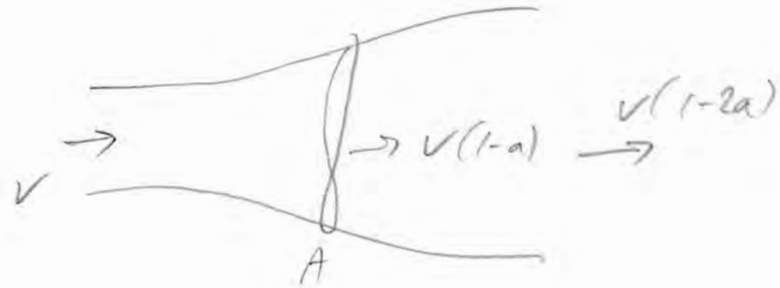
EGT1
Engineering Tripos Part IB

7 June 2022

Paper 8 Selected Topics **CRIB** (updated Dec 2022)

(a)



$\dot{m} = A\rho V(1-a)$

$$\text{Change in } KE = \frac{1}{2}\dot{m}\left(V^2 - (V(1-2a))^2\right) \qquad 1-4a+4a^2$$

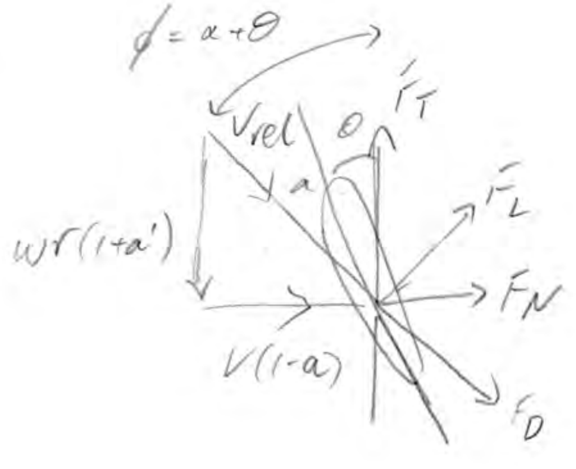
$$= \frac{1}{2}A\rho V^3(1-a)(4a)(1-a)$$

$$= 2A\rho V^3 a(1-a)^2$$

(c) - Aerodynamic profile is key for maximum efficiency. As the radius increases the peripheral speed increase so less twist needed to maintain the optimum angle of attack

Blades tend to get thicker at the root, as the bending loads increase nearer the root. The exact details depend on the geometric and material parameters.

(b)



$\sigma = \dfrac{1.1.3}{2\pi 12} = 0.0438$, $\theta = 10^\circ = 0.175$

$F_T = F_L \sin\phi - F_D \cos\phi$

$F_N = F_L \cos\phi + F_D \sin\phi$

$\tan\phi = \dfrac{V(1-a)}{\omega r(1+a')} = \dfrac{1}{3}\dfrac{(1-a)}{(1+a')}$

$\Delta T = B.r.F_T L$

$\Delta Power = \omega \Delta T = \omega B r F_T L$

$= 4.3.12.284 \times 0.1$

$= 4090\ W$

$a = \left(\dfrac{4\sin^2\phi}{\sigma C_N} + 1\right)^{-1}$

$a' = \left(\dfrac{4\sin\phi\cos\phi}{\sigma C_T} - 1\right)^{-1}$

Iterate starting with $a = a' = 0$

$\Rightarrow \phi = \tan^{-1}\left(\tfrac{1}{3}\right) = 0.322$

$\alpha = \phi - \theta = 0.147$

$C_L = 0.927 \quad C_D = 0.05$

$C_N = C_L\cos\phi + C_D\sin\phi = 0.895$, $C_T = C_L\sin\phi - C_D\cos\phi = 0.246$

$a = 0.089$, $a' = 0.0091$

Re-iterate: $\phi = 0.292$ $\alpha = 0.118$, $C_L = 0.740$, $C_D = 0.05$

$C_N = 0.723$, $C_T = 0.165$, $a = 0.087$, $a' = 0.0066$

Once more: $\phi = 0.294$, $\alpha = 0.119$, $C_L = 0.749$ $C_D = 0.05$

$C_N = 0.731$, $C_T = 0.169$, $a = 0.087$, $a' = 0.0067$ ✓ close enough

$\Rightarrow F_T = \tfrac{1}{2}\rho V_{Rel}^2 C_T C$

$= \tfrac{1}{2}.1.2.25.48.0.169.1.1 = 284\ N/m$ $\quad kg\,m^{-3}\,m^2 s^{-2}\,m$ $\quad N s^2 m^{-1} s^{-2} = N m^{-1}$ ✓

$V_{rel}^2 = (\omega r(1+a'))^2 + (v(1-a))^2 = (4.12.(1.0067))^2 + (16(1-0.087))^2 = 25.48$

Power (calc above in ☁) = 4.1 kW

(a)(i) - High power application with large speed increase needed.

- stringent weight (as tower head) and space (as taking up cross-section) requirements not as critical for stationary applications

- high efficiency as power-generating device (though this is typical

- reliability in challenging remote / offshore environment

(ii) Lots of potential for cyclic loading with varying speed of wind → varying stresses. Enhanced by potential for resonant response in tower or blade leading to over-loading.

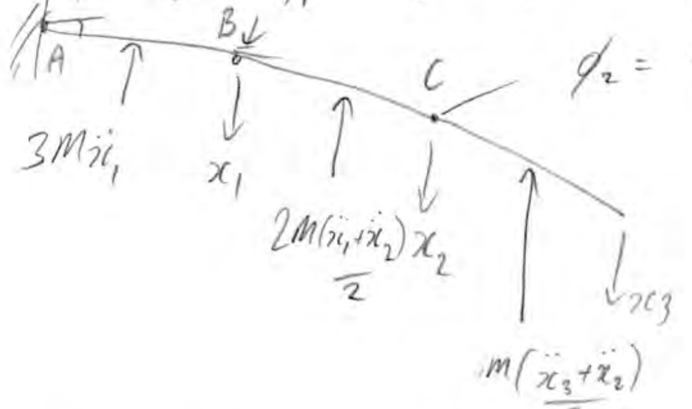
(iii) Blades have demanding weight and strength / stiffness requirements ⇒ composites. CFRP would be ideal but more costly. May get away with GFRP, particularly where its lower E is not so critical (eg in smaller blades)

(iv) Various places can generate noise, especially the gearbox and generator and air flow off the blades. Varying impact of noise, depending on site and perception. Particularly challenging with persistent single notes from mechanical components.

(b)

$$\phi_0 = \frac{x_1}{L} \qquad \phi_1 = \frac{x_2 - x_1}{L} - \frac{x_1}{L} = \frac{x_2 - 2x_1}{L}$$

$$\phi_2 = \frac{x_3 - x_2}{L} - \frac{x_2 - x_1}{L} = \frac{x_3 - 2x_2 + x_1}{L}$$



M_C ↺ $\quad k\phi_2 + M\frac{(\ddot{x}_3 + \ddot{x}_2)}{2}\frac{L}{2} = 0$

M_B ↺ $\quad 2k\phi_1 + 2M\frac{(\ddot{x}_1 + \ddot{x}_2)}{2}\frac{L}{2} + M\frac{(\ddot{x}_3 + \ddot{x}_2)}{2}\frac{3L}{2} = 0$

M_C ↺ $\quad 3k\phi_0 + 3M\ddot{x}_1 \cdot \frac{L}{2} + 2M\frac{(\ddot{x}_1 + \ddot{x}_2)}{2}\frac{3L}{2} + M\frac{(\ddot{x}_3 + \ddot{x}_2)}{2}\frac{5L}{2} = 0$

$$\frac{k}{L}\underbrace{\begin{pmatrix} 1 & -2 & 1 \\ -4 & 2 & 0 \\ 3 & 0 & 0 \end{pmatrix}}_{\overset{k}{\sim}}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \frac{ML}{4}\underbrace{\begin{pmatrix} 0 & 1 & 1 \\ 2 & 5 & 3 \\ 9 & 11 & 5 \end{pmatrix}}_{\overset{m}{\sim}}\begin{pmatrix} \ddot{x}_1 \\ \ddot{x}_2 \\ \ddot{x}_3 \end{pmatrix} = 0$$

Solutions of form $\underline{m}^{-1}\underline{k}\,\underline{x} = \omega^2 \underline{x}$

Using calculator $\underline{m}^{-1}\underline{k} = \frac{4k}{ML^2}\underbrace{\begin{pmatrix} 6.33 & -4.67 & 1.33 \\ -9.83 & 8.67 & -2.83 \\ 10.83 & -10.66 & 3.83 \end{pmatrix}}_{A}$

Find eigenvalue & eigenvector by iteration. Invert A to find lowest eigenvalue

$$B = A^{-1} = \begin{pmatrix} 3 & 3.67 & 1.67 \\ 7 & 9.83 & 6.83 \\ 11 & 17 & 9 \end{pmatrix}$$

$B \times \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 8.34 \\ 21.66 \\ 37 \end{pmatrix}$ (using calculator)

$B^2 \times \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 166 \\ 450 \\ 792 \end{pmatrix} \rightarrow \begin{pmatrix} 1 \\ 2.71 \\ 4.77 \end{pmatrix}$

$B^3 \times \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3474 \\ 9417 \\ 16620 \end{pmatrix} \rightarrow \begin{pmatrix} 1 \\ 2.71 \\ 4.78 \end{pmatrix}$

$\rightarrow$ converged $\lambda = 3474/166 = 1/0.048$

Smallest λ for $A = 0.048$

$\Rightarrow \omega = \sqrt{0.048 \times \frac{4k}{ML^2}} = \frac{0.44}{L}\sqrt{\frac{k}{M}}$

Mode shape

3(b) (i) The DFIG enables variable-speed operation over a range of wind speeds. This enables the turbine rotor to operate at optimal tip-speed ratio and hence optimum power coefficient. Although there are other technologies that achieve variable speed operation DFIGs enable the use of fractionally-rated converters. They are also able to generate reactive power. They operate at high power density (through the use of a gearbox) and they are robust & reliable

(ii) $P = \frac{1}{2}\rho C_p A v^3$

$\rho = 1.23\ kg/m^3$
$C_p = 0.4$
$v = 12\ m/s$
$P = 5\ MW$

$\therefore A = 11762\ m^2$
$= \frac{\pi d^2}{4}$ $\therefore d = 122\ m$ (rotor diameter)
$R = 61\ m$ (radius)

