Engineering Tripos Part IIA THIRD YEAR

Module 3F5: Computer and Network Systems

Solutions to 2016 Tripos Paper

Authors: Andrew Gee and Tim Wilkinson

1. **Carry-lookahead adders**

(a) Amdahl's Law says that it is of paramount importance to optimise the speed of those components that are most frequently used in a computer system ("make the common case fast"). For most instruction set architectures, ALUs are used at least once in the execution of just about every instruction. Faster ALUs would therefore have a significant effect on the speed of the overall system. [10%]

(b) Carry lookahead can be used to determine the carry inputs to each full adder without using ripple carry. For each bit i of the adder, we define two signals, generate gi and propagate pi. Bit i generates a carry if the two bits it is adding are both 1, and propagates a carry if either of the two bits it is adding is 1:

$$g_i = a_i \cdot b_i \qquad p_i = a_i + b_i$$

c1, the carry into bit 1, will be 1 if either bit 0 generates a carry or c0 is 1 and bit 0 propagates a carry:

$$c_1 = g_0 + p_0 \cdot c_0$$

Likewise for c2 and c3:

$$c_2 = g_1 + p_1 \cdot g_0 + p_1 \cdot p_0 \cdot c_0$$
$$c_3 = g_2 + p_2 \cdot g_1 + p_2 \cdot p_1 \cdot g_0 + p_2 \cdot p_1 \cdot p_0 \cdot c_0$$

These expressions show how the carry-in signals can be obtained after $1 + 2 = 3$ gate delays. The full adders themselves require a further two gate delays, giving a total of five gate delays. [25%]

(c) (i) $p_1 = a_1 + b_1$ $g_1 = a_1 \cdot b_1$ [5%]

(ii) $P_0 = p_0 \cdot p_1$ $G_0 = g_1 + g_0 \cdot p_1$ [10%]

(iii) $c_2 = G_0 + P_0 \cdot c_0$ [5%]

(iv) $c_3 = g_2 + p_2 \cdot c_2$ $c_4 = G_1 + P_1 \cdot c_2$ [15%]

c3 and c4 are ready in $1 + 2 + 2 + 2 = 7$ gate delays. The full adders themselves require a further two gate delays, giving a total of nine gate delays. [10%]

(d) At first sight it appears that (b) wins on all counts: it is apparently faster, and since each level of carry lookahead requires more gates, it is also more compact. However, note how the expressions in (b) require gates with a fan-in of four, whereas those in (c) require a fan-in of only two. Suppose that the 4-input gates are implemented by chaining together three 2-input gates, for example p2.p1.p0.c0 = p2.(p1.(p0.c0)). This would make the 4-input gates three times slower than the 2-input gates. Leaving aside the full adders, which are common to both designs, the design in (b) would have all the carries ready after $1+2\times 3=$ seven 2-input gate delays, which is the same time it takes the design in (c) to calculate c3 and c4. More generally, the larger the fan-in the slower the gate, so it is not inconceivable that the design in (c) is as fast as the design in (b). [20%]

