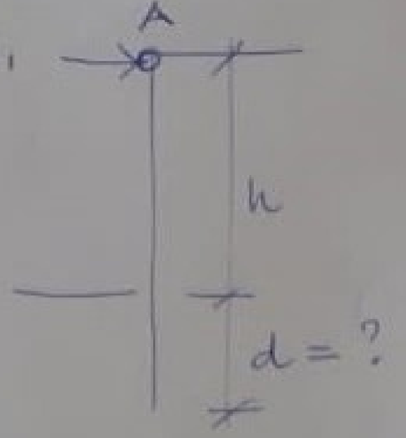
1.



$G_s = 2{,}65$

$e = 0{,}85$

$\varphi' = 32°$

$\delta = \frac{\varphi'}{3} = 10{,}67°$

a. $\gamma_d = \gamma_w \frac{G_s}{1+e} = 9.81 \; \frac{2.65}{1.85} = 14.05 \; kN/m^3$

$\gamma_s = \gamma_w \frac{G_s+e}{1+e} = 9.81 \; \frac{2.65+0.85}{1.85} = 18.56 \; kN/m$

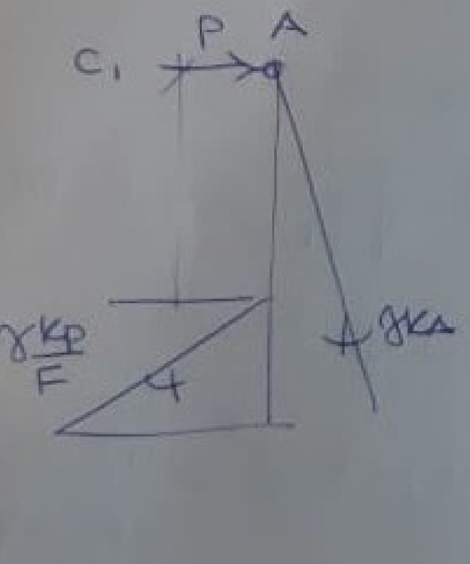
b. $K_A = \frac{1-\sin 32°}{1+\sin 32} = 0{,}307$

$2\theta = \sin^{-1}\left(\frac{\sin 10.67}{\sin 32}\right) + 10{,}67 = 31.11° = 0{,}543$

$K_p = \frac{\cos 10.67}{1-\sin 32}\left[\cos 10.67 + \sqrt{(\sin 32)^2 - (\sin 10.67)^2}\right] e^{0{,}543\,tg} =$

$= 4.342$

c.



$F = 1.6$ take moments about A

$M_R = \frac{\gamma K_A}{3}(h+d)^3$

$M_S = \frac{1}{2}\gamma \frac{K_p}{F} d^2\left(\frac{2}{3}d + h\right)$

d	M_R	M_S
3	1049	1373
2	737	559
2,5	884	914
2,45	868	874

depth of embedment, d

$d = 2,44$ m

compute real value of $\bar{F}$

$$\frac{1}{2}\gamma \frac{k_p}{\bar{F}} d^2 \left(\frac{2}{3}d + h\right) = \gamma \frac{k_a}{3}(h+d)^3$$

$$\bar{F} = \frac{3}{2}\frac{k_p}{k_a}\frac{d^2}{(h+d)^3}\left(\frac{2}{3}d+h\right) =$$

$$= 1,5\ \frac{4,342}{0,307} \cdot \frac{2,44^2}{(8,44)^3} \cdot (0,67 \cdot 2,44 + 6) =$$

$= 1,604$ (very close to required)

prop force P

$$P = \frac{1}{2}\gamma k_a (h+d)^2 - \frac{1}{2}\gamma \frac{k_p}{\bar{F}} d^2 =$$

$$= 0,5 \times 14.05 \times 0,307 \times (8.44)^2 - 0,5 \times 14,05 \times \frac{4.342}{1.604} \times 2,44^2 =$$

$= 40,33$ kN/m

maximum bending moment, M_{max}

find z where shear force is $= 0$

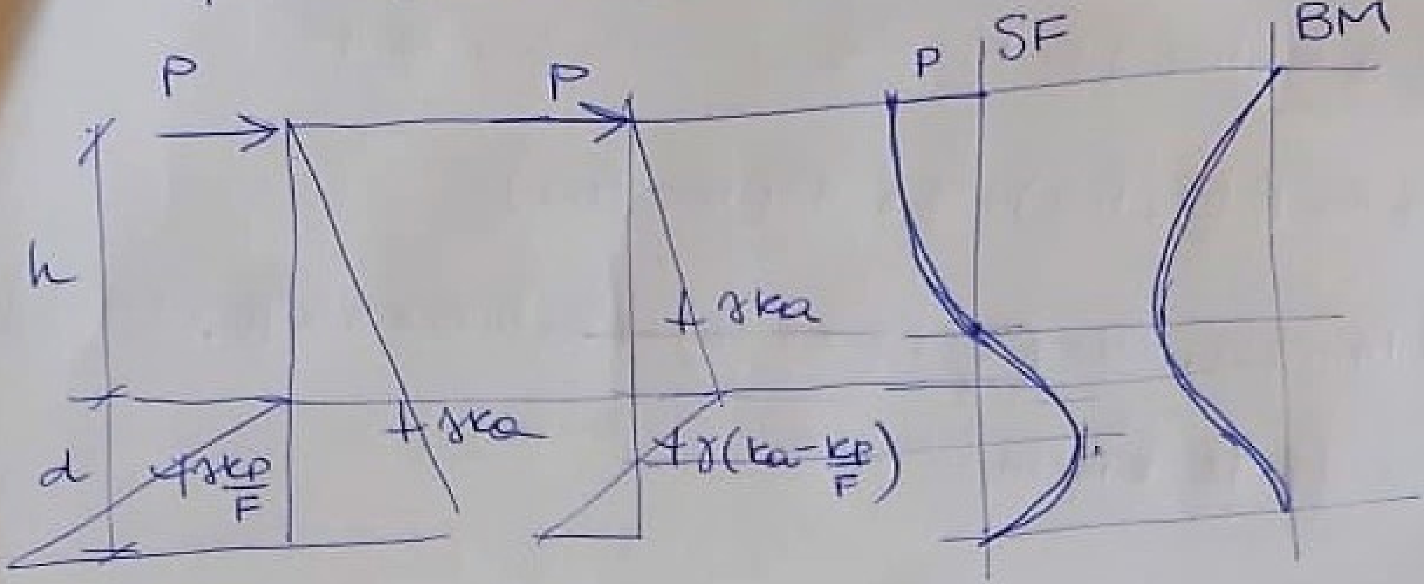
$$T(z) = P - \frac{1}{2}\gamma k_a z^2 = 0 \qquad \bar{z} = \sqrt{\frac{2P}{\gamma k_a}} = 4,32 \text{ m}$$

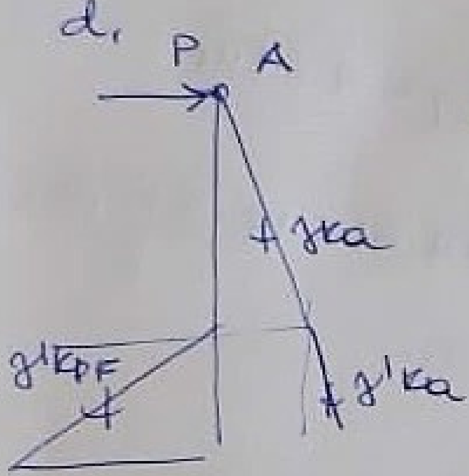
$\bar{z} < h$ M_{max} above dredge level

$$M_{max} = 40,33 \times 4,32 - \frac{14,05 \times 0,307 \times 4,32^3}{6} = 116,20$$

kNm/m

qualitatively:





$$M_R = \frac{1}{2}\gamma K_a h^2 \cdot \frac{2}{3}h + \gamma K_a h d\left(\frac{d}{2}+h\right) + \frac{1}{2}\gamma' K_a d^2\left(\frac{2}{3}\right.$$

$$= K_a\left[\gamma\left(\frac{h^3}{3} + \frac{hd^2}{2} + h^2 d\right) + \gamma'\left(\frac{d^3}{3} + \frac{d^2 h}{2}\right)\right]$$

$$M_S = \frac{1}{2}\gamma' \frac{K_p}{F} d^2\left(\frac{2}{3}d + h\right)$$

d	M_R	M_S	
2,5	845	569	
3.0	991	855	
3,4	1118	1134	
3,35	1101	1097	
3.36	1104	1104	✓

$d = 3{,}36$ m

$\bar{F} = 1.60$

prop force P

$$P = \frac{1}{2}\gamma K_a h^2 + \gamma K_a h d + \frac{1}{2}\gamma' K_a d^2 - \frac{1}{2}\gamma' \frac{K_p}{F} d^2$$

$$= \gamma K_a \left(\frac{h^2}{2} + hd\right) + \frac{1}{2}\gamma' d^2 \left(K_a - K_p/F\right) =$$

$$= 14.05 \times 0{,}307 \times \left(\frac{36}{2} + 6 \times 3{,}36\right) + 0{,}5 \times 8.75 \times 3.36^2 \left(0{,}307 - \frac{4{,}34}{1.6}\right)$$

$$= 45{,}92 \text{ kN/m}$$

$$T(z) = P - \frac{1}{2}\gamma K_a z^2 = 0$$

$$z = \sqrt{\frac{2P}{\gamma K_a}} = \sqrt{\frac{2 \times 45{,}92}{14.05 \times 0{,}307}} = 4{,}61 \text{ m}$$

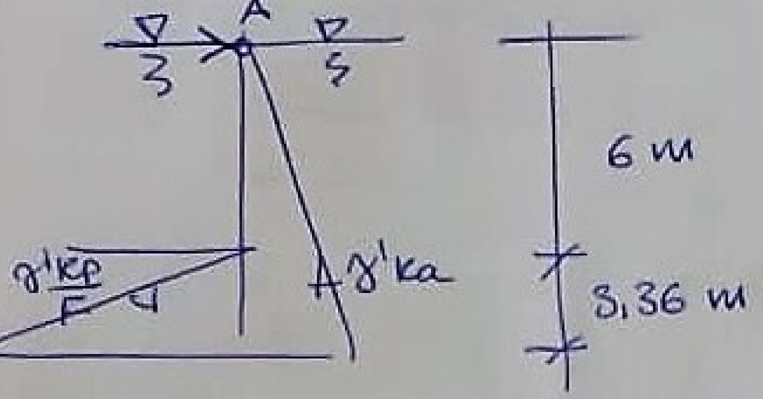
$$M_{max} = Pz - \frac{\gamma K_a z^3}{6} = 45{,}92 \times 4.61 - \frac{14.05 \times 0{,}307 \times 4{,}61^3}{6}$$

$$= 141{,}2 \text{ kNm/m}$$

Qualitatively, same as before

---

e



$M_R = \gamma' \frac{ka}{3}(h+d)^3$

$M_S = \frac{1}{2}\gamma' \frac{kp}{F} d^2 \left(\frac{2}{3}d + h\right)$

$M_R = M_S \quad F = \frac{3}{2}\,\frac{kp}{ka}\,\frac{d^2}{(h+d)^3}\left(\frac{2}{3}d+h\right) =$

$= 1.5\,\frac{4.342}{0,307} \cdot \frac{3,36^2}{9,36^2}\left(0,67 \times 3.36 + 6\right) =$

$= 2,40$

prop force, P

$P = \frac{1}{2}\gamma'\left(ka\,(h+d)^2 - \frac{kp}{F}d^2\right) =$

$= 0,5 \times 8,75\left(0,307\,(9.36)^2 - \frac{4,342}{2,40} \cdot 3,36^2\right) =$

$= 28,31$ kN/m

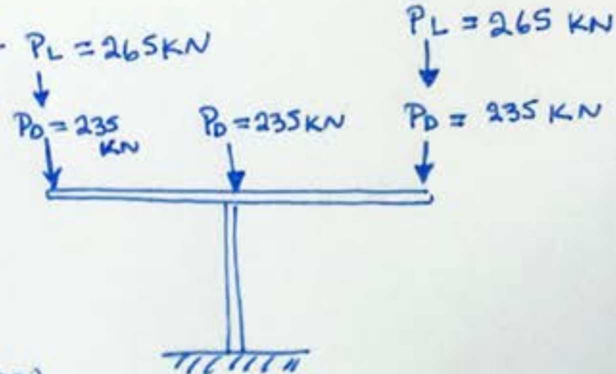
$\bar{z} = \sqrt{\frac{2P}{\gamma' ka}} = \sqrt{\frac{2 \times 28,31}{8,75 \times 0,307}} = 4,6$

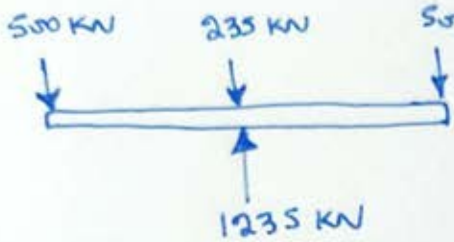
$M_{max} = 28,31 \times 4,6 - \frac{8,75 \times 0,307 \times 4,6^3}{6} =$

$= 86,65$ kNm/m

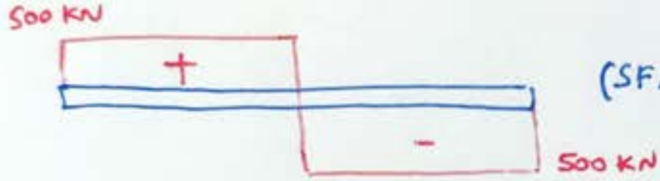
Renforced concrete question crib

a) Maximise the load at the two ends

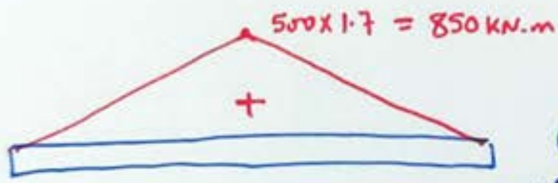




(FBD)

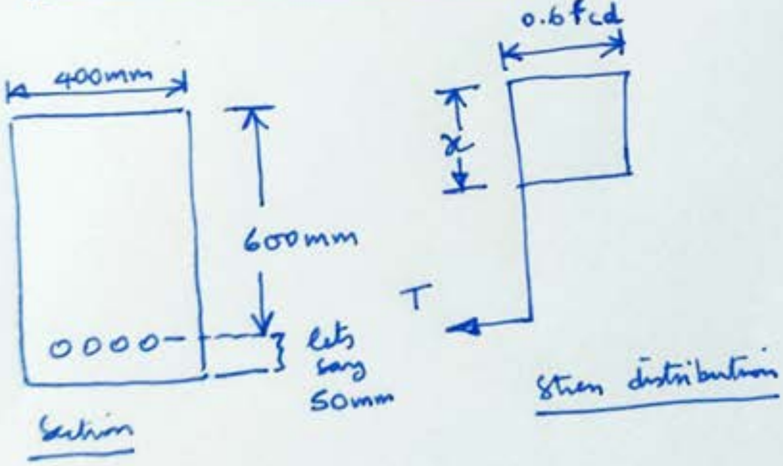


(SFD)



(BMD)

b)



(*) Lets assume the area of steel = 1.5% of the cross-section ⇒

$A_s = 0.015 \times 400 \times 600$

$= 3600 \text{ mm}^2$

Lets say 12 ϕ 20mm bars.

$= 3770 \text{ mm}^2$

(*) Assume the steel yields ⇒ Tension (T) = $3770 \times 460 = 1734200$ N

Tension (T) = compression = $1734200 = (0.6) \cdot 40 \cdot 400 \cdot X$

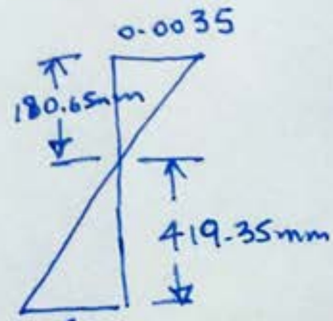
$\Rightarrow X = 180.65$ mm

(*) Need to check steel has yielded :

strain in steel $(\varepsilon_s) = \frac{0.0035}{180.65} \times 419.35$

$= 0.8\%$ (OK)

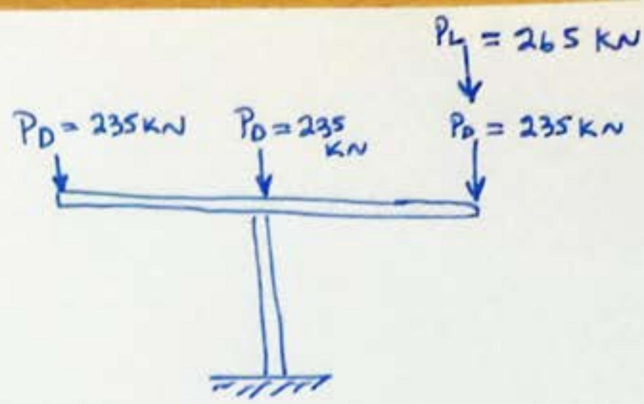
also x is less than 0.5d.