(iii) The DFIG rotor speed $\omega_s = \frac{2\pi f}{p}$

$f = 50\ Hz$, $p = 8$ $\quad \omega_s = 39.3\ rad/s$

The bladed rotor speed ω_t satisfies $\lambda = \frac{\omega_t R}{v}$

$\therefore 9 = \frac{\omega_t\, 61}{7}$ $\quad \therefore \omega_t = 1.03\ rad/s$

$\therefore$ need gearbox of ratio $N = \frac{\omega_s}{\omega_t} = \frac{39.3}{1.03} = 38.2$

The input power to the DFIG is $\left(\frac{7}{12}\right)^3 \times 5MW$

since power $\propto v^3$

$\therefore P = 0.992\ MW$

3(b)(iv) The torque at normal speed
$\omega_s = 39.3$ rad/s is found from

$T\omega_s = 0.992\,MW \quad \therefore T = 25.2\ kNm$

Assuming $\left|R_2'/s\right| \gg R_1$ and $\gg X_1 + X_2'$

then $$T = \frac{3V^2 s}{\omega_s R_2'}$$

Star connection $\therefore$ $V_{ph} = \dfrac{3.3\times10^3}{\sqrt{3}}$

Slip is negative $\therefore$ $$-25.2\times10^3 = \frac{3\left(\frac{3.3\times10^3}{\sqrt{3}}\right)^2 s}{39.3\times0.18}$$

$$\therefore s = -0.0163$$

check assumptions :

$\dfrac{R_2'}{s} = 11\Omega \quad \gg R_1$
$\gg X_1 + X_2'$

so simplified torque-slip equation is justified

Current $$I = \frac{V_{ph}}{R_2'/s} = \frac{\frac{3.3\times10^3}{\sqrt{3}}}{11}$$
$$= 173\ A$$

2022 2P8 Section D

Section D, Q1

(a) Increasing bypass ratio allows the same net thrust to be obtained from a larger air mass flowrate at a smaller jet velocity, this increases the propulsive efficiency $\eta_p = 2V/(V_j + V)$. As the bypass ratio increases, the installation drag of the engine increases (larger nacelle surface area) as does the weight of the engine and nacelle. Fan tip Mach numbers are approximately sonic, so increasing the fan tip diameter means reducing the low pressure shaft speed (unless a gearbox is used) and this means the low pressure turbine blade speeds are relatively low.

(b)

$$T_{02} = T_a(1 + 0.5(\gamma - 1)M^2) = 217(1 + 0.5 \times 0.4 \times 0.82^2) = 246.2 \text{ K}$$

$$p_{02} = p_a(T_{02}/T_a)^{\gamma/(\gamma-1)} = 22700(246.2/217)^{\gamma/(\gamma-1)} = 35300 \text{ Pa}$$

(c)

$$p_{023} = 1.5p_{02} = 1.5 \times 35300 = 52950 \text{ Pa}$$

$$T_{023s} = T_{02} \times 1.5^{(\gamma-1)/\gamma} = 246.2 \times 1.5^{(\gamma-1)/\gamma} = 276.4 \text{ K}$$

$$T_{023} = T_{02} + (T_{023} - T_{02})/0.9 = 246.2 + (276.4 - 246.2)/0.9 = 279.8 \text{ K}$$

(d)

$$T_{045} = T_{04} - (T_{03} - T_{023}) = 1500 - (800 - 279.8) = 979.8 \text{ K}$$

$$T_{045s} = T_{04} - (T_{04} - T_{045})/0.9 = 1500 - (1500 - 979.8)/0.9 = 922.0 \text{ K}$$

$$p_{04} = 30p_{023} = 30 \times 52950 = 1589000 \text{ Pa}$$

$$p_{045} = p_{04}/(T_{04}/T_{045s})^{\gamma/(\gamma-1)} = 1589000/(1500/922.0)^{\gamma/(\gamma-1)} = 289200 \text{ Pa}$$

(e)

$$c_p = \gamma/(\gamma - 1)R = 1004.5 \text{ J/kg/K}$$

$$T_{05} = T_{045} - \Delta T_0^{LPT} = 979.8 - 390.1 = 589.7 \text{ K}$$

$$T_{05s} = T_{045} - (T_{045} - T_{05})/0.9 = 979.8 - (979.8 - 589.7)/0.9 = 546.3 \text{ K}$$

$$p_{05} = p_{045}/(T_{045}/T_{05s})^{\gamma/(\gamma-1)} = 289200/(979.8/546.3)^{\gamma/(\gamma-1)} = 37440 \text{ Pa}$$

Isentropic expansion in core nozzle,

$$T_9 = T_{05}(p_a/p_{05})^{(\gamma-1)/\gamma} = 589.7 \times (22700/37440)^{(\gamma-1)/\gamma} = 511.1 \text{ K}$$

$$V_{jc} = \sqrt{2c_p(T_{05} - T_9)} = \sqrt{2 \times 1004.5 \times (589.7 - 511.1)} = 397.3 \text{ ms}^{-1}$$

(f)

$$V = M\sqrt{\gamma R T_a} = 0.82 \times \sqrt{\gamma R \times 217} = 242.1 \text{ ms}^{-1}$$

Bypass jet velocity = core jet velocity, $V_{jb} = V_{jc}$,

$$\Delta T_0^{fb} = (V_{jb}^2 - V_j^2)/(2c_p \times 0.9) = (397.3^2 - 242.1^2)/(2 \times 1004.5 \times 0.9) = 54.88 \text{ K}$$

Low pressure shaft energy balance:

$$\dot{m}_b \Delta T_0^{fb} + \dot{m}_c(T_{023} - T_{02}) = \dot{m}_c \Delta T_0^{LPT}$$

$$\text{bpr} = \dot{m}_b/\dot{m}_c = (\Delta T_0^{LPT} - (T_{023} - T_{02}))/\Delta T_0^{fb} = (390.1 - (279.8 - 246.2))/54.88 = 6.500$$

Section D, Q2

(a)

$$D = 500 \times 1000g/(L/D) = 500 \times 1000 \times 9.81/20 = 245300 \text{ N}$$

$$F_N = D/4 = 61310 \text{ N}$$

(b) Applying the SFME between nozzle exit (station 19) and downstream where $p = p_a$ and the jet velocity is V_j:

$$\dot{m}V_{19} + p_{19}A_N = \dot{m}V_j + p_aA_N$$

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

If we assume the propelling nozzle is choked (very likely due to pressure rise across fan, and in the engine intake due to flight speed) then the LHS is set by the engine operating point (determined by p_{02}, T_{02} and $\dot{m}_f$). Hence the expression for the non-dimensional group $\tilde{F}_G$ has a physical basis.

(c) Inlet conditions at cruise:

$$T_{02c} = 217(1 + 0.5(\gamma - 1)0.85^2) = 248.4 \text{ K}$$

$$p_{02c} = 22700 \times (248.4/217)^{\gamma/(\gamma-1)} = 36410 \text{ Pa}$$

and $V_c = 0.85\sqrt{\gamma R\, 217} = 251.0 \text{ ms}^{-1}$.

Inlet conditions at take-off:

$$M_{to} = 90/\sqrt{\gamma R\, 288.15} = 0.2650$$

$$T_{02to} = 288.15(1 + 0.5(\gamma - 1)0.2650^2) = 292.2 \text{ K}$$

$$p_{02to} = 101300 \times (292.2/288.15)^{\gamma/(\gamma-1)} = 106300 \text{ Pa}$$

(i)

$$\begin{aligned}\tilde{F}_G &= (F_{Gc} + p_{ac}A_N)/A_N p_{02c} = (F_{Nc} + \dot{m}_cV_c + p_{ac}A_N)/A_N p_{02c} \\ &= (61310 + 140 \times 251.0 + 22700 \times 1)/(1 \times 36410) = 3.273\end{aligned}$$

$$F_{Gto} = \tilde{F}_G A_N p_{02to} - p_{ato}A_N = 3.273 \times 1 \times 106300 - 101300 \times 1 = 246800 \text{ N}$$

(ii)

$$\tilde{\dot{m}} = \dot{m}_c\sqrt{c_pT_{02c}}/(A_N p_{02c}) = 140\sqrt{1004.5 \times 248.4}/(1 \times 36410) = 1.921$$

$$\dot{m}_{to} = \tilde{\dot{m}}A_N p_{02to}/\sqrt{c_pT_{02to}} = 1.921 \times 1 \times 106300/\sqrt{1004.5 \times 292.2} = 377.0 \text{ kgs}^{-1}$$

$$F_{Nto} = F_{Gto} - \dot{m}_{to}V_{to} = 246800 - 377.0 \times 90 = 212800 \text{ N}$$

(d)

$$F_{Nto} = mg(1/(L/D) + \sin\theta)/3 = 500 \times 1000 \times 9.81 \times (1/10 + 0.03)/3 = 212550 \text{ N}$$

This is less than the net thrust obtained in part (c)(ii) for when the engine operates at the same non-dimensional point at take-off as at cruise. This indicates that the engine can supply this thrust and that the engine can take off with only 3 engines.

(e) Aircraft typically spend most time at cruise so we want to minimise sfc and hence maximise overall efficiency here; this is the thermodynamic design point where we maximise the thermal efficiency by optimising the thermodynamic cycle (by choosing pressure ratio, TET, cooling flow rate and maximising turbomachinery isentropic efficiency). At take-off, the high ambient T means high engine temperatures, hence turbine cooling needs to be designed to cope with this challenge.

Section D, Q3

(a) Assuming level flight in cruise:

$$\dot{m}_f = dm/dt = V\,dm/ds = -\mathrm{sfc}\,F_N = -\mathrm{sfc}\,mg/(L/D)$$

$$-\mathrm{sfc}\,g/(VL/D)ds = dm/m$$

$$-\mathrm{sfc}\,gs/(VL/D) = \ln(m_2/m_1)$$

If $m_1 = m_0 + m_f$ and $m_2 = m_0$ (all fuel used in cruise):

$$-\mathrm{sfc}\,gs/(VL/D) = \ln(m_0/(m_0+m_f))$$

$$\exp(\mathrm{sfc}\,gs/(VL/D)) = (m_0+m_f)/m_0 = 1 + m_f/m_0$$

(b)

$$V = 0.82\sqrt{\gamma R\,220} = 243.8 \text{ ms}^{-1}$$

$$m_f = m_0 \exp(sg\,\mathrm{sfc}/(VL/D) - 1) = 250 \times \exp(8000 \times 9.81 \times 0.016/(243.8 \times 20) - 1) = 73.43 \text{ tonnes}$$

(c)

$$C_L = 2(m_0+m_f)g/(A\rho V^2) = 2(m_0+m_f)g/(Ap\gamma M^2)$$

$$p = 2(m_0+m_f)g/(A\gamma M^2 C_L) = 2 \times (250+73.43) \times 1000 \times 9.81/(550 \times 1.4 \times 0.82^2 \times 0.5) = 24510 \text{ Pa}$$

$$p/p_{sl} = 24510/101325 = 0.2419$$

interpolate using table in databook

$$h = 10000 + (0.2419 - 0.2615)/(0.2240 - 0.2615) \times 1000 = 10523 \text{ m}$$

(d) Advantages: reduces fuel burn; allows empty weight to be reduced as well (less fuel, lighter structure); aircraft can be used on more routes - better utilisation of fleet. Disadvantages: more time (and fuel burn) in climb and descent; location of airports may not be as desired; longer journey times (changing planes / refuelling).