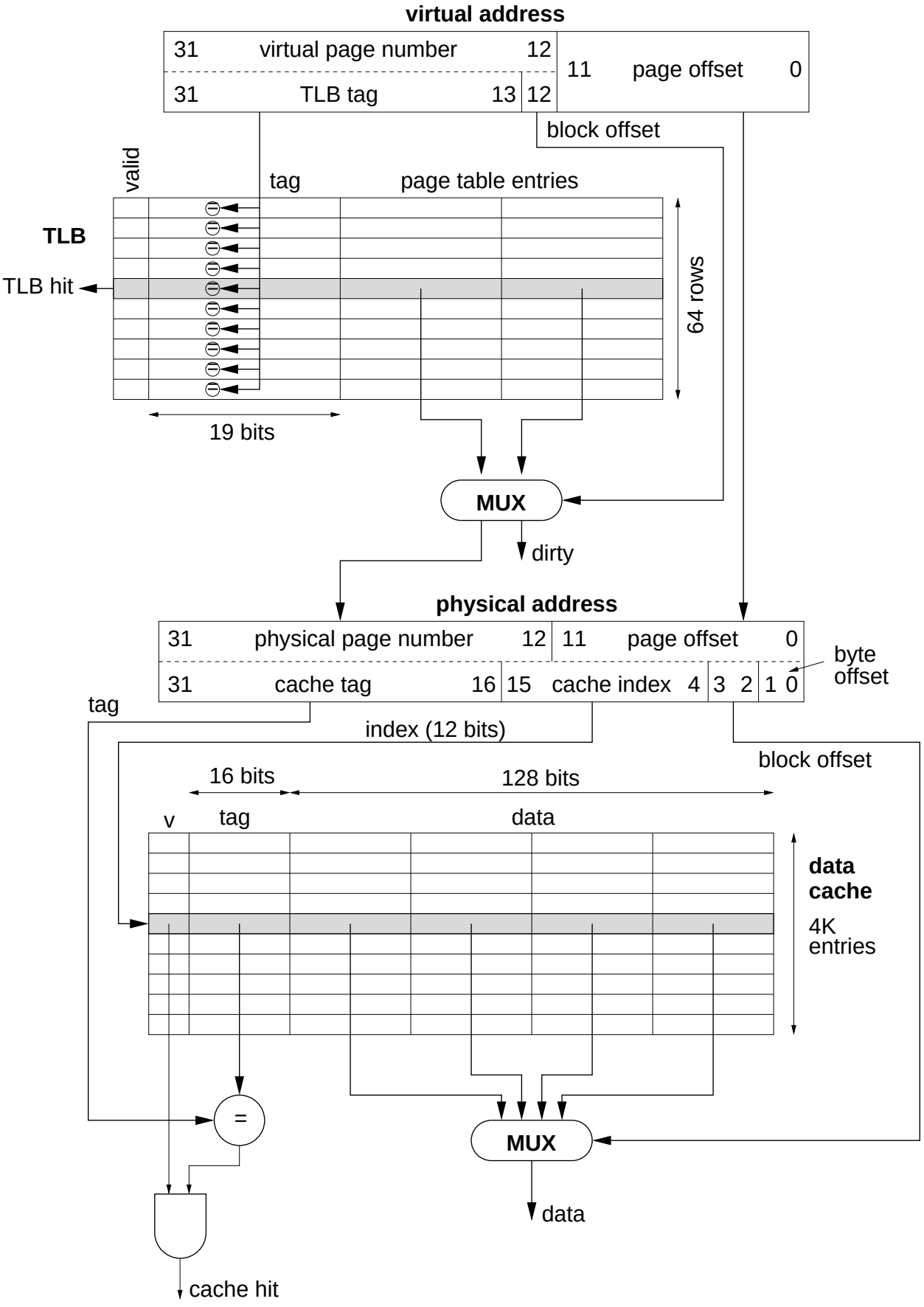
2. **Caches, TLBs and virtual memory**

(a) Cache access is faster than main memory access because of its physical proximity to the CPU and its construction out of static RAM, as opposed to slower dynamic RAM. Data still needs to be fetched from main memory to the cache, but this overhead is easily amortized through temporal locality of reference (so a fetched item is likely to be accessed again soon, and this time it will be in the cache) and spatial locality of reference (so a fetched item's neighbours are likely to be accessed soon, so fetch *blocks* of data at a time, paying the main memory latency price just once). [15%]

(b) Direct mapped caches have the lower hit time, since the index points to a unique block and no searching of the cache is required. However, since there is no choice of which block to replace on a miss, we might replace blocks that are going to be referenced again soon: this will lead to a high miss rate. In a set-associative cache, the index points to a (typically small) set of blocks that must be searched, increasing the hit time. But there is some flexibility as to which block to replace on a miss. We could, for example, consider temporal locality of reference and replace the least recently used (LRU) block, thereby reducing the miss rate. [20%]