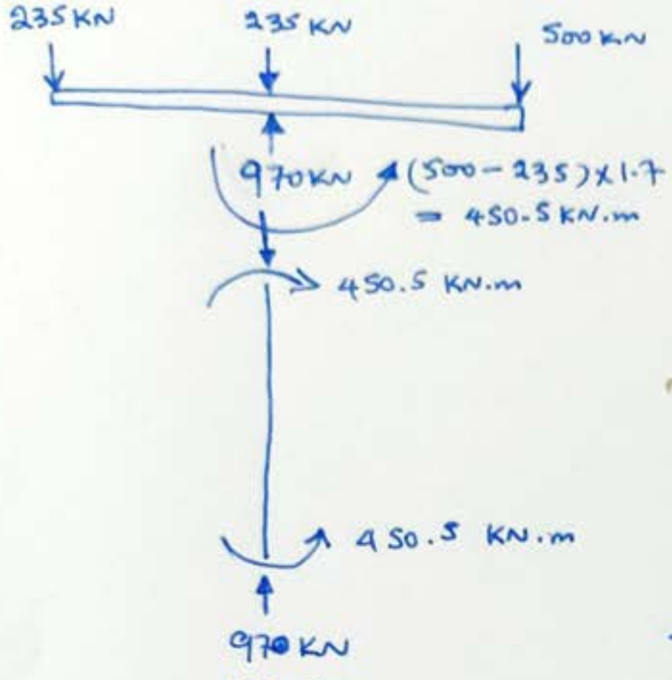


check

(*) $M = 1734200 \times (600 - \frac{180.65}{2}) \simeq 884$ KN.m (OK)

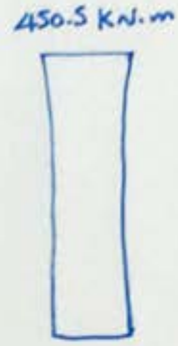
c) To maximise the bending moment in the column ⇒ maximise the load on either end (one end only).





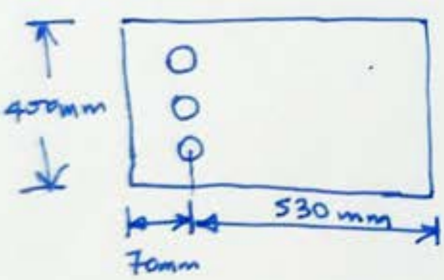
FBD

SFD



BMD

d)



$A_s = 3 \times \frac{\pi}{4}(32)^2 = 2412.74\,mm^2$

Tension $(T) = 460 \times 2412.74$

$= \underline{1109.860 \text{ KN}}$

Compression $(C) = 400.(0.6).40.X$

$= 1109.860 \times 10^3$

$\Rightarrow X = 115.61\,mm < 0.5d$ ✓

$\varepsilon_s = \frac{530 - 115}{115} \times 0.0035$

$= 1.2\% > \varepsilon_{yield}$ ✓

So, $M = 1109860\left(530 - \frac{115}{2}\right) \approx 524$ KN.m > 450.5 ✓

e) (i) Maximum compression load of 1500 KN should be considered

(ii) Reinforcement in the cross-section should be on both sides to account for a scenario of maximum load on either side ⇒ as such, double reinforcement should be taken into consideration in design.

(iii) any valid comment related to detailing (e.g. cover, spacing, etc)

3(a)

(i) The stability ratio is defined as:

$$N = \frac{\sigma_v - \sigma_T}{S_u}$$

where: σ_v is the total vertical stress at the tunnel axis, $\sigma_v = \gamma z$

σ_T is the support pressure applied at the tunnel face

S_u is the undrained shear strength of the clay

For open tunnelling $\sigma_T = 0$.

If $N > 5$ the tunnel is unlikely to be stable and will need face support, i.e. excavation will have to be carried out using a closed shield, either earth pressure balance (E… or slurry shield (SS).

@ city A $\quad N = \frac{20 \times 20}{200} = 2 < 5$ ✓

@ city B $\quad N = \frac{20 \times 20}{60} = 6.7 \gg 5$ ✗ not OK

(ii) tunnelling induced ground movements transmit to existing buildings as settlement, rotations, and distortions of their foundations. These can induce damage affecting apperance and aesthetics, serviceability or function, and in the most severe cases, stability.

Masonry buildings are particularly

that can cause tensile strains in the masonry leading to cracking.
Hogging deformations are potentially more damaging because they induce tensile strains in the top of the building whereas tensile strains caused by sagging are at foundation level.
A settlement trough induced by tunnelling has Gaussian shape: a ~~the~~ building directly above the centre line will deform in sagging whereas a building to one side of the tunnel will deform in hogging.

(iii) Compensation grouting is a mitigation measure where horizontal steel tubes are inserted in the ground at an intermediate level between the tunnel crown and the foundations of the building and cement grout is injected from the tubes to compensate for settlements caused by tunnelling. Instrumentation and monitoring of ground and building response are crucial to inform decisions on where and when to grout to prevent differential settlements, that are particularly damaging for masonry structures

3(b) if tunnelling in clay, the permeability is low enough for there to be no time for drainage and therefore the undrained strength is often enough to guarantee stability. If tunnelling in coarse grained materials (sand and gravel) water will flow into the face causing collapse. problems caused by tunnelling in permeable soils below the water table may be overcome by (1) lowering the water table by pumping from wells, (2) injecting grout to reduce the permeability, (3) using closed shields such as earth pressure balance (EPB) shields or slurry shields (SS).

3(c) Short term (undrained) conditions

use undrained shear strength to calculate horizontal stress (total) in active and passive state

active $\sigma_h = \sigma_v - 2s_u$

passive $\sigma_h = \sigma_v + 2s_u$

For sand, calculate horizontal effective stress σ_h' and add

$\sigma_h = \sigma'_h + u$

$\sigma'_h = k_a \sigma'_v$.

$k_a = \dfrac{1-\sin\varphi'}{1+\sin\varphi'}$ (assuming smooth wall) conservative

$\varphi = 35°$ $k_a = 0{,}271$

Backfilled side (active pressures)

@ surface $\sigma_v = \sigma'_v = 50\ kN/m^2$

$\sigma_h = \sigma'_h = 0{,}27 \times 50 = 13{,}5\ kPa$

@ 5m depth to water level $\sigma_v = 50 + 5 \times 17 = 135\ kPa$

$u = 0$ $\sigma_v = \sigma'_v$ $\sigma_h = \sigma'_h$

$\sigma_h = \sigma'_h = 0{,}27 \times 135 = 36{,}5\ kPa$

@ 7m depth sand/clay interface $\sigma_v = 135 + 2 \times 19 = 173\ kPa$

$u = 2 \times 10 = 20\ kPa$

$\sigma'_v = 153\ kPa$ $\sigma'_h = 41{,}5\ kPa$

$\sigma_h = \sigma'_h + u = 61{,}5\ kPa$

at top of clay $\sigma_v = 173\ kPa$

$\sigma_a = \sigma_v - 2s_u = 173 - 120 = 53\ kPa$

@ 10 m depth $\sigma_v = 173 + 20 \times 3 = 233\ kPa$

$\sigma_a = \sigma_v - 2s_u = 233 - 120 = 113\ kPa$

opposite side (passive pressure)

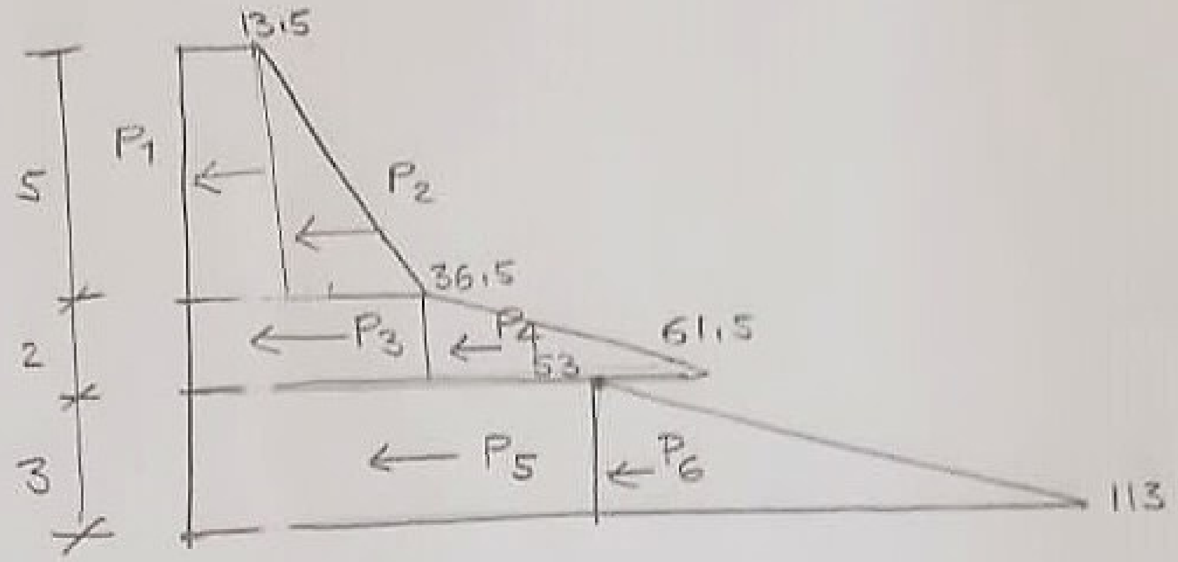
@ 7m depth (river bed) $\sigma_h = \sigma_v = u = 20\ kN/m^2$

$\sigma_h = \sigma_v + 2s_u = 140\ kN/m^2$

@ 10m depth $\sigma_v = 20 + (3 \times 20) = 80\ kN/m^2$

$\sigma_h = 80 + 2s_u = 200\ kN/m^2$

σ_h (kPa)



$P_1 = 13.5 \times 5 = 67.5$ kN/m

$P_2 = \frac{1}{2}(23 \times 5) = 57.5$ kN/m

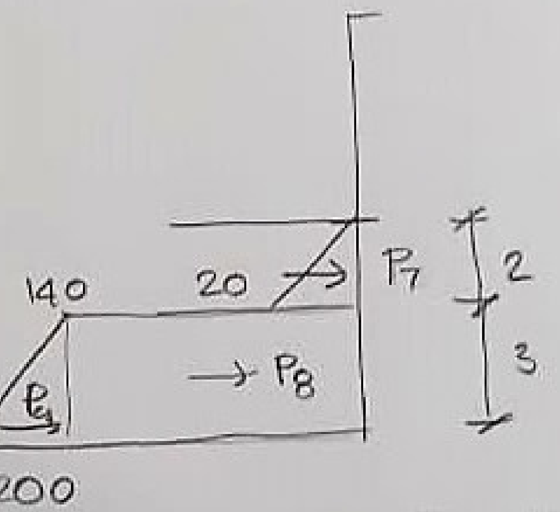
$P_3 = 36.5 \times 2 = 73$ kN/m

$P_4 = \frac{1}{2}(16.5 \times 2) = 16.5$ kN/m

$P_5 = 53 \times 3 = 159$ kN/m

$P_6 = \frac{1}{2}(60 \times 3) = 90$ kN/m

$$\sum_{1}^{6} P_i = 463.5 \text{ kN/m}$$



$P_7 = \frac{1}{2}(20 \times 2) = 20$ kN/m

$P_8 = 140 \times 3 = 420$ kN/m

$P_9 = \frac{1}{2}(60 \times 3) = 90$ kN/m

$$\sum_{7}^{9} P_i = 530 \text{ kN/m}$$

FS = Factor of safety = $\dfrac{\text{Total passive force}}{\text{Total active force}}$

$FS = \dfrac{530}{463.5} = 1.14$ This a very small value. Conservative

assumption (ignore base friction + smooth wall)

(a) A wind turbine utilises a 3-phase, star-connected induction generator connected directly to the 3.3 kV, 50 Hz grid. It has the following equivalent circuit parameters: R_1=1.1 Ω; R_2'=0.8 Ω; X_1=1.4 Ω; X_2'=1.1 Ω. The magnetising reactance and iron loss resistance are large enough to be ignored.

(i) At a slip of -0.03 find the generator phase current and hence its output real and reactive power. [4]

(ii) Derive an expression for the real part of the generator phase current in terms of the equivalent circuit parameters, the slip and the phase voltage. [4]

(iii) By considering that the generator output power is related only to the real part of the generator phase current, find the generator slip if the output power is 1 MW. [4]

CRIB

i) $V_{ph} = 3.3\text{kV}/\sqrt{3} = 1905\text{V}$

$$\bar{I} = \frac{V_{ph}}{R_1 + \frac{R_2'}{s} + j(X_1 + X_2')} = \frac{1905}{\left(1.1 + \frac{0.8}{-0.03}\right) + j(1.4 + 1.1)}$$

$$= \frac{1905}{-25.6 + j2.5} = 74.2\angle -174.4^\circ$$

$P = 3V_{ph}I_{ph}\cos\phi = -422\text{ kW}$ (so output power is + 422kW)

$Q = 3V_{ph}I_{ph}\sin\phi = 41.4\text{ kVA}$. [4]

ii) $$\bar{I} = \frac{\bar{V}}{R_1 + \frac{R_2'}{s} + j\overbrace{(X_1 + X_2')}^{X_T}} = \frac{V\left(R_1 + R_2'/s - jX_T\right)}{\left(R_1 + \frac{R_2'}{s}\right)^2 + X_T^2}$$

$$\therefore Re(\bar{I}) = \frac{V\left(R_1 + R_2'/s\right)}{\left(R_1 + \frac{R_2'}{s}\right)^2 + X_T^2}$$

[2]

iii) $P = 3V\,Re(\bar{I}) = -10^6 \Rightarrow Re\,\bar{I} = -175A$

$$-175 = \frac{1905x}{x^2 + 2.5^2} \quad \text{where } x = R_1 + \frac{R_2'}{s}$$

$$x^2 + 10.89x + 6.25 = 0$$

$$x = -0.608 \text{ or } x = -10.28$$

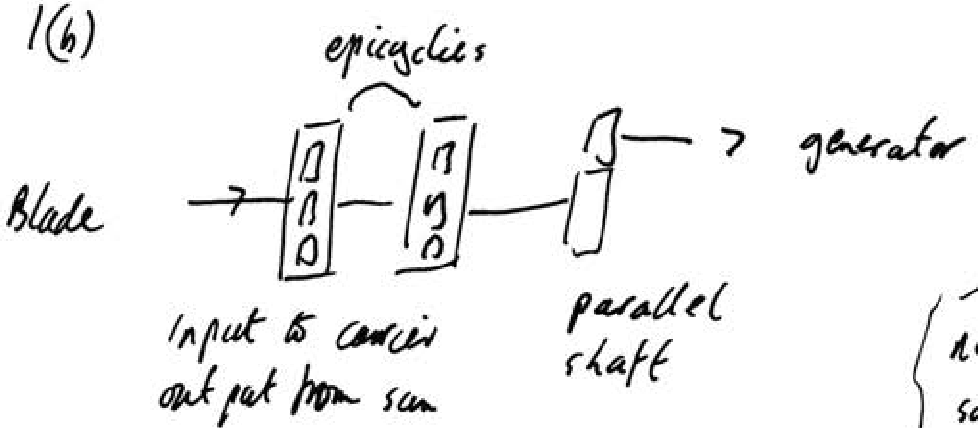
giving $s = -0.471$ or $s = -0.07$

$s = -0.07$ is the obvious choice, to a good approximation the slip will increase by $\dfrac{10^6}{0.422 \times 10^6} = 2.36$ i.e. -0.071

[6]

There were many excellent answers to the induction generator part of this question. Most candidates were able to work out the phase current and real and reactive power, the main errors being wrong use of factors of 3 and sqrt(3). Many candidates were able to derive an expression for the real part of the input current. The most common error was to express it as V/(Re(Z)). The final part stumped many candidates, who got lost in a tangle of algebra, which can be avoided by a simple substitution. Gratifyingly, those who did solve this part mostly went on to justify which of the two solutions was the correct one, demonstrating a good understanding of this topic.