(e) Neglecting fuel used in climb and descent.

$$m_{f2} = 2 \times 250 \times \exp(4000 \times 9.81 \times 0.016/(243.8 \times 20) - 1) = 68.71 \text{ tonnes}$$

(f) Evaluate new required sfc:

$$\ln(m_f/m_0 + 1) = (sg\,\mathrm{sfc})/(VL/D)$$

$$\mathrm{sfc} = \ln(m_f/m_0+1)VL/D/(sg) = \ln(68.71/250) \times 243.8 \times 20/(8000 \times 9.81) = 0.01509$$

Now, using overall efficiency,

$$\eta_{ov} = \eta_p \eta_{th} = V/(\mathrm{sfc}\,\mathrm{LCV})$$

$$\mathrm{sfc} = V/(\mathrm{LCV}\,\eta_{th}) \times (1/\eta_p)$$

So if reduction in sfc is caused only by increase in η_p,

$$\eta_p^{new}/\eta_p^{old} = \mathrm{sfc}^{old}/\mathrm{sfc}^{new} = 0.016/0.01509 = 1.061$$

1 a Quantum mechanics applies when the de Broglie wavelength ($\lambda = h/p$) is similar to the length scale of a system. This is the case for electrons on the nanoscale. The electron behaves as a wave rather than a particle. ψ must be continuous and so for an infinitely deep well must be zero at the edges, so the wave function must have an integer number of half wavelengths across the well (ie $x = n\lambda/2$). As each of these states will have a unique energy, we have energy quantisation.

b i. For $n=1$

$$E = \frac{\pi^2 \hbar^2}{2mL}$$

$$= \frac{\pi^2 (1.055 \times 10^{-34})^2}{2 \cdot 9.109 \times 10^{-31} \cdot (1 \times 10^{-9})^2}$$

$$E = 6.03 \times 10^{-20}\,\text{J} = 376\,\text{meV}$$

This energy is kinetic, so $E = p^2/2m$

$$\therefore p = \sqrt{2mE}$$

$$= \sqrt{2 \cdot 9.109 \times 10^{-31} \cdot 6.03 \times 10^{-20}}$$

$$p = 3.31 \times 10^{-25}\ \text{kg m s}^{-1}$$

ii. $$\Delta x \geq \frac{\hbar^2}{2\Delta p}$$

$$\geq \frac{(1.055 \times 10^{-34})^2}{2 \cdot 3.31 \times 10^{-25}}$$

$$\geq 1.68 \times 10^{-44}\,\text{m} \Rightarrow \text{much smaller than } L.$$

c The energy of the next lowest state is

$$E = 4\frac{\pi^2\hbar^2}{2mL^2}$$

so if the energy of the lowest state is E_1 then the energy difference is

$$\Delta E = 3E_1$$
$$= 3.376 \text{ meV}$$
$$\underline{\Delta E = 1.13 \text{ eV}}$$

At room temperature, thermal energy is $\sim kT$

$$kT = 0.862 \times 10^{-4} \cdot 298$$
$$kT = 25.7 \text{ meV}$$

This is much smaller than ΔE so the electron is almost certain to be in the $n=1$ state. We could estimate probability of being in $n=2$ using Boltzmann.

$$p(n=2) \approx e^{-1.13/0.025}$$
$$\approx 10^{-19}$$

d $\psi \sim A\cos(\pi x/L)$ where x is the distance from the centre. Hence the probability the electron is within x of the centre is

$$P = \int_{-x}^{x} A^2 \cos^2\left(\frac{\pi x}{L}\right) dx$$

$$\therefore P = \frac{A^2}{2}\int_{-x}^{x} 1 + \cos\left(\frac{2\pi x}{L}\right) dx$$

$$P = \frac{A^2}{2}\left[x + \sin\left(\frac{2\pi x}{L}\right)\cdot\frac{L}{2\pi}\right]_{-x}^{x}$$

For $x = L/2$

$$P = \frac{A^2}{2}\left\{\left[\frac{L}{2} + \sin(\pi)\cdot\frac{L}{2\pi}\right] - \left[\frac{-L}{2} + \sin(-\pi)\cdot\frac{L}{2\pi}\right]\right\}$$

$$P = \frac{A^2 L}{2} = 1 \Rightarrow A^2 = \frac{2}{L}$$

For $x = L/4$ (within $0.25\,nm$)

$$P = \frac{1\cancel{2}}{L}\left\{\left[\frac{L}{4} + \sin(\pi/2)\cdot\frac{L}{2\pi}\right] - \left[\frac{-L}{4} + \sin(-\pi/2)\cdot\frac{L}{2\pi}\right]\right\}$$

$$= \frac{\cancel{2}1}{L}\left\{\frac{L}{2} + \frac{L}{2\pi}(1+1)\right\}$$

$$= \frac{1}{2} + \frac{1}{\pi}$$

$\underline{P = 0.818}$

Classically, we would expect the probability to be the same everywhere, so a probability of 0.5 of being in the centre half of the well. Quantum mechanics means it ~~is~~ is much more likely for the electron to be close to the centre of the well.

2 a We need a TFT for each colour sub-pixel in the display, so the number of TFTs is
$n = 3 . 1920 . 1080 = 6\,220\,800$.

b The area of one sub-pixel's liquid crystal is

$$A = \frac{0.6 \times 0.3375 \times 0.9}{6220800} = 2.93 \times 10^{-8}\ m^2$$

Hence,

$$C_S = \frac{\varepsilon_0 \varepsilon_r A}{d} = \frac{8.854 \times 10^{-12} . 8 . 2.93 \times 10^{-8}}{10 \times 10^{-6}} = 0.208\ pF$$

c To get within 5% takes 3 time constants ($1/e^3 = 0.049$). Hence

$$R_{TFT} = \frac{t}{3C_S} = \frac{15 \times 10^{-6}}{3 . 0.208 \times 10^{-12}} = 24\ M\Omega$$

d In the linear regime

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

$$\frac{\partial I_{ds}}{\partial V_{ds}} = \mu \frac{W}{L} C \left[(V_{gs} - V_T) - V_{ds} \right]$$

Hence $R_{TFT} = \dfrac{L}{\mu W C[(V_{gs} - V_T) - V_{ds}]}$

$$24 \times 10^6 = \frac{L}{\mu W C [3 - 0.5 - 0.75]}$$

$$\mu \frac{WC}{L} = \cancel{42 \times 10^6}\ \frac{1}{42 \times 10^6}$$

$$\therefore \frac{WC}{L} = \frac{42 \times 10^6}{1 \times 10^{-4}} = \cancel{42 \times 10^{2}}\ 2.37 \times 10^{-4}$$

Take $W/L = 1$ then

$$C = \frac{\varepsilon_0 \varepsilon_r}{d}$$

$$d = \frac{\varepsilon_0 \varepsilon_r}{C}$$

$$= \frac{8.854 \times 10^{-12} . 7.5}{\cancel{42 \times 10^{-10}} \ 2.37 \times 10^{-4}}$$

$$\underline{d = 280\text{nm}}$$

e At 60 Hz, the TFT refreshes the voltage on the liquid crystal once every ~~8~~ 1/60 = 16.7ms. The time constant

$$\tau = 10^6 . R_{TFT} . C$$

$$= 10^6 . 24 \times 10^6 . 0.208 \times 10^{-12}$$

$$\tau = 5\text{s}$$

$$\therefore \exp\left(\frac{-t}{\tau}\right) = \exp\left(\frac{-\cancel{8}16.7 \times 10^{-3}}{5}\right) = 0.9967$$

$\therefore$ Percentage drop = 0.33%

3 a Best advice is to use either a fuming nitric acid clean or a solvent clean. Fuming nitric is good if there might be organic or metal contaminants. It is quite aggressive so solvent is probably best with 10 min ultrasound in each of acetone, IPA and DI water.

b Thermal evaporation is cheap and could cope with the area. However, 1 μm is quite thick for this technique.
Sputtering could also do the area and 1 μm is a viable thickness too. It is more expensive than evaporation, but widely used for displays.
Electroplating can also do the area, is low cost and 1 μm is certainly viable. However, it requires a conducting substrate. Therefore it needs an additional seed layer deposited first which increases complexity.

Sputtering is therefore probably best. It is not the cheapest, but is simplest for this application and requires no other ~~processing~~. Key process conditions will be:

- Base pressure for low contamination
- RF power for sufficient growth rate so deposition in reasonable time but not too high as this will cause stress.
- Ar gas flow rate and pressure.

c Positive tone resist as $2\mu m$ is closer to the limit for negative tone.

d
$$R = \frac{3}{2}\sqrt{(\lambda(s + z/2))}$$

Contact printing so $s = 0$.

$$\therefore \quad R = \frac{3}{2}\sqrt{\frac{\lambda z}{2}}$$

$$R^2 = \frac{9}{4} \cdot \frac{\lambda z}{2}$$

$$\therefore z = \frac{8R^2}{9\lambda}$$

$$= \frac{8 . (\cancel{2}^{0.5} \times 10^{-6})^2}{9 . 365 \times 10^{-9}}$$

$$z = \cancel{} \, 610 \text{ nm}.$$

As the Cu layer is $2\mu m$ thick, we need a selectivity of $2/0.61 = 3.28$ at least and probably >5.