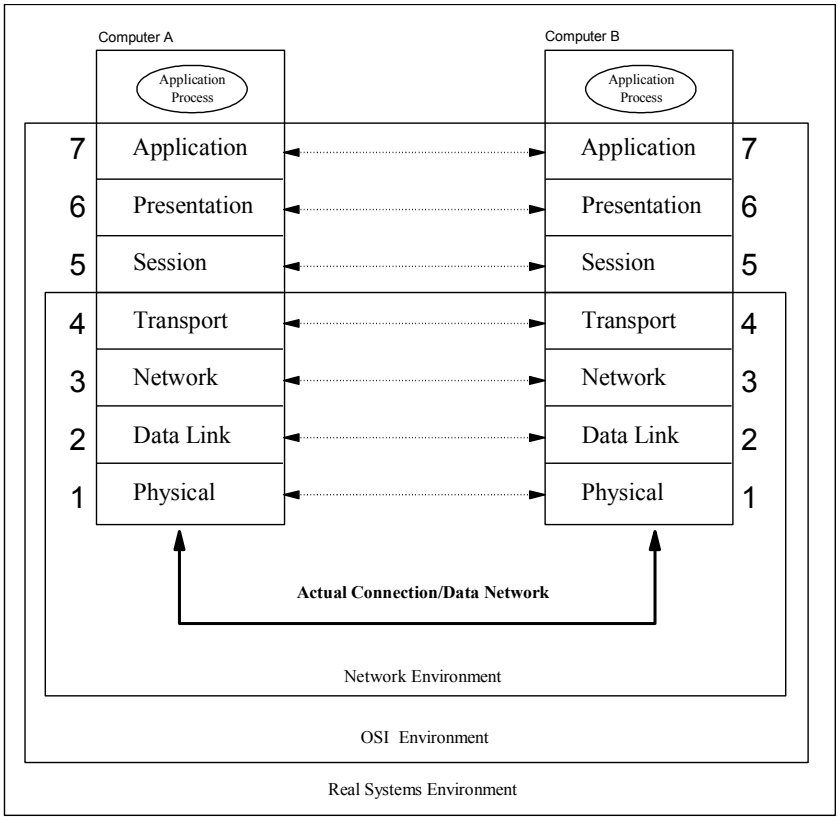
(c) There are two main requirements that motivate the adoption of a virtual memory system. The first is the desire to be able to write programs without having to worry about the amount of physical memory installed in the computer. The second is the need for the CPU to execute multiple processes separately: each process should be unaware of, and protected from, the others. [15%]

(d) If the page size is 4 KBytes ($=2^{12}$), the page offset must be 12 bits, leaving 20 bits for the virtual and physical page numbers. The TLB is fully associative so there is no index: all 64 rows must be searched. Each TLB block contains two page table entries, so there is a 1-bit block offset at the end of the virtual page number. The remaining 19 bits of the virtual page number comprise the TLB tag. The data cache is 64 KBytes large, which is 16K words or 4K ($=2^{12}$) blocks. The index into the cache is therefore 12 bits. There is a block offset of 2 bits and a byte offset of 2 bits, leaving 16 bits for the cache tag. A hardware schematic of the TLB and cache can be found on the next page. [50%]


virtual address
31
virtual page number
12
11
page offset
0
31
TLB tag
13
12
block offset
valid
tag
page table entries
TLB
TLB hit
64 rows
19 bits
MUX
dirty
physical address
31
physical page number
12
11
page offset
0
byte offset
31
cache tag
16
15
cache index
4
3
2
1
0
tag
index (12 bits)
block offset
16 bits
128 bits
v
tag
data
data cache
4K entries
=
MUX
data
cache hit

Question 3

a) For successful data communication across a network, appropriate operating procedures must be established. They must be specified in detail and strictly adhered to by the sending data terminal (or computer) and any intervening switching centres. These procedures are called protocols. Many local area networks (LANs) interconnect data terminals (or computers) from the same manufacturer and operate using proprietary protocols. The need arose for communication between computers and terminals from different manufacturers. Open systems interconnect (OSI), to enable networks to be machine independent.
Computers may use different languages, data formats and operating systems; hence, the interface between user (application) programs and underlying communications services may be different. The ISO standards are based on a seven layer protocol known as the ISO reference model for OSI. Both the network dependent and application oriented (network independent) components of the OSI model are in turn implemented in the form of a number of protocol layers. The boundaries and processes for each layer have been selected based on experience gained from other standards in the past.



The OSI model can be broken up within this system into 7 separate layers, each with a clearly defined purpose and protocol.

- Each layer is a service user to the layer below and a service provider for the layer above.
- Each layer is specified independently of the other layers;
- Each layer has a defined interface with the layer above and below.

