Optical properties of earth-abundant semiconductors for renewable energy

Max Birkett <**max.birkett@gmail.com**>

Stephenson Institute for Renewable Energy, Department of Physics,

Chadwick Building, University of Liverpool, Peach Street, Liverpool, L69 7ZF, UK

September 30, 2016

Contents

Declarations and Acknowledgements

I am most grateful for the generosity of the following workers, who provided materials, time or assistance necessary for the completion of this research

- **Adam W. Welch**[a] and **Andriy Zakutayev**[a] for providing Cu_3N and $CuSbS_2$ material for investigation, and Hall carrier concentration estimates.

- **Rob Treharne**[b] for kindly providing ellipsometry spectra and CdTe material.

- **Mohana K. Rajpalke**[b,e] and **Wojciech M. Linhart**[b,f] for providing FTIR spectra of $CuSbS_2$; and to Mohana for assisting with SEM and EDS of Cu_3N and $CuSbS_2$.

- **Thomas J. Whittles**[b] for providing XPS spectra of Cu_3N and for assisting with $CuSbS_2$ FTIR spectra.

- **James T. Gibbon**[b] for assisting with x-ray diffraction of $CuSbS_2$.

- **Paul Thompson**[d] for kindly providing temperature-dependent XRD patterns of Cu_3N.

- **Christopher Muryn**[c] and **Vin R. Dhanak**[b] for providing patterns and assisting with temperature-dependent XRD of Cu_3N.

- **Christopher Savory**[g] and **David O. Scanlon**[g] for sharing HSE06 band structures and density of states for Cu_3N and $CuSbS_2$.

- **Tim D. Veal**[b] for guidance and illuminating discussions throughout.

I declare that all work within this thesis is the independent work of the author, and to the best of my knowledge, has not been published previously, except where indicated in the text. Where raw data has been provided by collaborators, all reduction, analysis and development is solely the work of the author.

<div align="right">Max Birkett,   30th June 2016</div>

Abstract

The research described within primarily addresses the experimental determination of optical properties in emerging photovoltaic (PV) materials. Work proceeds with two specific aims: to consolidate and clarify experimental practice on the determination of optical properties in polycrystalline systems, then to apply any findings in investigations of the relatively unstudied semiconductors copper nitride and copper antimony sulphide, which fulfil many of the requirements for next generation PV materials: earth-abundance, scalability, bipolar doping, near-optimal band gaps, strong absorption, and beneficial transport properties.

While the literature already offers extensive theoretical treatments of optical phenomena and the propagation of light, somewhat less-discussed is the task of practical determination of optical properties via the inversion of measured spectra. In the most general cases inversion may be non-trivial even for simple systems: no global solutions may exist. Furthermore, emerging thin-film photovoltaic technologies may utilise material which, whilst commercially attractive, may not be suited for study by reflection/transmission spectroscopy, while researchers often choose rather elementary spectral reduction approaches where more desirable alternatives exist. After reviewing various models, methods and issues, a self-consistent code is described which determines absorption spectra with improved accuracy.

Practical work on copper nitride Cu_3N and copper antimony sulphide $CuSbS_2$ comprises experimental and first-principles investigations. Optical studies via FTIR and spectroscopic ellipsometry establish absorption and refractive index spectra for both materials; in each case, strong absorption (exceeding $6 \times 10^4 \, cm^{-1}$) is found just beyond the absorption onset. Direct band gaps, average phonon energies and Bose–Einstein electron-phonon interaction strengths are determined by fitting the temperature-dependence of the absorption edge. Atypically small temperature-dependence of the direct gap is found in Cu_3N (along with an optically-active TO-phonon mode), whilst in $CuSbS_2$ a possible excitonic state is seen at low-temperature just above the absorption onset; a symmetry analysis suggests distinctly enhanced absorption for this state: further work with oriented single crystals is proposed.

Structural investigations by x-ray diffraction and Rietveld or Pawley refinement find Cu_3N and $CuSbS_2$ geometries broadly consistent with prior findings. Credible thermal expansion is finally established in Cu_3N between 4.2 and $280 \, K$ by temperature-dependent XRD; very little expansion is seen below $100 \, K$: further synchrotron work is proposed. A quasi-harmonic model estimates the Cu_3N zero Kelvin lattice parameter, Debye temperature and average Gruňeisen parameter. Density-functional theory calculations on Cu_3N and $CuSbS_2$ suggest band structures and symmetries, band gap evolution, selection rules for optical dipole transitions, valence-band density of states (supported via x-ray photoelectron spectroscopy), and evaluate potential structural distortions: such as perovskite rigid-unit modes.