RC1

Paper 8 – Section F (2022)

SMOOTHING

1(a) (i) Smoothing need – reduce high frequency noise which is amplified by differentiation

– scale selection (remove freq above ω_c)

(ii) Gaussian is a low-pass filter. Show by considering Fourier transform

$$g_\sigma(x) \underset{}{\overset{FT}{\rightleftharpoons}} G(\omega)$$

$$G(\omega) = e^{-\frac{\omega^2\sigma^2}{2}}$$

$G(\omega)$ is a scaled-gaussian with $\sigma' = \frac{1}{\sigma}$

ie. cut-off frequency $\omega_c \propto \frac{1}{\sigma}$

(iii). Consider a family of smoothed images, $S(x,y,\sigma) = I(x,y) * G_{\sigma_i}(x,y)$

Efficiency from Image Pyramid

– do smoothing as 2 × 1D convolutions $G_\sigma(x,y) = g_\sigma(x) * g_\sigma(y)$

– sample scale-space logarithmically

$$\sigma_i = 2^{i/s}\sigma_0 \qquad i = 1 \ldots\ldots N$$

– apply small incremental blurs to get a larger σ

$$g(\sigma_{i+1}) = g(\sigma_i) * g(\sigma_K)$$

small kernel

1a(cont)
iii

- after s convolutions

$$\sigma_{i+s} = 2\sigma_0$$

so can <u>sub-sample</u> to $\frac{1}{4}$ size. (Nyquist sampling)

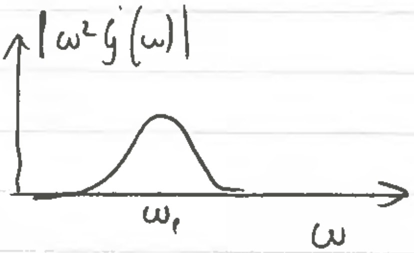
- repeat for next octave

- apply same family of σ_k to smaller image
(equivalent to large σ on full-size image)

(b) BAND-PASS FILTERING

(i) Laplacian of gaussian is a band-pass (see Fourier Transform)

$$\frac{d^2 g_\sigma(x)}{dx^2} \rightleftharpoons -\omega^2 \mathcal{G}_{\sigma_i}(\omega)$$

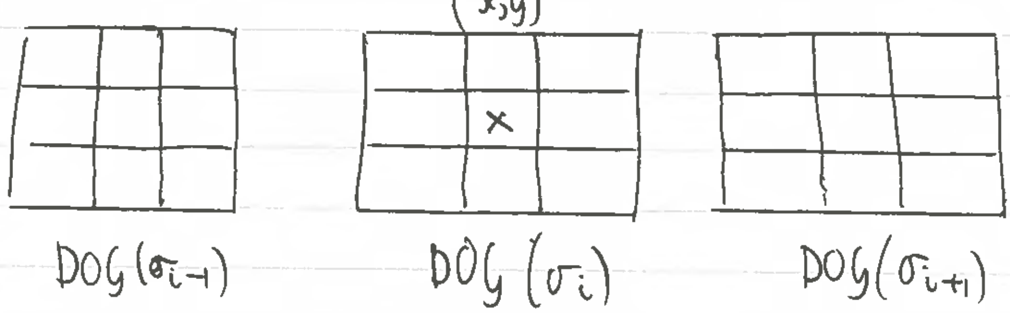


Difference of gaussian is a good approximation to $\sigma^2 \nabla^2 \mathcal{G}_\sigma(x,y)$

$$\underline{S(x,y,k\sigma) - S(x,y,\sigma) \simeq \sigma^2 \nabla^2 S}$$

Alternative: gaussian is a LP. Difference is a B.P.
Efficiently done by subtracting blurred images in same octave (i.e. same size)

(ii) Find min/max of DOG image over position (x,y) and scale (σ_i)
For each pixel test $DOG(x,y,\sigma_i)$ and 26 neighbors for local max/min.



(c) Descriptor and matching:

(i)

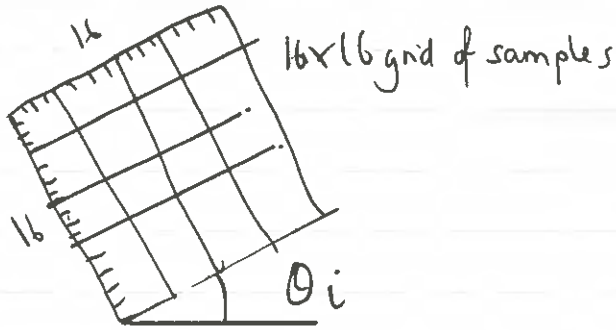
Sample 16x16 pixels at correct scale from $S(x_f, y_f, \sigma_i)$

Compute gradients $\nabla S(x_f, y_f, \sigma_{fi})$ and form a histogram of gradient magnitudes for each orientation in 10° bins, accumulating $|\nabla S|$ in each bin orientation $\angle \nabla S$.

Smooth.

Find maximum $\longrightarrow$ this is the dominant orientation, θ_i

Resample 16x16 at correct scale σ_i and orientation θ_i



(ii) SIFT

- 4x4 grid of smoothed pixels $S(x, y, \sigma_i)$ – 16 cells
- compute ∇S at each pixel
- weight by $G_\sigma(x_i, y_i)$ around blob centre.
- HOG for each cell, 8 orientations $\pm 0, \pm 45^\circ, \pm 90^\circ$

- concatenate into 128D vector
- normalize and truncate all elements ≥ 0.2 to 0.2
- renormalize to give a unit 128D vector

1c(ii) Invariance to - scale — blob centre and scale normalizes size
- orientation — dominant orientation to sample 16x16
- lighting — gradients and normalisation
- perspective — HOGs and 16 cells (not complete) and will fail with big viewpoints
- occlusion — local descriptor avoiding boundaries

Q1c(iii) Matching

Compute SIFT descriptor for each keypoint in both images

For each keypoint in viewpoint 1 we have, $\underline{x}$, find the best match and second best match $\underline{x}_1$ and $\underline{x}_2$ in viewpoint 2 using D, euclidean distance and nearest neighbour.

Accept if $\dfrac{D(\underline{x}, \underline{x}_1)}{D(\underline{x}, \underline{x}_2)} < 0.7$ to avoid ambiguous matches.

Efficiently done with a binary k-d tree structure.

(d) 3D shape — use camera geometry (projection matrix) and triangulation

Pose — need at least 3 pt correspondences.

Q2

(a) Describe a convolutional layer of a convolutional neural network (CNN).
(i) A convolutional layer computes the convolution of # output channels feature maps, each of which is a 2D tensor of shape (kernel size, kernel size) with the input from the previous layer, e.g. at the lowest layer, this could be the pixel intensities of an image. The input is a 3D tensor, with the first two dimensions corresponding to vertical and horizontal location in a 2d grid; the last dimension corresponds to the input channels (e.g. RGB for a color image). The convolution operator corresponds to computing the inner product of the feature map with each (kernel size, kernel size) square of activations (e.g. pixel intensities) in the input.

(ii) We simply count the number of such squares to get the height and width of the output, and the final dimension is defined by the number of feature maps. So the shape is
(32- kernel size + 1, 32- kernel size + 1, # output channels).

(iii) Because convolution is a linear operation, we need to introduce a nonlinearity between subsequent convolutional layers to increase the representational capacity of the CNN. This nonlinearity is called the activation function, and a typical example is the rectified linear (ReLU) activation function: max(0, x).

(iv) Each feature map is defined by a set of weights and a bias. We compute the inner product of the weights with a square of input values at a given location and add the bias. This pre-activation is used as input to the activation function to compute the final activation of that particular feature map of the convolutional layer at that particular location.

(b) Discuss the role of pooling layers in CNNs.

(i) Pooling combines the activations of a given channel across multiple adjacent locations (typically a 2x2 square), to produce a single activation. Max pooling outputs the max of the input values, average pooling outputs the average.

(ii) Pooling performs dimensionality reduction, which can reduce the computational requirements. Another reason to use pooling is to encode translation invariance.

(iii) Pooling destroys information about location, which might be useful when the precise location of features is important for solving a task.

(c) A dataset of images is collected from a car driving on roads in Canada. Images are labeled based on whether they contain a pedestrian or not.

(i) Supervised learning amounts to training a machine learning model to accurately predict the labels assigned to some input data (e.g.) images, by comparing the predictions to the labels and updating parameters to produce more accurate predictions. Labels are typically generated before training via human annotation.

(ii) The images can be used as input and the labels (1 = pedestrian, 0 = no pedestrian, or vice-versa) as targets. We train the model to make accurate predictions on this dataset and hope it will generalize to unseen images.

(iii) A typical loss function is binary cross entropy, i.e. the negative log-likelihood of the true label.

(iv) A typical learning algorithm is stochastic gradient descent, which computes the gradient of the loss function with respect to the parameters on a given subset (called a “mini-batch”) of the input examples, and takes a step in the opposite direction.

(v) We should split the dataset into (randomly selected, non-overlapping) training, validation, and test subsets. The training set is used to compute the gradient updates, which change the models parameters. The validation set is used to tune the hyperparameters (e.g. kernel size or learning rate) to achieve good generalization performance, and the test set is (only) used to estimate how well different models will perform on unseen data once we have finished experimenting.

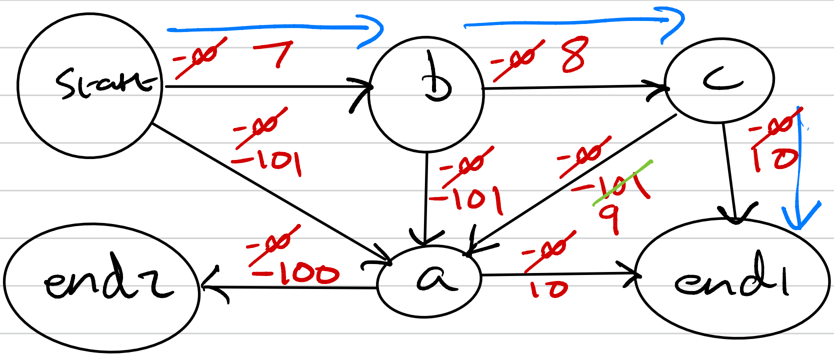
(vi) The model likely fails because of distributional shift -- the training and test data are not from the same distribution. For instance, there might be more snow in the images from Canada, and the cars will be on the other side of the road. We could address this by including images from the UK in the dataset, and/or evaluating test performance separately on Canadian and UK examples, to determine whether the model will generalize well to data from the United Kingdom. We can also describe this as overfitting to the training distribution, but regularization is not expected to solve out-of-distribution generalization problems like this -- we need more diverse data.