As far as users are concerned, communication appears to take place across each layer. Each data exchange passes down to the bottom layer (the physical layer) at the sending terminal, crosses the network to the receiving terminal and then passes up again.

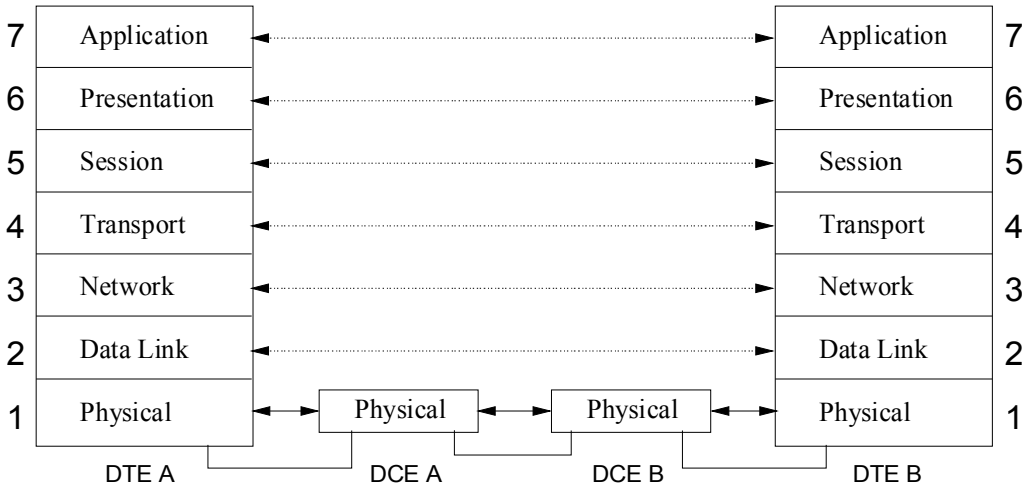
Data communication between layers is done through the addition and reading of headers on the data

The reference model gives rise to three different environments:

(1) **The network environment**. The protocols and standards relating to the different types of underlying data communications networks.
(2) **The OSI environment**. Adds additional application oriented protocols and standards to allow end systems to communicate with one another in an open way.
(3) **The real systems environment**. Builds on the OSI environment and is the manufacturers own proprietary software and services.

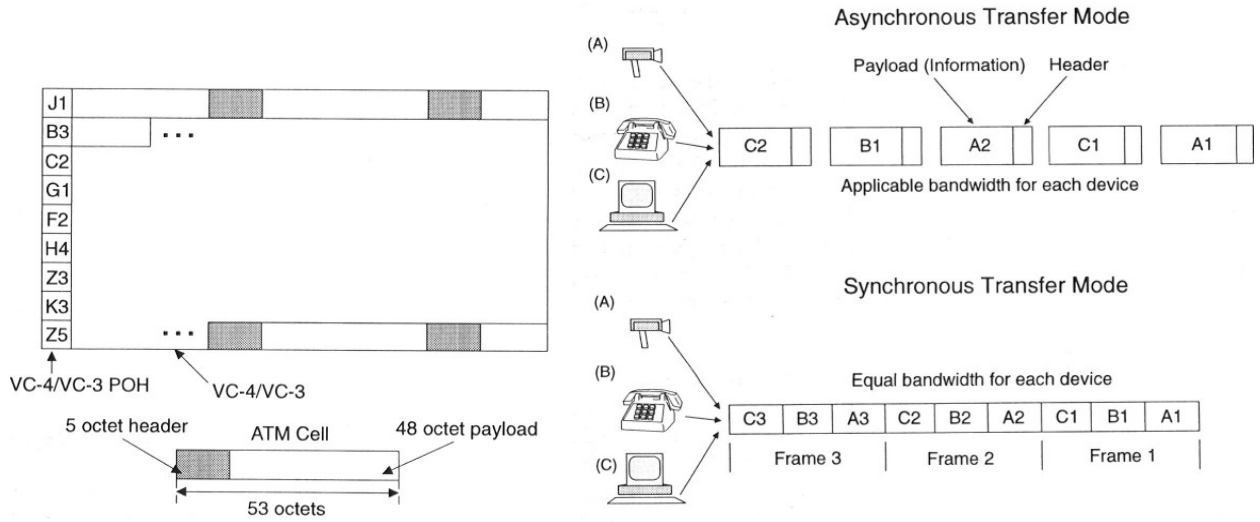
b) When working with the lowest level of the OSI model (physical layer), it is often useful to add a common communications device such as a modem to utilise the available communications channel. It is useful to split the lower end of the OSI model into two parts:

- The data terminal equipment (DTE)
- The data circuit equipment (DCE) or data circuit terminating equipment.