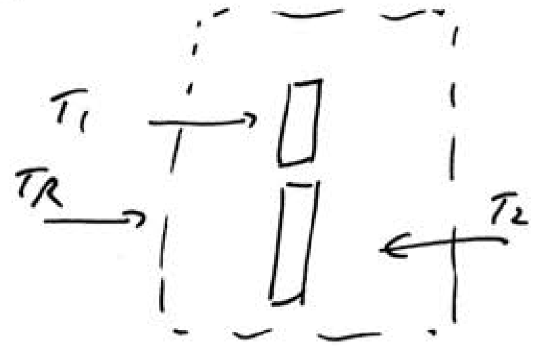
1(b)



need to include some explanation

- need to reduce speed by ~60–100
- epicyclics reduce size and weight
- multi-stage needed for large speed change
- final stage less highly loaded (as faster) so size not so critical

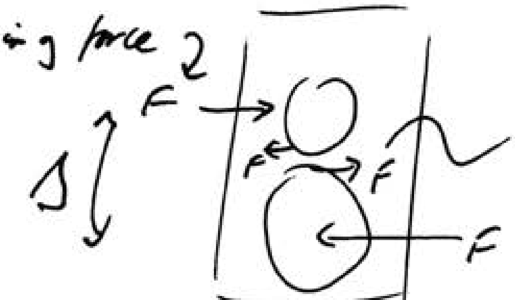
(c)



Since the speeds change through the gearbox, the torques T_1 and T_2 are not the same. Overall equilibrium requires a reaction torque T_R as well as the shaft torques. $(T_1 + T_R - T_2 = 0)$

or

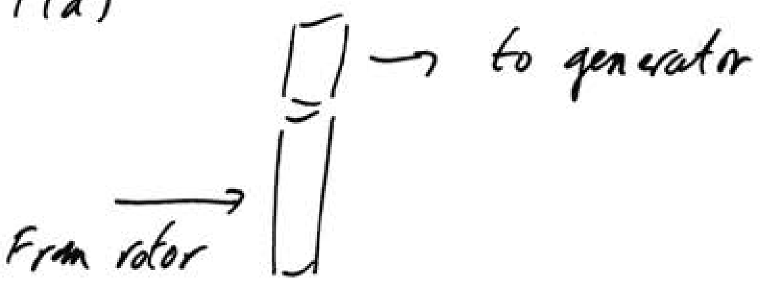
Bearing force ?

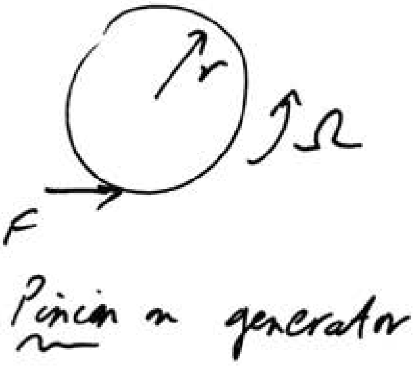


this point important

These Fs are the same, so bearing forces F generate a torque $F\Delta$ which needs to be equilibriated.

1 (d)





For pinion drive Power $P = F.r.\Omega = \frac{F m N \Omega}{2}$

$$\boxed{F = \frac{\sigma_b . w m}{2.08}}$$

$$\Rightarrow N = \frac{2P}{\Omega m} \cdot \frac{2.08}{\sigma_b w m}$$

$$= \frac{4.16 \cdot 2 \times 10^6}{\left(\frac{1800 \times 2\pi}{60}\right) \times (8 \times 10^{-3})^2 \times 400 \times 10^6 \times 60 \times 10^{-3}} = 28.7 \rightarrow 29 \text{ th teeth}$$

Ω in rad/s

m^2 - convert to SI

prime number is good

For 6:1 speed reduction need 29 × 6 = 116 teeth on wheel

this part not always included, otherwise mostly well done

2 (a) (i) Key driver is cost, apparently economies of scale in design and installation have made a big difference to this. More reliable and higher winds offshore help. Technological improvements in larger wind turbines have driven this, with significant investment from big players. Onshore development is hindered by societal push-back, but the public's opposition to wind farms has softened more recently. Perhaps power transmission costs are also relevant, with generation nearer major conurbations. The government push for sustainability, following policy commitments, has allowed helped the environment for long-term investment. [Technology points were well covered, the influence of government policy and measures was sometimes omitted.]

(ii) Sustainable development is about meeting current needs without damaging the needs of future generations. There is no guarantee that technology will respect the needs of future generations. For example using up rare metal resource limitations have an impact on whether future generations can continue to benefit from permanent magnet power generation. But also other resources, such as land, could be taken away. The effect of developments on jobs, and the potential of development financial sustainability is important too. The balance of sustainability for wind turbines is favourable, because of their contribution to renewable energy and so global warming mitigation, but nevertheless these other factors are important. [Answers which defined what was meant by sustainable development gave more directed answers. Some students failed to address the question, in not relating things to wind turbines.]

(iii) Lightweighting is important when there are moving parts. The fast-moving blade structure at the top of the mast is subject to dynamic loading and self-weight loading, so reducing weight helps keep loads and hence costs down. The blades are also a very significant part of the tower-head mass, so reducing their weight helps with tower design associated with vibration. For these reasons composites are a good design choice for the blades. The tower-head mass is important, so it would make sense to reduce the gearbox mass, but the cost of this could be prohibitive. The very highly loaded gears themselves need a high strength material, with composites more difficult to manufacture and make strong enough. The tower mass is less critical, as it is not moving, though a reduction would help with vibration prevention. But the cost of this will be prohibitive. [Good answers included the importance of weight for moving parts, but emphasised the cost-performance trade-off for the blades and tower.]

(iv) Because of the need to reduce cost and mass, it is important not to use too much material in structures like towers and blades. Their consequent relatively slender nature means that bending deflections can become important, both in their own right and as a consequence of the vibration induced, which in turn leads to noise and fatigue loading. Resonant frequencies need to avoid the 1P and 3P driving frequencies. The vibration can be induced by a range of drivers, including the wind spectrum, wind shear and out-of-balance forces. So modelling the interaction between these effects and the structural response is critical to ensure an adequate design life and performance. [Good answers included the vibration and noise issues, but also the effect on fatigue performance.]

2 b) $D = 20\,m$, $D_{hub} = 2\,m$

$\Rightarrow R = 10$, $r_{min} = 1\,m$

$$c = c_0\left(1 - \frac{r}{D}\right) \qquad c_0 = 1 \qquad \rho = 1.2$$

(i) $C_d = 1.5$

Max moment at root of the blade

$$F_D = \frac{1}{2}\rho V^2 c\, C_D$$

$$M = \int_{r_{min}}^{R} F_D\,(r - r_{min})\,dr$$

$$= \frac{1}{2}\rho V^2 C_D \int_1^{10} \left(1 - \frac{r}{20}\right)(r-1)\,dr$$

$$= \frac{1}{2}\rho V^2 C_D \int_1^{10} \left(r - \frac{r^2}{20} - \frac{r}{20} - 1\right)dr$$

$$= \frac{1}{2}\rho V^2 C_D \left[\frac{r^2}{2} \times \frac{19}{20} - \frac{r^3}{60} - r\right]_1^{10}$$

Common mistakes were integrating M =int rF dr rather than M =int (r-r_{min})F dr. Many students confused diameters and radius, sometimes for both limits, sometimes for just one. Some students didn't realise the maximum moment would always be at the root, so put r as the upper limit for the integral, then differentiated the resulting equation, looking for roots of the equation.

$$= 59.2 \text{ kNm}$$

(ii) $$\sigma > \frac{M y_{max}}{I}$$ $r = 0.1$, $t = 0.01$

$$I = \pi r^3 t \quad \text{or} \quad I = \frac{\pi}{4}\left(r^3 - (r-t)^3\right)$$

(assume thin walled)

$$y_{max} = r = 0.1$$

$$\sigma_{SL} = 188.5 \text{ MPa}$$

(iii) $$M_{sw} = \int_1^{10} 50g(r-1)\,dr$$

$$= 50g\left[\frac{r^2}{2} - r\right]_1^{10}$$

$$= 50g\left(50 - 10 - \left(\frac{1}{2} - 1\right)\right)$$

$$= 50g \times 40.5 = 19.8 \text{ kNm}$$

$$\sigma_{sw} = \frac{M_{sw} y_{max}}{I} = 63.2 \text{ MPa}$$

Most students got this part, although there was confusion among some as to what y_{max} should be and whether the r given represented an inner, outer or average radius. All consistent answers were given full marks.

$$\sigma = \sigma_{SW} \frac{y}{r_c} + \sigma_{SL} \frac{x}{r_c}$$

$$x = r_c \cos\theta \qquad y = r_c \sin\theta$$

$$\sigma = \sigma_{SW} \sin\theta + \sigma_{SL} \cos\theta$$

$$\frac{d\sigma}{d\theta} = \sigma_{SW} \cos\theta - \sigma_{SL} \sin\theta = 0$$

$$\sigma_{SW} \cos\theta = \sigma_{SL} \sin\theta$$

$$\Rightarrow \tan\theta = \frac{\sigma_{SW}}{\sigma_{SL}}$$

$$\Rightarrow \theta = 18.5^\circ$$

$$\sigma_{max} = \sigma_{SW} \sin\theta + \sigma_{SL} \cos\theta$$

$$= 198.9 \text{ MPa}$$

$$\Rightarrow 5\% \text{ increase}$$

Very few students were able to get this part. Most were able to calculate the extra stress due to the self-weight of the horizontal blade, but then added the two stresses together. Marks were given for any answer that included a suggestion that the student understood that the two forces were acting in different directions.

Q 3 Calculation of I

$$I = \frac{1}{12} b d^3 - \frac{1}{12}(b-2t)(d-2t)^3$$

$$\approx \frac{1}{12}\left(2td^3 + 6bd^2 t\right) \qquad \text{for } t \ll b, d$$

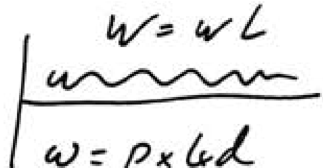
$$= \frac{t}{6}\left(d^3 + 3bd^2\right)$$

$$I_{edgewise} = \frac{t}{6}\left((4d)^3 + 3.d.(4d)^2\right) = \frac{56}{3} t d^3$$

$$I_{flapwise} = \frac{t}{6}\left(d^3 + 3.4d.d^2\right) = \frac{13}{6} t d^3$$

Also simpler approximations OK

(a)



$$\delta = \frac{wL^4}{8EI} \quad \text{from data book}$$

$$= \frac{4\rho d . L^4 \quad 6}{8 \; E \quad 13 t d^3} = \frac{3}{13} \frac{\rho L^4}{E t d^2}$$

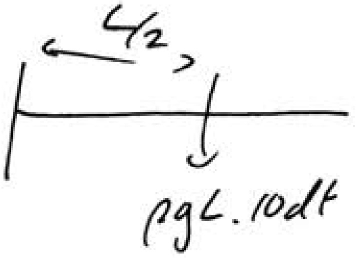
mass = $\rho L d t$

Eliminate $t \Rightarrow$ mass $= 10 \rho L d . \dfrac{3\rho L^4}{13 E \delta d^2} = \dfrac{30 \rho\rho L^5}{13 E \delta d}$

To minimise mass, minimise $\dfrac{\rho}{E}$

Not as well done as expected.

3 (b)



Moment $M = \frac{L}{2} \cdot 10\rho g L d t$

(i) $\sigma = \frac{My}{I} = 5L^2 \rho g d t \times 2d \times \frac{3}{56 t d^3} = \frac{15}{28} \frac{L^2 \rho g}{d}$

Note that ts cancel out so the strength constraint puts a constraint on the materials $\sigma = \sigma_f \Rightarrow \frac{\sigma_f}{\rho} > \frac{15}{28} \frac{L^2 g}{d}$

(ii) Now to consider mass : $m = 10\rho L d t$ as before.

To minimise mass need to minimise ρ

This part not answered well.

(c) With the combined constraints, the strength limited design implies minimising t, which is given by the deflection constraint $t = \frac{3}{13} \frac{\rho L^4}{5Ed^4}$

So the strength constraint becomes $m = 10\rho L d \cdot \frac{3}{13} \frac{\rho L^4}{5Ed^4} \Rightarrow$ minimise $\frac{\rho}{E}$ as for the deflection constraint.

But with also the constraint $\frac{\rho}{\sigma_f} < \frac{28 \cdot d}{15 L^2 g} = 1.6 \times 10^{-4}\ kg\,m^{-3}/Pa$

$= 0.16\ \frac{Mg/m^3}{MPa}$

This logic was generally missing

From the data book this eliminates some dense materials and foam. Ceramics not suitable.

So minimise $\rho/E \rightarrow$ CFRP, then steel, wood.

Sensible approaches following on from answers to (a) & (b) scored marks

(d) Increasing L will make the strength constraint more important, so could change the material choice.

Also expect deflection constraints to become more important leading to CFRP. Manufacturing methods need to change for larger blades, but also introduce options for more complex and structurally efficient designs.

2021 2P8 Section D Solutions

1(a)

$$\frac{p_{02}}{p_a} = \left(1 + \frac{\gamma - 1}{2} M^2\right)^{\frac{\gamma}{\gamma-1}} \qquad \Rightarrow \qquad M = \sqrt{\left(\left(\frac{p_{02}}{p_a}\right)^{\frac{\gamma-1}{\gamma}} - 1\right)\frac{2}{\gamma - 1}}$$

$$M = \sqrt{\left(\left(\frac{45}{29}\right)^{\frac{0.4}{1.4}} - 1\right)\frac{2}{0.4}} = 0.818$$

$$T_{02} = T_a\left(1 + \frac{\gamma - 1}{2} M^2\right) = 220 \times (1 + 0.2 \times 0.818^2) = 249.4\ K$$

(b) M=0.6

$$\frac{\dot{m}\sqrt{c_p T_{02}}}{A p_{02}} = \frac{\gamma}{\sqrt{\gamma - 1}} M \left(1 + \frac{\gamma - 1}{2} M^2\right)^{-\frac{\gamma+1}{2(\gamma-1)}} = 1.0781$$

$$\gamma = \frac{c_p}{c_p - R} \quad \Rightarrow \quad c_p = \frac{R\gamma}{\gamma - 1} = 1004.5$$

$$A = \frac{\dot{m}\sqrt{c_p T_{02}}}{1.0781 \times p_{02}} = \frac{50\sqrt{1004.5 \times 249.4}}{1.0781 \times 45000} = 0.516$$

Mean radius

$$r = \frac{A}{2\pi h} = \frac{0.516}{2\pi \times 0.3} = 0.274\ m$$

(c) (i)

$$\eta = 0.9 = \frac{\left(\frac{p_{03}}{p_{02}}\right)^{\frac{\gamma-1}{\gamma}} - 1}{\frac{T_{03}}{T_{02}} - 1} \qquad \Rightarrow \qquad T_{03} = 249.4 \times \left(\frac{(35)^{\frac{\gamma-1}{\gamma}} - 1}{0.9} + 1\right) = 737.6\ K$$

$$\dot{m}c_p(T_{03} - T_{02}) = \dot{m}c_p(T_{04} - T_{05})$$

$$T_{05} = T_{04} - (T_{03} - T_{02}) = 1600 - (737.6 - 249.4) = 1111.8\ K$$

(ii) Combustor enthalpy rise is the energy flow rate in the fuel

$$\dot{m}_f LCV = \dot{m}c_p(T_{04} - T_{03})$$

$$\dot{m}_f = 50 \times 1004.5 \times \frac{(1600 - 737.6)}{43 \times 10^6} = 0.99\ kg/s$$

(iii)

$$\frac{p_{05}}{p_{04}} = \left(1 - \frac{1 - \frac{T_{05}}{T_{04}}}{\eta}\right)^{\frac{\gamma}{\gamma-1}} = \left(1 - \frac{1 - \frac{1111.8}{1600}}{0.9}\right)^{\frac{\gamma}{\gamma-1}} = 0.235$$

$$p_{04} = 35 \times 45 = 1575\ kPa$$

$$p_{05} = 1575 \times 0.235 = 369.7\ kPa$$

$$0.5 \times V_j^2 = c_p T_{05} \left(1 - \left(\frac{p_a}{p_{05}}\right)^{\frac{\gamma-1}{\gamma}}\right) = 1004.5 \times 1111.8 \times \left(1 - \left(\frac{29}{369.7}\right)^{\frac{\gamma-1}{\gamma}}\right) = 5.77 \times 10^5\ J/kg$$

$$V_j = 1074\ m/s$$

$$\eta_p = \frac{2V}{V + V_j} = \frac{2 \times 0.818\sqrt{1.4 \times 287 \times 220}}{0.818\sqrt{1.4 \times 287 \times 220} + 1074} = 0.37$$

(d) (i) $W_x = 50 \times 0.9 \times 5.77 \times 10^5 = 26.0MW$

(ii) Advantages – this is a way to increase the effective bypass ratio, lower the jet velocity and increase propulsive efficiency. It is possible to have many fans. Integration with airframe can have benefits such as reduced nacelle area, or boundary layer ingestion.