3) a) Dijkstra can be used to minimize the sum of positive costs, or maximise the sum of negative rewards, but here there are positive rewards. 4

b) $V(a) \leftarrow \max(10, -100) = 10$, $V(c) \leftarrow \max(-1+10, 10) = 10$
$V(b) \leftarrow \max(-1+10, -1+10) = 9$, $V(\text{start}) \leftarrow \max(-1+9, -1+10) = 9$ 5

c) i) problem is deterministic, so there is no need for averaging. 3

ii)



greedy path, shown in blue, has an actual reward of 8 vs expected reward of 7 (ie $Q(\text{start}, \to b)$) 6

iii) start, $\to a$ (as was visited too early) 3

iv) start, $\to a$ now updated to 9, so is the greedy first action.

next action is $\to$ end1 probability $1-\epsilon = 0.9$
or random, with probability $\epsilon = 0.1$
ie $\to$ end1, probability 0.05
$\to$ end2, probability 0.05

$\Rightarrow$ expected reward is $-1 + (0.95 \times 10 + 0.05 \times -100)$
$= \underline{3.5}$ 4

[Many candidates wrote that the probability of taking the "wrong" action was ϵ rather than $\frac{\epsilon}{2}$]

SECTION G

Bioengineering

1 (a) Explain what affects the visibility of different structures in 2D images of the back of the eye, *other than resolution*, when imaged both with a Fundus Camera and a Scanning Laser Ophthalmoscope (SLO). How can this visibility be improved? [4]

Answer: Visibility depends on both overall light levels and contrast. For overall light levels, the Fundus camera uses an anulus to transmit light in and receives light in a central area, which is not very efficient. For an SLO, the input light is at the central spot (from the laser) and light is collected at all other locations around this, which is much more efficient. Since the eye can only tolerate certain light levels, this limits the noise level (and hence visibility) in the image, but up to a point, light levels can be improved by using eye drops which enlarge the pupil.

To improve contrast, the light wavelengths can be controlled by use of a filter so that certain structures (which may be more reflective at that wavelength) are highlighted more than others. It is also possible to use a contrast agent, i.e. a chemical which increases reflectivity of, for instance, the blood vessels.

(b) A light beam is focused by a lens. The beam cross-section is a disc of radius r given by:

$$r^2(z) = {r_0}^2 \left(1 + \left(\frac{\lambda z}{\pi {r_o}^2}\right)^2\right)$$

where z is the axial distance from the focal point, r_0 is the radius at the focal point, and λ is the wavelength of light. Assume that all components are in air.

(i) Derive expressions for the lens axial resolution Δz_l (along the z-direction) and radial resolution Δx, each in terms of λ and r_0. [4]

Answer: The radial resolution Δx is defined as the minimum diameter of the beam, i.e. the diameter at the focal point. Hence:

$$\Delta x = 2r_0$$

The axial resolution of the lens Δz_l is defined as the range in z over which the beam has at most twice the area as at the focus. Hence we need to find the point at which

$r(z) = \sqrt{2}r_0$, at which point $z = \frac{\Delta z_l}{2}$, so:

$$\begin{aligned}
\left(\sqrt{2}r_0\right)^2 &= {r_0}^2\left(1 + \left(\frac{\lambda\frac{\Delta z_l}{2}}{\pi {r_0}^2}\right)^2\right) \\
2 &= \left(1 + \left(\frac{\lambda \Delta z_l}{2\pi {r_0}^2}\right)^2\right) \\
\left(\frac{\lambda \Delta z_l}{2\pi {r_0}^2}\right)^2 &= 1 \\
\Delta z_l &= \frac{2\pi {r_0}^2}{\lambda}
\end{aligned}$$

(ii) Show that the numerical aperture, NA, of the lens is given by NA $\approx \frac{\lambda}{\pi r_0}$. [4]

Answer: NA, for a lens in air, is defined as $\sin\theta$ where θ is the maximum angle subtended by a parallel beam of light passing through the lens. If the lens has diameter D and focal length f, then (taking small angle approximations):

$$\mathrm{NA} \approx \frac{D}{2f}$$

Using the provided equation, $r(z) = \frac{D}{2}$ at $z = f$, hence:

$$\left(\frac{D}{2}\right)^2 = {r_0}^2\left(1 + \left(\frac{\lambda f}{\pi {r_o}^2}\right)^2\right)$$

Since $D \gg r_0$:

$$\begin{aligned}
\left(\frac{D}{2}\right)^2 &\approx {r_0}^2\left(\frac{\lambda f}{\pi {r_o}^2}\right)^2 \\
\frac{D}{2} &= \frac{\lambda f}{\pi r_0} \\
\mathrm{NA} &\approx \frac{D}{2f} \\
&= \frac{\lambda}{\pi r_0}
\end{aligned}$$

(iii) Hence derive an expression for NA just in terms of Δz_l and Δx. [3]

Answer: From the expression for Δx, we have:

$$r_0 = \frac{\Delta x}{2}$$

and also using the expression for Δz_l:

$$\lambda = \frac{2\pi\left(\frac{\Delta x}{2}\right)^2}{\Delta z_l} \tag{1}$$

Substituting these into the expression for NA gives:

$$\begin{aligned}
\text{NA} &= \frac{\lambda}{\pi r_0} \\
&= \frac{\frac{2\pi\left(\frac{\Delta x}{2}\right)^2}{\Delta z_l}}{\pi\frac{\Delta x}{2}} \\
&= \frac{\Delta x}{\Delta z_l}
\end{aligned}$$

(c) The lens in (b) is used in an SLO by adding a confocal aperture with diameter d, located close to the focal point. The axial resolution of the SLO is Δz_s.

(i) If Δz_s is defined as the range over which at least half the beam area will pass through the aperture, derive an expression for Δz_s in terms of λ, r_0 and the aperture diameter d. [3]

Answer: The axial resolution of the system is given by the point where the aperture cuts off half of the beam area, i.e. where $r(z) = \frac{d}{\sqrt{2}}$, in which case $z = \frac{\Delta z_s}{2}$:

$$\begin{aligned}
\left(\frac{d}{\sqrt{2}}\right)^2 &= {r_0}^2\left(1 + \left(\frac{\lambda\frac{\Delta z_s}{2}}{\pi {r_0}^2}\right)^2\right) \\
\left(\frac{\lambda \Delta z_s}{2\pi r_0}\right)^2 &= \frac{1}{2}d^2 - {r_0}^2 \\
\Delta z_s &= \frac{2\pi r_0}{\lambda}\sqrt{\left(\frac{1}{2}d^2 - {r_0}^2\right)}
\end{aligned}$$

(ii) Bearing in mind your answers to (b) and (c)(i), what is the approximate minimum achievable axial resolution of the system in practice, just in terms of r_0? What value of d would achieve this? [5]

Answer: On first inspection it would appear that a value of $d = \sqrt{2}r_0$ would result in a system axial resolution $\Delta z_s = 0$. However, that is simply because only half the light is passing through the aperture even at the focus. In reality the axial resolution can not be better than that of the lens Δz_l. In that case:

$$\begin{aligned} \Delta z_s &= \Delta z_l = \frac{2\pi {r_0}^2}{\lambda} \\ \left(\frac{1}{2}d^2 - {r_0}^2\right) &= r_0 \\ d &= 2r_0 \end{aligned}$$

To get Δz_s in terms or r_0, consider that the maximum NA we can achieve in air is (slightly less than) 1. Hence, from (b)(iii), $\Delta z_l \approx \Delta x$, and therefore:

$$\Delta z_s \approx \Delta x = 2r_0$$

(iii) Other than resolution, what are the constraints on the size of the confocal aperture d? [2]

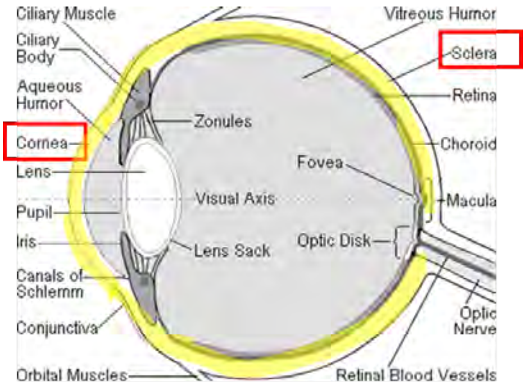
Answer: The main other constraint is the amount of illumination: all the light beam should pass through the confocal aperture if it is focused at the correct point. Given the aperture has to have a finite thickness, d will always need to be somewhat larger than the value given in (c)(ii).

Examiner's Note: There were some very good answers to this question, but many answers suffered from a lack of careful reading of what the question was actually asking. (a) was generally well answered, but some students simply wrote down anything they knew about both techniques. In (b) several students presumed this was a repeat of the exact derivation from lectures, when it was actually simpler than that. Multiple answers started with the required result $\Delta x = 2r_0$ and then went on to needlessly derive this in terms of other parameters. The proof for NA was surprisingly poorly answered given that this was exactly as presented in the handout. Many students did successfully express NA in terms of resolutions in (iii) though. Part (c) was clearly more difficult (which was the intention). Though there were some good attempts at (i), this was rarely followed through into (ii), with several students stating numerical values rather than (as was asked) the resolution in terms of r_0. (iii) was answered very well.

2 (a) Delivery of liquid drug formulation to the retina is usually achieved through the sclera instead of through the cornea of the eye.

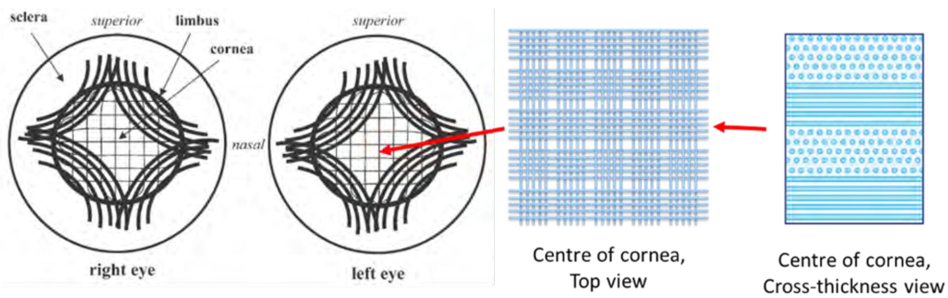
(i) Describe the anatomy, functions and extracellular matrix organisation of the cornea and the sclera. Use sketches to assist your answers. [5]

Answer:

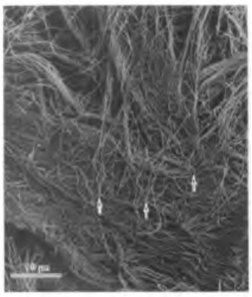


The eyeball has a globe shape where the exteriorly accessible areas are covered by the cornea and the sclera. The anatomical positions of cornea and sclera are highlighted in the eye diagram above.