This makes for a common DTE/DCE interface standard which can be kept OSI compliant through its defined physical properties. Within this standard interface it is possible to allow a DCE/DCE connection which is non OSI based. This allows extra proprietary features to be added within the physical layer which may only be compatible with a single equipment manufacturer or telecommunication service provider. These services can be quite complicated (even including higher layer processes) such as traffic monitoring, statistics etc as long as they stay wholly within the DCE/DCE interface. However the overall DTE/DTE physical layer process must remain entirely transparent.

c) ISDN is the integration of both telephony and data into the same network. But this network will eventually have to cope with extreme data rates and bursty systems hence there is a need for a network which is capable of emulating both packet switching and circuit switching and to emulate both packet and circuit performance to provide both bandwidth on demand as well as constant bandwidth services.



The solution was the section of the B-ISDN standard called asynchronous transfer mode or ATM which was designed to be specifically compatible with SDH and run at the same base data rate of 155Mb/s. Hence a single ATM cell will occupy 53 bytes per SDH frame. Cell relay and its modern equivalent Asynchronous Tranfer Mode (ATM) is similar to frame relay but with a much smaller cell. In ATM, the cell is limited to 53bytes, a 48 byte payload and a 5 byte header. ATM is defined at a transmission rate of 155.52Mbits/sec to be compliant with SDH and SONET.

The mapping of ATM onto the OSI model is difficult, as there are layer splits.

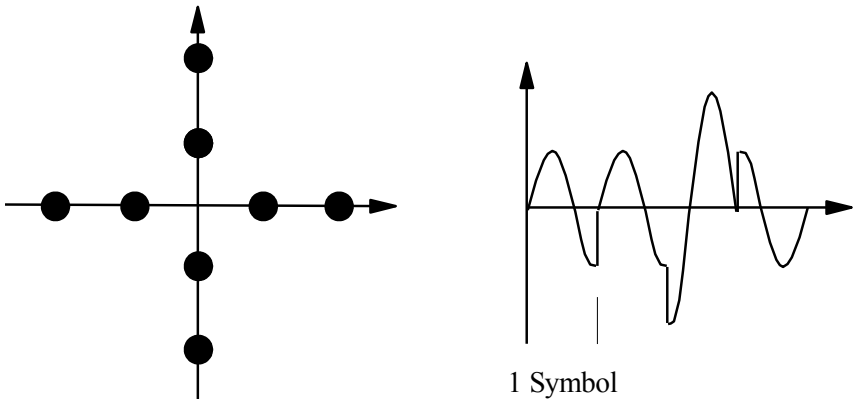
OSI	ATM	ATM Sublayer
3/4	AAL	SAR / CS
2/3	ATM	ATM
2	Physical	TC
1		PMD

The ATM layer deals with cells and cell transport. It defines the header layout and functionality. The ATM adaptation layer (AAL) takes packets that are larger than cells and segments them. The AAL is split into the segmentation and reassembly (SAR) sublayer and the convergence sublayer (CS). The SAR/CS split allows different network services like file transfer and video on demand that have different error correction requirements and quality of service.

The physical layer has more functionality than in the OSI model and is split into the physical medium dependent (PMD) sublayer and the transmission convergence (TC). The short packet length means that in theory each cell could be individually routed. This is not very efficient as the header is too small. In reality a cell is sent initially to set up all the ATM switches in readiness for a sequence of cells to be transmitted.

Question 4

a) A dial-up modem allows a data rate that exceeds the 3.4kHz data rate set by twisted pair wiring and digital telephone network. This is done using clever modulation schemes other than direct binary amplitude or phase down the modem channel. A common technique is quadrature amplitude modulation (QAM). By combining amplitude and phase modulation it is possible to build up a constellation of data points that are transmitted by the modem at a physical baud rate less that 3.4kHz.