Chapter 1

Renewable energy and photovoltaics

1.1   Renewable energy and climate change

While scientific research is inherently valuable, researchers working in renewable energy areas are fortunate to contribute to a field which is building to tackle perhaps the most important problem of our age. **The role of anthropogenic (human caused) greenhouse gas (GHG) emissions in climate change is clear and warming is unprecedented**[1]: the last 30 years have been the warmest for 1400 years, while the concentrations of CO_2, CH_4 and N_2O have reached levels unseen for $800\,000$ years. Even under moderate mitigation strategies*, mean global surface temperatures are *more likely than not* to warm by over $2.0\,°C$ by 2100 (relative to the pre-industrial age). Significant negative impacts are envisaged globally for people and ecosystems, including[1]: droughts, leading to drinking water shortages, agricultural damage and migration; coastal erosion and sea-level rises; terrestrial and marine ecosystem damage; increased incidence of extreme weather events, e.g. heat waves and heavy precipitation; and unknown risks due to loss of polar ice and permafrost. Substantial reductions in GHG emissions are necessary to limit damage and irreversible change; even if GHG emissions could be halted today, impacts are likely to persist for centuries. Yet there is still time to act: emission reductions over the next few decades can reduce impacts over future centuries[1].

The agreement[2] of 195 nations in Paris to limit global warming to $2.0\,°C$ relative to pre-industrial times (pursing efforts towards $1.5\,°C$), sets significant technological goals which must be met if the impacts of climate change are to be mitigated. While doubts may once have been expressed on the capacity of renewables to meet global energy demand, the pace of recent developments is encouraging. In the UK and Germany in 2015, renewables respectively met $25\,\%$[3] and $33\,\%$[4] of electricity demand (outstripping that from coal in the UK). Unprecedented milestones for electricity generation were then seen: zero coal was burnt in the UK on several occasions[5], while Portugal ran entirely on renewables for 107 hours[6]. Elsewhere, after a 15 year de-carbonisation programme, Uruguay has transformed from a net electricity importer to a net exporter, with renewables now meeting $\sim 95\,\%$ of demand ($\sim 55\,\%$ of total energy needs after including transport fuel)[7]. Costa Rica, Iceland, Paraguay

---

* Such as the intermediate mitigation path RCP4.5, as described in the IPCC AR5[1].
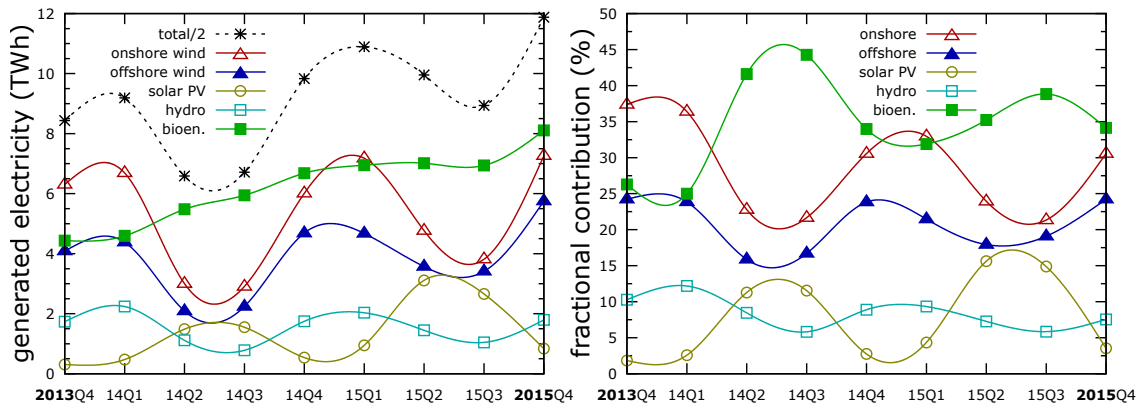
Figure 1: Renewables contribution to UK electricity generation. Data[3] cover the period from the fourth quarter 2013 (Q4) through 2015 Q4. The markers correspond to the actual datapoints; the curves are added to depict the trend. The *total/2* trace shows half of the total for all renewable technologies. The complementary nature of wind and PV technologies is illustrated by the cyclical trend in the figure. The fractional contribution of each PV technology is shown in right-hand figure.

and Lesotho offer further examples of good practice[7]. Problems such as energy storage, seasonal balancing and transport still await effective solutions; yet with $147\,\text{GW}$ of renewable capacity installed in 2015, comparable to the total generating capacity of Africa (despite government subsidies favouring[8] fossil fuels over renewables by 4 to 1), it seems feasible that the goals set in Paris may yet be achieved if the political consensus can be sustained.

Even if the clear imperative to adopt renewable technologies to tackle climate change could be set aside, the limited supply of fossil fuels and the growing demand for energy necessitates new solutions which will ensure energy security and enable socio-economic development, particularly in regions stricken by extreme poverty[9]. Renewable technologies such as solar photovoltaics (PV) are well-suited in this regard; the radiant energy delivered by the Sun is the Earth's most abundant resource, with as much energy delivered in one hour as is consumed by all human activities in an entire year[9].