Cornea provides $\frac{2}{3}$ of the optical focus power. It 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 provides support for the eyeball and helps to maintain its shape. 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).

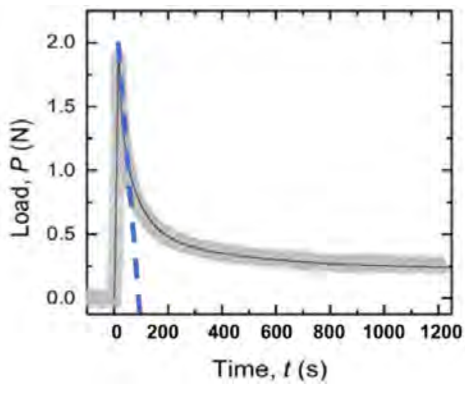


(ii) Based on your answer in (i), briefly explain why a poroelasticity model could be useful in describing fluid transport behaviour as a result of intraocular pressure (IOP) for the sclera, but not for the cornea. [2]

Answer: As the cornea is tightly packed by crystalline collagen fibres, the cornea has a low permeability (in many cases, can be considered as impermeable). The more open network packing structure of sclera affords higher permeability, of which effective network porosity could be evaluated by a poroelastic model.

(iii) Figure 1 below shows the time dependent response of a sclera which was subjected to an indentation test. Based on the information given in this figure, estimate the hydraulic permeability κ of the tested sclera specimen, assuming the material time constant $\tau = \frac{h^2}{E\kappa}$, where the specimen thickness h is 1 mm and the sclera material Young's modulus E is 2 MPa. [4]

Answer:



Based on the figure the relaxation time scale, $\tau \approx 100$ s. Thus $\kappa \approx 5 \times 10^{-15}$ m^{4}Ns.

(iv) Discuss the conditions and assumptions under which the estimation of hydraulic permeability of sclera in (iii) is valid. [3]

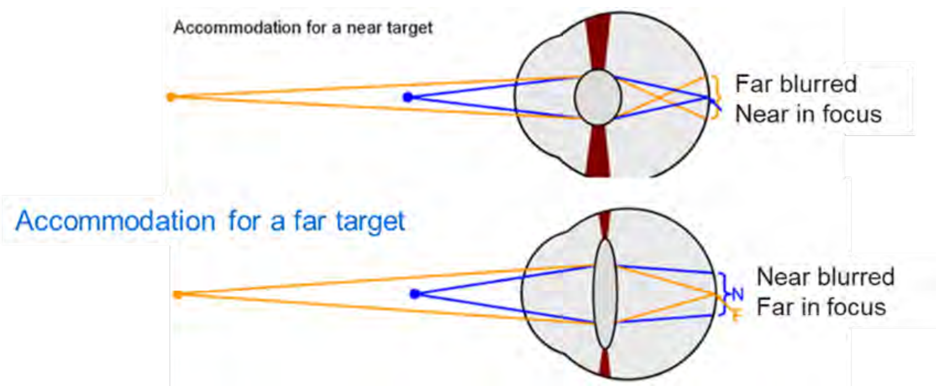
Answer: 1-D convectional flow across the sclera, thus the sclera should have uniform thickness where the sample area is large with no edge effects. Also, the sclera specimen is assumed to behave in the linear poroelasticity regime. In vivo and in vitro measurements the same.

(v) The diffusion constants D for different drug molecules in sclera were found to range from $4 \times 10^{-13}\,\mathrm{m^2s^{-1}} \lesssim D \lesssim 2 \times 10^{-10}\,\mathrm{m^2s^{-1}}$. Along with the information from (ii) to (iv), comment on whether drug transport in sclera is dominated by diffusion or convection at IOP=20 mmHg. (1 mmHg = 133.3 Pa). [3]

Answer: Based on dimensional analysis, transport by convention due to IOP is estimated as to be at least $\kappa \times \mathrm{IOP} \approx 1.3 \times 10^{-11}\,\mathrm{m^2s^{-1}}$ (based on the poroelasticity model). This value is much larger than the transport driven by diffusion.

(b) Aided by a sketch, describe lens accommodation in young, normal, healthy eyes. [3]

Answer: Lens curvature is controlled by ciliary muscles. By changing curvature, the eye focuses on objects at different distances. This process is called accommodation.



(c) Intraocular lenses, and daily disposable contact lenses, can both be used to augment the eye for correcting 'near-sightedness' (or, myopia). For each case, discuss the material selection criteria and list the typical materials used. Comment on the advantage(s) and disadvantage(s) for using intraocular lenses versus daily disposable contact lenses for eyesight corrections. [5]

Answer: For both cases, the material needs to be transparent, biocompatible for the duration of use, and can be made into a concave lens shape for correcting near-sightedness.

For intraocular lens, it is implanted in the eye as a permanent structure, thus materials need to have long-term biocompatibility, anti-fouling property, and be highly durable. Materials used are e.g. PMMA (early version), and acrylates silicone rubber (modern version to enable minimally invasive surgery).

For daily disposable contact lens, since the lens is placed on the surface of the eye, and can be disposed easily, it can be made of lower cost materials, which are flexible, less durable but more for comfortable wearing. Material examples are such as silicone rubbers and hydrogels.

Intraocular lens is a permanent structure, where the procedure require surgical insertion; while contact lens needs to be changed daily and could be more prone to infection.

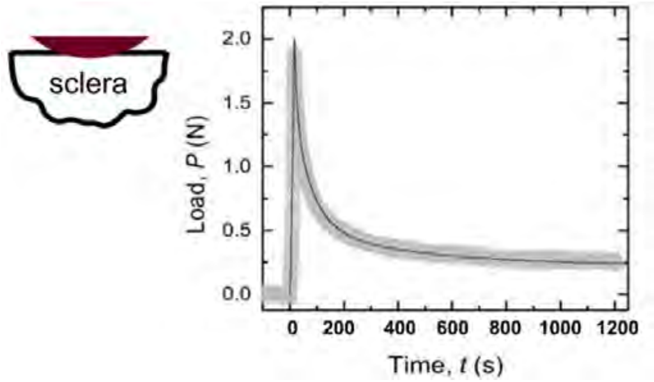


Figure 1.

Examiner's Note: For mostly descriptive questions of (a, i), (b) & (c), high marks were awarded for completeness. For part (a,ii) to (a,v), students were required to comprehend and grasp contents taught in the course to answer those questions. Nonetheless, there were many good answers and hence the average was fairly high.

3 This question is about efficient coding. A neuroscientist discovers a neuron in the brain of an adult primate that responds linearly to the scalar activities s_{L} and s_{R} of two retinal ganglion cells, from the left and right eyes respectively. For simplicity, the activities s_{L} and s_{R} are modelled as correlated Gaussian variables, each with zero mean and variance σ^2, and with correlation $\rho > 0$. The response of the neuron is modelled as $x = w_{\mathrm{L}}s_{\mathrm{L}} + w_{\mathrm{R}}s_{\mathrm{R}} + \epsilon$, where w_{L} and w_{R} are some weight parameters and ϵ is independent Gaussian noise with zero mean and unit variance.

(a) What condition(s) on the two ganglion cells' receptive fields would justify the assumption of a positive correlation $\rho > 0$ between s_{L} and s_{R}, under natural visual stimulation? Additionally, why is it unrealistic to envisage a perfect correlation, $\rho = 1$? [5]

Answer: For the two ganglion cells to respond similarly (i.e. in a positively correlated manner), their receptive fields must have the same polarity (on-center/off-surround vs. off-center/on-surround) as well as the same spatial location. Moreover, incompressible noise in their responses should naturally limit ρ – a perfect correlation of $\rho = 1$ would only arise if the two cells were entirely noiseless.

(b) For each of these statements, explain, with reason, whether it is likely wrong or could well be true:

(i) The neuron was discovered in LGN, $w_{\mathrm{L}} > 0$, and $w_{\mathrm{R}} > 0$;

(ii) The neuron was discovered in V1, $w_{\mathrm{L}} > 0$, and $w_{\mathrm{R}} > 0$;

(iii) The neuron was discovered in V1 of the right hemisphere, $w_{\mathrm{L}} = 0$, and $w_{\mathrm{R}} > 0$.

[5]

Answer: (i) Likely wrong: retinal inputs from the two eyes to LGN neurons tend to become segregated over development, such that one is unlikely to find a binocular neuron in the adult LGN. (ii) Could well be true: many cortical neurons in V1 receive binocular inputs and could therefore have the response profile of x as modelled here with both weight parameters . (iii) Could well be true, provided the RF of the right-eye ganglion cell is the left hemifield.

(c) Express the variance of the neuron's response x in terms of ρ, σ^2, w_{L} and w_{R}. [3]

Answer: The marginal variance of x is by definition $\mathbb{E}[x^2]$ (as x has zero mean), which can be expanded as $\mathbb{E}[\epsilon^2 + w_{\mathrm{L}}^2 s_{\mathrm{L}}^2 + w_{\mathrm{R}}^2 s_{\mathrm{R}}^2 + 2w_{\mathrm{L}}w_{\mathrm{R}}s_{\mathrm{L}}s_{\mathrm{R}} + 2\epsilon(w_{\mathrm{L}}s_{\mathrm{L}} + w_{\mathrm{R}}s_{\mathrm{R}})] = \mathbb{E}[\epsilon^2] + w_{\mathrm{L}}^2\mathbb{E}[s_{\mathrm{L}}^2] + w_{\mathrm{R}}^2\mathbb{E}[s_{\mathrm{R}}^2] + 2w_{\mathrm{L}}w_{\mathrm{R}}\mathbb{E}[s_{\mathrm{L}}s_{\mathrm{R}}] + 2\mathbb{E}[\epsilon(w_{\mathrm{L}}s_{\mathrm{L}} + w_{\mathrm{R}}s_{\mathrm{R}})]$. The last term in the last expression is zero (because ϵ is independent of both s_{L} and s_{R} and has zero mean), and the rest can be written as $\mathrm{Var}[x] = 1 + \sigma^2(w_{\mathrm{L}}^2 + w_{\mathrm{R}}^2 + 2\rho w_{\mathrm{L}} w_{\mathrm{R}})$.