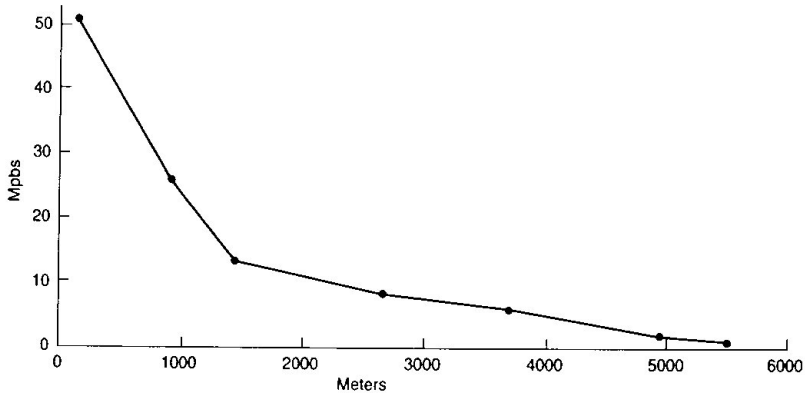


8 Symbols, 2 amplitudes
4 phases

1 Symbol

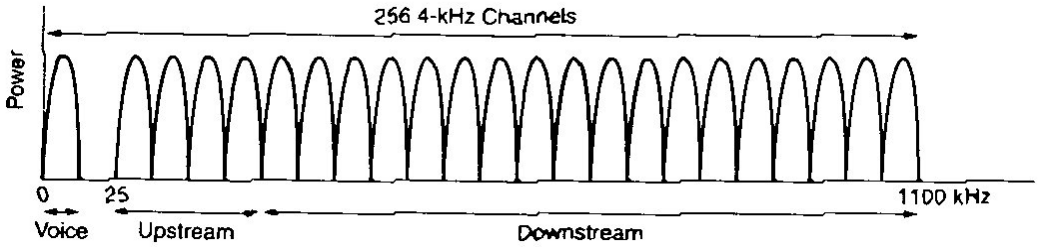
The choice of constellation must balance out the number of phase steps and amplitude levels. A strong constellation should have a minimum number of amplitude levels and a quantised even phase step. Much higher transmission rates can be achieved using more complex constellations in conjunction with error correction and compression algorithms.

The main limitation on the bandwidth of a Cat 3 UTP telephone line is filter placed at the exchange end before the voice signal is digitised. This limitation means that a normal modem must squeeze all of the data through a 3.7kHz wide band. The Cat 3 UTP is in fact capable of transmitting data over short distances at frequencies in excess of 1.1MHz, which is a property exploited by digital subscriber line modems which bypass the filter in the line card at the exchange and access all of the available line bandwidth.



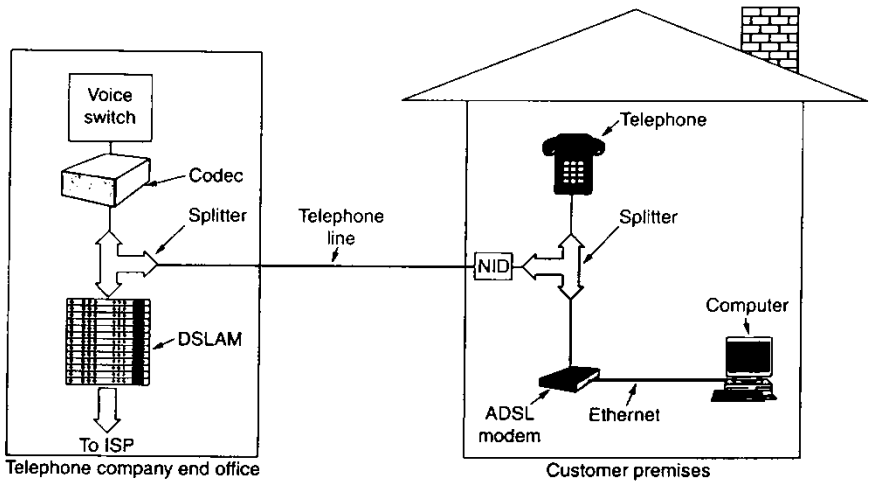
A further enhancement on modem technology is the digital subscriber line (DSL) which allows the transmission digital data at a very high rate over standard twisted pair cables up to a distance of around 5km. This is often referred to as a ‘broadband’ service by Telcos such as NTL and BT. There