After a £10 billion subsidy programme in the UK[4], the latest statistics find an interesting mix of renewable technologies[3]. Electricity generation data is shown in Figure 1 for on- and off-shore wind, various bio-energies, hydroelectric[†] and solar PV technologies. While bio-energy and on- and off-shore wind are presently the leading generation technologies, a seasonal cyclical trend is clearly evident in the data, with wind and hydroelectric technologies performing better in the winter (quarters Q4 and Q1) due to greater winds and precipitation, while solar PV does better in the summer, due to more hours of direct sunlight. The cyclical behaviour seen in Figure 1 illustrates the need for an **integrated renewable energy system incorporating a mixture of complementary technologies**: when no sun is shining, the weather may well be windy or raining, whilst bio-energies offer fallback strategies for zero-carbon generation when there is neither sun nor wind. The specific figures for 2015 show wind and hydroelectric generation benefiting from above average wind and rain, with a corresponding drop in solar PV generation offset by increased installed capacity.

[†] Bioenergy comprises: gas from landfill, sewage and municipal waste; plant and animal biomass; anaerobic digestion and co-firing with fossil fuels. Hydro includes wave, river and storage reservoir/dam use (excluding pumped storage[10]).

1.2 Fundamental concepts in photovoltaics

This section develops an appreciation of fundamental concepts in photovoltaics, including the pn-junction, cell efficiency, fill-factor, open-circuit voltage and short-circuit current; the text of Nelson[11] constituted a significant reference. All photovoltaic (PV) devices convert light into electrical work via some variation of the following basic mechanism

$$
\begin{aligned}
\text{photon} &\to \text{electron-hole pair} \\
&\to \text{charge separation (asymmetry)} \\
&\to \text{difference in CB/VB quasi-Fermi levels (open circuit voltage)} \\
&\to \text{external electrical work (with load connected)} \tag{1.1}
\end{aligned}
$$

Historically, inorganic semiconductors have been ideal media in which to realise such processes: as these materials permit the creation and separation of the electron and hole (the absence of an electron) and the technology of these materials is well developed. New PV technologies employing organic or solution-processed media or nano-technologies are currently under investigation. This thesis will focus upon new inorganic semiconductors.

When photons with sufficient energy are absorbed in materials electrons are excited to higher energy levels; in semiconductors electrons may be excited across the band gap which separates the valence and conduction bands (respectively VB and CB). Without some built-in material asymmetry, these electrons would relax back to the valence band within $\sim 1\,\mu s$; however, semiconductors may be prepared with asymmetries called pn-junctions during growth: which comprise high-quality crystallographic junctions between n- and p-type material. When such junctions are formed, majority carriers (electrons in n-type material and holes in p-type material) seeing few carriers in their band on the other side of the junction, diffuse across the junction where the carriers neutralise with each other and leave a depletion region: a region depleted of carriers. As a consequence of this majority carrier flow, the n-type region becomes positive and the p-type region becomes negative and a *built-in* electric field develops across the junction. Net diffusion of majority carriers ceases when the electric field is sufficient to counteract further diffusion, with the bands setup as Figure 2a.

In thermodynamic equilibrium in the dark, the electron and hole populations in the valence and conduction bands are described by the Fermi level E_{F}. When light falls upon a
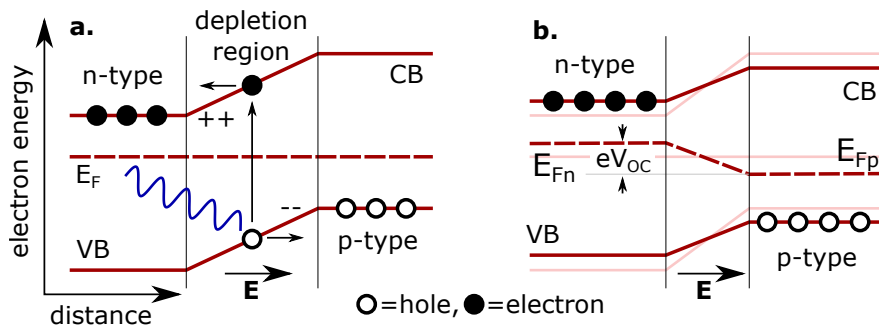
Figure 2: Electrons, holes and bands in the solar cell under illumination. (**a**) shows the bands around the junction at the instant when a photon is absorbed: the carriers are separated by the built-in field **E**; at some later time (**b**) shows the separation of the quasi-Fermi levels and the resultant open-circuit voltage V_{OC}.

pn-junction (see Figure 2a), electron-hole pairs generated in the depletion region are rapidly separated by the electric field before they have time to recombine: holes moving in one direction and electrons in the other in a *drift current*; pairs generated outside the depletion region will also contribute to this current if the minority carrier is within a diffusion length of the depletion region edge. Soon after, charge carriers start piling up at each end of the junction (see Figure 2b) and the new electron and hole populations are described by the quasi-Fermi levels $E_{F,n}$ and $E_{F,p}$. An open circuit voltage $V_{OC} = (E_{F,n} - E_{F,p})/e$ develops at the terminals of the cell[12,13]. Electrical work is then performed if a load is connected to the cell terminals.

