SECTION A

Introductory Business Economics B1 (Paper 8)

2013-2014

CRIBS

Exam setter: Dr. Andrea Mina

SECTION A *Introductory Business Economics B1 (Paper 8)*

CRIBS

1 (a) Use an appropriate diagram or diagrams to represent the industry demand and supply curves, and the individual firm's demand and cost curves in a perfectly competitive market. On this basis:

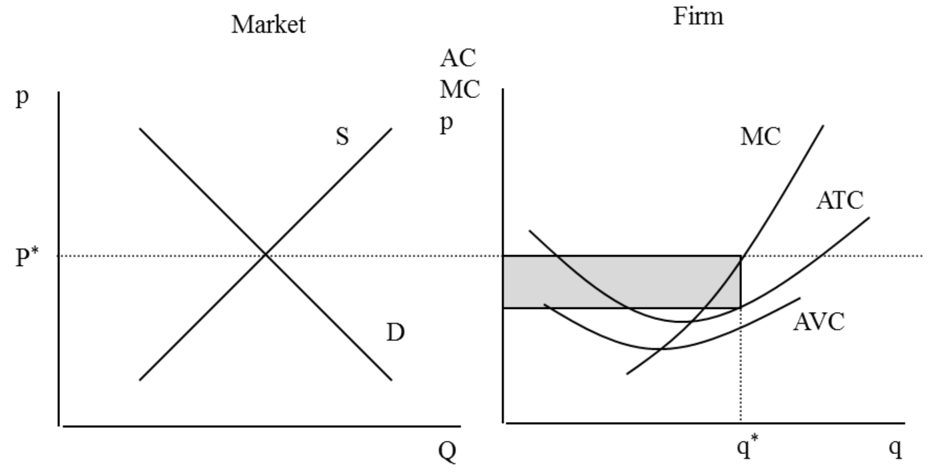
(i) Identify the point at which the firm's profits are maximised and explain your answer; [5]

(ii) Define the firm's shut-down condition in the short and in the long run. [5]

A perfectly competitive market has the following theoretical characteristics:

- Many suppliers each with an insignificant share of market;
- Each firm is too small to affect price via a change in market supply. Each individual firm is therefore a price taker;
- Each firm produces homogeneous undifferentiated products (Consumers perceive the products to be identical);
- Consumers have perfect information about the prices all sellers in the market charge and there are no transaction costs;
- All firms have equal access to resources (such as technology and inputs) and there are no barriers to entry and exit of firms.

The market demand curve is represented according to the standard 'law of demand', while the demand curve of the individual firm reflects the fact that the firm has no influence on price, which is set by the market at equilibrium. The students should use the following diagrams, complete with the curves that describe the cost structure of the firm, to represent these concepts.



The point at which profits (grey area in the graph) are maximised reflects the standard MR=MC condition, bearing in mind that in a perfectly competitive market MR=AR=P. Therefore the profit maximising output (q*) is identified at the intersection between the price level and the MC curve.

In the sort run – where the firm incurs both fixed and variable costs – the firm will shut down when it is unable to cover its variable costs (TR < VC, equivalent to P < AVC). The firm has to cover its fixed costs even it is making negative profits. It can minimise losses by producing as long as it can cover its variable costs. In the long run – where all costs are variable – the firm will shut down a P < AC.

(b) Under what circumstances may a monopoly be more desirable for consumers than perfect competition? [5]

The microeconomic case against monopoly is that by taking the market demand curve as its own, the monopolist is able to exert power over the setting of price and/or output. In so doing it can earn abnormal profits by appropriating consumer surplus at the expense of economic efficiency, which also implies that the product is being under-consumed. However, there are cases when monopolies can generate advantages for consumers that a perfectly competitive market may not be able to deliver. These include:

- The ability to incur large R&D expenditures that a perfectly competitive firm is not able to afford;
- The possibility to avoid duplication of fixed costs (an inefficient use of resources) under the constraints of a natural monopoly
- The exploitation of economies of scale, which reduce unit costs, and these cost advantages can be reflect in prices;
- International competitiveness (a degree of local monopoly power may help the firm to be very competitive on international markets).

(c) Compare and contrast the concepts of 'comparative' and 'competitive' advantage in international trade. [10]

The theory of comparative advantage explains why economies may improve their welfare by trading with each other even if one country has absolute cost advantages. The fundamental principle is that trade depends on the possibility and advantages (cost and benefits) of *exchange*. The availability of natural resources, costs of production factors or know-how generate differences in the productive capacities of different economies. Although specialisation may lead to absolute advantages in production we should consider the opportunity costs of production and identify activities where disadvantage is least or advantage is best, so that all countries will benefit from trade. Candidates may present a simple two-country, two-good case (as discussed in lectures with reference to the production and trade of wine vs. computers).

The theory of comparative advantage is a good demonstration of the benefits of trade, but also has limitations.

- Not all countries will be equally well off as gains from trade may not be evenly distributed across all trading countries. In addition, it is possible that the terms of trade may deteriorate over time;
- Not all citizens may benefit to the same degree within countries;
- There are risks of excess specialisation, especially in the context of a developing economy.

Contrary to the theory of comparative advantage, the concept of 'competitive' advantage emphasises that 'advantage' cannot be considered as a static property of an economy, a sector or a firm: it cannot be taken for granted and can be deliberately created, for example by investing in new knowledge and technology. For the advantages of trade to persist it is important that an economy fosters opportunities for growth by making strategic choices in the development and use of distinctive resources (natural, human, and technological). The better students may be able to provide appropriate historical and or/comparative examples.

2 (a) What factors can determine the degree of market power of a firm? [5]

The degree of market power is a key dimension of market structure. It specifically involves the ability of firms to make decisions about prices. It derives from the existence of entry barriers, and is highest when there is only one supplier on the market and no close substitutes to a product or service.

Barriers can be:

- Structural (or 'Innocent'), due to differences in production costs;
- Statutory, that is to say entry barriers given force of law (National Lottery or television and radio broadcasting licences)
- Strategic. These may be deliberately looked for and designed to block potential entrants from entering a market (making the market 'less contestable'), and to protect the monopoly power of the incumbent and therefore its supernormal profits. They may involve pricing decisions (e.g. limit o predatory pricing), advertising, marketing or R&D investments.

The better students will be able to identify other interesting examples such as the control of some scarce resources, patents, vertical integration or brand development.

(b) With reference to the concept of price discrimination:

(i) Identify and explain the different types of price discrimination; [5]

(ii) Explain the conditions that make it possible for a firm to apply price discrimination. [5]

In a standard monopoly model it is assumed that a firm sells all units of output at the same price. However, differences in consumer's willingness to pay (both between different consumers and between different units) mean that a monopolist often seeks to appropriate as much consumer surplus as possible by applying price discrimination.

The three types of price discrimination are:

- First-degree (or perfect) price discrimination: the monopolist sells different units of output for different prices and these prices may differ from person to person; this enables the monopolist to extract as much surplus as possible from each consumer, for example through a bargaining process (this enables the monopolist to discover the reservation price of each consumer).
- Second-degree price discrimination: the monopolist sells different units of output for different prices, but each individual who buys the same amount of the good pays the same price (e.g. bulk discounts for large purchases);
- Third-degree price discrimination: the monopolist sells output to different consumers for different prices, but every unit of output sold to a given person sells for the same price (e.g. student discounts at the cinema, subscription prices to academic journals).

The better students will be able to produce and discuss the diagrams for each type of price discrimination as treated in the lectures.

For a firm to apply price discrimination techniques it has to be able to exert some degree of market power (no flat demand curve). Secondly, markets must be separate and consumers must not be able to resell the product across market segments. Finally, demand elasticity must differ between markets for second- and third-degree price discrimination. For third-degree price discrimination it is also necessary that the firm is able to identify *ex ante* the two submarkets (contrary to second-degree p.d. where consumers self-select into different demand curves).

(c) Illustrate the idea of the 'circular flow' of the macroeconomy and identify the relationship between injections and withdrawals at equilibrium. [10]

The students should start by explaining that the 'circular flow' is a way of representing how resources and associated spending move in the macroeconomy. To begin with, households own and supply factors of production to firms (namely their labour), receive income in return, and spend income on goods and services. In turn, firms use factors of production supplied by households to produce goods and services, pay households for these factors, and sell goods and services to households. The part of income not spent goods and services constitutes saving. Savings exit the circular flow, and are channelled back into the system by financial institutions (e.g. through loans to firms) and result in investment (new capital goods). Imports and exports complete the framework in an open economy.

Having distinguished between direct and indirect taxes, and defined the concept of transfer payment, all 'withdrawals' (or 'leakages', i.e. net savings, net taxes and import expenditure) and all 'injections' (i.e. investments, government expenditure and export expenditure) are identified and can be located in the context of the circular flow.

At equilibrium injections should be equal to withdrawals. Therefore, $S + T + Z = I + G + X$. Injections, however, do not always match withdrawals and it is possible that the economy will present budget surpluses or deficits (compare T with G), trade surpluses or deficits (compare Z with X), and positive or negative net capital outflows (compare S and I). The better students will be able to define and briefly discuss each of these concepts.

3

Solution

(a)

(i) A tunnel's stability in clay is determined by its *stability ratio*, N, defined as

$$N = (\sigma_v - \sigma_t)/s_u$$

where σ_v = total vertical pressure at tunnel axis level = γ z
(γ = unit weight, z = depth)
σ_t = tunnel support pressure (if any, = 0 if open face)
s_u = undrained strength of the clay at tunnel axis level

If the value of N is *less than about 5* the tunnel face will be stable.

In zone A s_u = 200 kPa at a depth of 25m. Total vertical pressure σ_v = 25 x 20 = 500 kPa. Hence for open face tunnelling (σ_t = 0) N = 500/200 = 2.5. This is less than 5, hence open face tunnelling feasible.

In zone B s_u = 50 kPa at a depth of 25m. Total vertical pressure σ_v = 25 x 20 = 500 kPa. Hence for open face tunnelling (σ_t = 0) N = 500/50 = 10. This is much greater than 5, hence open face tunnelling is not feasible because tunnel would not be stable.

(ii) Low permeability soils are *clays*, high permeability soils are *sands and gravels*, with silts in between. If tunnelling in clays, the permeability is low enough for there to be *no time for drainage* (unless tunnelling is halted) and therefore the *undrained strength* is often high enough to ensure temporary stability of the tunnel face. However, if tunnelling in sands and gravels below the water table, the *water will flow into the face*, causing collapse and de-stabilising the tunnel.

Potential problems in tunnelling below the water table in sands and gravels can be overcome by (a) lowering the water table by *pumping from wells* installed for the purpose (b) *injecting grout* into the ground in advance of tunnelling – usually chemical grouts – to reduce the permeability (c) using *compressed air* in the tunnel – all of (a), (b) and (c) enable open face tunnelling to proceed.

(b) *Closed face* tunnelling machines exert a pressure on the face as the soil is being excavated. The two types of closed face tunnelling machines are *slurry* machines or *earth pressure balance* machines. Both tunnelling machines cut the soil by a rotating wheel. *Slurry* machines involve a *bentonite* slurry that is circulated under pressure into the cutterhead: the excavated soil and slurry are taken by *pipes to the ground surface*, where the soil is separated from the slurry which is then re-circulated. The pressure on the soil face is maintained by the pressurised slurry. *Earth pressure balance* (EPB) machines maintain a pressure on the soil face by means of the excavated soil being mixed with *conditioning agents* and confined to a pressurised chamber. The pressurised soil is removed from the pressurised chamber by means of a *screw conveyor*, which allows the pressure to drop to atmospheric pressure so that it can taken out of the tunnel by a belt conveyor.

(c) *Segmental linings*. These are commonly used for lining circular tunnels, constructed with tunnelling machines, either open-face or closed-face. The segments are usually made out of *pre-cast concrete*, but sometimes from *SGI (Spheroidal Graphite Iron)*. Advantages: made in factory under carefully controlled conditions, relatively easy to handle, erected within tunnelling machine, *robust*, very rare for collapse to occur. Disadvantages: usually only OK for circular tunnels, therefore *lack of flexibility on shape*, difficult to vary thickness

Sprayed concrete linings(SCL). Sometimes known as *NATM (New Austrian Tunnelling Method)*. Concrete sprayed onto excavated soil surface, accelerators added, *hardens rapidly*, usually with *steel fibres added or light reinforcement mesh.* Advantages: very *versatile*, particularly for *station* construction (eg Crossrail), can easily change thickness, *excavated shape*. Disadvantages: needs careful quality control, *susceptible to poor workmanship*, collapse of tunnels has occurred.

(d) Masonry buildings are particularly susceptible to *differential* settlement and cracking is associated with *tensile strain*. Buildings subjected to *hogging* deformation are more susceptible than those subject to *sagging*, because the tensile strains tend to be induced in the top of the building whereas in the sagging zone they are in the foundations. *Compensation grouting* involves injection of grout into the ground between the tunnel and the building foundation in a *controlled* manner. The principal aim is to reduce the potential *differential* settlement of the building, thereby limiting damage. The grout is injected from *tube-à-manchettes* (TAM's) which are installed in the ground *before tunnelling*, usually from an adjacent shaft. *Instrumentation* is installed on the building (levelling and/or electrolevels) and in the ground (extensometers) to monitor settlement and ground movements, and the grout is injected *in response to the measurements*.

