdot 10^{-7}$ s	\simeq	3 n	ninutes
Read 1 MB sequentially from memory	250000.0 ns	=	$2.5 \cdot 10^{-4}$ s	\simeq	5	days
IP packet round-trip (local area network)	500000.0 ns	=	$5.0 \cdot 10^{-4} \text{ s}$	\simeq	11	days
Disk seek	10000000.0 ns	=	$1.0 \cdot 10^{-2}$ s	\simeq	231	days
Read 1 MB sequentially from disk	20000000.0 ns	=	$2.0 \cdot 10^{-2}$ s	\simeq	462	days
IP packet round-trip (internet)	1500000000.0 ns	=	$1.5 \cdot 10^{-1}$ s	\simeq	3472	days

estimate the associated latencies (i.e., how long they take)?

- Answer: these operations stem from standard components;
 - knowing what they are, plus
 - understanding how they work, i.e., why the absolute and relative latencies are as listed, is an important step toward using them effectively, or improving their design or

is an important step toward using them effectively, or improving their design or implementation.

Part 3: unit motivation, i.e., the "why" (9)

Low-level applications of computer architecture

Quote

Where a calculator on the ENIAC is equipped with 18,000 vacuum tubes and weighs 30 tons, computers in the future may have only 1,000 vacuum tubes and perhaps weigh 1.5 tons.

- Popular Mechanics

Component	Then	Now
Processor	thousands of instructions/sec	billions of instructions/sec
Memory	hundreds of bits	gigabytes
Storage	thousands of bytes	terabytes
Input/Output	paper tape	anything you can imagine
Software	hand wired	high-level languages

Question: these are valid C programs for x86-32 processors; what do they do, and why write them like this?

```
Listing (C)

1 int weight( uint32_t x ) {
2   int t = 0;
3
4   for( int i = 0; i < 32; i++ ) {
5     if( (x >> i) & 1 ) {
6     t += 1;
7   }
8   }
9
10   return t;
11 }
```

```
Listing (C)

1 int weight( uint32_t x ) {
2   int t;
3
4   asm (" movl $0, %0; movl $32, %%ecx; "
5   "0: decl %%ecx ; bt %%ecx,%1 ; "
6   " adcl $0, %0; test %%ecx,%%ecx; "
7   " jnz 0b ;
8   : "=&r" (t) : "r" (x) : "%ecx", "cc" );
10
11   return t;
12 }
```

Question: these are valid C programs for x86-32 processors; what do they do, and why write them like this?

```
Listing (C)

1 int weight( uint32_t x ) {
2 int t = 0;
3
4 for( int i = 0; i < 32; i++ ) {
5 if( (x >> i ) & 1 ) {
6 t += 1;
7 }
8 }
9 }
10 return t;
11 }
```

```
Listing (C)

1 int weight( uint32_t x ) {
2   int t;
3   4   asm ( " mov1 $0, %0 ; mov1 $32, %%ecx ; "
5       "0: decl %%ecx ; bt %%ecx,%1 ; "
6       " adcl $0, %0 ; test %%ecx,%%ecx ; "
7       " jnz 0b ; "
8       : "=&r" (t) : "r" (x) : "%ecx", "cc" );
10
11   return t;
12 }
```

- ▶ Answer: both compute the Hamming weight of a 32-bit integer x;
 - C doesn't provide you with (direct) access to some things a processor can do,
 - two examples here are adcl (or "add-with-carry") and bt (or "bit test") instructions which means understanding
 - what a processor can do irrespective of the language, and
 - use of any non-standard language features

are important aspects of writing better programs.