With a cell illuminated by a photon flux $b(E)$, with photon energy $E = \hbar\omega$, a short-circuit current J_{SC} flows if the terminals are connected together

$$J_{SC} = e \int_0^\infty b(E)\, Q(E)\, \mathrm{d}E \tag{1.2}$$

where e is the electron charge and $Q(E)$ is the quantum efficiency: the probability that a photon of energy E will contribute an electron to the external circuit. The quantum efficiency depends upon material and device factors, such as absorption strength and the efficiency of charge separation and collection. As charge separation is achieved by the electric field of a pn-junction, the device is essentially an ideal diode with instantaneous current J given by

$$J_{dark} = J_0 \left(e^{eV/k_B T} - 1 \right) \tag{1.3}$$

where J_0 is the diode constant and J_{dark} is the usual diode current which would flow in the dark under some voltage bias V. Under the *superposition approximation* the instantaneous current under illumination J is assumed to be the difference of the short-circuit and dark currents, $J = J_{SC} - J_{dark}$; at open-circuit $J = 0$, which gives an open-circuit voltage V_{OC}

$$V_{OC} = \frac{k_B T}{e} \ln \left(\frac{J_{SC}}{J_0} + 1 \right) \tag{1.4}$$

The instantaneous power delivered by a cell is given by $P = JV$. Maximum power P_{max} is delivered at some particular operating point $\{J_m, V_m\}$. This is often written in terms of a quantity called the fill factor FF, which is frequently used in the PV literature

$$FF = \frac{J_m V_m}{J_{SC} V_{OC}} \tag{1.5}$$

The efficiency of power conversion η is defined with the incident optical intensity P_s as

$$\eta = \frac{\text{power at some operating point}}{\text{incident optical intensity}} = \frac{JV}{P_s} \tag{1.6}$$

so that the maximum efficiency $\eta_{max} = J_m V_m / P_s = FF\, J_{SC}\, V_{OC}/P_s$. See Figure 3a.

1.2.1 Radiant energy flux from the Sun and ambient

The Planck radiation function[14] (see equation (A.5) in Appendix A) gives the spectral energy density $\tilde{u}(E)$ for a blackbody in the energy interval from E to $E + dE$ at a temperature T

$$\tilde{u}(E) = \frac{\mathrm{d}u}{\mathrm{d}E}\, \mathrm{d}E = \frac{8\pi}{h^3 c^3} \cdot \frac{E^3}{e^{E/k_B T} - 1}\, \mathrm{d}E \qquad \left[\mathrm{J\,m^{-3}} \right] \tag{1.7}$$
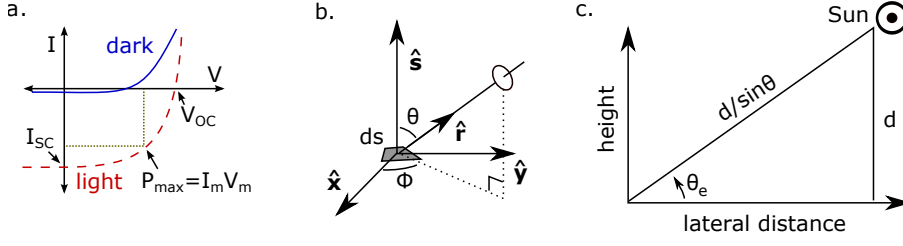
Figure 3: current-voltage relation, spherical polar coordinates and air mass. (**a**) shows the I-V relation, open circuit voltage V_{OC}, short circuit current I_{SC} and maximum power point $P_{\max}$ for a solar cell in the dark and light. (**b**) depicts the area element $ds = r^2 \sin\theta \, d\theta \, d\phi$. (**c**) the solar atmospheric path length is proportional to the cosecant of the elevation angle θ_e.

where c, E and h are respectively the speed of light, photon energy and Planck constant. For a blackbody of constant temperature, the photon number flux radiated from the surface per unit area and per unit solid angle is $\beta(E) = \frac{c}{4\pi E} \, \tilde{u}(E) \, dE$. The flux from a surface area element $\mathbf{ds} = ds \, \hat{\mathbf{s}}$ (with unit normal vector $\hat{\mathbf{s}}$) into a solid angle $d\Omega$ around the unit propagation vector $\hat{\mathbf{r}}$ is then

$$\beta(E, \mathbf{s}, \theta, \phi) = \frac{2}{h^3 c^2} \cdot \frac{E^2}{e^{E/k_B T} - 1} \, dE \, d\Omega(\theta, \phi) \, \mathbf{ds} \cdot \hat{\mathbf{r}} \qquad [\mathrm{s}^{-1}] \qquad (1.8)$$