4 Solution

(a) Diaphragm walls have two principal advantages for this application: they provide a water tight barrier to keep the excavation dry (which is critical when undertaking excavations in sand below the water table) and if propped near or at the top of the wall they cause minimum subsidence behind the wall. They are formed from reinforced concrete and become part of the main basement walls of the new building. The process involves excavation of alternating panels along the proposed wall using bentonite slurry to prevent the sides of the excavation collapsing. Construction starts with the installation of shallow concrete or steel guide walls. The excavation is then undertaken using a clamshell grab (in the case of soils). Bentonite slurry is pumped into the trench to provide temporary support and a prefabricated reinforcing cage is lowered to the bottom of the completed trench. Concrete is then placed at the bottom of the trench using a tremie pipe, displacing the bentonite slurry, until the entire panel is full of concrete. The sequence is repeated with the next panel.

(b)

Vertical stresses

Total vertical stress at the ground surface $\sigma v = 30$ kPa.

Total vertical stress at the water table (2m depth) $\sigma v = 30 + 2 \times 16 = 62$ kPa

Total vertical stress at the sand-clay interface (8 m depth) $\sigma v = 30 + 2 \times 16 + 6 \times 18$
$= 170$ kPa

Total vertical stress at 12 m $\sigma v = 30 + 2 \times 16 + 6 \times 18 + 4 \times 17 = 238$ kPa

Effective vertical stress at the ground surface $\sigma'v = \sigma v - u = 30 - 0 = 30$ kPa

Effective vertical stress at the water table (2m depth) $\sigma'v = \sigma v - u = 62 - 0 = 62$ kPa

Effective vertical stress at the sand-clay interface (8 m depth) $\sigma'v = \sigma v - u = 170 - 6 \times 10$
$= 110$ kPa

Effective vertical stress at 12 m $\sigma'v = \sigma v - u = 238 - 10 \times 10 = 138$ kPa.

Active side in sand

Active pressure coefficient, Ka = $(1-\sin\phi)/(1+\sin\phi) = (1-\sin35°)/(1+\sin35°) = 0.27$

Pressure at the ground surface $\sigma_h = Ka\sigma'_v + u = 0.27 \times 30 + 0 = 8.1$ kPa

Pressure at the water table (2m depth) $\sigma_h = Ka\sigma'_v + u = 0.27 \times 62 + 0 = 16.74$ kPa

Pressure at the sand-clay interface in the sand (8 m depth) $\sigma_h = Ka\sigma'_v + u$

$= 0.27 \times 110 + 6 \times 10 = 89.7$ kPa

Active side in clay

Pressure at the sand-clay interface in the clay (8 m depth) $\sigma_h = \sigma_v - 2cu = 170 - 2 \times 75$

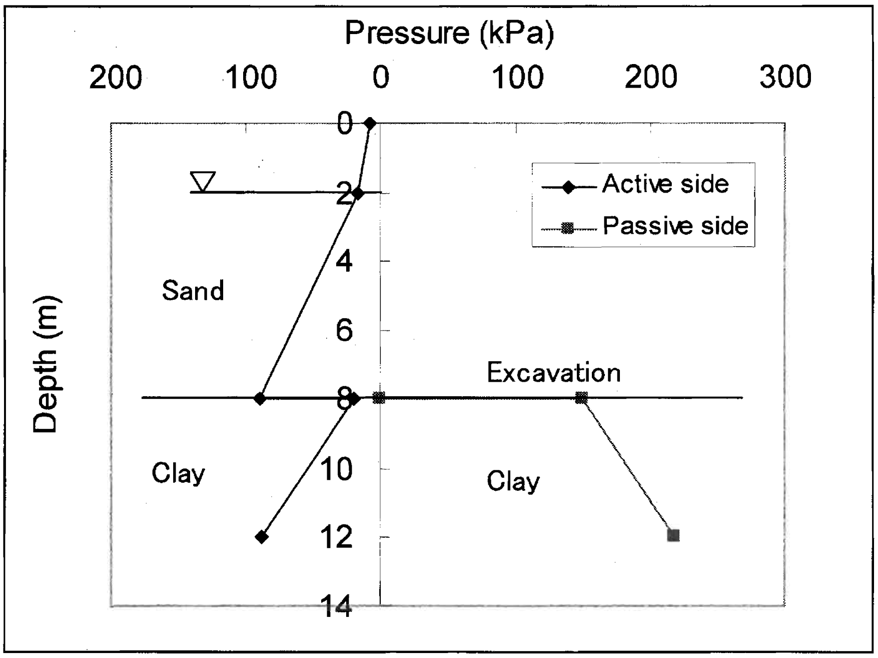
$= 20$ kPa

Pressure at 12 m $\sigma_h = \sigma_v - 2cu = 238 - 2 \times 75 = 88$ kPa

Passive side in clay

Pressure at the excavation level (8 m depth) $\sigma_h = \sigma_v + 2cu = 0 + 2 \times 75 = 150$ kPa

Pressure at 12 m $\sigma_h = \sigma_v + 2cu = 17 \times 4 + 2 \times 75 = 218$ kPa



(c)

Taking moments about the prop:

Active moment (driving) = (8.1 x 2 x 1) + [0.5 x (16.7-8.1) x 2 x (2 x 2/3)] + (16.7 x 6 x 5)

+ [0.5 x (89.7 – 16.7) x 6 x (2 + 2/3 x 6)] + (20 x 4 x 10) + [0.5 (88 -20) x 4 x (8 + 2/3 x 4)] kN-m/m

= 16.2 + 11.5 + 501 + 1314 + 800 + 1450

= 4093 kN-m/m

Passive moment (resisting) = (150 x 4 x 10) + [0.5 (218 -150) x 4 x (8 + 2/3 x 4)] kN-m/m

= 6000 + 1450

= 7450 kN-m/m

Factor of safety = total resisting moment/ total driving moment = 7450/4093 = 1.82

(d)

Under drained conditions, passive pressure coefficient Kp in clay = (1+ sin25°)/(1-sin25°)

= 2.45

Horizontal passive pressure in clay zero at bottom of excavation increasing linearly with depth

Total vertical stress at depth of 4m σv = 4 x 17 = 68 kPa

Water pressure at depth of 4m, u = 4 x 10 = 40 kPa

Effective vertical stress at depth of 4m $\sigma' v = \sigma v - u$ = 28 kPa

Horizontal passive pressure at depth of 4m = (Kp x $\sigma' v$) + u = (2.45 x 28) + 40 = 108.6 kPa

Passive moment (resisting) provided by clay therefore reduced to:

0.5 x 108.6 x 4 x (8 + 2/3 x 4) = 2317 kN-m/m

Factor of safety = 2317/4093 = 0.57

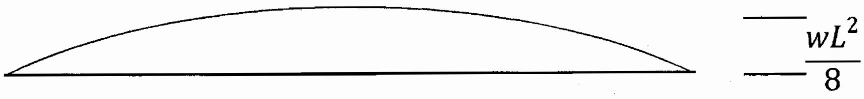
This shows that the wall would fail (by rotation about the prop) if the project was delayed long enough for the clay on the passive side of the wall to drain via the sand layers thereby reducing its passive resistance.

5 **Solution**

a)

Condition 1: Simply Supported

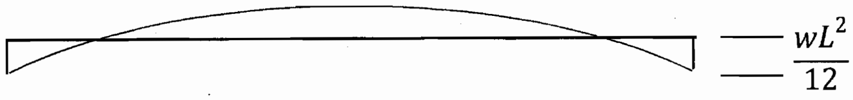
Moment in middle = $\frac{wL^2}{8} = \frac{180 \times 14^2}{8} = 4{,}410\ KNm/m$



Condition 2: Fixed on both sides

Moment at supports: $\frac{wL^2}{12} = \frac{180 \times 14^2}{12} = 2{,}940\ KNm/m$

Moment in middle = 4,410 – 2,940 = 1,470 KNm/m



b)

At C need to design for singly reinforced with moment 4,410 KNm/m

Maximum moment for singly reinforced = $0.15 \times f_{cu} \times bd^2 = 4{,}410\ KNm$

Given $b = 1{,}000\ mm\ and\ f_{cu} = 40\ \ => \ d^2 \geq \frac{4{,}410 \times 10^6}{40 \times 0.15 \times 1{,}000} => d \geq 857\ mm$

c)

Overall depth = 1,800 mm

Cover = 50 mm

Rebar assume 40 mm

Effective depth = 1,800 – 50 – 40/2 = 1,730 mm

Assume x = 0.5

Then $A_s = \frac{4{,}410 \times 10^6}{0.87 \times 460 \times 1{,}730\left(1 - \frac{0.5}{2}\right)} = 8{,}493\ mm^2/m$

Steel force = $0.87 \times f_y \times 8{,}493 = 3{,}399\ KN/m$

Concrete stress = $0.4 \times 40 = 16\ MPa$

Depth at neutral axis = d_n, $d_n \times 1000 \times 16 = 3{,}399 \times 10^3 => d_n = 212\ mm$

Corrected lever arm = 1,730 – 212/2 = 1,624 mm

A_s required = $\frac{4{,}410 \times 10^6}{0.87 \times 460 \times 1{,}624} = 6{,}785 \frac{mm^2}{m} < 8{,}493 \frac{mm^2}{m}$

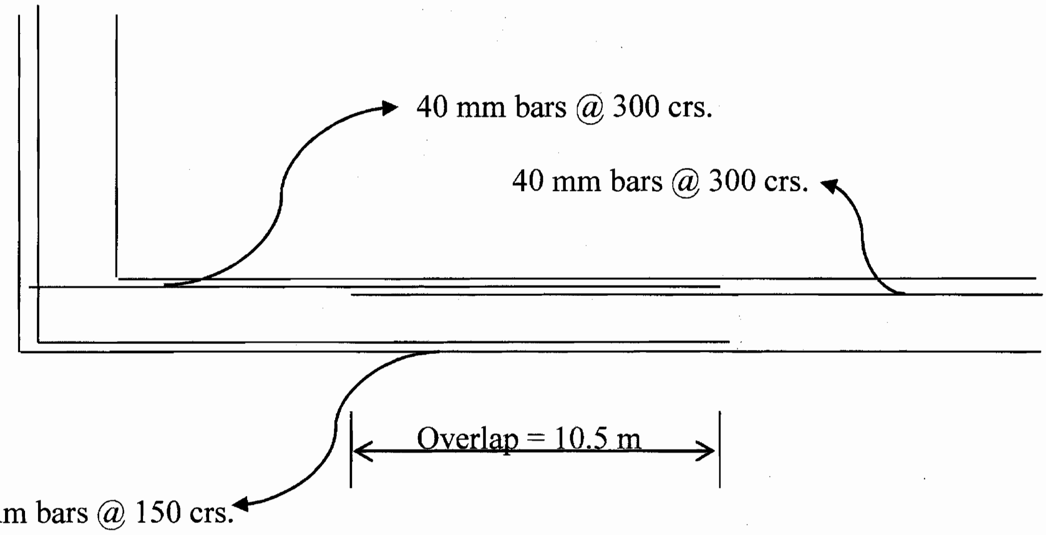
Could iterate again, but not worth it.

At A & B use 2/3 of this (4,523 mm^2/m)

40 mm bars, area $=\frac{\pi d^2}{4} = 1{,}256\ mm^2, need\ \frac{6{,}785}{1{,}256} = 6\frac{bars}{m} = 40mm\ bars\ @\ 150\ crs.$ at C

At A & B could use 32 mm bars (area = 64% of 40 mm bars) at same crs.

d)



e) Structures in ground need to have very little cracking. Minimum thickness only ensures strength. Extra thickness will reduce cracking.

Engineering Tripos Part 1B

Paper 8, Selected Topics, Section C

Crib 2013/14 (M Sutcliffe)

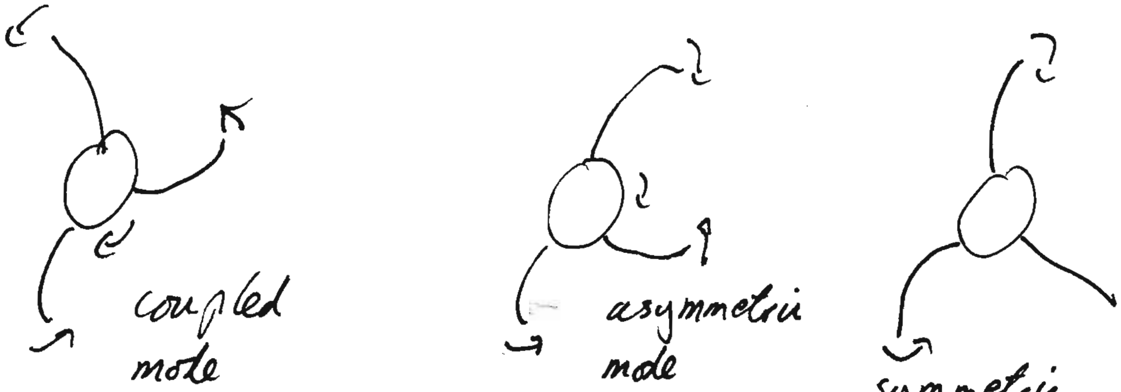
Q6 (a) Loading - need to consider fatigue loads associated with bending, vibration and self-weight
Materials - bonding, eg of composites to metals might be a source of weakness, in fatigue.
Detailed design - need to check load paths, stress concentrations (as fatigue is very sensitive to stress raisers)

(b)

→ Displacements, forces → Equilibrium for each blade and overall

- model blades as single-degree-of-freedom systems linking blade loads to hub loads via a spring element

Construct $\underline{m}\ddot{\underline{x}} + \underline{k}\,\underline{x} = 0$ matrix for free vibration, and use eigenvalue / vector analysis $[\underline{m}^{-1}\underline{k}\,\underline{x} = \omega^2 \underline{x}]$ to get resonant frequencies and modes.



coupled mode

asymmetric mode

symmetric mode

6(c) No. of cycles in stress range $S \to S + dS$

$$= N_{TOT}\, \phi(S)\, dS$$

Use Miner's rule so that lifetime used up sums to 1

proportion of life used up

$$\Rightarrow 1 = \int_0^\infty \frac{N}{N_f}\, dS = N_{TOT} \int_0^\infty \phi(S)\, S^m S_0^{-m}\, dS$$

These steps to start the question missed by many

$$= N_{TOT}\, S_0^{-m} \int_0^\infty \frac{S^m}{\bar{S}} \exp\left(-\frac{S}{\bar{S}}\right) dS$$

Put $t = S/\bar{S}$

$dt = \frac{1}{\bar{S}}\, dS$

$$\Rightarrow 1 = N_{TOT}\, \frac{S_0^{-m}}{\bar{S}} \int_0^\infty t^m \bar{S}^m \exp(-t)\, \bar{S}\, dt$$

$$= N_{TOT} \left(\frac{\bar{S}}{S_0}\right)^m \int_0^\infty t^m \exp(-t)\, dt$$

$$\Rightarrow N_{TOT} = \left(\frac{S_0}{\bar{S}}\right)^m \frac{1}{10!} = \frac{(30)^{10}}{10!}$$

$$\text{Lifetime} = N_{TOT}/10^7 = 16.3 \text{ years}$$

(d)
- obtain turbine loading, eg from site wind loads or standard wind spectrum plus dead weight loading
- need structural model to infer bolt loading from external loads
- use rain flow analysis to derive pdf from resulting time history.