(d) Express the mutual information (in bits) between x and the stimulus vector $\vec{s} = (s_{\mathrm{L}}, s_{\mathrm{R}})$ in terms of ρ, σ^2, w_{L} and w_{R}. [Hint: do this by first finding expressions for the marginal and conditional entropies of the response: H_x and $H_{x|\vec{s}}$.] [5]

Answer: Conditioned on s_{L} and s_{R}, the response variance is given by the noise variance which is 1. So the conditional entropy of the Gaussian variable x conditioned on $\vec{s}$ is $H_{x|\vec{s}} = \frac{1}{2}\log_2(2\pi e)$. Given the marginal variance calculated in the previous part, the marginal response entropy is given by $H_x = \frac{1}{2}\log_2[2\pi e(1 + \sigma^2(w_{\mathrm{L}}^2 + w_{\mathrm{R}}^2 + 2\rho w_{\mathrm{L}} w_{\mathrm{R}}))]$. The mutual information is thus $I[x;\vec{s}] = H_x - H_{x|\vec{s}} = \frac{1}{2}\log_2[1 + \sigma^2(w_{\mathrm{L}}^2 + w_{\mathrm{R}}^2 + 2\rho w_{\mathrm{L}} w_{\mathrm{R}})]$.

(e) Now suppose the weight vector $(w_{\mathrm{L}}, w_{\mathrm{R}})$ is constrained to have unit length, but otherwise has unconstrained direction in this two-dimensional parameter space. Which direction yields maximal coding efficiency? State your answer in terms of the angle θ such that $w_{\mathrm{L}} = \cos\theta$ and $w_{\mathrm{R}} = \sin\theta$. [5]

Answer: To obtain an efficient encoder, we should maximise the mutual information obtained in the previous part, or equivalently, the response variance obtained in the first part. This is given by $1 + \sigma^2(1 + 2\rho\cos\theta\sin\theta) = 1 + \sigma^2(1 + \rho\sin 2\theta)$. Since ρ is positive, this is maximised when $\sin 2\theta$ is maximised, which happens for $\theta = 45°$ or $\theta = 225°$, and yields a mutual information of $\frac{1}{2}\log_2[1 + \sigma^2(1 + \rho)]$.

(f) Assume $w_{\mathrm{L}} > 0$. With the weight vector $(w_{\mathrm{L}}, w_{\mathrm{R}})$ constrained to have unit length as in part (e), we now add the further constraint that the response variance should not exceed the upper bound $b > 0$. Describe qualitatively how the optimal direction of the weight vector (*i.e.*, the angle θ that yields maximal coding efficiency under the new additional constraint) changes as b is lowered from very large values. [2]

Answer: Given the expression obtained in part (d), maximising the mutual information is equivalent to maximising the response variance. The latter is given by $\mathrm{Var}[x] = 1 + \sigma^2(1 + \rho\sin 2\theta)$ which has a maximum of $1 + \sigma^2(1 + \rho)$ obtained at $\theta = 45°$, and a minimum equal to $1 + \sigma^2(1 - \rho)$ obtained for $\theta = -45°$. Thus for b larger than the maximum $1 + \sigma^2(1 + \rho)$ the optimal direction remains at $\theta = 45°$. But as b is lowered below this value, imposing a smaller upper bound on the response variance, the optimal direction deviates from $\theta = 45°$, and moves towards $\theta = -45°$, and reaches it when b approaches the minimal value of response variance, $1 + \sigma^2(1 - \rho)$. (For values of b below this value the problem does not have a feasible solution.)

Examiner's Note: Part (a) was well answered by all candidates who attempted this question. Part (b) was generally well answered; a few students incorrectly answered (b.ii), wrongly invoking ocular dominance as meaning that neurons only respond to one eye (as opposed to responding more strongly to input from one eye compared to the other). Only a couple of students correctly answered Part (c), even though it was a relatively simple variance calculation. For some reason, many students ended up with a subtractive (instead of positively additive) contribution of the pairwise correlation between sL and sR. Most who attempted part (d) gave a correct general formula for the mutual information, but propagated their mistakes from (c). Only a couple of students answered (e) and (f) in some detail, and even then they did not provide a convincing derivation.

END OF PAPER

Section H: Manufacturing and Management (3 questions, 7 pages)

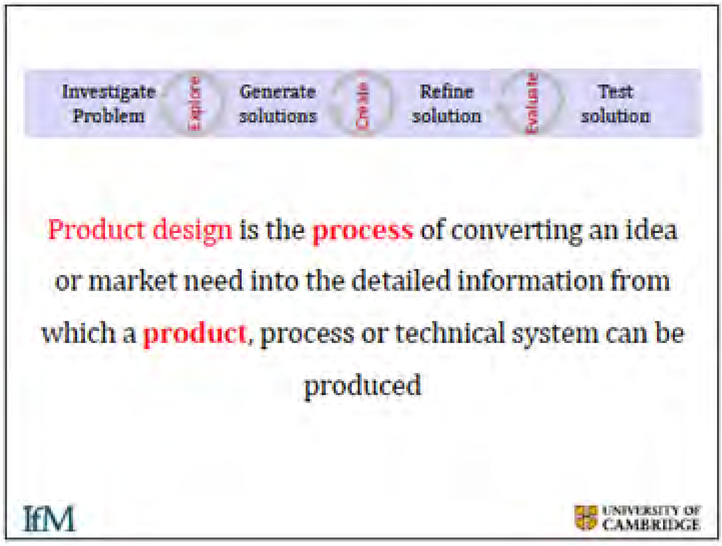
Question 1

(a) There are five types of market pull, students will choose four of these and describe what they mean and provide some examples:

- Experience in use: Oxo good grips, experiencing his wife having arthritis, Sam Faber designed good grips to be an easier to use peeler. His wife could still do the cooking. Yeah!
- Lead user ideas: lead users are those that adopt new ideas and maybe adapt their products before others do. A classic example is the BMX bike.
- Fashion: new trends are often seasonal and these variations drive consumer demand. This translates to consumer goods, when a brand might release a new colour of phone.
- Legistlation/context: responding to changes in the law or in social or economic contexts can be a good source of market pull. For example, changes in petrol pricing and availability is generating a pull for electric vehicles.
- Luxury to common: the first electric vehicles were highly priced luxuries, but as there is growing demand, these are becoming more commonplace.

(b) There are many (hundreds) of representations of the design process. Most have a starting point (idea, problem) and a finishing point (delivered solution) and break this into a series of manageable stages. Any representation will be helpful.

The slide from the notes is:



Investigate problem: doing some research to understand what is needed in a new bike, how effective current solutions are and to identify a gap in the market. This could also include research on new technologies or options.
Generate solutions: a divergent phase, where multiple possible solutions are considered.
Refine solution: combining the best of the concept designs and refining this to a fully resolved solution.
Test and implement: testing assumptions at each stage of the design process to determine whether the solution fits the intended need.

(ii)

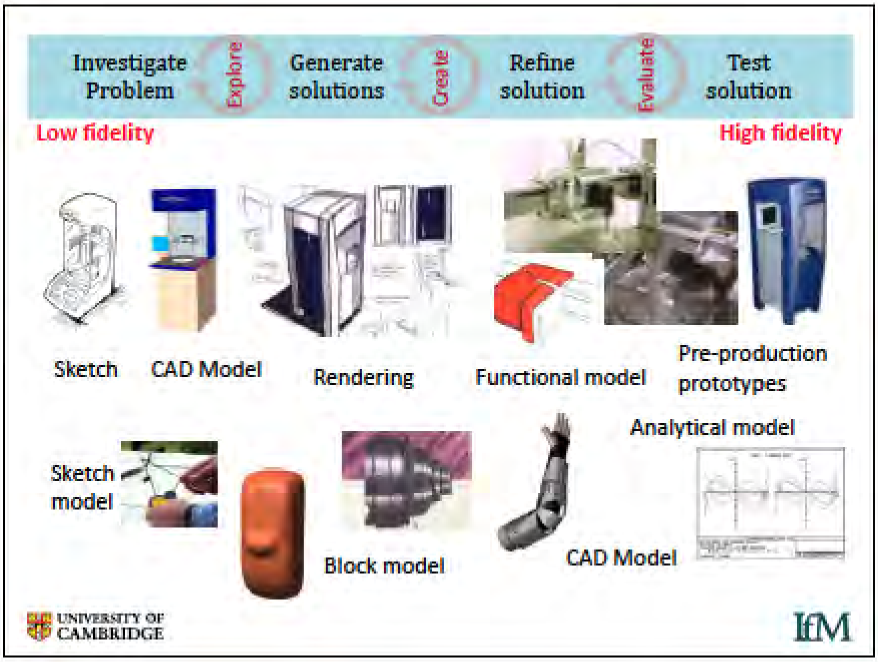
Design is essentially a process governed by prototypes, from low-fidelity early on in the design process, through to more expensive higher fidelity prototypes later on.

Prototypes are always there to help generate new insights or help answer unknowns in the design process. This might be related to technical risks (e.g. will a mechanism work) or to market risks (i.e. which colour would consumers like).

For a bicycle, low fidelity prototypes could include sketches of mechanisms, sketches of layouts, cardboard models of different features, etc. High fidelity prototypes could include fully working pre-production prototypes.

Very best answers might include potential pitfalls of prototyping with discussion in relations to the constraints that the student in the question situation has in using some of these techniques for the assignment (e.g. time/cost) .

Students could give different examples of low-fidelity and high-fidelity prototypes that might be used at different stages of the bike design process.



(c)

(i) Patents

Patents are IP rights that protect technical inventions. Answers should mention the three patentability criteria: Novelty, inventive step and industrial applicability. They would mention that patents are exclusive rights, i.e. they can be used by the owner to prevent others from using the invention protected by the patent, respectively its claims (claims can be narrow or broad or something in between). However, it is up to the owners to take action to enforce its patents. Students might mention the specific character of computer-implemented inventions, i.e. that protect certain kind of software that clearly demonstrate a “further technical effect”. Renewal fees typically have to be paid to keep patents “alive” for a maximum duration of 20 years (with renewal fees typically increasing over time). Patents can be licensed in all kind of ways, such as exclusively to selected licensees or non-exclusively charging the licensees royalties. Patents are national rights, so need to be applied in countries separately (with some exceptions, eg the EPO member states) and the inventions can be freely used in countries where no patents exist. These bundles of national patents are called patent families.

(i) Copyrights

In contrast to patents, copyrights do not protect an invention, but the expression of creative content, traditionally such as literary works, drawings, computer generated illustrations and – importantly today – also software code. In most countries, copyrights do not need to be registered, but arise automatically when the content is created. In most countries, copyrights persist for about 70-80 years until after the death of the creator. Similarly, to patents, copyright protected content can be licenses through a variety of options.