Writing the solid angle $d\Omega$ in spherical polar coordinates $\sin(\theta) \, d\theta \, d\phi$ (see Figure 3b) and the dot product $\hat{\mathbf{s}} \cdot \hat{\mathbf{r}} = ds \, \cos(\theta)$, then integrating just the angular terms in (1.8) over a hemisphere

$$\int_0^{2\pi} d\phi \int_0^{\pi/2} d\theta \cdot \underbrace{\sin(\theta) \cos(\theta)}_{1/2 \sin(2\theta)} = \pi \qquad (1.9)$$

Thus the radiative (number) flux crossing a blackbody surface element ds is

$$b_{\mathrm{amb}}(E) = \frac{2\pi}{h^3 c^2} \cdot \frac{E^2}{e^{E/k_B T} - 1} \, dE \, ds \qquad [\mathrm{s}^{-1}] \qquad (1.10)$$

The radiative photon flux from the Sun at the Earth's location is reduced by the geometric factor $R_{\mathrm{Sun}}^2 / d_{\mathrm{Earth}}^2$, where R_{Sun} and d_{Earth} are respectively the radii of the Sun and that of the Earth's orbit. Thus the solar flux at the surface of a solar cell is

$$b_{\mathrm{Sun}}(E) = \frac{2\pi}{h^3 c^2} \left(\frac{R_{\mathrm{Sun}}}{d_{\mathrm{Earth}}} \right)^2 \frac{E^2}{e^{E/k_B T} - 1} \, dE \, ds \qquad [\mathrm{s}^{-1}] \qquad (1.11)$$

Multiplying (1.11) by the photon energy E and integrating over energy and unit solar cell area gives the power received from the Sun (per unit area)

$$P_{\mathrm{Sun}}(E) = \frac{2\pi}{h^3 c^2} \left(\frac{R_{\mathrm{Sun}}}{d_{\mathrm{Earth}}} \right)^2 \int_0^\infty \frac{E^3}{e^{E/k_B T} - 1} \, dE \qquad (1.12)$$

The integral on the right-side has the form of the Riemann zeta function $\zeta(s)$, defined as

$$\zeta(s) = \frac{1}{\Gamma(s)} \int_0^\infty \frac{x^{s-1}}{e^x - 1} \, dx \qquad (1.13)$$

where $\zeta(4)$ has the value $\pi^4/90$ and $\Gamma(s)$ is the gamma function, defined by $\Gamma(s) = (s-1)!$. Changing variables $x = E/kT$ and integrating gives the solar power flux

$$P_{\mathrm{Sun}}(E) = \sigma \left(\frac{R_{\mathrm{Sun}}}{d_{\mathrm{Earth}}} \right)^2 T^4 \qquad \text{with} \qquad \sigma = \frac{2\pi^5 k^4}{15 h^3 c^2} \qquad (1.14)$$

12
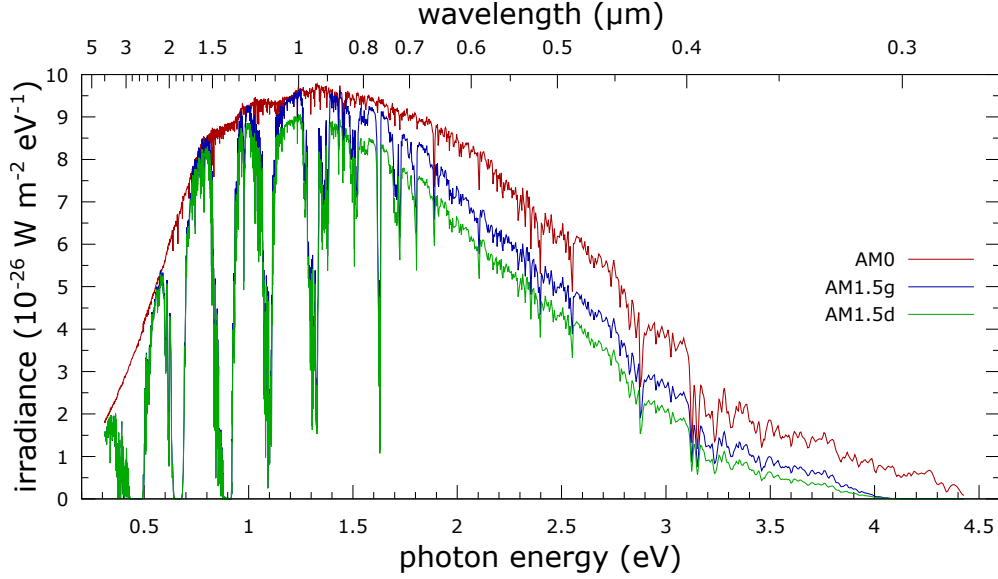
Figure 4: Standard solar AM0, AM1.5g and AM1.5d spectra (ASTM G-173-03 A): incorporating molecular extinction lines and used respectively for extra-terrestrial, unconcentrated and concentrated PV roles.