7 (a) Economic and political drivers - e.g. government incentives, changes in tax regime, uncertainty in investment, increased pressure for C02 emissions reductions, increase in fuel costs.

Technical drivers - e.g. improvements in small and large wind efficiency, large offshore reliability, transmission costs, need for good wind, change in other energy sources (e.g. nuclear).

Societal - acceptance of small wind farms. Movement off-shore of large installations. Acceptance of climate change argument.

Overall the factors are likely to lead to larger offshore or Scottish developments. Not clear about home production, not likely to make much impact so perhaps won't be incentivised.

b) Capacity factor is defined as Annual energy production/Annual energy production operating continuously at rated output. Since the revenue generated by a wind turbine is related to its actual energy production, whereas its capital cost is typically proportional to its rated output, capacity factor essentially measures revenue/capital cost. Hence a high capacity factor suggests improved economic viability.

c) (i) Putting the numbers into $P = 0.5C_p\rho Av3$ gives for the rated power, and $\lambda = \omega R/v$ to obtain the angular speed:

Option 1: $P = 0.5\times0.35\times1.23\times\pi\times40^2\times12^3 = 1.87$ MW and $\omega = 8\times12/40 = 2.4$ rads^{-1}
Option 2: $P = 0.5\times0.35\times1.23\times\pi\times40^2\times16^3 = 4.43$ MW and $\omega = 8\times16/40 = 3.2$ rads^{-1}

(ii) Complete the table below for both options.

Option 1

Wind speed (ms^{-1})	No of days	Hours	Power (MW)	Energy (MWhr)
<3	20	480	0	0
8	200	4800	0.554	2659
12	90	2160	1.87	4039
16	40	960	1.87	1795
>20	15	360	0	0

Total energy = 8493 MWhr.

7(cont)

Option 2

Wind speed (ms^{-1})	No of days	Hours	Power (MW)	Energy (MWhr)
<3	20	480	0	0
8	200	4800	0.554	2659
12	90	2160	1.87	4039
16	40	960	4.43	4253
>20	15	360	0	0

Total energy = 10951 MWhr.

(iii) For option 1 the capacity factor is 8493/(365×24×1.87) = 0.518.

For option 2 the capacity factor is 10951/(365×24×4.43) = 0.282.

Option 2 has just over half the capacity of option 1, suggesting that the return on capital would be similarly reduced. Thus option 1 has the greater economic viability.

c) Doubly-fed induction generators have the following advantages: they allow the system to operate at continuously variable speeds over a limited range, thereby optimising the energy harvested from the wind; they only require a fractionally-rated bi-directional converter to achieve this; they are robust, efficient and cheap to manufacture.

[A well answered question, with a high average. Typical issues were listing factors but not commenting on changes in installation in (a), and failing to note the no-power regimes and the cap on power produced as the rated power in (c).]

8 (a)(i) ρ - density of air

A - swept area (πR^2) of turbine

V_0 - free-stream air speed (relative to turbine)

(ii) $P = N u$, where $u = V_0(1-a)$ is the velocity at the rotor plane

$$C_{NA} = \frac{N}{\frac{1}{2}\rho A V_0^2} = \frac{P/u}{\frac{1}{2}\rho A V_0^2}$$

$$= \frac{\frac{1}{2}\rho A V_0^3 \times 4a(1-a)^2}{\frac{1}{2}\rho A V_0^2 \times V_0(1-a)}$$

$$= 4a(1-a)$$

Many candidates derived this using a control volume analysis, which was more involved.

(b) No losses, driving force = axial force on turbine

$\Rightarrow P = N v_{car}$ ← this step not always spotted

$\therefore \frac{1}{2}\rho A V_0^3 C_p = \frac{1}{2}\rho A V_0^2 C_{NA} V_{CAR}$

$\Rightarrow V_0 C_p = C_{NA} V_{CAR}$

Motion of car into wind $\Rightarrow V_0 = V_{WIND} + V_{CAR}$

$\Rightarrow (V_{wind} + V_{car})\, 4a(1-a)^2 = 4a(1-a)\, V_{CAR}$

$$\left[1 + \frac{V_{CAR}}{V_{wind}}\right] (1-a) = \frac{V_{CAR}}{V_{WIND}}$$

$$\frac{V_{CAR}}{V_{WIND}} = \frac{1-a}{a}$$

8 (c) (i) $a = 1/3 \Rightarrow \dfrac{V_{CAR}}{V_{WIND}} = \dfrac{2/3}{1/3} = 2$

$\Rightarrow V_0 = V_{WIND} + V_{CAR} = \dfrac{V_{CAR}}{2} + V_{CAR} = V_{CAR} \cdot \dfrac{3}{2}$

Tip speed ratio $\lambda = \dfrac{\omega R}{V_0} = 8$ (NB V_0, not V_{wind})

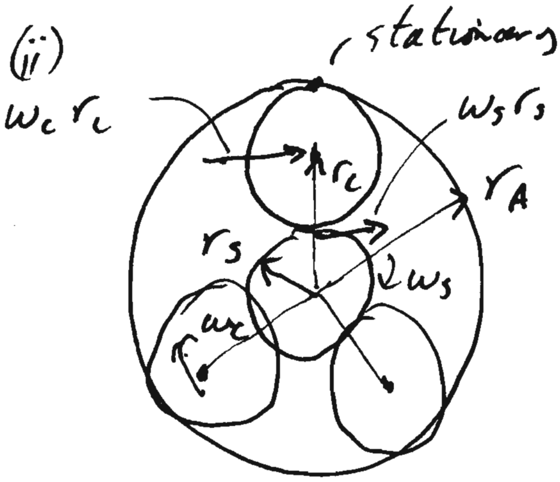
wheel speed $V_{CAR} = \Omega r$ (Ω = wheel angular velocity)

Thus $\dfrac{\omega}{\Omega} = \dfrac{\lambda V_0 / R}{V_{CAR}/r} = \lambda \dfrac{V_0}{V_{CAR}} \cdot \dfrac{r}{R} = 8 \cdot \dfrac{3}{2} \cdot \dfrac{1}{3}$

$= \underset{\sim}{4}$

[Required gear ratio]

(ii)



Epicyclic, turbine $\omega = \omega_s$

wheel $\Omega = \omega_c$

Kinematics (on top planet)

$\omega_c r_c = \omega_s r_s / 2$

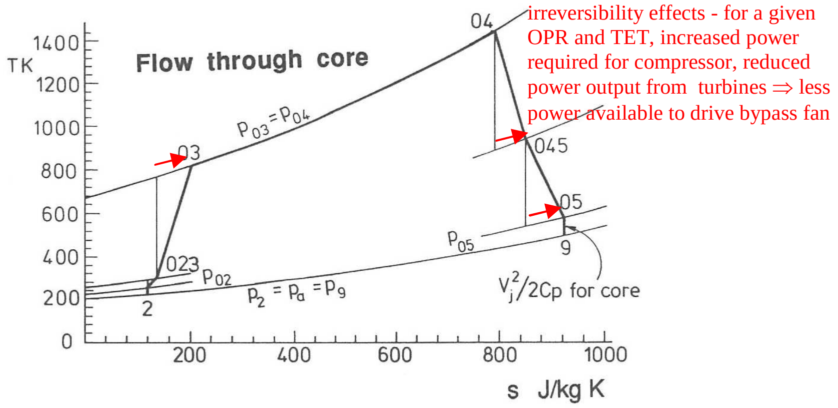
Geometry: $r_c = \frac{1}{2}(r_s + r_A)$

Combine: $\dfrac{\omega_s}{\omega_c} = \dfrac{\omega}{\Omega} = \dfrac{2 r_c}{r_s} = \dfrac{r_s + r_A}{r_s} = 1 + \dfrac{r_A}{r_s}$

But gear sizes are in proportion to tooth number:

$\dfrac{N_A}{N_s} = \dfrac{r_A}{r_s} = \dfrac{\omega}{\Omega} - 1 = \underset{\sim}{3}$

9. a)



Efficiency improvements (for given OPR, T_{04}) will increase the specific work output available from the LPT (T_{045} - T_{05}). Hence, more power to the fan for a fixed core mass flow. Given the jet velocities are fixed (implies fixed fan stagnation temperature rise), the fan must get larger and drive a higher mass flow. Hence, the bypass ratio and fan diameter will both *increase*. The *sfc* will *decrease* but only because the thermal efficiency of the engine improves (the propulsive efficiency would be unchanged as the jet velocities are fixed). The net thrust will *increase* as the total engine mass flow will increase and the jet velocity is the same, $F_N = \dot{m}\left(V_j - V\right)$.

[8]

b)

$$\eta_{is,c} = 0.89 = \frac{\left(\frac{p_{03}}{p_{023}}\right)^{\frac{\gamma-1}{\gamma}} - 1}{\frac{T_{03}}{T_{023}} - 1} \Rightarrow \frac{T_{03}}{T_{02}} = \frac{(28)^{\frac{\gamma-1}{\gamma}} - 1}{0.89} + 1 = 2.787 \Rightarrow T_{03} = \underline{827.94\ \text{K}}$$

$$\psi = \Delta h_0 / U^2 = \frac{c_p\left(T_{03} - T_{023}\right)}{n_{stage} \times (r\Omega)^2} = \frac{1010 \times (827.9 - 297)}{10 \times (0.35 \times 10000 \times \pi/30)^2} = \underline{0.399}$$

A stage loading of 0.4 is typical for a compressor design

[4]

c) Work balance for HP shaft, $\dot{m}c_p\left(T_{03} - T_{023}\right) = \dot{m}c_p\left(T_{04} - T_{045}\right)$

$$\Rightarrow T_{045} = T_{04} - \left(T_{03} - T_{023}\right) = 1700 - (827.94 - 297) = 1169.1\ \text{K}$$

$$\psi = \Delta h_0 / U^2 = \frac{c_p\left(T_{04} - T_{045}\right)}{n_{stage} \times (r\Omega)^2} = \frac{1010 \times (1700 - 1169)}{2 \times (0.38 \times 10000 \times \pi/30)^2} = \underline{1.693}$$

$$\eta_{is,t} = 0.85 = \frac{1 - \frac{T_{045}}{T_{04}}}{1 - \left(\frac{p_{045}}{p_{04}}\right)^{\frac{\gamma-1}{\gamma}}} \Rightarrow \frac{p_{045}}{p_{04}} = \left(1 - \frac{1 - \frac{T_{045}}{T_{04}}}{0.85}\right)^{\frac{\gamma}{\gamma-1}} = \left(1 - \frac{1 - \frac{1169.1}{1700}}{0.85}\right)^{\frac{1.4}{.4}} = 0.201$$

$$\Rightarrow p_{045} = 68 \times 28 \times 0.201 = \underline{383.3\ \text{kPa}}$$

[5]

d) Work balance for the LP shaft:

$$\dot{m}c_p\left(T_{023}-T_{02}\right)+BPR.\dot{m}c_p\left(T_{013}-T_{02}\right)=\dot{m}c_p\left(T_{045}-T_{05}\right)$$

$$T_{05}=T_{045}-\left[\left(T_{023}-T_{02}\right)+BPR\left(T_{013}-T_{02}\right)\right]=1169.1-\left[\left(297-255\right)+10\times 45\right]=677.1\text{K}$$

For the LP Turbine,

$$\frac{p_{05}}{p_{045}}=\left(1-\frac{1-\frac{T_{05}}{T_{045}}}{0.9}\right)^{\frac{\gamma}{\gamma-1}}=\left(1-\frac{1-\frac{677.1}{1169.1}}{0.9}\right)^{\frac{1.4}{.4}}=0.1101 \;\Rightarrow\; p_{05}=383.3\times 0.1101=42.2\text{ kPa}$$

For the exhaust jet, given the nozzle is isentropic:

$$0.5V_9^2=c_p\left(T_{05}-T_9\right)=c_pT_{05}\left(1-\left(p_9/p_{05}\right)^{(\gamma-1)/\gamma}\right)$$

$$V_9=\sqrt{2c_pT_{05}\left(1-\left(p_9/p_{05}\right)^{(\gamma-1)/\gamma}\right)}=\sqrt{2\times 1010\times 677.1\times\left(1-\left(26.4/42.2\right)^{.4/1.4}\right)}$$

= <u>414 m/s</u>

[8]

10. a) Lift coefficient is defined as $C_L = L/0.5\rho AV^2$.
The aerodynamic performance of an aircraft (flying at high subsonic Mach numbers) depends on both lift coefficient and Mach number. For minimum fuel burn M and C_L need to be chosen carefully such that $ML/D = f(M, C_L)$ is a maximum.