Question 2

(a) IPR systems were put in place to create an incentive for inventors to invest in technical R&D (i.e. in the case of patents) or creators more broadly (e.g. artists, musicians in the case of copyrights or designers). IPR systems provide them with a means to protect their creations/inventions and exclude others from using/copying (without permission) them for a certain period of time. Through this kind of "monopolistic"-effect the creators can monetize their inventions and recoup their investments through monopoly prices (higher than if competition would exist). Without IPR systems – theory goes – any potentially commercially successful invention/creation would instantly be copied by imitators (second movers/followers), which have not to recoup any upfront development investments. In other words, without IPR systems well-functioning markets hardly provide incentives for anyone to invest in potential risky development efforts. Different IPR systems have been developed for different IP rights, such as patents and trademarks, which are targeted at different industries or rather different types of inventions/creations. For instance, patents typically are used in R&D intensive industries (e.g. aerospace, automotive, ICT), while design rights are prominent in industrial sectors such as furniture, textiles and copyrights in artistic sectors, such as music, publishing and art. All IPR systems essentially assign ownership rights to their creators, who can then decide to what extent they want to utilize these ownership rights themselves in their own products and services or whether they want to monetize them / exploit them externally (e.g. via licensing). Globally IPR systems differ, with the patent systems being most harmonized (i.e. their three patentability criteria a mostly accepted in developed countries).

b(i)

The answers to this question might first explore the different types of IP might that can be involved in this case. The question states ("use of standard and bespoke components"). Standard components might indicate that there is not necessarily any novel IP involved. However, bespoke components could relate to potential inventive activity which might result in potential patentable matter, copyright protected software, or possibly design rights. Also, the integration of bespoke and standard components could require development efforts, of which the results could be possible protectable, i.e. involve some IP (e.g. energy management systems that increase the efficiency of a system). Also, the product as such ("optical camera system") could be branded, which would involve one or multiple trademarks, which the owners could choose to register.

Selling the IP, as one business model, would mean that the ownership of all (or part of the) IP will be transferred from the inventors to the buyer of the IP. The ownership transfer through a sale of the IP is a one-off transaction. The inventors would be likely to charge a higher price for the IP compared to the licensing model, however, would permanently loose the rights to the IP (unless they negotiate a use license for themselves (there are various options in any sales contract). This model might be an advantage if the inventors lack the resources to bring the product to the market themselves. If the inventors have applied for patents in multiple countries, they could also decide to commercialise their IP in certain countries themselves but sell the respective patent family members (the same argument goes for trademarks) for the countries in which they do not wish to / lack the resources to commercialise their IP.

Licensing IP the inventors would have to find at least one licensee and negotiate a licensing contract. Licensing contracts allow all kinds of options. This could be an exclusive licensing contract that would limit what the licensee could do (e.g. only use the IP in a certain geography or for certain applications/markets). If the technology has a wide application potential in various industries, for various markets, the inventors could build a licensing models that utilised non-exclusive licensing. An example for this is ARM, which licenses its chipset design to 500+ companies around the world. In such a model, the inventors would have to have resources to negotiate each licensing contract and also monitor correct royalty payments (such audits can be very costly), but would possibly generate a constant income from their IP over many years (in contrast to the one-off payment when selling the IP). However, as they would remain the IP owners, it would be their responsibility to maintain their IP and possibly enforce it against imitators so to protect their licensees against competition. This could potentially be risky and the inventors could attempt push this risk to the licensees in the respective licensing contracts. However, this would possibly result in a negative impact on the royalty rate the inventors would want to negotiate to maximise their royalty income.

b(ii)

1. The design of the process for making the product is dependent on the nature of the product. Consideration needs to be give to the interaction of job design, process technology, balancing of supply and demand, and supply chains.
2. Given the nature of the product (a security system), quality assurance will be a key consideration. As such, the inventors will need to think carefully about how much they choose to make themselves, versus getting someone else to make. On the other hand, if the inventors find that they are not very good at running a production facility, they may be wise to outsource to a more specialist organisation. They may find that there are specific bits of the production process that do wish to keep in house, while letting other elements be outsourced.
3. Given the volumes, it is likely that the production process will be at the high end of of batch production, or the low end of mass production.
4. As they plan to produce in quite high volumes, and plan for even higher volumes in the future, they are going to need to consider whether they will be needing to have an automated approach, or to use manual labour. This decision will also be highly influenced by availability

of key staff either to set up and run an automated approach, or to set up and run the team to do the assembly manually.
5. Consideration would also need to be given to serviceability issues – will the product be made in such a way that will allow for maintenance and repair (i.e. using screws for assembly) or replacement (i.e. use glue and clips for assembly). This choice will also affect the choice of approach used in the job design.

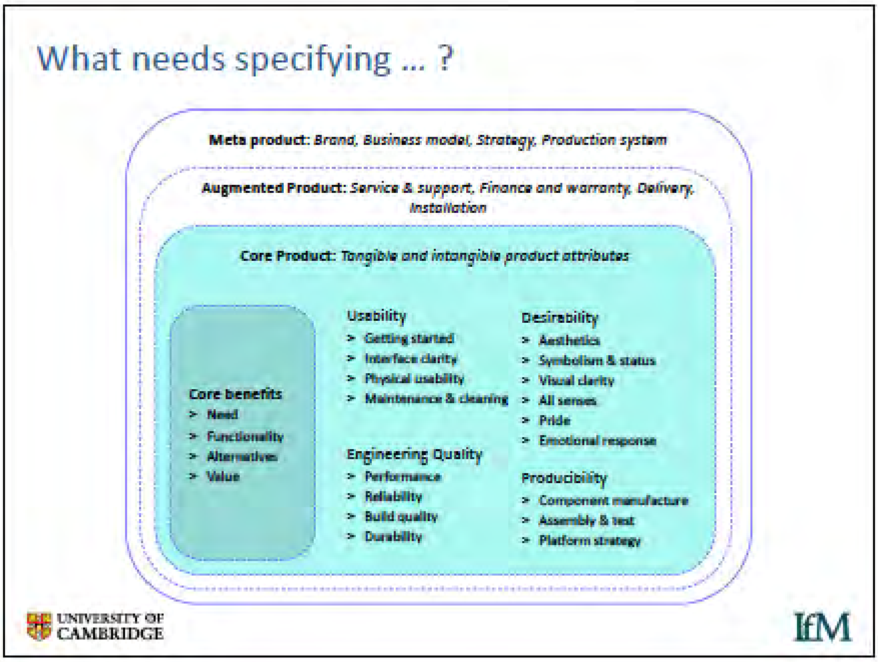
Question 3

(a) The role of the product specification in the design process.

- Defines what needs to be designed
- Expresses customer needs in the 'language of the customer'
- Does not limit ways in which the requirements are met
- Provides design targets
- Sets design constraints
- Provides precise, unambiguous, measurable detail about what the product must do
    - Quantifies and qualifies
- Enables the evaluation of solutions
- Evolves as new information is learnt.

The specification is a document that would evolve throughout the design process, starting with imprecise and incomplete information and developing to have increased precision and accuracy as the process progresses.

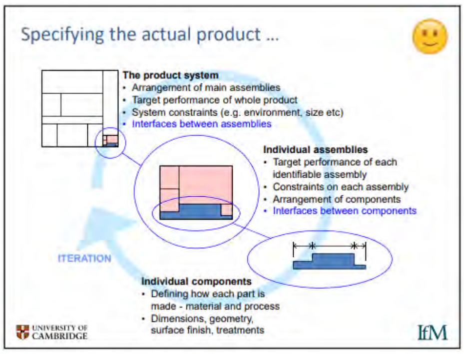
Different aspects of the product should be specified. This is known as the 'design mix':



- Core benefits: the benefits of the product as a whole, easily captured in an 'elevator pitch'.
- Core product: the features of the product itself, including issues of engineering performance, manufacturing cost, aspects of ergonomics and usability etc.

- Augmented product: any product related services that are critical for the product to be a commercial success. For example, delivery etc.
- Meta product: the underlying business strategy or production system that might be needed. For example, if a new product requires a new production line.

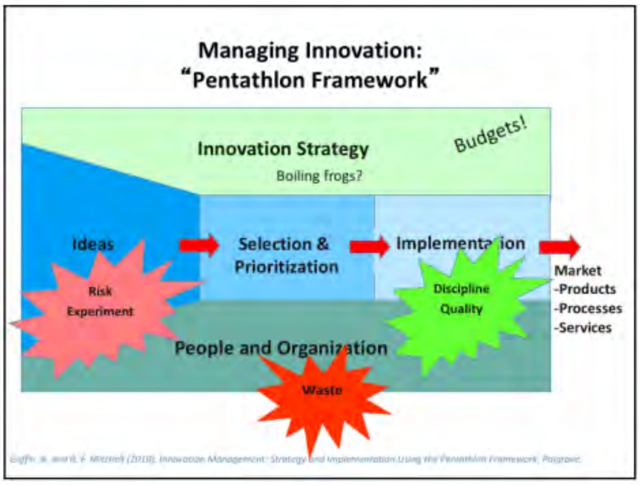
The specification should ideally quantify critical features and where this is not possible, qualify them, in order that the statements are not vague and performance can be evaluated.



(b).

Sell a product plus consumables: Pro = Long-term revenues on the back of the sale of each product; Con = Need infrastructure to deliver consumables. Examples: Nespresso machines, printers, razors and blades.
Sell services enabled by a product: Pro = Long-term revenues on the back of the sale of each product, Con = Need infrastructure to provide services. Examples: mobile phones and internet access/apps; Rolls-Royce ‘power by the hour’.
Sell a service: Pro = No manufacturing costs; Con = Can be hard to scale up the business (in many cases can only grow through recruiting and training lots of people). Examples: Consultancy services; AirBnB; advertising agencies.

(c)



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. For example, communicating internally an innovation strategy becomes harder as companies increase in size. Companies should design processes to make sure that everyone involved in innovation is aware of where the company wants to go.
2. Ideas: This is the raw material for innovation. Idea generation needs to be sEupported 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. Examples of how this can be done include technques for identifying a market for a new technology (e.g. market surveys, observation, interviews, personas) and techniques such as brainstorming.
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. Examples: Bubble charts to decide on the proportion of incremental vs radical innovation.
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. For instance if any of the projects is in collaboration between large and small companies, there have to be attention on how to make sure that across the partners there is an agreement on how the technology/project is brought to maturity (who does what etc).
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. Depending on the aspects, different types of people might be required: for example for setting up partnerships with s firms, we should be looking into people who are aware of the challenges and of the approaches to help collaborations of this type succeed.