Challenges – weight, cooling and size (of motors and generators). Transmission efficiency losses (related to cooling challenge).

2(a) The expression $F_G + p_a A_N$ only depends on the conditions inside the engine.

$$F_G = \dot{m}V_{19} + (p_{19} - p_a)A_N$$

So

$$F_G + p_a A_N = \dot{m}V_{19} + p_{19}A_N$$

For a choked engine core the right hand side of the equation is fixed by the engine conditions.

$\dot{m}_f LCV$ is the flow of energy in the fuel. This is non-dimensionalised by the rate of doing work of the engine which is a force x velocity. The force is $p_{02} \times A_N$. The velocity is the speed of sound based on the inlet conditions $\sqrt{\gamma R T_{02}}$ which is proportional to $\sqrt{c_p T_{02}}$. Dividing the two we get

$$\tilde{\dot{m}}_f = \frac{\dot{m}_f LCV}{A_N p_{02}\sqrt{c_p T_{02}}}$$

$\tilde{F}$, can be written as a function of non-dimensional fuel mass flow, $\tilde{\dot{m}}_f$, alone. This is because the engine condition is fixed by $\dot{m}_f$, p_a, T_a, V_{19}. For a choked exit nozzle, and fixed environmental conditions, the only degree of freedom is the fuel flow rate $\dot{m}_f$. One non-dimensional parameter therefore fixes the engine condition. In this case $\tilde{\dot{m}}_f$ has been chosen.

(b)

$$p_{01} = p_a\left(1 + \frac{\gamma - 1}{2}M^2\right)^{\frac{\gamma}{\gamma-1}} = 29\left(1 + \frac{\gamma - 1}{2}0.8^2\right)^{\frac{\gamma}{\gamma-1}} = 44.2\ kPa$$

$$T_{01} = T_a\left(1 + \frac{\gamma - 1}{2}M^2\right) = 220\left(1 + \frac{\gamma - 1}{2}0.8^2\right) = 248.2\ K$$

$$V = M\sqrt{\gamma R T_a} = 0.8 \times \sqrt{1.4 \times 287.1 \times 220} = 237.9\ m/s$$

$$\eta_p = \frac{2V}{V_j + V} = 0.8 \quad \Rightarrow \quad V_j = V\left(\frac{2}{\eta_p} - 1\right) = 237.9 \times \left(\frac{2}{0.8} - 1\right) = 356.8\ m/s$$

$$\dot{m} = \frac{F_N}{V_j - V} = \frac{50000}{356.8 - 237.9} = 420.4\frac{kg}{s}$$

$$F_G = \dot{m}V_j = 420.4 \times 356.8 = 150\ kN$$

(c)

$$\tilde{F} = \frac{F_G + p_a A_N}{p_{02}A_N}\ , \qquad \tilde{\dot{m}} = \frac{\dot{m}\sqrt{c_p T_0}}{p_{02}A_N}$$

$$\left(\frac{\dot{m}\sqrt{c_pT_0}}{p_{02}A_N}\right)_{flight} = \left(\frac{\dot{m}\sqrt{c_pT_0}}{p_{02}A_N}\right)_{Test}$$

$$\left(\frac{420.4\sqrt{248.2}}{44200}\right)_{flight} = \left(\frac{\dot{m}\sqrt{288}}{100000}\right)_{Test}$$

$$\dot{m}_{test} = 882.8\ kg/s$$

$$\left(\frac{F_G + p_aA_N}{p_0A_N}\right)_{flight} = \left(\frac{F_G + p_aA_N}{p_0A_N}\right)_{test}$$

$$F_{G\,test} = A_Np_{0test}\left(\frac{F_G + p_aA_N}{p_0A_N}\right)_{flight} - p_{a\,test}A_N$$

$$F_{G\,test} = 100000\left(\frac{150000 + 29000}{44200}\right)_{flight} - 100000 = 305\ kN$$

The gross thrust at take off is far higher than the net thrust available in cruise. The low M at take off means that the gross thrust is close to the net thrust available at take off. Even allowing for the climb angle, there is enough extra thrust to allow the plane to take off, even with one engine out.

(d)

$$\eta_o = \eta_p\eta_t = 0.8 \times 0.42 = \frac{F_NV}{\dot{m}_fLCV}$$

So

$$sfc = \frac{\dot{m}_f}{F_N} = \frac{V}{0.8 \times 0.42 \times LCV}$$

$$sfc = \frac{\dot{m}_f}{F_N} = \frac{237.9}{0.8 \times 0.42 \times 43 \times 10^6} = 1.646 \times 10^{-5}$$

$$sfc_{flight} = 16.46\ gkN^{-1}s^{-1}$$

$$\dot{m}_{f\,flight} = \frac{50000 \times 237.9}{0.8 \times 0.42 \times 43 \times 10^6} = 0.823\ kg/s$$

$$\left(\frac{\dot{m}_fLCV}{A_Np_{02}\sqrt{c_pT_{02}}}\right)_{flight} = \left(\frac{\dot{m}_fLCV}{A_Np_{02}\sqrt{c_pT_{02}}}\right)_{test}$$

$$\dot{m}_{f\,test} = \left(p_{02}\sqrt{T_{02}}\right)_{test}\left(\frac{\dot{m}_f}{p_{02}\sqrt{T_{02}}}\right)_{flight} = 100000 \times \sqrt{288}\left(\frac{0.823}{44200\sqrt{248.2}}\right)_{flight} = 2.01\ kg/s$$

On test

$$F_{G_{test}} = F_{N_{test}} = 305\ kN$$

$$sfc_{test} = \frac{2.01}{305} = 6.58\ gkN^{-1}s^{-1}$$

(e) At the top of climb the altitude in the same as cruise. This means that the ambient conditions are the same and the Mach number is the same. This means that the stagnation conditions at engine inlet are the same. The thrust required is greater and so the dimensionless thrust is greater at the top of climb. Because the dimensionless thrust is a function of $\widetilde{\dot{m}}_f$, this must also rise. Since the dimensionless thrust is highest at top of climb, so is the dimensionless air mass flow and hence the Mach numbers in the engine.

3. (a) Aerothermal engineering influences range via:

L/D via airframe aerodynamic design.
sfc via overall efficiency: propulsive efficiency (bypass ratio) thermal efficiency (core engine design).
W1/W2 via structure design and material choice.

(b) Start of cruise

$$C_L = \frac{mg}{0.5\frac{p}{RT}AV^2} = \frac{mg}{0.5pA\gamma M^2}$$

$$M = \sqrt{\frac{mg}{0.5pA\gamma C_L}} = \sqrt{\frac{100000 \times 9.81}{0.5 \times 0.2 \times 10^5 \times 250 \times 1.4 \times 0.5}} = 0.749$$

(c)(i)

$$V = M\sqrt{\gamma RT} = 0.749\sqrt{1.4 \times 287 \times 220} = 222.6 \text{ m/s}$$

$$\frac{W_1}{W_2} = \exp\left(\frac{s\ g\ sfc}{V\frac{L}{D}}\right) = \exp\left(\frac{4000 \times 9.81 \times 0.014}{222.6 \times 20}\right) = 1.131$$

$$\therefore W_2 = 88.4 \text{ tonnes}$$

$$C_L = \frac{L}{0.5\rho AV^2} = \frac{mg}{0.5\frac{p}{RT}AV^2}$$

This implies that

$$p = \frac{mgRT}{0.5AV^2C_L} = \frac{mg}{0.5C_L\gamma AM^2} = \frac{88.4 \times 10^3 \times 9.81}{0.5 \times 0.5 \times 1.4 \times 250 \times 0.749^2} = 17.7 \text{ kPa}$$

(c)(ii) To keep the cruising at constant altitude means that the pressure in the lift coefficient is constant.

$$C_L = \frac{mg}{0.5pA\gamma M^2}$$

If the lift coefficient is kept optimal ~0.5 then as the mass of the aircraft drops the Mach number of the flight must be dropped. This will increase the flight time which is often undesirable.

If the Mach number is kept constant, then the lift coefficient drops through flight and *L/D* drops. This means that the range of the aircraft is reduced.

(d)

$$\eta_{ov} = \frac{VF_N}{m_f LCV} = \frac{V}{sfc\ LCV}$$

$$s = \frac{L/D}{g}\,\eta_{ov}\,LCV\ln\left(\frac{W_1}{W_2}\right)$$

Assuming constant L/D and W1/W2 :

$$\frac{d\,s}{d\,\eta_{ov}} = \frac{L/D}{g}\,LCV\ln\left(\frac{W_1}{W_2}\right) = \frac{20}{9.81}43\times10^6\ln(1.131) = 10.8\times10^3\text{km}$$

$$\Delta s = \frac{d\,s}{d\,\eta_{ov}}\Delta\eta_{ov} = 10.8\times10^3\times0.01 = 108\text{ km}$$

(e)

$$\eta_{ov} = \eta_p\eta_{th}$$

If the propulsive efficiency is fixed, we need to increase the thermal efficiency of the engine and this can be achieved by:

- Higher engine pressure ratio (challenge is increased number of compressor stages, smaller blade heights at compressor exit, larger tip clearance, higher gas temperature at compressor exit leading to higher specification materials)
- Raising turbine entry temperature (challenge is cooling of turbine)
- Higher isentropic efficiencies of the compressor and turbine (challenge is these are already high)

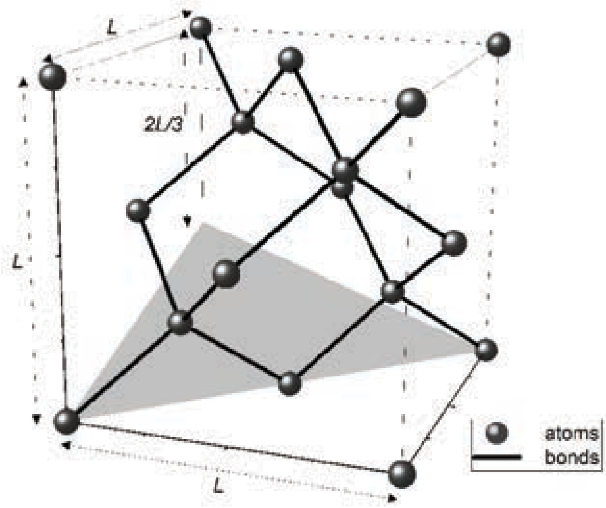
SECTION E: *Electrical Engineering*

1 (a) (i) Diamond structure [1]

(ii) Carbon, silicon and germanium. [2]

(iii) (1 1 1) [2]

(b) (i)



[3]

(ii) The general formula for plane spacing d based on a lattice of side lengths a, b and c and Miller indices h, k and l is

$$\frac{1}{d^2} = \frac{h^2}{a^2} + \frac{k^2}{b^2} + \frac{l^2}{c^2}$$

For a cubic lattice where $a = b = c = L$, this becomes

$$d = L/\sqrt{h^2 + k^2 + l^2}$$

$$d = 0.543/\sqrt{1^2 + 1^2 + 3^2}$$

$$d = 0.164 \text{ nm}$$ [4]

(iii) We need to work out the wavelength of the X-rays with a photon energy of 8.04 keV. As photons are massless, we must use $E = hc/\lambda$, so (remembering to use h with units of eV s)

$$\lambda = \frac{hc}{E} = \frac{4.136\times10^{-15}\times2.998\times10^{8}}{8040} = 1.542 \times 10^{-10} \text{ m}$$

Diffraction peaks (from the Bragg Law) occur for $n\lambda = 2d\sin\theta$. So taking the lowest order peak when $n = 1$ gives

$$\theta = \sin^{-1}\left(\frac{\lambda}{2d}\right) = \sin^{-1}\left(\frac{1.542\times10^{-10}}{2\times1.64\times10^{-10}}\right) = 28.0° \qquad [4]$$

(iv) At the minimum X-ray photon energy, $\theta = 90°$, and so $\sin\theta = 1$, meaning that the photon wavelength must be

$$\lambda = 2d = 2 \times 0.164 = 0.328 \text{ nm}$$

The corresponding photon energy is then

$$E = \frac{hc}{\lambda} = \frac{4.136\times10^{-15}\times2.998\times10^{8}}{3.28\times10^{-10}} = 3.78 \text{ keV} \qquad [4]$$

(c) As we know the conductivity of the semiconductor and that it is intrinsic, so $n = p = n_i$, we can work out the intrinsic carrier concentration from

$$\sigma = ne\mu_e + pe\mu_h = n_i e(\mu_e + \mu_h)$$

$$n_i = \frac{\sigma}{e(\mu_e+\mu_h)} = \frac{5\times10^{-4}}{1.602\times10^{-19}(1400+450)} = 1.68 \times 10^{20} \text{ cm}^{-3}$$

As it is a pure semiconductor in group IV, it will ionise with either a single positive or single negative charge. Therefore, we need to calculate the number density of atoms *N*, divide this by the intrinsic carrier concentration, and this will give us an estimate of probability of any one atom being ionised. We can calculate N as each unit cell has eight atoms attributed to it and is a cube of known side length, so (working in units of cm)

$$N = \frac{8}{(0.543\times10^{-7})^3} = 5 \times 10^{22} \text{ cm}^{-3}$$

Hence

$$P_{ionised} \sim \frac{n_i}{N} = \frac{1.68\times10^{20}}{5\times10^{22}} = 3.4 \times 10^{-3} \qquad [5]$$

2 (a) (i) Active matrix liquid crystal display. [2]

(ii) The TFT uses a thin film semiconductor for the channel and is fabricated on an insulating substrate, whereas the MOSFET uses the bulk substrate semiconductor for the channel. The TFT is operated in an accumulation mode whereas the MOSFET is operated typically in inversion mode. Hence, the MOSFET as implanted wells of the inversion carier at the source and drain. The MOSFET usually uses a top gate structure whereas the TFT employs a bottom gate most commonly. [4]

(b) An electronic engineer needs to specify a TFT to control the current flow through an Organic Light Emitting Diode (OLED) pixel in a display. The OLED needs to draw a current of 1 μA when the voltages on the TFT are V_{ds} = 1 V and V_{gs} = 2 V. The threshold voltage V_t = 0.5 V. The engineer has fabrication facilities which allow them to choose from the materials shown in Table E.1 (over page) and the minimum channel length which can be fabricated is 1 μm. They have been asked to ensure that the area occupied by the TFT in the pixel is minimised.

(i) We know that a current of 1 μA when the voltages on the TFT are V_{ds} = 1 V and V_{gs} = 2 V and the threshold voltage V_t = 0.5 V. Therefore, substitution into the MOSFET equation gives

$$I_{DS} = \mu_{FE} \frac{W}{L} C_{ox}[(V_{GS} - V_T - V_{DS}^2/2)]$$

$$10^{-6} = \mu_{FE} \frac{W}{L} C_{ox}[(2 - 0.5 - 1^2/2)]$$

$$\frac{W}{L} = \frac{10^{-6}}{\mu_{FE} C_{ox}}$$

To minimise area, we must minimise W for a fixed L (the minimum allowable). Therefore, we should use the semiconductor with the greatest mobility, which is amorphous indium gallium zinc oxide and the dielectric with the greatest permittivity (to maximise C_{ox} for the same physical dimensions) which is aluminium oxide. We know that

$$C_{ox} = \frac{\varepsilon_0 \varepsilon_r}{t}$$

remembering that C_{ox} is the capacitance per unit area. Hence substitution gives

$$\frac{W}{L} = \frac{10^{-6} t}{\mu_{FE} \varepsilon_0 \varepsilon_r}$$

Substituting for L = 1 µm and $\varepsilon_r = 9.1$ gives an explicit relationship between W and t for our design

$$t = W \times 80.6 \times 10^{-3}$$

Therefore, if we also choose W = 1 µm to keep area low, then we end up with t = 80.6 nm, which is reasonable. Indeed, any solution which keeps 0.5 µm < W < 2 µm and therefore 40 nm < t < 160 nm would be excellent. [14]

(ii) IN the scenario where V_{ds} = 1 V and V_{gs} = 0 V, the TFT is 'off' and so current is just the conductivity of the semiconductor in the channel to a good first approximation (assuming no gate leakage). Thereforem we need to first know the conductivity of the semiconductor, which is amorphous indium gallium zinc oxide for our design.

$$\sigma = ne\mu_e = 10^{13} \times 1.602 \times 10^{-19} \times 10$$
$$\sigma = 1.6 \times 10^{-5}\ \Omega^{-1}\ \mathrm{cm}^{-1} = 1.6 \times 10^{-3}\ \Omega^{-1}\ \mathrm{m}^{-1}$$

Then the resistance of the channel is

$$\frac{V_{DS}}{I_{DS}} = \frac{L}{\sigma W t_{semi}}$$

where t_{semi} is the thickness of the channel semiconductor. Therefore

$$I_{DS} = \frac{V_{DS}\sigma W t_{semi}}{L} = \frac{1 \times 1.6 \times 10^{-3} \times 1 \times 100 \times 10^{-9}}{1} = 160\ \mathrm{pA}$$ [5]

NOTE: For the TFT the equation governing current flow is

$$I_{ds} = \mu \frac{W}{L} C_{ox}\left[\left(V_{gs} - V_t\right)V_{ds} - \left(V_{ds}^2/2\right)\right]$$

3 (a) (i) The deposition rate is given by

$$R = k\frac{\cos\alpha\cos\beta}{d^2}$$

where k is a constant of proportionality. At point A, $\alpha = \beta = 45°$ and so the deposition rate R_A is

$$R_A = k\frac{\cos 45°\cos 45°}{2L^2} = \frac{1}{4}\frac{k}{L^2}$$

At point B,

$$\cos\alpha = \frac{L}{L\sqrt{5}} = \frac{1}{\sqrt{5}}$$

$$\cos\beta = \frac{2L}{L\sqrt{5}} = \frac{2}{\sqrt{5}}$$

$$R_B = k\frac{\frac{1}{\sqrt{5}}\frac{2}{\sqrt{5}}}{5L^2} = \frac{2}{25}\frac{k}{L^2}$$

Hence,

$$\frac{R_A}{R_B} = \frac{25}{8}$$ [4]

(ii) Electron beam evaporation uses a beam of high energy electrons to heat the metal to be evaporated which is in a small crucible. In thermal evaporation, by contrast, the metal to be evaporated is either placed in a small metal filament made of a material with a higher melting temperature or a wire of a similar high melting temperature metal is coated in the metal to be evaporated and a current passed through the filament to heat and evaporate the lower melting temperature metal. [2]

(iii) The advantages of electron beam evaporation include lower contamination as the crucible is not in contact with the molten evaporating metal and less residual heating of the sample compared with thermal evaporation. Disadvantages of electron beam evaporation include high cost and complexity, and the risk of X-ray generation. [3]

(b) (i) The key advantages of hard disk drives are high density of data storage and good reliability. As a result, they are particularly good for data centres.