where σ is the Stefan–Boltzmann constant. Using a solar surface temperature of $5760\,\mathrm{K}$ and radii of $R_{\mathrm{Sun}} = 6.96 \times 10^8\,\mathrm{m}$ and $d_{\mathrm{Earth}} = 1.50 \times 10^{11}\,\mathrm{m}$ suggests a flux of $1.35\,\mathrm{kW/m^2}$ at the Earth's location[11]. Practically, this solar flux is diminished by absorption (from H_2O and CO_2 bands) and Rayleigh scattering (primarily from N_2 and O_2 molecules) in the Earth's atmosphere, see Figure 4.

The atmospheric path length depends on the Sun's angle of elevation θ_e above the horizon: with a minimum path when the Sun is directly overhead, see Figure 3c. The peak elevation angle changes throughout the year and with geographic location. To quantify the atmospheric path length, the *air mass* (AM) is defined as follows

$$AM = \frac{\text{atmospheric path length to Sun}}{\text{atmospheric path length when Sun directly overhead}} = \frac{1}{\sin\theta_e} = \operatorname{cosec}\theta_e \quad (1.15)$$

For comparative work the ASTM G-173-03 reference solar spectrum (see Figure 4) is used. This standard specifies two separate AM 1.5 spectra each with the Sun at an elevation angle of $41.8°$ (chosen for mid-latitude population centres). For flat terrestrial solar cells the AM 1.5g (global) spectrum has an integrated intensity normalised to $1\,\mathrm{kW/m^2}$ and incorporates both a direct solar flux and that due to diffuse scattering over a hemisphere; for concentrating PV applications, the AM 1.5d (direct & circumsolar) spectrum has an integrated intensity of $900\,\mathrm{W/m^2}$ and includes only radiation in a $\pm 2.5°$ cone centred on the sun. For space applications, an AM 0 reference spectrum (ASTM E-490) is defined with an integrated intensity of $1366\,\mathrm{W/m^2}$, which is similar to the value found with (1.14) for a blackbody spectrum.

Note in Figure 4 the growing impact of Rayleigh scattering with increasing photon energy and the significant valleys caused by the molecular absorption bands. For maximum performance, the absorption characteristics of any PV device must be matched to the appropriate spectrum. Additionally, note the significant energy contributions from the diffuse scattering of light: with a $\sim 15\,\%$ difference between AM 1.5g and AM 1.5d.

1.2.2 Detailed balance: radiative recombination, band gap and efficiency

Consider a material in thermal equilibrium with its surroundings, also considered as a black-body with temperature T_{amb}. The material absorbs some fraction of radiation from the surroundings according to its absorbance $a(E)$; in return, the material radiates an identical fraction given by its emittance, $e(E)$. The equality $a(E) = e(E)$ is the condition of *detailed balance*: the requirement that the matrix elements for absorption and emission are equal. Under illumination, the quasi-Fermi levels in the solar cell separate by a factor $\Delta\mu$, which (as a non-equilibrium condition) appears in (1.10) with a negative sign[11], giving a photon flux

$$b_{amb}(E, \Delta\mu, T) = \frac{2\pi}{h^3 c^2} \cdot \frac{E^2}{e^{\frac{E - \Delta\mu}{k_B T}} - 1} \, \mathrm{d}E \, \mathrm{d}s \qquad [\mathrm{s}^{-1}] \qquad (1.16)$$

So the net flux is the difference between the absorbed and emitted flux, using (1.10) and (1.16)

$$b_{net}(E, \Delta\mu) = b_{Sun}(E) + \left\{ 1 - \frac{R_{Sun}^2}{4 \, d_{Earth}^2} \right\} b_{amb}(E, 0, T_{amb}) - b_{amb}(E, \Delta\mu, T_{amb}) \qquad (1.17)$$

where the factor in braces is a geometric correction for the fraction of ambient flux which is replaced with solar flux (given the areas $4\pi d_{Earth}^2$ and πR_{Sun}^2). Equation (1.17) **requires a radiative recombination flux** (above thermal equilibrium) for any potential $\Delta\mu$. As cells may only absorb photons with energies greater than their band gap, this gives a current density

$$j(V) = e \int_{E_g}^{\infty} \left[1 - R(E) \right] b_{net}(E, eV) \, \mathrm{d}E \qquad (1.18)$$

where $\Delta\mu = eV$ (for a potential bias V), the term $\left[1 - R(E) \right]$ accounts for transmission of light through the front face, and an absorbance $a(E) = 1$ for $E \geq E_g$ else $a(E) = 0$ is assumed. For a fixed incident spectrum, equation (1.18) allows the calculation of the instantaneous power $P = jV$ as a function of potential V, with the maximum power generated at the point where $\mathrm{d}P/\mathrm{d}V = 0$. With these assumptions, the efficiency depends only on the ambient temperature, incident spectrum and semiconductor band gap E_g. Additionally, optics may be used to *concentrate* the light, increasing the angle subtended by the sun; under maximum concentration (4π steradians or $46\,400$ suns) the (approximate) factor in braces in (1.17) vanishes.