[3]

b) Find the lift coefficient at the flight condition:

$$C_L = \frac{L}{0.5\rho AV^2} = \frac{mg}{0.5\gamma p_a M^2 A} = \frac{160\times 9.81}{0.5\times 1.4\times 28.7\times 0.8^2\times 300} \quad \Rightarrow C_L = 0.407$$

Hence, the lift-to-drag ratio is

$$L/D = \frac{C_L}{0.014+0.046C_L^2} = \frac{0.407}{0.014+0.046\times 0.407^2} = \underline{18.82}$$

[4]

c) First find the lift coefficient for maximum *L/D* by differentiation

$$\frac{d(L/D)}{dC_L} = \frac{1\times\left(0.014+0.046C_L^2\right) - C_L\times 2\times 0.046C_L}{\left(0.014+0.046C_L^2\right)^2} = 0$$

$$\therefore 0.046C_L^2 = 0.014 \qquad \Rightarrow C_L = \sqrt{\frac{0.014}{0.046}} = 0.552$$

$$L/D_{\max} = \frac{0.552}{0.014+0.046\times 0.552^2} = \underline{19.70}$$

$$M_{req} = \sqrt{\frac{mg}{0.5\gamma p_a C_L A}} = \sqrt{\frac{160\times 9.81}{0.5\times 1.4\times 28.7\times 0.552\times 300}} = \underline{0.687}$$

[6]

d) Rearrange the Breguet equation to get it in terms of fuel burn

$$s = \frac{V\ L/D}{g\ sfc}\ln\left(\frac{W_{start}}{W_{end}}\right) = H\ln\left(\frac{W_{start}}{W_{start}-W_{fuel}}\right)$$

$$\frac{W_{start}-W_{fuel}}{W_{start}} = e^{-\frac{s}{H}} \qquad \Rightarrow W_{fuel} = W_{start}\left(1-e^{-\frac{s}{H}}\right)$$

For flight at M=0.8, use subscript 1,

$$H_1 = \frac{V_1\,L/D_1}{g\ sfc} = \frac{0.8\sqrt{1.4\times 287\times 226.7}\times 18.82}{9.81\times\ 0.016} = 28950\text{ km}$$

For flight at max L/D, use subscript 2,

$$H_2 = \frac{V_2\,L/D_2}{g\ sfc} = \frac{0.687\sqrt{1.4\times 287\times 226.7}\times 19.7}{9.81\times\ 0.016} = 26023\text{ km}$$

From above,

$$\Rightarrow \frac{W_{fuel,2}}{W_{fuel,1}} = \frac{1-e^{-\frac{s}{H_2}}}{1-e^{-\frac{s}{H_1}}} = \frac{1-e^{-\frac{7000}{26023}}}{1-e^{-\frac{7000}{28950}}} = \frac{0.2359}{0.2148} = 1.098$$

Hence, the fuel burn increases for the lower Mach number by $\underline{9.8\%}$.

[8]

e) Comparing the range parameters, $H_2 < H_1$. The amount *L/D* increases is proportionally less than the amount the flight speed *V* reduces in order to get the C_L required. A better strategy would be to maximise *ML/D*. The optimum C_L should be set by changing the flight altitude at the Mach number required (rather than compromising the flight speed). Given the *L/D* equation is only valid up to M=0.8, could choose M=0.8 and increase flight altitude until $C_L = 0.552$. It may be possible to have higher *ML/D* at higher *M*, but more information on the aircraft aerodynamics would be needed.

[4]

11 a)

$$\eta_{ov} = \frac{\text{Propulsive power to aircraft}}{\text{Rate of heat release from fuel}} = \frac{V \times F_N}{\dot{m}_f LCV} = \frac{VF_N}{\Delta KE} \times \frac{\Delta KE}{\dot{m}_f LCV} = \eta_p \eta_{th}$$

[3]

b)

$$V = M\sqrt{\gamma R T_a} = 0.85 \times \sqrt{1.4 \times 287 \times 223} = 254.4\,\text{m/s}$$

$$\eta_{ov} = \frac{V \times F_N}{\dot{m}_f LCV} = \frac{254.4 \times 40}{0.65 \times 43000} = \underline{0.364}$$

$$\eta_p = \frac{2V}{V + V_j} = \frac{254.4 \times 2}{254.4 + 360} = \underline{0.828}, \qquad \eta_{th} = \frac{\eta_{ov}}{\eta_p} = \frac{0.364}{0.828} = \underline{0.440}$$

[4]

c)

The mass flow rate of air is a significant mass flow and has a high momentum. It is therefore non-dimensionalised as a mass or momentum flux ($\dot{m}\sqrt{c_p T_{02}}$ represents a rate of change of momentum). The mass flow rate of fuel, as a mass or momentum flow, is insignificant. However, it represents a large rate of energy release and is therefore non-dimensionalised as a power ($A_N\, p_{02}\sqrt{c_p T_{02}}$ = force × distance).

$$T_{02} = T_a\left(1 + \frac{\gamma - 1}{2}M^2\right) = 223\left(1 + 0.2 \times 0.85^2\right) = \underline{255.2\text{ K}}$$

$$p_{02} = p_a\left(1 + \frac{\gamma - 1}{2}M^2\right)^{\frac{\gamma}{\gamma-1}} = 26\left(1 + 0.2 \times 0.85^2\right)^{3.5} = \underline{41.7\text{ kPa}}$$

$$F_N = \dot{m}_a\left(V_j - V\right), \qquad \dot{m}_a = \frac{F_N}{\left(V_j - V\right)} = \frac{40000}{360 - 254.4} = \underline{378.8\,\text{kg s}^{-1}}$$

$$\tilde{\dot{m}}_a = \frac{378.8 \times \sqrt{1010 \times 255.2}}{41.7 \times 10^3 \times 2.6} = \underline{1.774}$$

$$\tilde{\dot{m}}_f = \frac{0.65 \times 43 \times 10^6}{41.7 \times 10^3 \times 2.6 \times \sqrt{1010 \times 255.2}} = \underline{0.5078}$$

[8]

d) Use subscript TO for takeoff condition

$$T_{02,TO} = T_a\left(1 + \frac{\gamma - 1}{2}M^2\right) = 288\left(1 + 0.2 \times 0.23^2\right) = 291.0\text{ K}$$

$$p_{02,new} = p_a\left(1 + \frac{\gamma - 1}{2}M^2\right)^{\frac{\gamma}{\gamma-1}} = 101\left(1 + 0.2 \times 0.23^2\right)^{3.5} = 104.8\text{ kPa}$$

$$\dot{m}_{a,TO} = \tilde{\dot{m}}_a \frac{p_{02,TO} A_N}{\sqrt{c_p T_{02,TO}}} = 1.774 \frac{104.8 \times 10^3 \times 2.6}{\sqrt{1010 \times 291}} = 891.6\text{ kg/s}$$

$$\dot{m}_{f,TO} = \tilde{\dot{m}}_f \frac{p_{02,TO} A_N \sqrt{c_p T_{02,TO}}}{LCV} = 0.5078 \frac{104.8 \times 10^3 \times 2.6\sqrt{1010 \times 291}}{43 \times 10^6} = 1.7446\text{ kg/s}$$

$$V_{TO} = M\sqrt{\gamma R T_a} = 0.23\times\sqrt{1.4\times 287\times 288} = 78.24\,\text{m/s}$$

$$\eta_{ov,TO} = \frac{V\times F_N}{\dot{m}_f\, LCV} = \frac{78.24\times 180}{1.7446\times 43000} = \underline{0.1877}$$

$$V_{j,TO} = \frac{F_N}{\dot{m}_a} + V = \frac{180000}{891.6} + 78.24 = 280.12\, m\, s^{-1}$$

$$\eta_{p,TO} = \frac{2V}{V+V_j} = \frac{78.24\times 2}{78.24+280.12} = \underline{0.4367}, \qquad \eta_{th,TO} = \frac{\eta_{ov}}{\eta_p} = \frac{0.1877}{0.4367} = \underline{0.4299}$$

The thermal efficiencies at takeoff and cruise are almost the same (as the non-dimensional operating point is the same, but nozzle/duct losses may differ). The propulsive efficiency at takeoff is very low because of the low flight speed and this leads to a low overall efficiency and high fuel flow rate.

[10]

Dr C. A. Hall

CRIB SHEET (Paper 8) (AN/1)

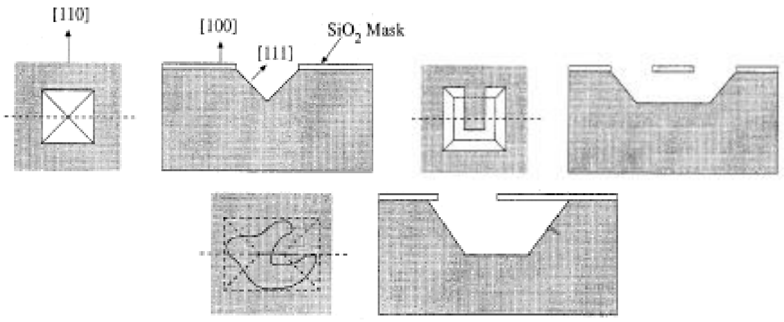
<u>1 (a) (i) Si and ICs</u>
- Lower leakeage current compared to its predecessor, Ge.
- High quality (low interface state density) silicon dioxide compared to the water soluble germanium oxide.
- Well-established/mature processing steps lending themselves well to scalability to large areas.
- Readily abundant silica and silicates comprises 25% of the earth's crust.
- Low cost silicon substrates

<u>(a) (ii) Si and MEMS</u>
- Integration with CMOS signal processing/readout electronics
- Has selective etch rates for different crystal orientations – e.g. (100) and (111)
- High mechanical strength
- Offers good selection of masking materials – e.g. SiO_2, Si_3N4
- Mature processing steps hence low cost

(b) Two thermal oxidation schemes used in IC processing: dry and wet oxidation
<u>Dry oxidation:</u> used for formation of gate oxide – under which a conducting layer is formed between S and D.
<u>Wet oxidation:</u> used for formation of field oxide to provide isolation from other device structures.

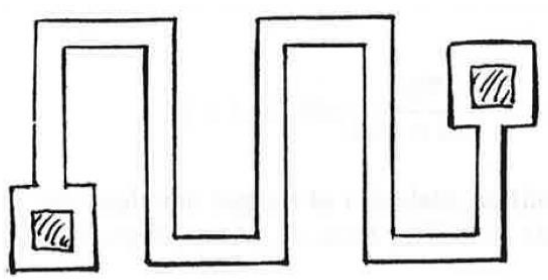
Oxidation rate relates to the rate of incorporation of silicon atoms into the silicon dioxide network – rate depends on the surface density of silicon atoms constant and hence dependent on crystal orientation. The oxidation rate for (111)-silicon is larger because the density of silicon atoms in the (111)-plane is larger than that in the (100)-plane.

(c) Final anisotropically-etched profiles with each type of pattern illustrating the <100>, <110>, and <111> crystal planes and the orientation-dependence of the etching.



(d) For a 200 $\Omega/\square$ diffusion, the aspect ratio (L/W) of a 10 kΩ resistor is $R = R_S\,(L/W)$, yielding $L/W = 10k\Omega/200\Omega/\square$. Thus <u>aspect ratio = 50</u>.

For a layer width of 5 um, L = 250 um. For minimal area, layout is as shown on the right.



2(a) Book work

(b) Accumulation increases the current I from a low value by increasing carrier concentration, so that it is normally off.
Depletion reduces and ultimately switches off the current and carrier density in channel from an existing value, so that it is normally on.

(c) SiO2 is used as gate oxide, dielectric between metal lines, passivant, etch mask, diffusion mask. Gate oxide use is most important for this course.

Si uses MOSFET design with SiO2 as an excellent gate dielectric - good.
The best aspect of Si is that it is most commercially useful design - good.
Si has reasonable band gap and moderate mobility and thus speed – not so good.
GaAs does not have native oxide to use as gate dielectric - bad.
Also cannot use SiO2 on GaAs – v bad.
GaAs uses MESFET design.
GaAs has higher mobility and thus higher speed - good.
GaAs has higher cost of material - bad.

(d) Gauss law applied to a cube shape of height x and dopant density N (show diagram) equate field out of top face with enclosed charge density to give

$$\varepsilon E = N.e.x$$

Integrate this to give

$$-\varepsilon V = N.e.x^2/2 \qquad or \qquad V_g = Nex^2/(2\varepsilon)$$

(e) 10^{19} cm^{-3} = 10^{25} m^{-3} !

$V = (\varepsilon E^2 /2eN) = ([2.10^7]^2\ 10^{-10} /2\ \ 1.6.\ 10^{-19}\ .\ 10^{25}) = 0.012\ V$

3)

(a) In a doped semiconductor – the ionised dopants supply all the carriers.
Free carriers act like conduction electrons in a metal.
Conductivity is limited by scattering of carriers with vibrating atoms.

(b) mobility = velocity / electric field.

Conductivity = mobility x charge x electron density, $\sigma = N.e.\mu$ [not in data book]

Effective mass is a parameter defined as force on electron for example due to an electric field divided by the resulting electron acceleration. It has dimensions of mass, but need not equal the free electron mass.