(ii) The three conditions are:

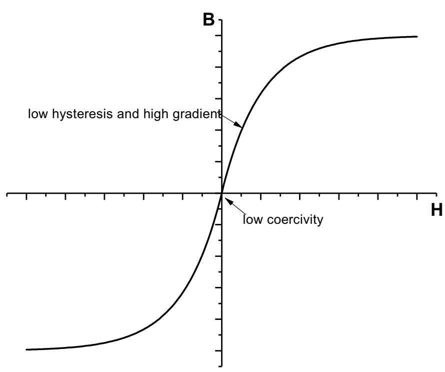
1. The system must stably exist in two distinct states. This is achieved by there being 'easy directions' in the ferromagnetic crystal lattice with a high energy penalty for pointing in other directions. As a result, magnetisation is forced to be in the same direction as the writing head. Domain size must be sufficiently large to ensure stability for more than 10 years.
2. It must be possible to switch between the two states. Therefore, the energy required to rotate the dipoles must be low compared with the energy required to move domain walls.
3. There must be a measurable difference between the two states. This requires a high remnance flux. [3]

(iii)

→	←	→	←	←	→	→	→	←

[4]

(iv) We need to be able to complete a magnetic circuit, so an underlayer allows a lateral flux between the perpendicular grains where the flux is passing to or from the read/write head. A soft magnetic material is required for this which requires little energy to magnetise, so there must be low hysteresis (small area inside the B-H curve), a high relative permeability (a high gradient around the origin of the B-H curve) and a low coercivity.



[6]

Numerical Solutions

1. (a) (iii) (1 1 1)

(b) (ii) 0.164 nm

(iii) $\theta = 28.0^{\circ}$

(iv) 3.78 keV

(c) 3.4×10^{-3}

2 (b) (i) Amorphous indium gallium zinc oxide, aluminium oxide, $L = 1$ μm with other parameters depending on design and good values being $W = 1$ μm and $t = 80.6$ nm.

(ii) 160 pA (based on above design)

3 (a) (i) 25/8

1 (a) Factors: the position of the camera, the properties of the lens and the CCD, the shape of the structures in the scene, the nature and distribution of light sources, the reflectance properties of the visible surfaces.

Edge detection: Provides significant data reduction while preserving useful information content (i.e. it is possible to recognise objects from line drawings). Most of the discarded information is not useful for recovering scene structure and motion.

(b) (i) In order to remove high frequency noise which is amplified by differentiation.

Fourier transform of a Gaussian shaped pulse g_σ is another (unnormalised) Gaussian shaped pulse $G_{\sigma'}$ in a frequency domain (i.e. $g_\sigma(x) \rightleftharpoons G_{\sigma'}(f)$), where $\sigma' \propto \frac{1}{\sigma}$.

(ii) Perform two discrete 1D convolutions (kernel size $2n+1$) to obtain a smoothed image and two discrete 1D convolutions for differentiantion (kernel $[0.5, 0, -0.5]$). In total $2 \cdot (2n+1+2n) + 2 \cdot (3+2) = 8n+12$ operations (additions and multiplications) per pixel. Alternatively, differentiating a smoothed signal is equivalent to convolving the signal with a differentiated smoothing kernel: $\nabla S = \left[\frac{\partial(G_\sigma * I)}{\partial x}, \frac{\partial(G_\sigma * I)}{\partial y}\right]^T = \left[\frac{\partial G_\sigma}{\partial x} * I, \frac{\partial G_\sigma}{\partial y} * I\right]^T$.
The convolution with $\frac{\partial G_\sigma(x,y)}{\partial x}$ can be decomposed into two 1D convolutions as follows: $\frac{\partial G_\sigma(x,y)}{\partial x} * I(x,y) = g'^T_\sigma(x) * (g_\sigma(y) * I(x,y))$. Here g'^T_σ is a filter approximating a derivative of a 1-D Gaussian kernel applied row-wise. $g_\sigma(y)$ is a filter approximating a 1-D Gaussian filter applied column-wise. Similarly: $\frac{\partial G_\sigma(x,y)}{\partial y} * I(x,y) = g^T_\sigma(x) * (g'_\sigma(y) * I(x,y))$. The number of per-pixel operations (multiplications and additions) required for convolving with one 1-D filter is $2n+1+2n = 4n+1$. The total amount of operations required is $2 \cdot 2 \cdot (4n+1) = 16n+4$. Note we can ignore the number of operations required in order to compute derivative of Gaussian kernels.

(c) (i) Edge detection is performed by finding zero-crossings in the response of the Laplacian convolved with the smoothed image. In blob detection the minimum and maximum points are found instead.

(ii) By examining first order Taylor expansions: $\frac{dI}{dx}|_{(x)} \approx I(x) - I(x-1)$.
Hence $\frac{d^2I}{dx^2}|_{(x)} \approx (I(x+1) - I(x)) - (I(x) - I(x-1)) = I(x+1) - 2I(x) + I(x-1)$.

(iii) $\nabla^2 I|_{(x,y)} = \frac{\partial^2 I}{\partial x^2}|_{(x,y)} + \frac{\partial^2 I}{\partial y^2}|_{(x,y)}$.

Hence the required filter is: $\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$.

Integer operations are more efficient and avoid issues with floating point precision.

(iv) Fourier transform of the Laplacian of the Gaussian $\frac{d^2 g_\sigma}{dx^2}$ is proportional to $-\omega^2 G_{\sigma'}(\omega)$ in angular frequency domain (i.e. $\frac{d^2 g_\sigma(x)}{dx^2} \rightleftharpoons -\omega^2 G_{\sigma'}(\omega)$). Hence it behaves as $-\omega^2$ near origin, and is close to zero for significantly large ω.

Scaled LoG can be approximated by a difference of two Gaussians: $G(x, y, k\sigma) - G(x, y, \sigma) \approx (k-1)\sigma^2\nabla^2 G(x, y, \sigma)$. This can be proven by Taylor series expansion around the standard deviation σ of the Gaussian.

Hence *blob-like* feature image position and scale can be found by looking at local minimum and maximum points between neighbouring Gaussian smoothed images (within same octave) in an image pyramid : $S(x, y, \sigma_{i+1}) - S(x, y, \sigma_i)$.
The point is selected as a blob for a particular position (x_i, y_i) and scale σ_i if it is a local minimum or maximum point in its imediate 3×3 neighbourhood on the same scale σ_i and neighbouring scale images σ_{i-1} and σ_{i+1} in an image pyramid.

(d) (i) Steps:
- Produce a 16×16 patch at selected scale and orientation.
- Compute ∇S at each pixel.
- Weight it by $G_\sigma(x, y)$.
- Split the patch into 16 (4×4) cells.
- For each cell compute orientation histograms in 8 directions.
-Produce a 128D vector by concatenating histograms of cells and normalize it.
-Threshold elements higher than 0.2.
-Renormalise the vector.

Invariance to lighting: (i) use of gradients makes SIFT invariant to constant change in brightness, (ii) normalising histogram of orientations allows to remain invariant to contrast (multiplicative change in illumination), (iii) thresholding of the orientation histogram (followed by re-normalisation) builds robustness to non-linear changes in illumination (e.g. specularities).

Invariance to viewpoint changes: (i) SIFT is translation invariant due to its local nature of descriptor (calculated only on a portion of image), (ii) invariant to an in-plane rotation (some tolerance to out-of-plane rotation) changes as it is calculated for a given orientation, (iii) invariant to scale as it is calculated at a pre-determined scale, (iv) Gaussian weighting builds some tolerance to small occlusions.

Limitations: (i) large viewpoint changes (e.g. producing out-of-plane rotations), (ii) occluding boundaries.

(ii) First, SIFT descriptors are computed for a dataset of images. If multiple dominant orientations are present, several descriptors for the same image point are stored. For each SIFT feature in a query image, two nearest neighbours (e.g. using K-D tree algorithm) are found using Euclidean distance. Match is accepted, if the ratio between the second closest distance and the closest distance is less than 0.7. Image with most matches is retrieved.

2 A company that makes autonomous vehicles is in the process of designing a convolutional neural network (CNN) for classifying road conditions. The specification of the CNN is only partly complete. Currently it takes images Z as input and returns an output $\mathbf{a}$ which is a vector with $D = 100$ real-valued elements a_d that lie in the range $-\infty \leq a_d \leq \infty$.

(a) Explain why a CNN is more suitable for this task than a single neuron or a multi-layer perceptron (MLP). [6]

Single neuron – only implements a linear classification decision boundary. Road condition classification will require non-linear decision boundaries

Multilayer perceptron – can implement non-linear decision boundaries but will have too many parameters if applied to images (will overfit / require lots of data to train / will be computationally too intensive)

CNN – parameter sharing cuts down the number of parameters required & pooling builds in invariance / high level coarser features

(b) The company has collected a dataset comprising images Z taken by front facing cameras and associated labels y that indicate whether the road is dry ($y = 1$), wet ($y = 2$), or icy ($y = 3$).

(i) Describe how to complete the design of the CNN by transforming the output vector **a** into a form that is suitable for classifying unseen images into the three categories of road condition. [4]

b) i) Currently we have $\underline{a} = CNN_\theta(Z)$

to make appropriate for 3 class classification

first apply a linear layer to get a 3 dimensional vector:

dimensions
$$\underset{3\times 1}{\underline{h}} = \underset{3\times 100}{\underline{\underline{W}}}\ \underset{100\times 1}{\underline{a}}$$

second apply a soft max function to produce probabilities for each class

$$p(y=k \mid Z) = \frac{\exp(h_k)}{\sum_k \exp(h_k)} = x_k(Z; \underline{\underline{W}}, \theta)$$

(ii) Mathematically define an objective function that measures the quality of the complete CNN's predictions on training data $\{Z_n, y_n\}_{n=1}^{N}$. Explain the rationale behind the form of the objective function. [4]

ii) $$G(\underline{\underline{W}}, \theta) = -\sum_{n=1}^{N} \sum_{k=1}^{K} y_{nk} \log\left[x_k(Z_n; \underline{\underline{W}}, \theta)\right]$$

should be minimised (lower better)

matrix w/ $y_{nk} = 1$ if data point n is class k
$y_{nk} = 0$ otherwise

This is equivalent to the negative log probability of the class labels given the parameters and input images i.e.

$$-\log p(y_{1:N} \mid Z_{1:N}, W, \theta)$$

(iii) Briefly describe how the objective function can be used to train the CNN.

Your answer should outline the general approach to performing learning, but detailed derivations are not required. [4]

iii) should include

- description of gradient descent

$$\theta^{(new)} = \theta^{(old)} - \nu \frac{dG}{d\theta}$$

learning rate

gradient of objective wrt parameters

new setting of parameters

old setting of parameters

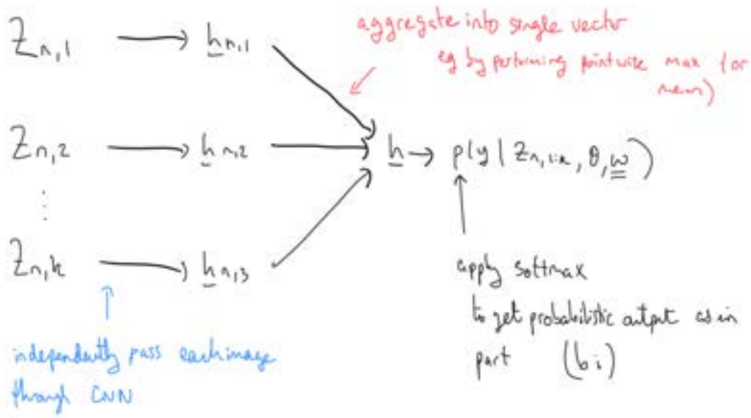
- use of minibatches to accelerate learning i.e. stochastic gradient descent

- mention regularisation / weight decay to avoid over fitting

(c) The company's autonomous vehicle fleet includes cars with different numbers of front facing cameras. This means that, in practice, several images are simultaneously collected of the same road from different positions, but the number of images K depends on the car. It is important to use all available cameras to classify the road conditions as patches of ice or water are often only visible to one camera. The company would therefore like to build a single system for all vehicles that can take a set of K images as input $\{Z_{n,k}\}_{k=1}^{K}$ (where K can vary) and return a single classification of the road conditions.

Design such a system using the trained CNN developed in part (b). Your answer should explain how aspects of your design reflect the need to handle (i) different numbers of input images K and, (ii) hard to spot road conditions that are visible to only one camera. [7]

c) Many different approaches here e.g.



- The aggregation step allows arbitrary numbers of images to be input int the system.

- If a max function is used to perform the aggregation then it can allow a single input image with evidence of eg. ice to dominate over the others

- Could aggregate predictions too (rather than on the hidden neuron activations) but it's hard to construct a scheme that yields a valid (normalised) probability & which allows evidence in a small subset of cameras to have a large effect on the classification.

3 Figure 1 depicts a map of locations with the road distance between each pair of locations, in km, indicated on the line joining them. The (x, y) location, in km relative to node C , is also given for each node.