Various efficiency limits are seen[15,16]: 33 % for the AM 1.5g spectrum of Figure 4 and a cell with a $1.4\,\mathrm{eV}$ gap; 31 % for the extraterrestrial AM0 spectrum approximated as a $5760\,\mathrm{K}$ blackbody using a cell with a $1.3\,\mathrm{eV}$ gap at $300\,\mathrm{K}$. The absolute limit was found by Shockley and Queisser[17], who reported a 44 % limit for a cell with a $1.1\,\mathrm{eV}$ gap at $0\,\mathrm{K}$ (and so no radiative recombination) with light incident from a $6000\,\mathrm{K}$ blackbody under full concentration.
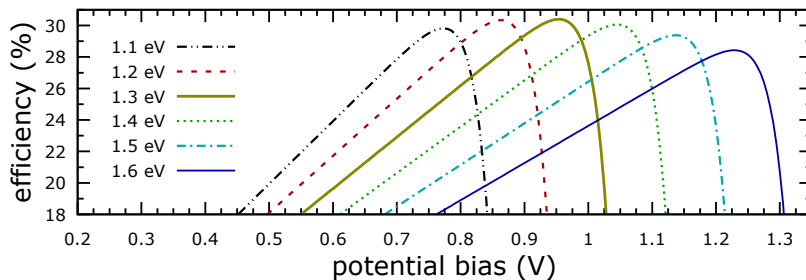
Figure 5: Efficiencies for cells with various gaps at 1 sun and $300\,\mathrm{K}$ using a $5760\,\mathrm{K}$ blackbody spectrum.
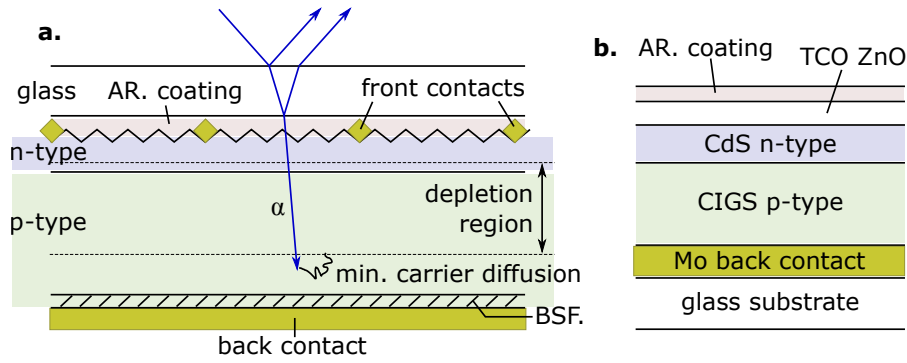
Figure 6: Typical solar cell layout and factors. (**a**) Shows a Si cell comprising electrical contacts, an anti-reflection (AR) layer and a pn-junction with back surface field (BSF). The n- upon p-type order is seen in Si cells, as the resistivity of n-type material is half that of p-type, while the latter has a longer minority carrier diffusion length[11]. (**b**) Shows a typical thin-film cell layout: with heterojunction and TCO front contact.

1.3 Solar cell layout and design factors

A typical silicon solar cell layout is seen in Figure 6a, which also demonstrates some important design factors for enhanced photovoltaic performance[18]. The design comprises a weather-proof glass front surface, front electrical contacts, an anti-reflection coating, a semi-conducting pn-homojunction comprising thin n-type and thicker p-type layers, with the later increasingly doped to introduce a back surface field (BSF). The front surface of the n-type layer has a textured geometry to further reduce reflectivity; a passivation treatment such as a native oxide (or even a separate semiconducting window layer) may also be used to reduce carrier recombination via front surface states. The back contact completes the cell.

Cell performance is enhanced significantly by minimising the fraction of light reflected from the front surface and any shadowing of the cell by the filamental front contacts[‡]. **Anti-reflection coatings are crucial**: the reflectivity of a thin gallium arsenide layer at the band edge is $\sim 40\%$. Figure 6b shows a typical copper indium gallium diselenide (CIGS) cell[11,19], in which the front contacts are replaced by a layer of transparent conducting oxide. A pn-heterojunction with a thin n-type window layer is necessary if satisfactory bipolar doping cannot be achieved within the absorbing material.