(c) constant field scaling. sizes shrink by a factor k.
transistors per chip goes as k^2,
lengths vary as 1/k,
power per chip = 1, stays constant.

(d) $v = \mu E,$ $E = V/d$ $t = d/v$

$so\ t = d^2 /\mu V$

mobility = $3.3\ m^2 /Vs.$

$t = 10^{-7}\ 10^{-7} /(3.3\ 0.8) = 3.8\ 10^{-15}\ s$

Comment - This is much faster than equivalent Si FET for which t ~ 10^{-12} s.

E From data book,
$V = Ned^2/2\varepsilon.$ $so\ N = 2\varepsilon V/ e\ d^2$

$V = 0.1\ V, eps = 1.5x10^{10}, d = 10^{-7}\ m.$
Substitute in gives

$N = 2\ x\ 1.5.10^{-10} /2\ .\ 1.6.10^{-19}\ (10^{-7})^2 = 1.875\ x10^{22}\ m^{-3}$

This is a moderate donor density (1.8x10^{16} cm^{-3}) so this makes an effective FET.

15 (a) **Probability of the data under uniform model:**

$$P(S|\text{uniform}) = \prod_{n=1}^{N}\prod_{c=1}^{3}\left(\frac{1}{100}\right) = 10^{-6N}$$

[8]

(b) Let $\theta = (\mu_1, \mu_2, \mu_3, \sigma_1, \sigma_2, \sigma_3)$ be the vector of model parameters. **The probability of the data under the Gaussian model:**

$$\begin{aligned} P(S|\theta) &= \prod_{n=1}^{N}\prod_{c=1}^{3}\mathcal{N}(x_{nc}|\mu_c, \sigma_c^2) \\ &= \prod_{n=1}^{N}\prod_{c=1}^{3}(2\pi\sigma_c^2)^{-1/2}\exp\{-(x_{nc}-\mu_c)^2/(2\sigma_c^2)\} \\ &= \prod_{c=1}^{3}(2\pi\sigma_c^2)^{-N/2}\exp\{-\sum_{n=1}^{N}(x_{nc}-\mu_c)^2/(2\sigma_c^2)\} \end{aligned}$$

[7]

Describe a procedure for fitting the parameters μ_c and σ_c^2 to the data:
At a high level, the maximum likelihood procedure could be used. This involves taking the derivative of the (log) likelihood with respect to the parameters, and setting to zero. This can be solved analytically (see below). Answers which suggest running an optimisation algorithm such as gradient descent or Newton's method get partial marks. The maximum likelihood solutions is:

$$\mu_c = \frac{1}{N}\sum_{n=1}^{N}x_{nc}$$

$$\sigma_c^2 = \frac{1}{N}\sum_{n=1}^{N}(x_{nc}-\mu_c)^2$$

The derivation is straightforward and follows from lecture notes. Intuitively, this sets μ_c to the sample mean of the color c features and σ_c^2 to the corresponding sample variance.
More sophisticated answers could be to add a log prior and do Maximum A Posteriori (MAP) point estimation of the parameters, regularised Maximum Likelihood, or to do full Bayesian inference on the parameters:

$$P(\theta|S) = \frac{P(\theta)P(S|\theta)}{P(S)}$$

Answers which just say use Bayes rule without giving more detail get partial marks. One motivation for Bayesian learning here is if N is small there is a great deal of uncertainty in θ and the ML method can overfit. [10]

(c) **What are some relative advantages and disadvantages of the models described above in parts (a) and (b)?**

The model in (a) is very simple and has no parameters. A major disadvantage of this is that it can't be fit to data. The model is (b) has parameters and can be fit to data, but a disadvantage is that it places probability mass over the whole real line, including $x_{nc} < 0$ and $x_{nc} > 100$. Thus it can predict "impossible" values. Both models assume independence.

Propose another model which you think might be interesting to use for the data S and describe why you think it might be a better model than the ones in parts (a) and (b).

Many options are possible here. A list of a few good suggestions:

- A multivariate Gaussian over the 3 variables in each data point. This can model the correlations in the three (RGB) channels.
- A model that uses beta distributions for each color scaled to cover the $[0, 100]$ range. This includes the uniform as a special case, is constrained to $[0, 100]$ unlike the Gaussian, but like the simple Gaussan it has 6 parameters that can be tuned to data.
- A simple or a multivariate Gaussian model constrained to the $[0, 100]$ intervals.
- A mixture of Gaussians, or mixture of betas as a more sophisticated model that allows for clusters in the data.

[8]

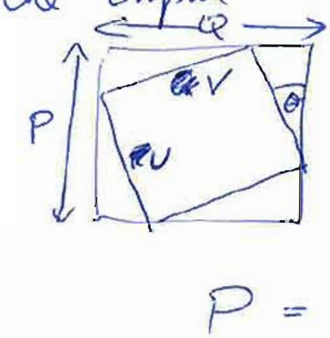
(a) The basic steps to derotate an image by 5 degrees are as follows:

(i) Calculate the size of the output image, so as not to lose information at the corners.

(ii) Calculate where each pixel in the output image needs to come from in the input image, as a pair of real (not integer) coordinates.

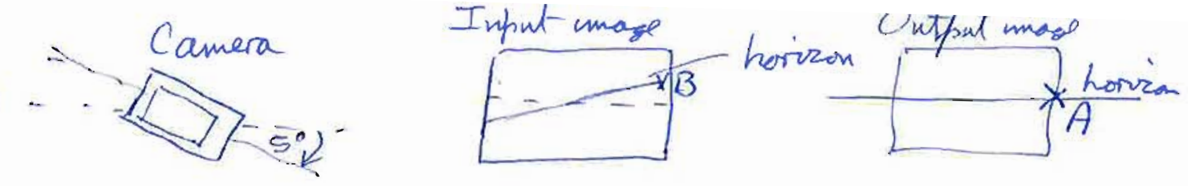
(iii) Fill each pixel in the output image with an interpolated pixel from the required location, using interpolation to deal with non-integer pixel ~~positions~~ locations.

If the output image is of size (P, Q), ~~then~~ and the input is of size (U, V), then to avoid cropping:



$$P = U|\cos\theta| + V|\sin\theta| \qquad \text{where } \theta = 5^\circ.$$
$$Q = V|\cos\theta| + U|\sin\theta|$$

If the camera is rotated 5° clockwise, the horizon will appear rotated 5° anticlockwise, so we will need to shift the pixels from positions that are 5° anti-clockwise ~~from the~~ from their positions in the output image.

(a) 

Let the rotation matrix be $R = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$

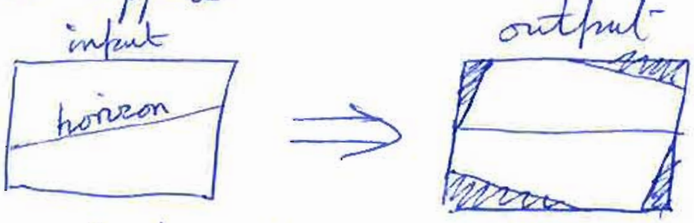
Since the pixel origin is at the centre of the images, We need to choose the correct polarity of θ such that

$$\begin{bmatrix} u \\ v \end{bmatrix} = R \begin{bmatrix} p \\ q \end{bmatrix} = \begin{bmatrix} p\cos\theta + q\sin\theta \\ -p\sin\theta + q\cos\theta \end{bmatrix}$$

If $p = 0$ & $q = 1024$, it will be at A in the output image. We need u to be small and negative, and v to be large and positive ~~so~~ at location B in the input image. Therefore $q\sin\theta$ must be small and negative, and so θ must be negative.

Hence $\theta = -5^\circ = \frac{-5\pi}{180} = -\frac{\pi}{36}$ rad in the above formula.

(b) If the output image is the same size as the input image, and a pure rotation is used then blank areas will appear as shown



To avoid having blank areas it is necessary to scale up the image so that the 1536×2048 rectangle fits within the rotated and enlarged input rectangle.

(b) (cont.) Let the scaling factor be K such that

$$\begin{bmatrix} u \\ v \end{bmatrix} = K \, . \, R \, . \begin{bmatrix} p \\ q \end{bmatrix}$$

For no blanks, the corners of the (p, q) rectangle must lie within the (u, v) rectangle, so that all output pixels come from points within the input image.

<u>Choose two adjacent corners of the output image</u>

At top-left of output image, $p = -768$ & $q = -1024$

$$\therefore \begin{bmatrix} u \\ v \end{bmatrix} = K \begin{bmatrix} -768 \cos\left(\frac{-\pi}{36}\right) - 1024 \sin\left(\frac{-\pi}{36}\right) \\ 768 \sin\left(\frac{-\pi}{36}\right) - 1024 \cos\left(\frac{-\pi}{36}\right) \end{bmatrix} = K \begin{bmatrix} -675.8 \\ -1087.0 \end{bmatrix}$$

At top-right of output image, $p = -768$, $q = +1024$

$$\therefore \begin{bmatrix} u \\ v \end{bmatrix} = K \begin{bmatrix} -768 \cos\left(\frac{-\pi}{36}\right) + 1024 \sin\left(\frac{-\pi}{36}\right) \\ 768 \sin\left(\frac{-\pi}{36}\right) + 1024 \cos\left(\frac{-\pi}{36}\right) \end{bmatrix} = K \begin{bmatrix} -854.3 \\ 953.2 \end{bmatrix}$$

From top-left, we see that $1024 \geqslant K \, . \, 1087.0$ for $|v| \leqslant 1024$

From top-right, we see that $768 \geqslant K \, . \, 854.3$ for $|u| \leqslant 768$

Hence the maximum K which satisfies both of these inequalities is $K = \frac{768}{854.3} = \underline{\underline{0.8990}}$

This will result in the central portion of the image being enlarged by $\frac{1}{K} = \underline{\underline{1.1124}}$

(c) Pixel interpolation will be needed because the ~~formulate~~ $\begin{bmatrix} u \\ v \end{bmatrix} = KR \begin{bmatrix} p \\ q \end{bmatrix}$ will result in non-integer u & v, ~~&~~ when p & q are the integer pixel locations (actually p & q will be odd multiples of 0.5).

To get a good quality output image, the input pixels surrounding each 'real' point (u, v) should be interpolated in 2-D to obtain the best estimate of the output pixel at (p, q).

The simplest 2-D interpolation strategy (other than nearest-neighbour) is bi-linear interpolation, which just uses the nearest 4 points surrounding (u, v).

The next step up in complexity and performance is bi-cubic interpolation, which uses the nearest 16 points in a 4×4 square surrounding (u, v). It requires at least twice as much computation per output pixel, ~~than~~ as the bi-linear method.

The figure, given in the question, shows that the frequency response of the cubic interpolator is significantly flatter out to approx $\frac{1}{3}$ of the sampling frequency than the linear interpolator, and hence will result in less visible blurring of the input image when it is derotated. It also has lower sidelobes and hence less aliassing distortion.

17 (a) For separable filters the 2D convolution can be written as two 1D filtering operations,

$$S(x,y) = \sum_{u,v} L(u,v)I(x-u, y-v) = \sum_{u} l_1(x-u) \sum_{v} l_2(y-v) I(u,v).$$

Here the filter can be expressed, for example, as the product of $l_1 = [2,3,2]$ and $l_2 = [2,4,2]$ and it is therefore separable and can be implemented as two 1D filtering operations. [4]

(b) The 2D convolution has computational cost $\mathcal{O}(NMK^2)$ where NM = number of pixels in the image, K^2 = number of pixels in 2D Gaussian filter. The optimised version is $\mathcal{O}(NMK)$ and so the 1D implementation is K times faster. [4]

(c) The Canny edge detection algorithm carries out the following computations on the smoothed image in order to locate the position and orientation of edges:

1. gradients: find gradient of smoothed image pixels $\nabla S(x,y)$
2. non-maximal suppression: place edgels where $|\nabla S(x,y)|$ greater than local values of $|\nabla S(x,y)|$ in directions $\pm \nabla S(x,y)$
3. threshold: only retain $|\nabla S(x,y)| \geq$ thresh
4. return: output edge positions (x_i, y_i) and orientations $\frac{\nabla S(x_i,y_i)}{|\nabla S(x_i,y_i)|}$

[6]

(d) The Harris corner detection algorithm carries out the following computations on the smoothed image in order to locate the position of corners:

1. compute gradients of smoothed image: $\nabla S(x,y) = (S_x, S_y)$
2. form outer product of gradients and smooth using another broader Gaussian low pass filter,

$$A = \begin{bmatrix} \langle S_x^2 \rangle & \langle S_x S_y \rangle \\ \langle S_x S_y \rangle & \langle S_y^2 \rangle \end{bmatrix}.$$

   Here $\langle f(x,y) \rangle = \sum_{u,v} G_{\sigma'}(u,v) f(x-u, y-v)$ and $\sigma' > \sigma$
3. locate corners and threshold: find locations where $\det(A) - \kappa \operatorname{trace}(A)^2 \geq$ thresh
4. return corner positions (x_i, y_i)

[6]

(e) Edges do not allow motion to be resolved in the direction of the edge (the so called aperture problem). For this reason, corners are superior to edges as interest points for tracking. However, corners are not scale invariant and will therefore perform poorly when used to match different views of the same object with large scale differences. This will occur when there is large depth variation in the image and when the frame rate is low. Blobs would be a more sensible feature in this case.

[5]

EGT1
ENGINEERING TRIPOS PART IB

?? 2014   ?? to ??

Paper 8: Section G Answer: CRIB

SELECTED TOPICS

STATIONERY REQUIREMENTS

SPECIAL REQUIREMENTS TO BE SUPPLIED FOR THIS EXAM

You may not start to read the questions printed on the subsequent pages of this question paper until instructed to do so.