are a variety of standards such as VDSL, HDSL and xDSL, but one of the more common is the asymmetric digital subscriber line (ADSL) which typically offers a 512bit/sec downstream and 64kbit/sec upstream or 1Mbit/sec downstream, 256kbits/sec upstream for premium services (eg BT and NTL). The problem is that the data rate or BW is limited the distance from the local exchange, hence the further away, the lower the BW and the more limited the available services.

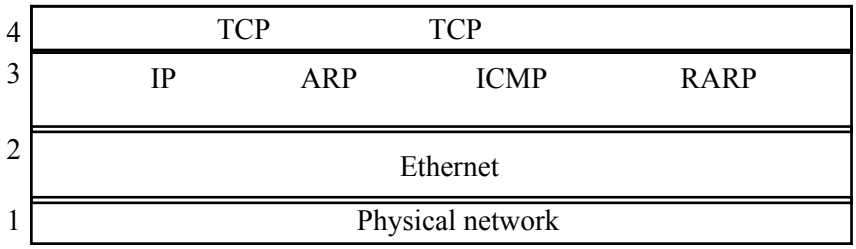


The transmission coding is very complex and often adaptive. In the discrete multi-tone (DMT) system the basic 1.1MHz bandwidth of the line is split into 245 4kHz wide channels. The first is for voice services, followed by a gap of 5 before the next 250 which are used for data. The mix of downstream and upstream channels can vary which gives the asymmetry in ADSL. This is because the majority of BW is used in downstream operations and uses the available BW more efficiently. A modulation scheme similar to V.34 (4000 baud) is used in each channel and each channel is monitored and the bitrate adapted to suite optimise the performance and compensate for and peculiarities in the line. With a high quality line a bitrate of over 10Mb/sec downstream is often possible.

b)



A typical ADSL system is shown above with a network interface device (NID) set at the boundary before a splitter filters off the voice services. The cable modem then connects to the computer via a standard such as Ethernet with a higher layer protocol like TCP/IP on top. The ADSL signals are passed through a digital subscriber line access multiplexer (DSLAM) which combines the channels and formulates packets for the data to be transmitted.



It the heart of the protocol stack is the internet protocol (IP) which is roughly equivalent to an OSI layer 3 protocol and the transport control protocol (TCP) an approximate equivalent of OSI layer 4.

The mapping of TCP/IP onto the OSI model is not ideal, as there are some layer 2 processes done in both TCP and IP and IP has no real provision for end to end connection management. IP and its addressing gives the potential for a software application or a user connected anywhere in the world to the internet to access any other computer or software application.

The transport control protocol (TCP) is usually used in conjunction with IP and the internet control message protocol (ICMP) to guarantee reliable transmission. On an end-to-end basis, TCP ensures correct sequencing of arriving frames of data and requests retransmission when necessary. The address resolution protocol (ARP) is used in association with IP to translate the IP addresses into physical hardware addresses. Thus, ARP is capable of determining the appropriate ethernet LAN address (MAC).

The most common layer 2 network is ethernet. The original design was based on a length of coaxial cable, with 'tee-offs' to individual work stations, with a maximum of around 500 stations. The Ethernet MAC sub layer frame can take two possible forms. The preamble/SFD, address and frame check sequence (FCS) are common to both types. They are referred to as *type* and *length encapsulation* frame formats. The key to MAC layer 2 protocols is a uniform globally unique hardware addressing structure so that LAN hardware can easily identify stations on the network and transmit packets between them. A MAC address is 48 bit long (6 bytes). Hence the Ethernet identifies the hardware structure of the communication link and is then mapped onto IP to perform the networking structure. This is then run through TCP to allow a basic form of external management to a fundamentally connectionless protocol system.

c) The main reason for the use of these protocols is the process of network evolution. Ethernet evolved out of the LAN expansions and uses a MAC address to define separate hardware configurations. It also evolved from the use of shared media which eventually collapsed into a switched backbone with structured cabling. However along with the MAC address comes a lot of redundant LAN features which leads to a long overhead of frame size and therefore less efficient use of BW. In a LAN this is to critical due to high data rates, but over ADSL this BW is limited by the properties of the phone line. Ethernet has been adopted by the majority of telcos as the layer 2 choice and there is a large degree of legacy equipment in place. It also allows for higher level address recycling through systems like NAT and DHCP.