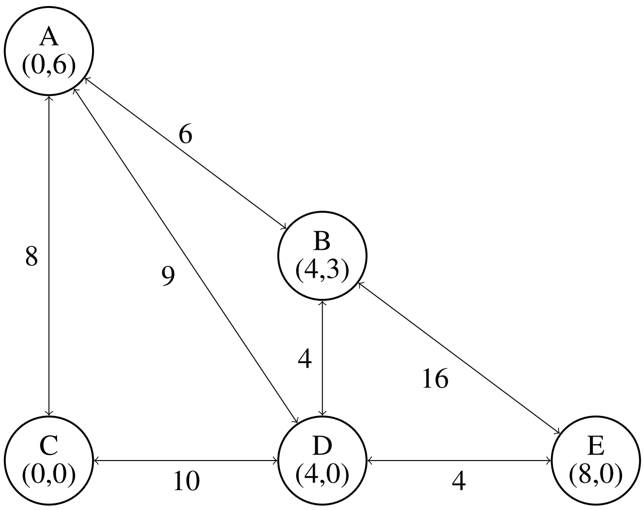


Fig. 1

(a) Use *Dynamic Programming* to label each node with the shortest path (i.e. distance along roads) to node E. Explain how you know this is the optimal solution. [25%]

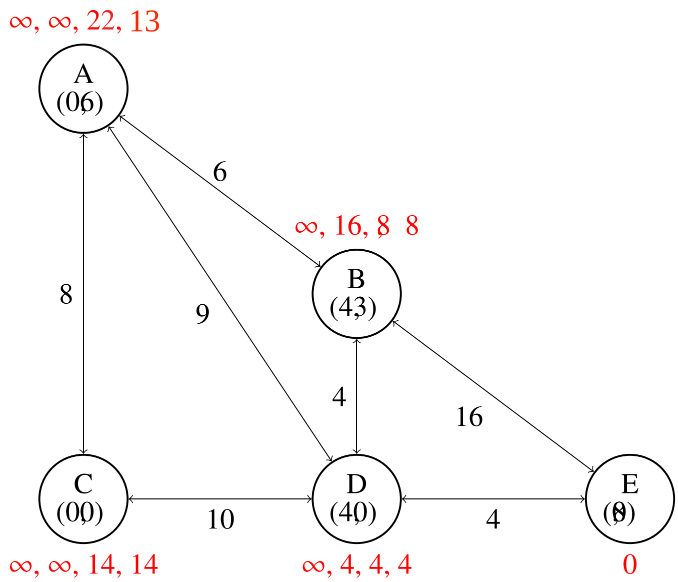
(b) Use Dijkstra's algorithm to find the shortest path between nodes C and node E, starting at C (i.e. labelling E with a distance of 0). Is it more efficient than Dynamic Programming for this example. [25%]

(c) Now use the A^* algorithm to find the shortest path between nodes C and E, starting at C. Does it help in this case? [25%]

(d) Now consider an agent trying to discover the shortest path from C to E using Q-learning with an ϵ-greedy action selection method. Set the initial Q for node E to 0, for all actions. Give a possible sequence of actions, starting at A and restarting at A whenever E is reached, and the corresponding estimates of Q at each stage, whenever it is changed for each state-action pair. [25%]

Crib:

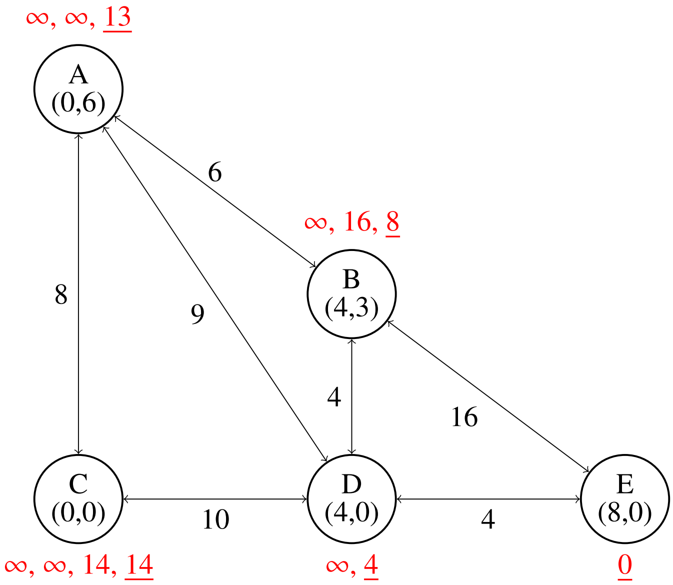
1 (a) If th o d r f lab lling nd s is $A \to E$ th n th lab lling will o cn as in th fb lowing d agram



Th each nd need to b v sited fo times b fo e th lab ls stp ch ng ng Th s is why it is known th t th lab lling is p imal, i.e. fo each nd

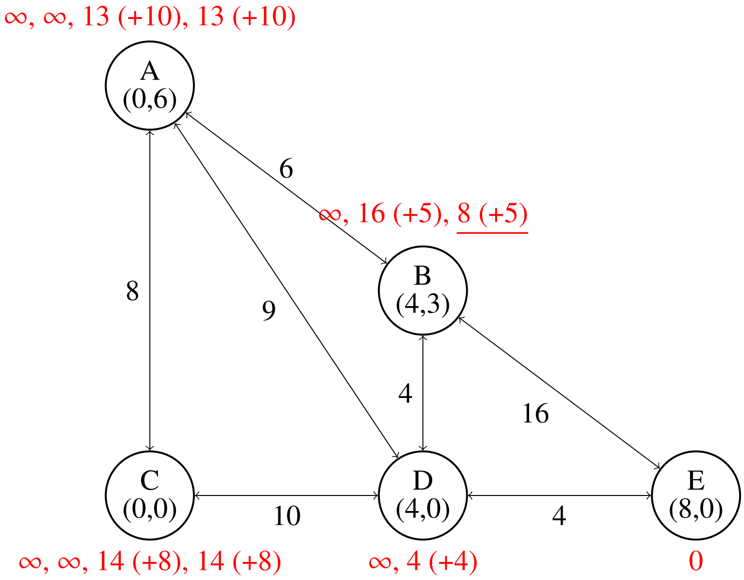
$$V(\cdot) = \min_{\text{neighbours}} \text{eg e leng h} + V(\text{neig} \quad)$$

(b In th fb lowing fig e th v sited nd at each stage is nd rlined



Every node has been visited, however they have only be visited once – and it is known at this stage that the solution is optimal. Thus the solution is significantly more efficient that dynamic programming even though no pruning of nodes has been achieved.

(c) A heuristic, here consisting of the Euclidean distance from each node to the destination, is added to the tentative distance from each node when selecting the next node to visit. (Note that the triangle sides are in the ratios 3:4:5).



The A^* algorithm here results in a slight saving as node A is never visited

(d) A Q-learning algorithm would wander aimlessly until stumbling upon node E. If the previous node was B, for example, as in

$CACDBE$

then the outgoing edge from B to E would be labelled with 16, so that $\min_a Q(B, a) = 16$. Restarting from C, for example, with all moves being random until B or E is encountered:

$CACDB$

now the outgoing edge from D to B is labelled with 4+16=20, and $\min_a Q(D, a) = 20$

The next action is likely, with probability $1 - \epsilon$ to be to

E

If greedy action selection is now followed, then sequences will end either BE or DBE.

Eventually, however, the action of moving to E will be taken from D (probability $\epsilon/3$ each time D is visited). Then the outgoing edge from D to E is labelled with 4, and $\min_a Q(D, a) = 4$.

Now $1 - \epsilon$ of paths that reach D will move straight to E. If a subsequent path is

$CADE$

then the outgoing edge from A to D will be labelled with 13, and $\min_a Q(A, a) = 13$.

Eventually, with probability ϵ each time, D will be chosen after C, and the outgoing edge from C to D will be labelled with 14, and $\min_a Q(A, a) = 14$. At this stage the optimal path has been found, the process however is far from optimal.

END OF PAPER

SECTION G: *Bioengineering*

*Answer not more than **two** questions from this section.*

1 (a) Explain what the term *tomography* means in Optical Coherence Tomography (OCT) of the eye. Describe how tomographic imaging is achieved in OCT, including how the resolution and the location of the image data are controlled. [6]

Answer: *Tomography* refers to the ability of these imaging modalities to return image data from a specific depth into the tissue, rather than just returning data from the first visible surface (like a photograph) or summed over a wide focal region (like a conventional microscope).

In time-domain coherence tomography, a laser pulse is used rather than a continuous laser. Tomography is achieved using a Michelson Interferometer. In this system, part of the laser pulse is sent to the eye, and part is sent to a reference mirror. The reflections from the eye and the reference mirror are then re-combined before being sent to a photo-detector. When the optical path lengths to the mirror and to the back of the eye are nearly identical, this causes interference fringes. These fringes show up as oscillations in the time-domain light signal, and are picked up by looking for high frequency content in this signal.

Moving the reference mirror changes the depth at which interference will occur, and hence also the depth at which the detector is sensitive to reflected light. The depth sensitivity is also slightly affected by the optics, i.e. the focal point of the objective lens.

The depth resolution in this system is controlled by the response of the interferometer, but is essentially the extent (length in time) of the laser pulse which is used. Resolution in the other directions is affected by the focal diameter of the lenses and also the spot size of the laser once it reaches the back of the eye. This may be larger than the achievable spot size in air since the laser will disperse as it passes through the eye.

(b) An OCT system illuminates the fundus with a short optical pulse E, of duration a (in s) and centre frequency ω_0 (in rad/s):

$$E = \begin{cases} e^{j\,\omega_0\,t} & -\frac{a}{2} < t < \frac{a}{2} \\ 0 & \text{otherwise} \end{cases}$$

What is the bandwidth B of this pulse (in Hz) ignoring all but the main lobe of the frequency response? [6]

Answer: In order to answer this, the Fourier Transform of the pulse E must be calculated:

$$\mathcal{F}(E) = \int_{-\frac{a}{2}}^{\frac{a}{2}} e^{j\omega_0 t} e^{-j\omega t}\, \mathrm{d}t$$

Combining the exponentials, integrating and using Euler's formula gives:

$$\begin{aligned}
\mathcal{F}(E) &= \int_{-\frac{a}{2}}^{\frac{a}{2}} e^{jt(\omega_0-\omega)}\,\mathrm{d}t \\
&= \left[\frac{e^{jt(\omega_0-\omega)}}{j(\omega_0-\omega)}\right]_{-\frac{a}{2}}^{\frac{a}{2}} \\
&= \frac{2}{(\omega_0-\omega)}\sin\left(\frac{a(\omega_0-\omega)}{2}\right) \\
&= a\ \mathrm{sinc}\left(\frac{a(\omega_0-\omega)}{2}\right)
\end{aligned}$$

Alternatively, this can be deduced from the data book, by noting that the given expression for E is just a rectangular pulse of duration a (a standard result) multiplied by a complex exponential (which shifts this result in frequency). The main lobe of a sinc function extends from $-\pi$ to π, hence:

$$\begin{aligned}
\frac{a(\omega_0-\omega)}{2} &= \pm\pi \\
\omega &= \omega_0 \pm \frac{2\pi}{a}
\end{aligned}$$

So the bandwidth B, in Hz, is:

$$B = \frac{4\pi}{2\pi a} = \frac{2}{a}$$

(c) A spectral OCT system uses a linear array of N photodiodes to detect the spectrum of the reflected pulse. These diodes are positioned such that they cover a frequency range Z (in Hz), which is always centred around ω_0. The speed of light in the fundus is c.

(i) Explain why the depth spacing between image samples, d, is given by $d = \frac{c}{Z}$. [3]

Answer: In spectral OCT, the frequency-domain measurements from this photo array are converted to time-domain measurements by use of the inverse Fourier Transform. In such a transform, if the sampling (maximum) frequency is S, then the difference between each time-domain measurement is the sampling period $\frac{1}{S}$. The diffence in depth is then $\frac{c}{S}$. Z is not quite the sampling frequency, since this frequency range does not necessarily start at zero. However, the only difference is a shift in frequency, which only affects the phase of the resulting inverse Fourier Transform (see data book). Hence $S = Z$ and $d = \frac{c}{Z}$.

(ii) If this OCT system uses the pulse E given in part (b), how does the spacing d relate to the pulse duration a? Explain any assumptions you make about the system design. [2]

Answer: We assume that the frequency range of the photodiode array should cover the same frequencies as are in the pulse, hence $Z = B$. Since $d = \frac{c}{Z}$ and $B = \frac{2}{a}$, then $d = \frac{c}{B} = \frac{ca}{2}$.

(iii) The pulse E and number of photodiodes N are fixed, but the frequency range Z is allowed to decrease, such that the total imaging depth increases. What limits the maximum imaging depth in this scenario, and what are the other consequences of increasing the imaging depth in this way? [4]

Answer: Since the sample spacing is $\frac{c}{Z}$, and there are N samples, the imaging depth is $\frac{Nc}{Z}$. Hence it can be increased aribitrarily by reducing Z. However, just because the spectral resolution is good enough to image to a certain depth does not mean there is actually any information at that depth. In practice the tissue attenuation will prevent the light being reflected by anything more than a few millimetres deep. Any imaging data deeper than this will just contain measurement noise.

Another consequence of reducing Z is that we are only measuring a small part of the actual bandwidth of the pulse B. We must, therefore, be losing some information from the signal. In fact, reducing Z also increases the sample spacing d. Since this spacing is usually set to $\frac{ca}{2}$, it does not take much to start sampling at worse than the optical resolution of the system (i.e. the pulse width a). This will reduce the effective resolution of the system and also introduce aliasing.

(iv) The pulse E and number of photodiodes N are again fixed, but now the frequency range Z is allowed to increase in order to improve the axial (depth) resolution. To what extent is this possible, and what would be the limiting factor in improving the resolution? What would eventually happen to the image data if Z was allowed to increase substantially? [4]

Answer: Since we are not changing the pulse, the resolution can not be better than the pulse width a. Hence any decrease in sample spacing d below $\frac{ca}{2}$ will only over-sample the data, rather than changing the resolution at all. The pulse is the limiting factor in resolution in this scenario.

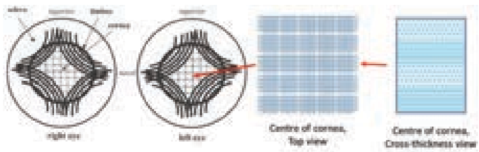
If Z is allowed to increase further, then the photodiode array will be detecting a frequency range which is broader than the bandwidth of the pulse B. Hence many of the diodes will only be measuring noise, and this noise will quickly swamp the inverse Fourier Transform, resulting in the loss of *any* useful imaging data at all.

2 (a) (i) Describe the functional roles of the cornea, sclera, and aqueous humour in the eye. [2]

Answer: Cornea provides provides a refractive index by bending light and provides 2/3 of focal power. Sclera provides structural support for the eye ball and maintains its shape. Aqueous humour provides lubrication to nourish the eye and acts as a stress and heat buffer.

(ii) Explain how the organisation of collagen and extracellular matrix fibres relates to the different functions of these eye components. [4]

Answer: Cornea consists of uniform-diameter collagen fibres organised with uniform spacing and packing in a layered fashion (see scheme below for the fibre organisation). In the centre of the eye, this provides good strength, toughness, and transparency. Collagen fibres on the edge of the eye are directed towards muscle attachment points (i.e. along the dominating tension axis).



Sclera consists of random web of fibres, and there is a low level of interweaving fibres between layers (see below). The random fibre packing induces an opaqute nature of sclera (but transparency is not needed), but improves the toughness, and compliance of the sclera for accomondating IOP (intra-occular pressure).

Vitreous humour has a low density random gel network with collage fibres, which gives modest structural properties and transparency but allows high permeability for nutrient transport, and acts as a heat and stress buffer due to its high water content and low fibre content.

(iii) Based on the microstructure of the cornea and sclera, explain why it is more convenient to deliver liquid drug formulation to the retina through the sclera than through the cornea. [2]

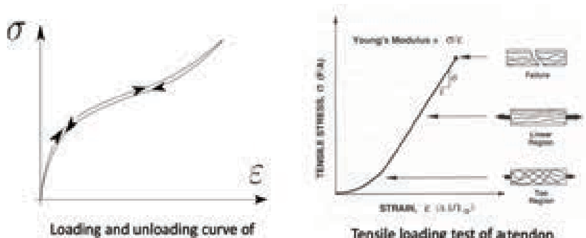
Answer: The eyeball has a globe shape where the exteriorly accessible areas are covered by the cornea and the sclera. As the cornea is tightly packed by crystalline collagen fibres, the cornea has a low permeability. The more open network packing structure of sclera affords higher permeability which supports a liquid drug transport across the sclera into the entre of the eye to the retina.

(b) Explain how ageing could alter the mechanical properties of soft biological tissues. [2]

Answer: In general, ageing increases the cross-linking density of extracellular matrices and reduces the protein turnover in biological tissues. Thus, one would expect the soft biological tissue to have higher Young's modulus (stiffer), less ductility (more brittle), but lower strength.

(c) (i) Explain what the term 'non linear elasticity' means when referring to the mechanical properties of materials. Describe the origin of non linear elasticity of soft biological tissues, and use appropriate sketches to aid your answer. [4]

Answer: In non-linear elasticity, the stress-strain curve is reversible (loading and unloading curves overlap, see below), but the curve deviates from a linear stress-strain relationship. Non-linear elasticity can be arisen as the internal molecular arrangement of the materials evolves with the deformation. The molecular arrangement is then recovered after the deformation is released. For example, tensile stretching of collagen fibres, typically shows a toe-linear shape of stress-strain curve. This shape results from the orientation and sequential 'recruitment' of collagen fibres. As fibres reorient, they can support more forces. The fibres can be re-coiled upon the force is released (i.e. the deformation is recoverable, and the process is elastic).



Loading and unloading curve of an arbitrary material showing non=linear elasticity

Tensile loading test of a tendon sample, made of predominately collagen fibres.