SECTION G

Bioengineering

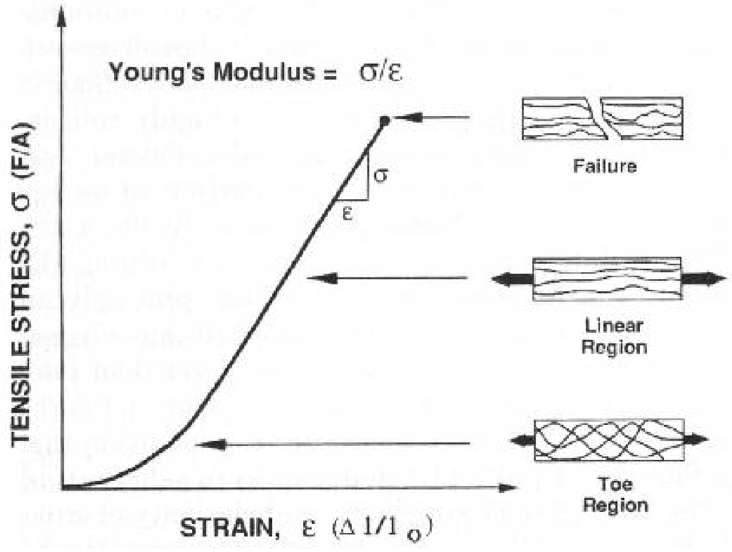
18 Two thirds of the total focussing power of the eye comes from the cornea.

(a) Describe the composition and structure of the cornea and how it differs from the adjacent sclera. [5]

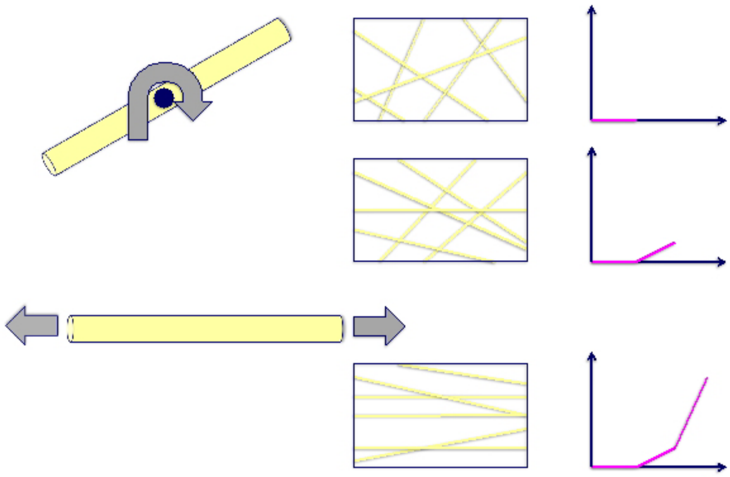
Answer: The cornea has cell layers on both surfaces and an interior that is largely made up of the fibrous protein collagen in a sugary and watery ground substance. Collagen in the cornea is crystalline: it is organized into very regular perpendicular lamellae with uniform spacing between individual collagen fibrils. The lamellar structure is important both mechanically and optically. Mechanically, the lamellar organization provides resistance to intraocular pressure and allows the cornea to serve its critical function of providing two thirds of the optical power of the eye overall. Optically, the regular crystalline structure allows for corneal transparency, again required for unconstructed vision. The collagen in the sclera is much less well-organized and this gives rise to the "white" appearance of the "whites" of the eyes as light is scattered rather than transmitted as in the regular structure of the cornea. In disorders and diseases of the cornea, the tissue can become cloudy due to changes in the collagen organization resulting in a loss of transmissibility of light.

(b) Explain the three aspects of the mechanical response of a soft tissue such as cornea that differentiate such tissues from traditional engineering materials. Use graphs to illustrate each difference and describe what aspect of the tissue microstructure gives rise to each aspect of the response. [6]

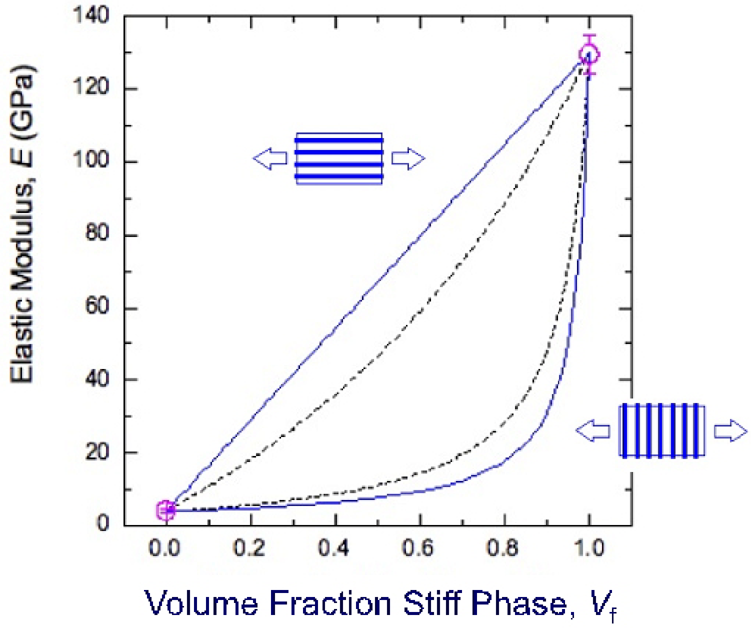
Answer: The response of soft tissues differs in three critical ways.



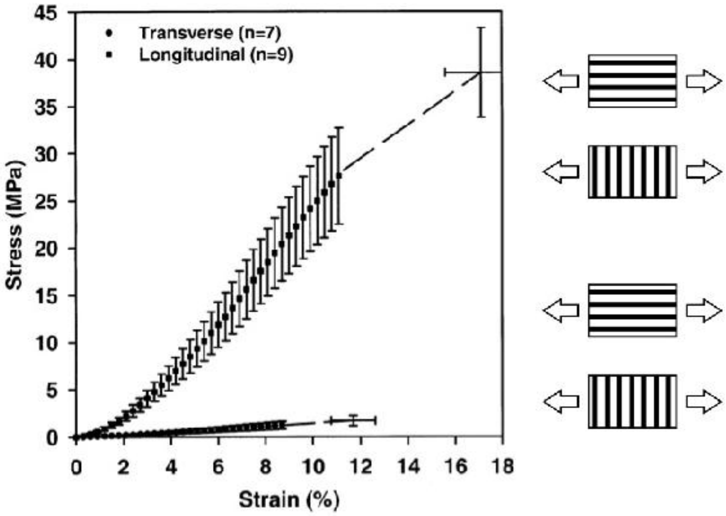
•Nonlinearity due to collagen reorientation and "recruitment".



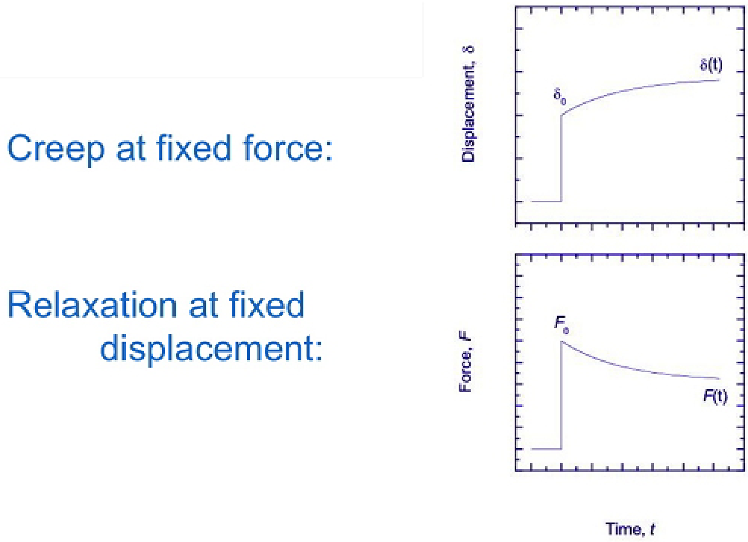
•Anisotropy — stiffness is greatest in the direction of the collagen fibres and least perpendicular to the dominant collagen fibre orientation.

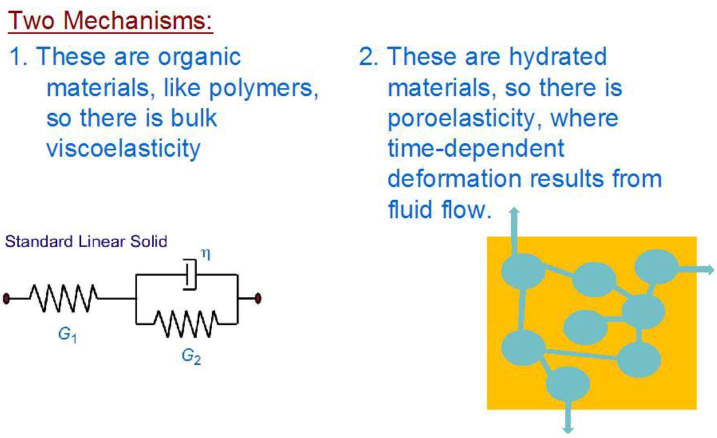


•Time-dependence due to both intrinsic viscoelasticity of the collagen fibres themselves and poroelasticity where there is fluid flow through the porous collagen network.



 (cont.



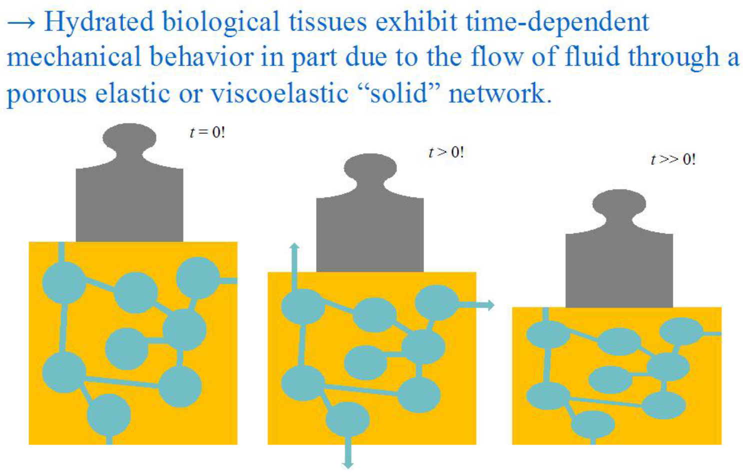


(c) Explain LASIK surgery and describe why LASIK patients are unsuited to be cornea donors. [4]

Answer: LASIK stands for Laser-ASsisted In-situ Keratomileusis, a minimally invasive surgical procedure used to correct vision defects, and falling into the general category of refractive surgeries. The basic procedure modifies corneal curvature, essentially changing the effective "lens" shape by which the cornea controls $\frac{2}{3}$ of optical focussing power. The procedure has three steps. First, a flap is created in the cornea. The corneal bed (stroma) beneath the flap is then reshaped by excising tissue with the laser. The flap is then repositioned, naturally adheres to the modified stroma and the tissue is allowed to heal. Because the cornea has been modified (importantly, thinned), a person who has had LASIK surgery is no longer eligible to be a cornea donor, leading to an increasing donor shortage worldwide.

(d) (i) Describe the transport mechanism in poroelasticity. [3]

Answer: Poroelasticity is the pressure-induced flow of fluid through a porous network, like a mechanical analog of chemical diffusion.



(ii) If a cornea has a thickness of 350 μm, an elastic modulus of 0.3 MPa, and an intrinsic permeability of 8.2×10^{-17} m^2, what is the time constant for poroelastic transport through the cornea? Assume that the viscosity of water is 1 mPa s. [3]

Answer:
The time constant for transport is:

$$\tau = \frac{h^2}{E\kappa} = \frac{h^2}{E\frac{k}{\eta}}$$

Plugging in the values gives:

$$\tau = \frac{h^2}{E\frac{k}{\eta}} = 4.98\text{s}$$

(iii) If the cornea thickness decreases to 200 μm, with all other parameters the same as in (ii), what effect does this have on the transport behaviour? [2]

Answer: A decreased thickness would decrease the time constant since h is on the top of the equation and squared, such that the time constant becomes:

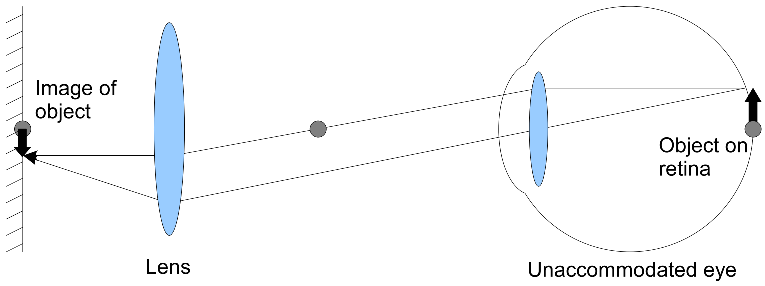
$$\tau = \frac{h^2}{E\frac{k}{\eta}} = 1.63\text{s}$$

(iv) With the thickness the same as in (iii) and the permeability unchanged, how would the modulus have to change to restore the original transport response in (ii)? [2]

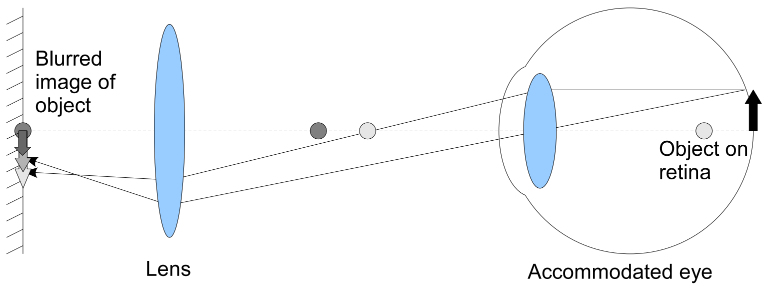
Answer: In order to compensate, the modulus would have to decrease as $E_2 = E_1\frac{{h_2}^2}{{h_1}^2}$ to a value of 0.098 MPa (roughly 0.1 MPa) to restore the transport time constant to ≈ 5 s.