IPv4 has also evolved over many decades as the layer 3 protocol of choice. The IP address has become standardised and is globally acknowledged as the best means of network routing and

interconnection. The main limitation of IPv4 is that it is fundamentally a connectionless protocol and so there is very little control of the connection at the packet level. ICMP was invented as a means of controlling packet connections, but it has not become globally accepted and remains a proprietary add-on to IPv4 and is often ignored or can lead to packets not being routed in the fats path. IPv4 also has a lot of other redundant features such as time to live, header checksums and fragmentation all of which reduces the BW efficiency of IP over line like ADSL. The lack of control at the IP level also head to the evolution of TCP which further adds to the length of the packets with extra fields, flows, windows etc and reduces the ADSL BW efficiency.

As a result of this long evolutionary process there is a lot of redundant features in the ethernet/IP/TCP combination. A better way to utilise ADSL BW efficiently would be to redesign the protocol stack to weed out all the unnecessary features.

i) Remove the use of ethernet and the MAC address. It is possible to switch directly with the IP address especially with DHCP. This is more efficient but will limit the compatibility with potential LAN products at the end of the ADSL connection. In most homes this is often wireless which can operate locally without the need for a layer 2 overhead beyond the router.
ii) Switch IPv6 which has more addresses and alleviates the need for MAC addresses and also has a much more sophisticated control process which could be used more efficiently than TCP. The problem is that there are still relatively few networks which operate fully at the IPv6 level and so there is a big risk of non-compatibility with other network nodes and connections.

iii) Design a whole new protocol from scratch or adapt an older point to point protocol to maximise the efficiency of the ADSL BW usage. The will be very difficult to do in reality due to a lack of compatibility and will threaten overall transparency

Assessors' remarks for question 1: This question tested the candidates' knowledge of carry-lookahead adders. All candidates could explain the importance of ALU latency in (a), and most offered good descriptions of single level carry-lookahead adders in (b). Marks dropped in (b) were generally down to careless slips when calculating gate delays (e.g. neglecting the latency of the full adders), though several candidates described multi-level carry-lookahead schemes and were marked down accordingly. Answers to (c) were pleasing, with most candidates able to formulate the hierarchical carry-lookahead expressions, with marks dropped mostly for gate delays as in (b). (d) was less well answered, though it was pleasing to see a handful of candidates identifying gate fan-in and latency as the crux of the matter.

Assessors' remarks for question 2: This question tested the candidates' knowledge of caches and virtual memory systems. (a)–(c) were very well answered, with almost all candidates able to explain the role of the cache, the relative advantages and disadvantages of direct mapped and set-associative caches, and the rationale for virtual memory. Answers to (d) were more variable, with a good number of perfect responses but others revealing fundamental misunderstandings, such as the presence of an index into a fully associative TLB. Attention to detail was also sometimes lacking when calculating the lengths of the various address fields.

Assessors' remarks for question 3: Overall a well answered question, with (a) being mostly book work. Many candidates failed to point out that the OSI system uses headers and footers for inter-layer communication. The DTE/DCE interface was understood well but few suggested reasons (such as MPLS) why it might be useful. The ATM section was not so well answered with only about half the candidates realising the difficulties in mapping it onto the OSI model.

Assessors' remarks for question 4: This question was only answered by a few candidates, perhaps because it strayed from the usual book work format. Most got the ADSL section right although only a few realised why the asymmetry leads to increased bandwidth. Answers to (b) were quite varied, but most candidates showed that the combination of TCP/IP and Ethernet involved a large overhead. Unfortunately, not many took this into (c) to conclude that there is a lot of bandwidth wasted on protocol management which does not apply to an ADSL link. Nobody spotted that it was in fact a point-to-point link system.

Andrew Gee & Tim Wilkinson
May 2016