Question: these programs sum the elements in an $(n \times m)$ -element matrix A; which one is faster, and why?

```
Listing (C)

1 int sum( int n, int m, int A[n][m]) {
2 int t = 0;
3
4 for( int i = 0; i < n; i++) {
6 for( int j = 0; j < m; j++) {
7 }
8 }
9
10 return t;
11 }
```

```
Listing (C)

1 int sum( int n, int m, int A[ n ][ m ] ) {
2   int t = 0;
3
4   for( int j = 0; j < m; j++ ) {
5      for( int i = 0; i < n; i++ ) {
6          t += A[ i ][ j ];
7      }
8    }
9
10   return t;
11 }</pre>
```

• Question: these programs sum the elements in an $(n \times m)$ -element matrix A; which one is faster, and why?

```
Listing (C)

1 int sum( int n, int m, int A[ n ][ m ] ) {
2   int t = 0;
3
4   for( int i = 0; i < n; i++) {
5      for( int j = 0; j < m; j++) {
6      t += A[ i ][ j ];
7   }
8   }
9   }
10   return t;
11 }
```

- ► Answer: the left-hand one;
 - C uses a row-major layout for 2-dimensional arrays,
 - row-major access to x results in sequential access in memory, which is more efficient if a cache memory is used

which means understanding

- how compilers work to produce low-level programs, and
- how the memory hierarchy works

are important aspects of writing better programs.

Part 3: unit motivation, i.e., the "why" (12) Low-level applications of computer architecture

Question: you want to print n integers held in an array A to the terminal, tab or new line delimitation is possible; which is better and why?

```
Listing (C)

1 void write( FILE* F, int* A, int n ) {
2 for( int i = 0; i < n; i++ ) {
3 fprintf( F, "%d\t", A[ i ] );
4 }
5 }

Listing (C)

1 void write( FILE* F, int* A, int n ) {
2 for( int i = 0; i < n; i++ ) {
3 fprintf( F, "%d\n", A[ i ] );
4 }
5 }
```

Part 3: unit motivation, i.e., the "why" (12) Low-level applications of computer architecture

• Question: you want to print *n* integers held in an array *A* to the terminal, tab or new line delimitation is possible; which is better and why?

```
Listing (C)

1 void write( FILE* F, int* A, int n ) {
2 for( int i = 0; i < n; i++ ) {
3 fprintf( F, "%d\t", A[ i ] );
4 }
5 }

Listing (C)

1 void write( FILE* F, int* A, int n ) {
2 for( int i = 0; i < n; i++ ) {
3 fprintf( F, "%d\n", A[ i ] );
4 }
5 }
```

- Answer: if better means faster the left-hand tab delimited one;
 - the C standard library is more complex than the interface suggests,
 - printf has a user-space buffer when stdout is connected to a terminal, which reduces the number of system calls (i.e., interaction with the kernel)

which means understanding

- how your program interfaces with the the kernel, and
- how the kernel interfaces with physical hardware are important aspects of writing better programs.

Part 3: unit motivation, i.e., the "why" (13) Linkage with historical lessons and effective communication

Question: a string data structure is important; Pascal and C take different approaches, i.e.,

and

but how was the choice made during development of C?

Part 3: unit motivation, i.e., the "why" (13) Linkage with historical lessons and effective communication

 Question: a string data structure is important; Pascal and C take different approaches, i.e.,

$$i = \ldots, 3, 4, 5, 6, 7, 8, \ldots$$
 $MEM = \langle \ldots, 5, 104, 101, 108, 108, 111, \ldots \rangle$
 $CHR(MEM[i]) = \ldots, ENQ, 'h', 'e', 'l', 'l', 'o', \ldots$

and

but how was the choice made during development of C?

- Answer: *many* reasons, e.g.,
 - memory footprint was a real issue, so allowing larger strings via a larger P-string length was unattractive, and
 - the PDP-11 had native support for ASCIIZ strings, meaning the choice of C-string was partly dictated by hardware,

suggesting that understanding *historical* design decisions can be relevant and useful in *modern* contexts!

Part 3: unit motivation, i.e., the "why" (14) Linkage with historical lessons and effective communication

Question: have a look at this 1956 advert for the UNIVAC computer

https://www.youtube.com/watch?v=Pd63MHGQygQ

and compare it with modern adverts of the same type.