19 (a) Describe, with appropriate diagrams, the key features of a fundus camera that allow a clear image of the retina to be created. [6]

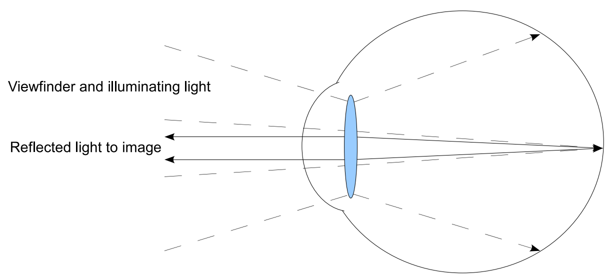
Answer: The most important features are the focussing technique and light transmission. Focussing is complicated by the light having to pass through the lens of the subjects eye as well as of the camera.



In order to create a clear picture of the retina, the focal point of the camera has to coincide with the focal point of the subject's eye, as in the diagram above. The patient also needs to be looking at a far object (not the fundus camera) so that their eye is unaccommodated.



If the patient is not looking at a far object, the image of the retina will be blurred, as shown above.



Light transmission is complicated by the cornea, which will reflect much of the light. As a result, the path of the transmitted light is kept separate from the path of the reflected light.

 BAD PAGEBREAK

This is achieved as in the diagram above, by transmitting light in through a disc-shaped aperture, but only receiving reflections from the centre of that disc. We therefore have to keep the pupil as dilated as possible, so more light can enter the eye: this is why eye-drops are administered before an examination.

(b) In spectral-domain optical coherence tomography, the interference pattern from the imaged tissue is passed through a diffraction grating and its spectrum is detected by a uniformly spaced linear array of photo-detectors. The inverse Fourier transform of this spectrum gives a measure of tissue reflectivity as a function of depth into the tissue. Assume that the laser light centre wavelength (in air) is 850 nm, the bandwidth is 100 nm and the photo-detectors are arranged to make the best possible use of this spectrum. The refractive index of the tissue is 1.36.

(i) If the diffraction grating spreads the light uniformly over the detector array with frequency, how many photo-detectors are required to image the tissue to a depth of 0.5 cm? [5]

Answer: The photo-detectors will be positioned to capture the useful spectrum of the laser, hence the effective sampling frequency f_s is the same as the bandwidth of the laser pulse (c is the speed of light in air):

$$f_s = \frac{c}{800 \times 10^{-9}} - \frac{c}{900 \times 10^{-9}} = 138,889c$$

The time for each sample is $\frac{1}{f_s}$. For N photo-detectors, and a speed of $\frac{c}{1.36}$, then the depth d is:

$$d = \frac{c}{1.36} \times \frac{N}{f_s} = \frac{N}{188,889}$$

If we want a useful depth of 0.5 cm, we actually need d to be 1 cm, since we only measure the magnitude of the Fourier transform:

$$N = 188,889 \times 0.01 = 1889$$

(ii) If the diffraction grating spreads the light uniformly over the detector array with wavelength, what are the minimum and maximum frequency differences between neighbouring photo-detectors, if the same number are used as in (i)? [3]

Answer: The wavelength spacing is $\frac{100}{1888} = 0.053$ nm. This gives:

$$f_{\text{min}} = \frac{c}{900 - 0.053} - \frac{c}{900} = 65.4c = 19.6\text{GHz}$$

$$f_{\text{max}} = \frac{c}{800} - \frac{c}{800 + 0.053} = 82.8c = 24.8\text{GHz}$$

 (cont.

(iii) How would you expect the tissue reflectivity measured in (i) to differ from that measured in (ii)? [3]

Answer: The uniform wavelength samples in (ii) have to be converted to uniform frequency in order to perform the inverse Fourier transform. Hence the effective image depth depends on how many samples are used in this conversion. If 1889 samples are used, the same depth is achieved, but the noise increases with depth due to the need to interpolate the higher frequency samples. The only way to avoid this noise increase is to use fewer samples, but that would also reduce the imaging depth.

(c) Data sampled in three dimensions, such as might be acquired from an optical coherence tomography system, is usually visualised on a two-dimensional display.

(i) Explain the term *reslicing* and outline the difficulties associated with reslicing sampled optical pulse-echo data. [4]

Answer: Reslicing describes the creation of an arbitrarily oriented cross-sectional slice through a 3D data set.
If displaying a 2D slice through 3D data, the regular 2D array of pixels on a typical display will not coincide with the location of the samples in the 3D volume. Hence interpolation or approximation is required to find values for the sampled data at the display points. The interpolation algorithm should be chosen dependent on how we believe the real data behaves, and affects what the displated slice will look like. Some algorithms generate very smooth data and others generate discontinous data. Either approach could be correct, depending on the circumstances.
Another factor in displaying a slice through a 3D data set is the orientation of the display. Some features of pulse-echo systems are aligned with the direction of travel of the light or the sound waves. For instance, a strong reflector of sound or light will not allow the signal to pass through to deeper tissue, and hence deeper tissue will appear very dark. Randomly orientated slices through 3D data are not aligned with the direction of travel and hence can make the appearance of such features confusing.
In addition, even objects with fairly simply geometry can look a lot more complex in reslices, since we do not normally see the cross-sections of objects.

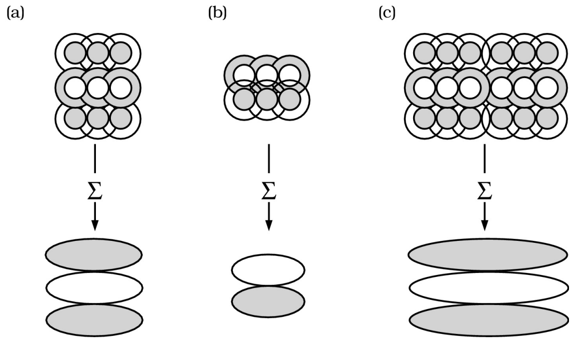
(ii) Briefly outline two techniques, other than reslicing, which can be used to visualise sampled three-dimensional data on a two-dimensional display. [4]

Answer: *Surface rendering* is the process of visualising a geometric surface so that it looks 3D on a 2D display. Starting with sampled three-dimensional data, we first need to extract an interesting surface from this data, which is typically an isosurface within the data. We can extract this by thresholding the data at some data level and using a mesh generation algorithm like Marching Cubes. A graphics card can then generate a surface which looks from this set of triangles by modelling a camera, light sources, and the geometry.

Volume rendering is the process of visualising the entire sampled 3D data set on a 2D display. Data values are maually assigned a colour and an opacity depending on their value, and these all contribute to the final image. It is also possible to define surfaces within the data at particular thresholds, which also have a given opacity. Combined with a camera and lighting model, these can also be blended into the final image.

20 (a) Sketch and describe the receptive fields of simple and complex cells in the primary visual cortex. Also sketch and describe a network architecture that could form simple cell receptive fields from the centre-surround receptive fields found in the lateral geniculate nucleus. [4]

Answer: Simple cells respond to oriented bars or gratings within a small region of the



visual field. Complex cells also respond to bars or gratings, but they are insensitive to small shifts in the position of the grating and often selective with respect to the direction of motion of the stimulus. Complex cells thus form an early level of position invariance in the visual system. Simple cells can be formed by summing over the output signals of several LGN cells, which all have ON (or OFF) center receptive fields with centers located on a straight line.

(b) Name four visual cues that can help to determine which of two objects is farther away. [2]

Answer: Any four of: relative size, occlusion, linear perspective, aerial perspective, shadow and illumination, texture gradients, motion parallax, binocular disparity, accommodation or vergence angle.

(c) Explain why, after fixating a waterfall for a minute, stationary objects you look at will appear to be moving upwards. [3]

Answer: Direction-selective neurons (e.g., complex cells in primary visual cortex) that respond to downward motion adapt to the extended stimulation and are thus less active after viewing the waterfall. Until the adaptation abates, (unadapted) neurons encoding upward motion are more active than those encoding downward motion, causing the perception of a net upward motion.

(d) Explain why, after accidents involving damage to the visual system, doctors check if the perceptual defect concerns a visual hemifield or an individual eye. [3]

Answer: Before the optic chiasm, each of the two optic nerves transmits information about one eye. At the optic chiasm, the fibres reorganize, so that information from each visual hemifield is relayed to the contralateral side of the brain. Whether a hemifield or an eye is affected therefore helps the doctor to determine the anatomical site of a potential brain damage. If the perceptual defect concerns an individual eye, the damage is in the eye or in the part of the optic nerve between the eye and the optic chiasm. If the perceptual defect concerns a hemifield, the damage is after the chiasm, potentially in the cortex.

(e) The following questions are about efficient coding in the (vertebrate) retina.

(i) Mutual information is a measure of the statistical dependency of two random variables. According to the efficient coding hypothesis, the mutual information of two specific random variables is maximized in retinal processing. Name these two variables and provide the formula for the mutual information of two continuous variables. [3]

Answer: The vector of pixels in the retinal image and the vector of ganglion cell responses.
There are different – equivalent – ways of defining mutual information:

$$\begin{aligned} \mathrm{I}(X,Y) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \mathrm{P}(x,y)\,\log\frac{\mathrm{P}(x,y)}{\mathrm{P}(x)\,\mathrm{P}(y)}\,dx\,dy \\ &= \mathrm{H}(X)-\mathrm{H}(X\mid Y)=\mathrm{H}(Y)-\mathrm{H}(Y\mid X) \end{aligned}$$

where

$$\begin{aligned} \mathrm{H}(X) &= -\int_{-\infty}^{+\infty}\mathrm{P}(x)\,\log\mathrm{P}(x)\,dx \\ \mathrm{H}(X\mid Y) &= -\int_{-\infty}^{+\infty}\mathrm{P}(y)\int_{-\infty}^{+\infty}\mathrm{P}(x\mid y)\,\log\mathrm{P}(x\mid y)\,dx\,dy \\ \mathrm{H}(Y) &= -\int_{-\infty}^{+\infty}\mathrm{P}(y)\,\log\mathrm{P}(y)\,dy \\ \mathrm{H}(Y\mid X) &= -\int_{-\infty}^{+\infty}\mathrm{P}(x)\int_{-\infty}^{+\infty}\mathrm{P}(y\mid x)\,\log\mathrm{P}(y\mid x)\,dy\,dx \end{aligned}$$

(ii) Describe the approximations and assumptions underlying the proposition that information maximisation in the retina can be achieved by whitening. [4]

Answer:

- The responses of retinal ganglion cells (RGCs) are nearly deterministic and so response noise entropy is zero, therefore information maximisation is equivalent to maximising the overall response entropy.

- The constraints acting on all RGCs are the same, therefore it will be optimal if all of the RGCs have the same overall response distribution.
- The RGC population response is approximately normally distributed, so that it is fully characterised by its first two moments (the mean and the covariance matrix), and in particular, its entropy is solely determined by the covariances.
- Each RGC acts as a simple linear filter on the input image.
- The statistics of natural images are translation (and rotation) invariant, and thus the optimal RGC filters will all be identical up to translation.
- The population of RGCs is large, and can be well approximated by a continuous "neural field".
- The input to the RGCs is a noiseless version of the retinal image.

(iii) Describe the shape of the Fourier spectrum of a whitening filter for natural images, and provide an explanation for this shape. [3]

Answer: The Fourier spectrum is monotonically increasing with frequency to compensate for the fact that the power spectrum of the input is monotonically decreasing with frequency because i) the power spectrum of natural images is decreasing, and ii) the optics of the eye act as a low-pass filter on the visual environment to create the retinal image.

(iv) Explain with reasons why pure whitening in the retina may not be advantageous. [3]

Answer: This is because RGCs receive a noisy version of the retinal image (due to upstream noise in retinal processing) and this noise part of the input should not be transmitted. However, as we saw above, the power of the true signal in natural images (seen through the optics of the eye) decays with frequency, so at high frequencies noise will dominate. This means that there should be a cut-off at high frequencies in RGC filters, preventing pure whitening of the input.

END OF PAPER

THIS PAGE IS BLANK

IB Paper 8 Section H Crib:

21 You have developed a very low-cost, wearable, wirelessly connectable device for monitoring a person's level of alertness.

(a) Discuss how you could identity and assess the scale of potential markets for this concept.

[9]

The first task is to define the market though segmentation. This could be achieved by considering:

- ***Benefits that are delivered**: what benefits do customers and users derive from the product?*
- ***Particular product attributes**: the easiest way to segment a market is to compare product attributes. This however tends to say little about the customers and is often the weakest approach.*
- ***Characteristics of the consumer**: this can be split to include demographics, and psychographics. Demographics relates to aspects such as social class, age, house size, sex etc. Psychographics relates to the user's attitudes and beliefs, what they feel, their lifestyle etc.*
- ***Product use**: describing ways in which a product is used. There might be strong customer loyalty, products might be used rarely or frequently, it could describe purchase behaviour (e.g. Distress purchase, seasonal patterns or regular upgrades etc).*

A good answer would then consider segmentation for this specific product from several of these perspectives.