(ii) Sequential recruitment of linear springs is a useful model for simulating the non linear elasticity of biological tissues. For example, the stress-strain data shown in Fig G1 can be modelled using a four-spring sequential recruitment model. Sketch an appropriate configuration for such a model, and estimate the appropriate parameter values for the model. [6]

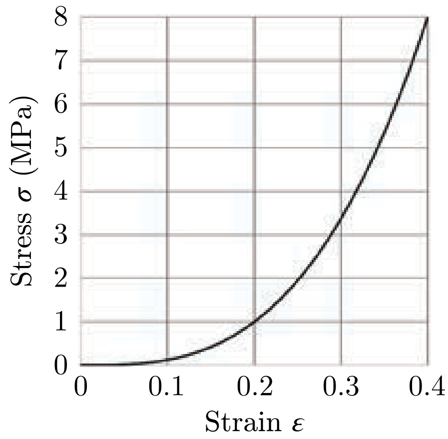
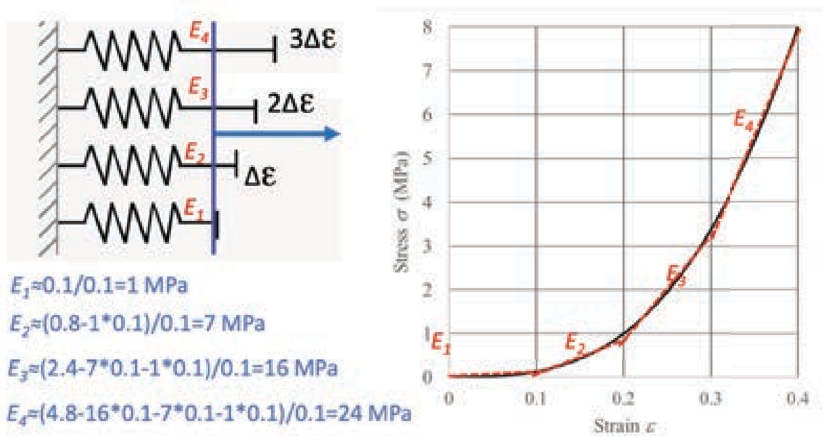


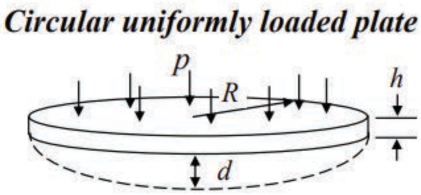
Fig G1

Answer: The figure below illustrates a suitable model, with $\Delta\epsilon = 0.1$. With this model, matching the stress-strain response with four straight-line segments as illustrated, which gives the spring stiffness values of E_1, E_2, E_3, and E_4 as shown below.



(d) A simplified elastic plate model can be used to model the deformation of *lamina cribrosa* resulting from the intraocular pressure of the eye. Describe such a model, explain the assumptions made, and justify the validity of the model. You may use sketch(es) to aid your answer. [5]

Answer: The elastic plate model assumes that the *lamina cribrosa* is a circular elastic membrane (e.g. of thickness h); this membrane is being clamped at its circumference. A uniform load of p, which is approximated to be a result of IOP is applied at the membrane. The model also assumed that the neural tissue at either sides of the *lamina cribrosa* have negligible mechanical interference to the system, and are omitted in the membrane deformation.



(p≈IOP; h: thickness of *lamina cribrosa*; R: radius of *lamina cribrosa*; d: displacement of the membrane due to pressure p)

One can then estimate the strain incurred to the membrane due to IOP. The model takes the material (or the *lamina cribrosa*) to be uniform and elastic (whereas in reality, the membrane has a non-uniform thickness, and exhibits viscoelastic properties). The elastic membrane model is better approximated when the diameter of the membrane is at least 10× the thickness of the membrane.

3 (a) (i) A dim star is more easily seen in peripheral vision than when the gaze is directed at it. Explain this phenomenon by referring to differences between the two types of photoreceptors in the retina. [2]

Answer: Mammalian retinas have many more rods than cones. However, cones are more numerous than rods in the fovea. Cones are less sensitive to light than rods: they need more photons than rods to get hyperpolarized. When the gaze is directed at the star, its image falls on the fovea and thus mostly on cones. By contrast, when looked at peripherally, the star's light hits more rods, which are more sensitive and thus detect the star more easily and at lower light intensities.

(ii) Explain how V1 cells can develop orientation selectivity despite receiving visual input from LGN cells that have circular (i.e. not orientation selective) receptive fields. [3]

Answer: Hubel and Wiesel's model of orientation selectivity provides one possible mechanism. This model stipulates that each V1 simple cell receives input from a collection of (excitatory) LGN neurons with neighbouring receptive fields (RFs) of a similar type (e.g. on-center/off-surround) but with locations tiling a straight line (arc segment) in the visual field. The stimulus that most strongly excites this simple cell is therefore a bright bar oriented so as to traverse all the on-regions of the LGN RFs (or a dark bar, depending on the polarity of the LGN RFs).

(iii) How can a point-like retina achieve two-dimensional vision? Provide an example for this from biology, including the name of the animal that achieves this, the components of its eye, and the principles of vision in it. [4]

Answer: The *Copilia* represents and example for this. The eye has two lenses, the 'objective' and the 'eyepiece', a retina (consisting of 5-7 receptors), and muscles to move the bottom apparatus (which is formed by the retina and the 'eyepiece' together). The small number of photoreceptors in the retina results in an extremely small field of view (3°) and an effectively 0-dimensional retina. However, the eye is oriented upwards, and the bottom apparatus moves along one dimension thus scanning the environment horizontally, while relevant objects (plankton) sink slowly downwards in the water thus adding a second effective dimension of scanning and hence effective vision.

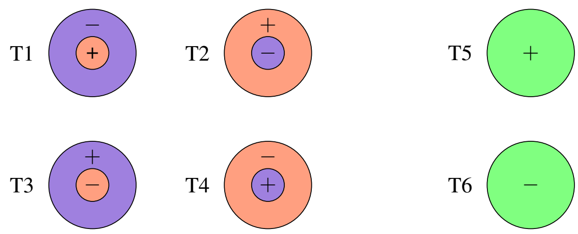
(b) In Fig. G.2, which cues allow the viewer to perceive depth? Provide a very brief explanation for each cue that applies. [3]

Answer: Most psychological cues discussed in lectures apply here: occlusion, texture gradients on the pavement, relative size e.g. of the people standing in front and back, linear perspective (buildings, lawns/paths), shadow and illumination (e.g. at the windows). Since this is a grayscale image, and since the scene has only a few tens of meters of depth, aerial perspective (differential scattering of light at different wavelengths) does not apply.

(c) This question is about efficient coding theory for color vision. The reflectance spectra of 20 different materials, found on recently discovered planet 2P8, are shown in Fig. G.3. What types of retinal ganglion cells would an inhabitant of this planet need in order to efficiently encode these reflectance spectra (in the sense of Principal Components Analysis, as shown in lectures)? Sketch their receptive fields and explain your reasoning. Note that the required ganglion cells need not be identical to those found in mammalian retinas. [5]

Answer: The first two principal components, i.e. the modes that explain most of the variance in the given reflectance spectra, can easily be seen to be: 1) a linear ramp over the whole visible spectrum, and 2) a bump around the greens. Thus, one would need the following types of RGCs:

- a red/violet opponent channel, i.e. a red-on-violet-off ganglion cell (responding strongly to red/orange wavelengths but suppressed by violet/blue), and because neural responses are positive (spikes), a similar type of GC with opposite polarity (violet-on-red-off). Together, these two types of cells would encode the amount of reflectance (positive or negative) contributed by PC1 for each surface. Because the on and off channels can correspond to either center or surround, this leaves us with four possible combinations of the RGC receptive fields shown in the figure below: { T1 and T3 }, { T1 and T4 }, { T2 and T3 }, { T2 and T4 }.
- a green-only channel, i.e. {T5 and T6} below (green-on and green-off); these would encode the PC2 contribution to each reflectance spectrum.



(d) This question is about efficient coding in the presence of sensitivity-dependent noise. A cell encodes a scalar stimulus variable, s, in its response, r. The encoding is linear with sensitivity $a \geq 0$, such that $r = a\, s + \epsilon$, where ϵ is noise. Let us assume that both s and ϵ are normally distributed (and, for simplicity, all quantities are dimensionless). Without loss of generality, the mean of both s and ϵ is zero. The variance of s is $\sigma_s^2 > 0$. Critically, the noise in the response grows supralinearly with the sensitivity of the cell, such that the variance of ϵ is $\left(a^4 + b\right)\, \sigma_r^2$ with $b > 0$ and $\sigma_r^2 > 0$.

(i) What quantity needs to be maximised for efficient coding? [1]

Answer: The mutual (Shannon) information between the stimulus and the response, $I_{r,s}$.

(ii) For $\sigma_s^2 = 1$, $\sigma_r^2 = 2$, and $b = 1$ what should be the sensitivity of the cell to achieve maximal coding efficiency? Ensure that you define all relevant intermediate quantities necessary for deriving the answer. [7]

Answer:

The definitions in the text imply the following:

$$\begin{aligned} s &\sim \mathcal{N}\left(0, \sigma_s^2\right) \\ r &= a\,s + \epsilon \\ \text{with } \epsilon &\sim \mathcal{N}\left(0, \left(a^4 + b\right)\, \sigma_r^2\right) \end{aligned}$$

and thus:

$$\begin{aligned} r|s &\sim \mathcal{N}\left(a\,s, \left(a^4 + b\right)\, \sigma_r^2\right) \\ r &\sim \mathcal{N}\left(0, a^2\,\sigma_s^2 + \left(a^4 + b\right)\, \sigma_r^2\right) \\ \mathrm{H}_{r|s} &= \log_2\left(\sqrt{2\,\pi\,e\ (a^4 + b)\ \sigma_r^2}\right) = \frac{1}{2}\log_2(2\,\pi\,e) + \frac{1}{2}\log_2\left(\left(a^4 + b\right)\, \sigma_r^2\right) \\ \mathrm{H}_r &= \log_2\left(\sqrt{2\,\pi\,e\ \left[a^2\,\sigma_s^2 + (a^4 + b)\ \sigma_r^2\right]}\right) = \frac{1}{2}\log_2(2\,\pi\,e) + \frac{1}{2}\log_2\left(a^2\,\sigma_s^2 + \left(a^4 + b\right)\, \sigma_r^2\right) \end{aligned}$$

The question requires finding the maximum of $\mathrm{I}_{r,s} = \mathrm{H}_r - \mathrm{H}_{r|s}$ w.r.t. a, which can be achieved by finding the roots of its derivative:

$$\begin{aligned} 0 = \frac{\mathrm{d}I_{r,s}}{\mathrm{d}a} &\propto \frac{2\,a\left(a^4 + b\right) - 4\,a^5}{(a^4 + b)^2 + a^2\ (a^4 + b)\ \frac{\sigma_s^2}{\sigma_r^2}} \\ 0 &= 2\,a^5 + 2\,a\,b - 4\,a^5 \\ &= -2\,a^5 + 2\,a\,b \end{aligned}$$

This equation has two roots:

$$\begin{aligned} a_1 &= 0 \\ a_2 &= \sqrt[4]{b} \end{aligned}$$

It is easy to check that a_1 and a_2 respectively correspond to a minimum and a maximum of $\mathrm{I}_{r,s}$. Therefore the value of the sensitivity that achieves maximal efficiency is:

$$a = a_2 = \sqrt[4]{b} = \sqrt[4]{1} = 1$$

Fig. G.2

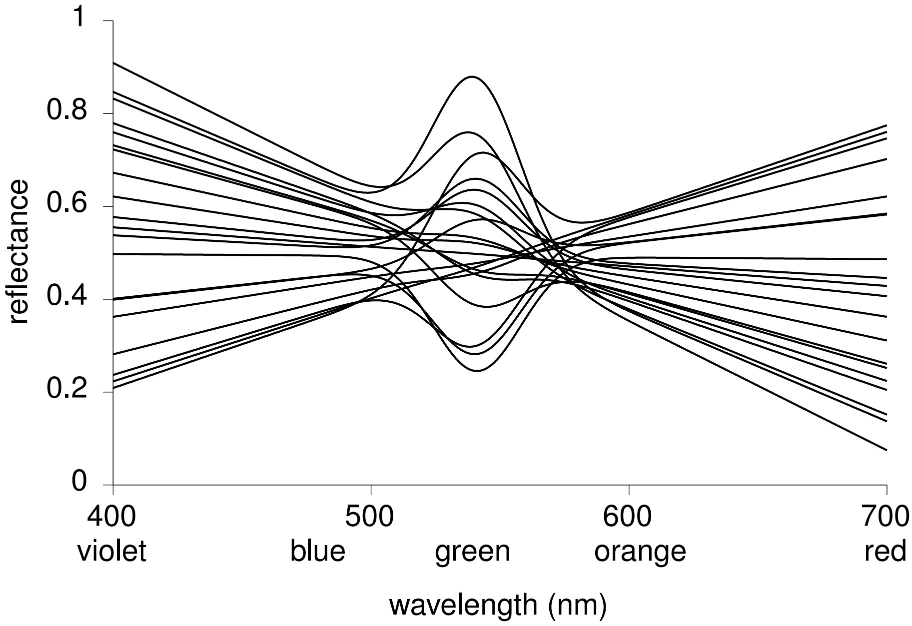

1
0.8
0.6
0.4
0.2
0
reflectance
400
violet
500
blue
green
600
orange
700
red
wavelength (nm)

Fig. G.3

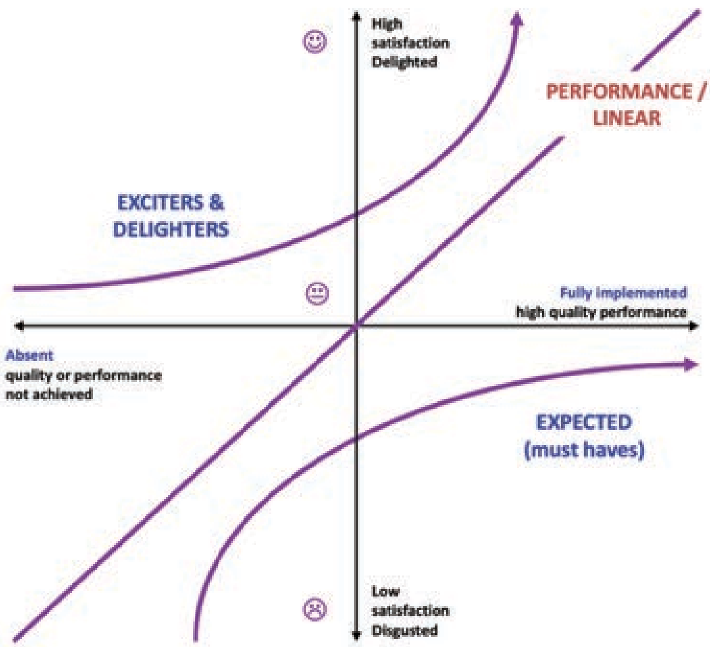
1 a, b, c

a) Sketch and describe the *Kano* model for televisions. **[5]**

Students might indicate the nature of each of the zones, as well as suggesting some features that would apply in the case of a new TV. At its minimum the students should be able to identify few attributes for each of the three categories for current televisions:

- **Expected:** these are must have features and not having them or if they are badly implemented is a secured failure. Today these features might be things such as having a remote control, auto-tuning, wall-fixing points, flat screen…..
- **Linear:** these are features where satisfaction increases in proportion to the extent (or effectiveness) to which they are implemented. In other words, the 'more the better'. For a new TV today, this could be number of channels, has diminishing power consumption, access to different types of broadcasting services, definition and picture quality …
- **Delighters:** features that will 'wow' customers, but not put them off if they are not present. For a new TV, these will be novel things that are not available on other TVs, such as novel aesthetics, integrated camera for face-time, a novel user interface, deep-sleep mode, Alexa built in, curved screen, looks like an art picture when turned off, 3D options, true black….. etc.

Best students will recognise the shift of the attributes towards the bottom of the diagram in time. today's exciters will become tomorrow's must-haves. For example, we were excited when we had 4 TV channels in 1984, but now this is not even sufficient as a must-have. Or the "Flat screen" used to be a delighter when LCD screens were first introduced for TVs. Today this is a must-have and there are new types of delighters.



b) Explain the link between the *Kano* and the *diffusion of innovation* models. [12]

The **link between the two models** is that both help reasoning about product attributes.

The **diffusion of innovation model explains** how people in a market adopt a technology, and what attracts each types of adopters group to an innovation. Innovators and early adopters at the early stages of the product appearing into a market would adopt it because it has exciting features, they see its potential. Instead, the majority of the market are attracted to an innovation only if it works extremely well and doesn't cause them any pain in adopting it. There are people who will not ever adopt the innovation or will do it after many others have.