Question: have a look at this 1956 advert for the UNIVAC computer

https://www.youtube.com/watch?v=Pd63MHGQygQ

and compare it with modern adverts of the same type.

Observation: users and terminology have changed a lot

U.	NΙ	VAC
----	----	-----

... takes business statistics from magnetic tape ... processing them at phenomenal speeds ... computes payrolls electronically and produce printed cheques, over 8000 cheques an hour.

Intel Core2 Duo

... combines two independent processor cores in one physical package ... processors run at the same frequency and share up to 6MB of L2 cache and up to 1333MHz Front Side Bus for truly parallel computing.

suggesting that effective communication via correct notation and terminology is important.

Conclusions

Take away points: this unit

- 1. is challenging and does demand hard work,
- 2. isn't exclusively about hardware, and
- 3. will equip you with the knowledge and skills required to
 - design better hardware,
 - design better software,
 - design better systems (e.g., combinations of hardware and software),
 - make better trade-offs,
 - understand complex behaviour,
 - debug complex behaviour,
 - think critically,
 - think "in parallel",
 - -

and, ultimately, solve important and interesting problems (in the short- and longer-term).

Additional Reading

- Wikipedia: Computer architecture. URL: https://en.wikipedia.org/wiki/Computer_architecture.
- A.S. Tanenbaum and T. Austin. "Section 1.1: Structured computer organisation". In: Structured Computer Organisation. 6th ed. Prentice Hall, 2012.
- A.S. Tanenbaum and T. Austin. "Section 1.2: Milestones in computer architecture". In: Structured Computer Organisation. 6th ed. Prentice Hall, 2012.
- A.S. Tanenbaum and T. Austin. "Section 1.3: The computer zoo". In: Structured Computer Organisation. 6th ed. Prentice Hall, 2012.
- W. Stallings. "Chapter 2: Computer evolution and performance". In: Computer Organisation and Architecture. 9th ed. Prentice Hall. 2013.
- W. Stallings. "Chapter 3: A top-level view of computer function and interconnection". In: Computer Organisation and Architecture. 9th ed. Prentice Hall, 2013.

References

- [1] Wikipedia: Computer architecture. URL: https://en.wikipedia.org/wiki/Computer_architecture (see p. 50).
- [2] W. Stallings. "Chapter 2: Computer evolution and performance". In: Computer Organisation and Architecture. 9th ed. Prentice Hall, 2013 (see p. 50).
- [3] W. Stallings. "Chapter 3: A top-level view of computer function and interconnection". In: Computer Organisation and Architecture. 9th ed. Prentice Hall, 2013 (see p. 50).
- [4] A.S. Tanenbaum and T. Austin. "Section 1.1: Structured computer organisation". In: Structured Computer Organisation. 6th ed. Prentice Hall, 2012 (see p. 50).
- [5] A.S. Tanenbaum and T. Austin. "Section 1.2: Milestones in computer architecture". In: Structured Computer Organisation. 6th ed. Prentice Hall, 2012 (see p. 50).
- [6] A.S. Tanenbaum and T. Austin. "Section 1.3: The computer zoo". In: Structured Computer Organisation. 6th ed. Prentice Hall, 2012 (see p. 50).
- [7] A. Huang and S. Cross. "Novena: a laptop with no secrets". In: IEEE Spectrum 52.11 (2015), pp. 40–56 (see pp. 15, 16).
- [8] P. Kocher et al. "Spectre Attacks: Exploiting Speculative Execution". In: IEEE Symposium on Security and Privacy. 2019, pp. 19–37 (see pp. 25, 26).
- [9] M. Lipp et al. "Meltdown: Reading Kernel Memory from User Space". In: USENIX Security Symposium. 2018, pp. 973–990 (see pp. 25, 26).