*Having defined this segments, could then undertake research onto size of these potential segments. This could be through **sampling** of potential customers, or considering markets for **similar products** for which data already exists.*

For sampling, you would need to be careful about the reliability of responses when assessing markets for 'new to the world' products.

For examination of similar products, you need to be very careful with markets that may be superficially similar but actually very different.

(b) Describe the steps that you could take to understand the requirements of potential customers for such a concept.

[8]

Need to describe the stages that need to be considered in undertaking any research or investigation of user needs.

*•**Focus**: The first stage is to be very clear about what it is that you are aiming to find out, the focus of the research. Is it research that is aiming to unearth new and previously unseen requirements, or is it research that is aiming to validate and qualify current perceptions.*

•***Stakeholders***: *who is it that is the subject of the research, and why are they being asked?*

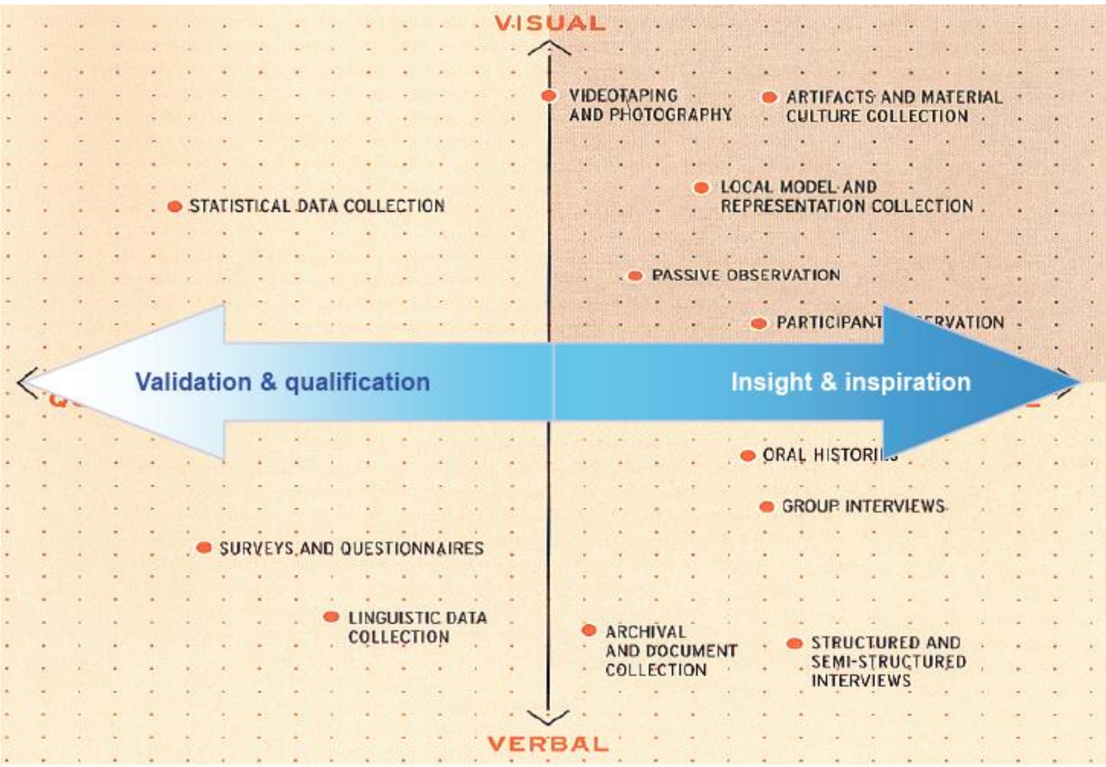
•***Researchers***: *Who will do the research? Ideally, this should be a range of people in the design team.*

•***Data collection***: *what will the mode of data collection be?*

•***Output***: *What will you do with the research when you have it? What decisions will it help make? How will it be presented and to who?*

Ways of gathering the data include those shown in the figure below. Broadly, you could:

- *ask potential users a set of* ***questions***
- *carry out some form of* ***observation*** *of behaviours in a relevant context*
- *develop a* ***prototype*** *and get potential users to provide feedback*



(c) Discuss the different business models you could use to capture value from this concept.

[8]

Options could include:

- ***Sell*** *the idea to someone else*
- ***License*** *the idea (once protected)*
- ***Partner*** *with some other organisation(s) that have resources required to get this concept to market.*

- *Attempt to **develop, make and sell** a product and/or deliver a service alone.*

Looking at this an alternative way:

- *Focus on the development of a **physical product** (which could be sold directly or indirectly via a channel) for single markets, or different products for different markets.*
- *Focus on the development of a **service enabled by this product** (e.g. could be good for employers of drivers / pilots / security guards etc to monitor behaviours)*
- *Could be product linked to collection of **accessories** (e.g. ways of adapting the single product for application in different segments).*

Better answers would then weigh up the relative merits of each of these business models in the context of this technology and this opportunity.

Stronger answers should also consider that the choice of business model will be influenced by the availability of resources, and this, in turn, is influenced by the funding available. You could discuss different funding options for each of the preferred business models, giving a very rough indication of which would need the most money, what the money would be spent on, and where it might be sourced. There may need to be iterations between the design of the preferred business model, the resources required to implement that business model, and the availability of funding to access the resources.

22 (a) Describe and give an example for each of the following types of innovation:

(i) product;

(ii) process;

(iii) placement; and

(iv) paradigm.

[8]

Product / service: What is offered to the customer (e.g. iPhone, Tesco home delivery)

*Process: How the operations of the business are managed (e.g. **production** process – float glass; and **business** processes on-line auction - eBay)*

Placement: How the product / service is positioned in the market (e.g. mobile phone moving from business productivity tool to consumer device; Lucozade moving from drink for those recovering from sickness, to sports energy drink).

Paradigm: A radical change in the way the business makes money (e.g. iTunes, NetFlix)

(b) Discuss why large, long-established firms tend to focus on incremental rather than radical innovations.

[9]

Key points to expand upon would be:

Large firms typically have invested large amounts of capital substantial resources designed to create value in a particular way (e.g. a factory built to make one type of product). They focus on making products in that factory as efficiently as possible, and developing the markets for those products. Making small improvements (incremental innovations) does not usually significantly disrupt the operations of the existing factories / supply chains, but does offer some improvement to their customers in terms of cost or quality. The company's stakeholders (suppliers, investors, customers, employees, etc) do not experience any dramatic change.

However, if the company tries to do something that is radically different from what they currently do, this is likely to cause disruptions for all their stakeholders. The firm's current **suppliers** *may find that they are no longer needed – and the firm has to develop whole new supply chains; the firm's current* **customers** *may not be the same ones that want the new product / service, and so the firm may lose current customers and have to build up new markets;* **investors** *may be very nervous about change that may, in the short term, reduce the value of their investment.*

(c) Discuss why technology-based start-up firms often form partnerships with larger, more commercially experienced firms.

[8]

The answer should expand on the following points:

***Problem**: Start-up firms often have part of complete commercial solution. Technology start-ups may have an exciting new technology, but not have the experience nor required resources to develop, make and sell a product or service based upon this technology. The start-up can attempt to overcome these shortcomings by raising investment to allow them to buy-in the things they need (specialised skills, equipment, routes to market, etc), but (a) they might fail to raise the money required and (b) it may take a long time to get the skills, equipment, etc.*

***Solution**: An alternative is to find a partner who already has the resources the start-up needs, and to find a way to work with them. Typically, such a partner would be a firm with substantial experience and the required development, production and marketing capabilities. There would also be advantages from the perspective of the large firm, that would be able to access the innovativeness of the start-up to address some opportunity area that the large firm might otherwise struggle to address.*

*A strong answer could also discuss the fact that the challenge for the start-up is not only to **find** the right partner, but to find a mutually beneficial way of **working together**. There are numerous pitfalls in both of these stages that could be discussed.*

23 (a) Describe, using examples, what is meant by:

(i) technology push; and

(ii) market pull.

[4]

Technology push = an invention/discovery which then seeks an application

Market pull = an application that requires an invention / discovery.

Technology push examples:

The Post-It note arose by accident

The DVD player arose by analogy and the Dyson was a transfer of an existing industrial technology to a new domestic applications

The domestic breadmaker arose from a structured search for new kitchen appliances

The 'inerter' arose due to a gap in an existing map of possibilities

New materials allowed the hair dryer to move from an expensive metal body to a cheaper plastic body.

Market pull examples:

The 'aural' thermometer for babies arose from the difficult experience of using conventional mercury thermometers measuring babies temperatures with

The chopper bicycle arose from modifications to existing bikes by enthusiastic users

Fridges and washing machines are now sold as fashion items as the kitchen has become the main entertaining room

The ink-jet printing industry around Cambridge has grown due to legislation on sell-by dates for food

The model-T Ford was successful because Ford found ways that by making cars cheaper he could turn a luxury product into a common one

(b) Describe the four tests that an invention must satisfy for it to be patentable.

[4]

Test 1 ***Novelty:*** *is the invention genuinely new, or has it been publically revealed before the filing date?*

Test 2 ***Inventive step:*** *asks whether there is a step involved in moving from what is already known (the prior art) to the invention that would not be obvious to someone who is quite skilled but completely unimaginative – that is the inventive step. Expert advice is often needed to establish if this is satisfied.*

Test 3 ***Practical application:*** *means that the invention must take the practical form of an apparatus or device, a product such as some new material or substance or an industrial process or method of operation. "Industry" is meant in its broadest sense as anything distinct from purely intellectual or aesthetic activity. It does not necessarily imply the use of a machine or the manufacture of an article.*

Test 4 ***Must not be excluded:*** *an invention is not patentable if it is: a discovery; a scientific theory or mathematical method; an aesthetic creation such as a literary, dramatic or artistic work; a scheme or method for performing a mental act, playing a game or doing business; the presentation of information, or a computer program.*

(c) Discuss the relative advantages and disadvantages of protecting an invention through:

(i) filing a patent; or

(ii) keeping the invention confidential.

[8]

- *A patent provides ensures that the nature of your invention is clear, and that if anyone attempts to copy your idea, you have a clear reference point around which to build a legal defence.*
- *There are costs associated with patenting (though the filing is free, the cost of legal advice and processing will be £5-10k. For international coverage, these costs can grow to >£100k. There is then the annual fee to be paid, and this will increase each year.*
- *Having the patent as the basis for legal action against someone suspected of infringing is just the starting point. Fighting a legal action can be extremely costly and time consuming. The Dyson and Kearns examples described in class provide examples that illustrate the time, costs and potential payouts possible. Recent cases between Samsung and Apple also provide interesting examples.*
- *A patent is basically a public disclosure of an invention, showing the world how to do something. Even if the idea is not directly copied, it can provide stimulation for innovation in similar areas. An alternative is to keep the idea confidential. The example in class of the Bessemer paint process is a good example of this. By simply ensuring that nobody new all the stages of the process, Bessemer was able to retain a monopoly in the 'gold' paint process.*
- *The use of confidentiality agreements for employees (and suppliers) can be very effective as a means of ensuring that an idea is not communicated to others, and the value creating potential of the idea is maintained within the organisation. However, confidentiality agreements only apply to those who have signed them.*
- *You can choose just to keep things secret, but this gives you no legal protection should someone start using your idea.*
- *If you chose to keep something secret, then someone then patents your idea, you may then find that you are liable to pay them a license fee if you wish to continue using that idea.*
- *For firms seeking to raise money from investors, not having any legal protection over your core invention may make them very nervous. For some sectors (e.g. advanced materials, drug development, etc) patents are critical, and any potential investor would want to know that you have clear registered ownership before investing any money. Analysis of the patent position of a start-up is often a key part of the due diligence process undertaken by would-be investors.*

A good answer should show understanding of issues such as these. The results could be presented as a comparative table to aid the discussion.

(d) Discuss the challenges that an inventor may face when seeking to commercially exploit potentially disruptive intellectual property in a sector dominated by large companies.

[9]

This question requires you to think about the specific issues of ***single inventor*** *(probably has* ***little money****, very specific skills, an idea probably at a* ***very early stage of commercialisation****), an idea that is* ***disruptive*** *(it will, in some way, change what is already being sold), facing competitors that have lots of* ***money****,* ***resources*** *and* ***experience****.*

The core of a good answer is that there is often significant ***inertia*** *in both markets and within individual firms. Large companies have become large by developing (usually) highly efficient and often complex systems for creating and capturing value from their assets, based around the current technology. There are often high barriers to entry for any new firm to enter that market, let alone one with a disruptive idea.*

An inventor with an idea for something that is better than whatever is currently being delivered to customers, but which is disruptive to the current players will face major resistance. ***Investors*** *may be very reluctant to back the inventor as they realise that it may require huge amounts of money to overcome the cynicism of customers. Few of the incumbent firms will want to back this new idea as it may destroy their current way of doing business.*

The incumbent firms will also often have made large investments in developing substantial IP in this area, and will have a very strong vested interest in continuing to generate returns from this IP. For example, Kodak had invested US$billions over the period of many years in film-based technology. For them, there was little incentive to switch to digital technologies which required a whole set of different skills, IP and resources.

Despite these setbacks, there are examples of small firms that have successfully disrupted markets with their new ideas: e.g. Apple with the Apple I and II kick-starting the very disruptive PC revolution; Napster with peer-to-peer music downloads that transformed the music industry; ARM and low power IP design; Autonomy and probabilistic search technology.

Strong answers could also reflect upon the strategies of such smaller companies that have managed to disrupt industry sectors successfully. Better answers should also consider the different strategies that incumbent firms may deploy to respond to a potential disruption (e.g., price undercutting, purchase and suppress, spread bad publicity, etc). The choice of these different strategies may depend upon the nature of the disruption, the firms involved, and the industry structure.