The Kano model works well to explain how consumers respond to products to which they are already familiar and how new innovations for products in a known category might be received. For more radical innovations, it can be argued that all features are 'novel' and therefore are delighters. So, the Kano model is less helpful in the early stages of the diffusion of innovation but much more helpful in the later stages, when there are more products competing and there is greater need for differentiation between them.

Students will ideally introduce **an example** such as when 'bread-makers' were first introduced. As a new category of product, there is no prior comparison and all features are essentially novel. The first customers are 'early adopters' who are obsessed by new things and want to be at the leading edge. They may even accept that products have failings, just to be at the forefront. Feedback from these early adopters can be crucial to help ensure greater take up of a radically new product. The lead users who might purchase this were keen on new kitchen gadgets or might have specific needs (e.g. being able to produce gluten free bread at home) that are not met by standard alternatives. So the delighters are initially very relevant for a small portion of the market (the innovators and lead users) who value the advantage either because of their personality or because of their needs (they understand the advantage given by the exciters and are ready to compromise on other more basic characteristics). If we use televisions as the example, when LCDs were introduced, the possibility to buy a flat screen TV gave people living in small apartments in cities the opportunity to install a television using less space compared to a Cathode Ray TV set. It is only when later generations of the LCD TVs are released that the Kano model becomes especially helpful in considering improvements or new features that might be future delighters.

c) Discuss, providing examples, how the diffusion of innovation model can be used by those who are trying to take their innovation to market. [8]

The discussion via a specific product(s) **example**, e.g. artificial intelligence search engine, televisions should help explaining the points.

The diffusion of innovation model can help to profile the different types of users of an innovation and develop mechanisms and value propositions to persuade them. Market research and analysis of different types of user and stakeholder can help a firm understand the needs and wants of people at different stages of the curve.

- in the early stages when a new innovation enters the market, users will be 'visionaries', who are actively seeking the latest things. The most adventurous early adopters are thrilled by the new performance offered by an innovation and are ready to overlook its limits and are ready to spend to obtain the innovation. However, these two groups represent a small segment of the market. At this point, product price might be too high for the majority. Example. If a new technology such as AR glasses are created, companies need to convince and excite the innovators to appreciate and test the innovation so that they can attract a group of early adopters who might pay a premium to buy the technology.
- To reach the full market diffusion, i .e. cross the Moore's chasm in the curve, companies need to make sure that the majority of the users adopt the innovation (some early, some later), by overcoming their resistance to change their behaviour. These users are convinced by the advantages of the innovation, but only if the problems, which adopting a new innovation causes, are resolved and it is not too hard to change what they did before. For example, for AR glasses to be adopted by many people, it will be useful that there is enough content to be overlaid in AR to be useful to a large group of people and that the AR doesn't cause people to feel sick whilst wearing AR glasses, as well as having to spend moderately and compatibly with their means to adopt the innovation. Those taking a technology to market might need specifically to devise mechanisms to bridge the 'Moore's chasm'. One of these is to identify lead users amongst the visionaries, who have a traction with majorities (i.e. influencers).
- There might be a group of users which will adopt innovations very late, when everyone else has already adopted it. Some might even not be able to adopt it at all. As a product heads towards the later stage of the lifecycle, late-adopters or sceptics might be enticed, but at this point, to keep sales high, new versions, new features and new technologies are needed, to maintain growth by starting the cycle again. Hence companies need a portfolio of innovations to maintain and grow the market.

2: a, b, c

A group of university students has developed an innovative augmented reality (AR) technology, combining new sensors and cameras, and new software, and is considering setting up a start-up.

a) **What different types of Intellectual Property (IP) could apply to the AR technology and how could they be protected?** [8]

It is expected that the answers would emphasise the different **ways to protect the main technological elements given**. The inventors will have to carry out some **prior-art search** to determine what part of the IP they think they have developed is really new.
It is important that the students consider the various applications of the technology to best design the protection strategy (e.g. considering different uses in different industries and different geographies)**.** Examples include:

- **Copyright** for the software code, including the code for the processing software, but possibly also the software that runs in the sensors and cameras or is needed to connect these to the processing software.
- **Copyright** for any other text that the students might create, such as any brochure, slide deck, software manuals, but also website content for the start-up webpage
- **Trademarks**, such as for the start-up name, but also the product's brandname. Trademarks could also be secured for components of the system, if applicable, such as for any particular sound that is played when the software is initiated.
- **Design rights** could be applicable if the sensor or camera (or in fact the software or the product that embed them (e.g. glasses) has some distinct appearance.
- **Patents:** could be filed Depending on what is really new with regards to the hardware (camera, sensor) and possibly even the software as long as there might be a "further technical effect". Patents are expensive, particularly due to the need of applying in different geographies to protect the innovation in different markets, but they are particularly good in communicating the value of a technology to potential partners and investors.
- **Trade secrets** could be mentioned. In software you could have those related to the software as well, if the software code is going to be kept secret and only the machine-readable code be made available at all.

b) Based on your understanding of the pros and cons of possible business models, and by providing suitable examples of possible business models, advise the start-up on the business model choice. [10]

The options for BM are shown **below, with possible examples:**

- Sell a product: **Example: New AR kit which embodies cameras/sensors/software (a specific product like Google Glasses)**
- Sell a service: **Example: sell consultancy – helping companies identify applications for AR technology in their products/services (e.g. using this AR for surgery robots, or embed AR in new phones)**
- Sell a product plus services: **AR kit + AR experiences (e.g. AR events for specific customers such as fashion designers) or AR glasses + help in tailoring the AR technology for customers' applications (e.g. medical)**
- Sell a product plus consumables: **Example: AR glasses + AR content (e.g. an equivalent of PokemonGO game)**

All these are possible and interesting, the students in the question should be ready to **modify/experiment with different business models** depending on the opportunities over time as the technology is matured (e.g. in the early stages as the product is not fully developed the service-based business models might be interesting to capitalise early on the IP).

A lot of resources are needed to develop appropriate manufacturing for **complex products like AR kits, made of various component**s and it is particularly important since we are talking about a **very new firm**, founded by **young students** (hence likely to be inexperienced and under resourced). As the students are at university, and at the beginning of their start-up business, it is advisable maybe to identify

- **a potential acquirer of the idea**, or
- **a licensor of the technology or**
- **a partner with whom to develop specific applications for different industries and the appropriate business model**.

The challenge is **finding buyers/ licensors/partners**. Asking help from the university technology transfer office or hiring specialised people to identify partners could be a way to do it.
Finding a buyer for the idea, could be a good way to minimise risks and capitalise immediately on the innovation. A company like Google or Apple might be interested in buying the technology outright. However, the advice should emphasise that this choice could limit the prospects of earning from the idea.

Licensing out a technology may be a good idea in terms of making money soon, but, beyond the challenge of the identification of the key licensor, the success of the licensing agreement depends on the capability of the licensor to use the technology which most likely needs adapting or needs know-how transfer. As it is possible to design a different strategy for commercialising various IP rights **and use different models with different potential partners,** they could plan a variety of approaches with different partners. This would give a breadth of opportunity to obtain value from the technology.

Identifying **partners** could help the start-up build on experience and resources, and be faster in reaching the market, and to share responsibility and risks for taking the technology to market. The main challenge is that it can be hard for a small company such as a start-up to work with (and 'get-on with') another established business, especially if this latter has very little experience of working with a start-up.

C) Based on your replies to (a) and (b), advise the students on how to pitch their idea to Venture Capital investors. [7]

Let's imagine the students chose to aim to sell a product plus services: AR glasses + help in tailoring the AR technology for customers' applications in the medical industry and that they will be looking for a partner to develop the technology towards that application.

The students should be advised to present this idea to VCs and **design an elevator pitch illustrating the core benefits of the technology. For example**:

"The AR technology we developed, will be useful **for** surgeons **who** are performing complex operations. The AR Technology will be embedded in **(category**) surgical visors and glasses and will **(benefit)** provide surgical guides in real time. **(unlike)** There are no equivalent technologies on the market which would allow **(differentiator)** the reduction of operations time, the anticipation and fast solution of complex surgery issues. The technology will minimise accidents in operating theatres."

Further, it would be necessary to prepare a **business plan** which illustrates how they intend to reach their goals. The following aspects should be covered:
• **Technology**: e.g. the technology is highly innovative. A broad set of protection mechanisms including patents have been already considered and applied for (see question a). The technology requires future development which we plan to carry out thanks to collaboration with important companies in the medical equipment sector.
• **Market.** E.g. There is a high need of such a technology on the market as (insert number) of operations are carried out daily on a national basis with an average of (percentage) errors.
- Management: e.g. We have a track record and demonstrated our management competence and capability on several accounts. For example: technical competence, problem-solving and leadership skills, building relationships with suppliers and customers, reputation in the market. We have demonstrated some of these skills (e.g. leadership demonstrated in university societies, during an hackathon etc).
• **Access to markets:** e.g. we are seeking specific partners in the medical instrument sector to also develop a more detailed marketing strategy
• **Regulation**: We need to research the regulatory aspects of this technology.
• **Exit:** We are planning to identify a partner who is potentially able/interested in acquiring our business, after initial development of our idea. The Technology Transfer of the university is already helping us.
Funding: We are seeking XXXX £ which we think would be necessary for

- perfecting and manufacturing small batch of beta products
- developing an initial database of content for AR and
- employing a business developing manager to identify a key partner in this field who will help us to understand the market (and regulation) and will support the development of the product and manufacturing solution.

POST EXAM COMMENTS: **In the three parts of question 2, the answers which concentrated on providing examples which demonstrated the application of the generic concepts to the specifics of this technology/situation were the minority.**

3 a, b, c

a) Using suitable examples, explain what is meant by *radical* and *incremental* innovation. Discuss how companies pursue different types of innovation. [5]

The terms refer to the degree of novelty or changes in any of the 4 Ps (products and services, processes, paradigm and positioning).
- Radical innovation = Significant changes: 'do what we do differently'. Using cryptocurrency (bitcoin) rather than traditional currency (credit, or banknotes)
- Incremental innovation = Small improvements: 'doing what we do but better'. Example = improve quality of tablets (e.g. lighter, better screens, some further features)

Any of the elements below could answer the 'how' part of the question:

- larger companies try to **balance innovation projects across costs and risk** to satisfy their current customers via increasing product features incrementally and to invest a little in more riskier projects (portfolio approach or bubble chart).
- The companies need to align the overall strategy of the business with the product level strategy and New Product Introduction (NPI) processes. This latter is **managed via a model** which comprises 5 aspects
    - 1.Innovation strategy: Developing and achieving the goals of the innovation strategy is the responsibility of top management and this requires a focus on a number of issues including (1) assessing market trends and how these drive the need for innovation in the company (2) the role of technology, the opportunities it can provide, and what expertise needs to be acquired, and (3) communicating the role of innovation within the company. There needs to be recognition that there are limited budgets – you can't do everything!
    - 2.Ideas: This is the raw material for innovation. Idea generation needs to be supported at the individual and team level, and from within and outside the company. These ideas should draw upon both technical possibilities and market opportunities. This is the time for experimenting with risky ideas – the cost of failure at this early stage is very low.
    - 3.Selection and prioritization: There must be an efficient process for ensuring that the best ideas are selected and pushed forward for development. Tools are used to consider the relevant merits of different projects, and for individual ideas to be considered as part of the portfolio of projects within the company. Often the hardest part is deciding what not to do.
    - 4.Implementation: The focus here is on developing new products and services as quickly and efficiently as possible. Here the focus must be on discipline and quality, not on creativity and change. As projects move closer to market, the cost of changes becomes higher.
    - 5.People and organization: Successful innovation is the result of having the right company culture, recruiting, training and motivating the best people, and providing organizational structures that enable rather than hinder innovation and collaboration.
- **There is difference between focus on innovation type depending on company type:** Large firms typically have invested substantial resources to create value in a particular way (e.g. a factory built to make one type of product, expertise….). Their focus has been on improving the efficiency of their manufacturing, distribution and marketing based on the product(s) they offer. Small improvements (incremental innovations) would not involve much changes in their operations in factories / supply chains, with the advantage of satisfying their customers better (e.g. providing new features, better costs, reduced costs, better quality). For suppliers, investors, customers, employees,

etc it is not a change which implies big transformations. However, doing something that is radically different from what is currently done, would impact on all the firm stakeholders. The company's current suppliers may be out of a job as the firm might need whole new ones (change in supply chains); the firm' might need to convince new customers to buy the radical innovation, so the firm may compete with itself for market and potentially, eventually lose current customers or have to cater for both old and new at the same time. Small firms, particularly if they are new, do not have this type of legacy and they are much less constrained in the type of innovation they pursue by their operations and market. A radical innovation would clearly set them aside from established companies and they could become more interesting in the light of investors. They would not have a reputation/recognition for a particular type of products and hence be less concerned of tarnishing their reputation by proposing products which are initially understood only by small niche markets.

POST EXAM COMMENT: Several responses were just straight from the notes, without evidence of personal understanding. Further, the majority of responses did not address the second part of the question.

b) **Discuss the principles and challenges for balancing supply and demand for a company manufacturing a domestic robot to be used to assist with various tasks in a kitchen.** [12]

Companies would like the demand for their products to grow in a predictable enough way so that the operations can be designed and scaled more easily. Balancing supply and demand requires thinking about the different perspectives of the business owner, the factory manager, the suppliers, and the customers.
Owner – wants to see all resources utilised as much as possible, no stock sitting around
Suppliers – want to deliver in bulk, ideally all at once
Customers – want immediate delivery, right price, right quality
It is really hard to balance these conflicting demands.

Demand: We can assume that:
- kitchen robots might be sold internationally – so there is a very large market- and there is modest request for changes across locations
- as this is a 'luxury' non-essential product, this is likely to be a product bought as a gift. Thus holiday seasons might provide surges in demand which could be a problem, but these periods are known.

- Therefore for managing demand, the firm could consider ;
  - amplifying seasonal aspects of the product when there is lesser demand. Robot attachments for seasonal needs e.g. making things with fresh fruits / ice-creams in summer etc.
  - Discounts could be offered for times outside the holiday season, and premium pricing used around Christmas / New Year.

- They could also consider offering different products ('basic' model for general use that might be bought more as a 'workhorse' product, and a 'premium' model that makes for a generous gift.).
- Could use product placement on TV/internet shows to boost demand.

Managing supply:
For the case of the domestic kitchen robot, we can assume the following:

- This product is likely to be a very good candidate for large employment of robots and automation. The production will be designed to produce a very large number of uniform items, but this requires much investments upfront and the producer needs to be sure that the demand will be constant for a long enough time to get a return of their investment in infrastructure.
- It is an assembled product, made up of electrical, electronic (hardware and software) and mechanical subsystems
- It is unlikely that the firm will be making all (or even any?) of these subsystems and their main task is assembly, integration, testing and shipping of the multiple subsystem

- What happens if product placement suddenly proves successful and all reserve stocks are needed immediately (peak of demand)? Also given the importance of the holiday season market, the firm cannot afford not to be able to deliver at this time, so they could be using the quieter season during the summer to stock up – but where would this stock be kept in the supply chain → it is necessary to have an agreement of who takes the risk in the supply chain.
- A task can only begin when the previous task has finished so that all inputs are available, and the relevant operator and tooling are available - so the people and equipment must be scheduled; Even only one missing component means that no kitchen robot can be completed. The cost of such a halt is quite expensive. operations hence need coordinating the various part of the supply chain.
- just sticking to just one model would allow maximum simplicity for working with suppliers. But if demand management requires multiple robot models to be manufactured (e.g. countercyclical product), this will become more complicated if different subsystems are needed. It usually takes time to switch from producing one type of product to another.

c) Demonstrate how you would define and segment the market for personal computers.

The **definition** and segmentation of the market are closely linked. We need to establish (define) the market we are trying to address, i.e. the set of potential buyers who might be interested in our product and hence also the possible competitors for the product. We can use different criteria:

- Narrow or broad
- goods which provide similar benefits (e.g. entertainment electronics).
- goods which compete for our money .

Example For PCs,

- (Broad) Market for Electronics goods, which include amongst competitors also the producers of tablets and smart phones as well as calculators.
- (Narrow) Market for PCs, could include all the types (laptop and desktop) of products or could be limited to one of these categories

For The segmentation Typically perceptual maps are used. These can be drawn according to different criteria:

- The benefits that the users/customers derive from their use. (E.g. ease of use, multi-purpose, status …)
- Characteristics of the user/consumer (demographic/psychographics) (e.g. Younger/older; professional/programmer – basic skills)
- How product are used / purchased (e.g. Office work, fun – free time)
- The attributes of the product (e.g. computational speed, aesthetics..)

A visual description of the market segmentation via perceptual maps could be added

POST EXAM COMMENT: Several answers did not address the ‘define’ part of the question.