Ideal materials should be earth-abundant, non-toxic, dopable both n- and p-type for high-quality crystallographic homo-junctions, possess an optimal band gap for the incident spectrum and a strong absorption onset to trap light using as little material as possible; few midgap states should be present leading to negligible losses from non-radiative recombination. Carrier collection is improved with long minority carrier lifetimes and diffusion lengths. To support this, surfaces should either have low densities of recombination states or some surface passivation process should be available. Use of a back surface field (an increasing carrier concentration gradient towards the rear surface) may also aid minority carrier confinement: as this feature creates a built-in electric field in a similar manner to the pn-junction.

[‡] Improving contact shadowing is complex as the contacts already have small surface areas and carry large current densities; further reducing surface area increases current densities and ohmic losses through series resistance in adjacent n-type material[11]. Because resistivity goes as $\rho = 1/\sigma = \left(|e|\{\mu_n n + \mu_p p\}\right)^{-1}$, material is usually highly doped near the front contacts and high majority carrier mobilities are desirable; junctions may be set near the front surface due to enhanced recombination in heavily doped material.

1.4 Overview of thesis structure and composition

At this point it is beneficial to briefly summarise the contents of the following four chapters :

- **Experimental and theoretical methods**

  This chapter describes the fundamental principles of the experimental and theoretical methods instigated during the research period in some detail. These include Fourier-transform infrared spectroscopy (FTIR), spectroscopic ellipsometry (SE), x-ray diffraction (XRD), x-ray photoelectron spectroscopy (XPS), scanning electron microscopy (SEM), energy dispersive spectroscopy (EDS), and density functional theory (DFT).

  The chapter assimilates existing knowledge on each technique, presenting this in a form suitable for a new researcher. Additional material is featured which goes beyond what is found in the literature, e.g. the FTIR interferogram-to-spectrum conversion code, or the rather simple illustration of the ellipsometric angles Δ and Ψ for elliptically polarised and compensated light. The level of coverage provided is weighted towards the extent of importance or significance of each technique to the research work, e.g. optical and structural studies have particularly detailed coverage.

- **The determination of optical parameters in absorbing multilayers**

  This chapter starts by setting out the theory for various optical phenomena of importance to researchers in photovoltaics, then develops a toolbox for modelling optical propagation. While much of this work is an assimilation of existing knowledge (e.g. light and electromagnetism, dielectric phenomena, perturbation theory, interband absorption, excitons, and the transfer matrix method), nowhere are these foundational topics covered in a manner suitable for workers in photovoltaics. Shortcomings in various traditional approaches for determining optical parameters, i.e. absorption spectra, are then analysed and a simple self-consistent code is found which achieves improved accuracies. The section concludes with an analysis of some well-known materials.

- **Temperature-dependent structural and optical properties of copper nitride Cu_3N**

  Each materials chapter has a structure rather similar to a journal article: opening with an introduction and research objectives, then detailing experimental and theoretical methods, before describing results and closing with a discussion and conclusions.

  Highlights for Cu_3N include the finding of atypically small temperature dependence of the direct band gap, and the determination of credible thermal expansion via a quasi-harmonic model: which is then used to evaluate lattice contributions to the band gap.

- **Temperature-dependent optical properties of copper antimony sulphide $CuSbS_2$**

  Structured as the previous chapter, specific highlights for $CuSbS_2$ include the possible observation of an excitonic state at low temperature, the evaluation of band symmetries, and the determination of selection rules for dipole transitions.

Chapter 2

Experimental and theoretical methods

2.1 Fourier transform infrared spectroscopy

Fourier-transform infrared (FTIR) spectroscopy enables the efficient optical study of material properties via high resolution reflection and transmission measurements. The technique is invaluable in diverse areas including: analytical and physical chemistry[20,21], biology, materials science, geophysics and astrophysics, clinical research, industrial process control and environmental monitoring[22]. Further information is revealed by varying angle of incidence, polarisation or sample temperature. The key benefits of FTIR derive from use of a broadband source, which necessitates a two-beam (e.g. Michelson) interferometer and distinguishes the technique from dispersive (grating) spectroscopy; spectra are then obtained from the measured interferograms by a Fourier-transform. This section discusses the technique and various experimental aspects. The definitive reference is the tome of Griffiths and de Haseth[23].

An FTIR spectrometer is depicted schematically in Figure 7. The source **S** emits a continuous black body spectrum. The beamsplitter **B** (which could be a half-silvered mirror) transmits 50% of the light towards the movable mirror M_1, with the remainder reflected towards the fixed mirror **F**. When the moving mirror is at the equilibrium position M_0 the paths BM_0 and BF have equal length; with the mirror at M_1 the optical path BM_1 is longer than BF by the length $\delta/2$ giving a total path difference (or retardation) of δ.
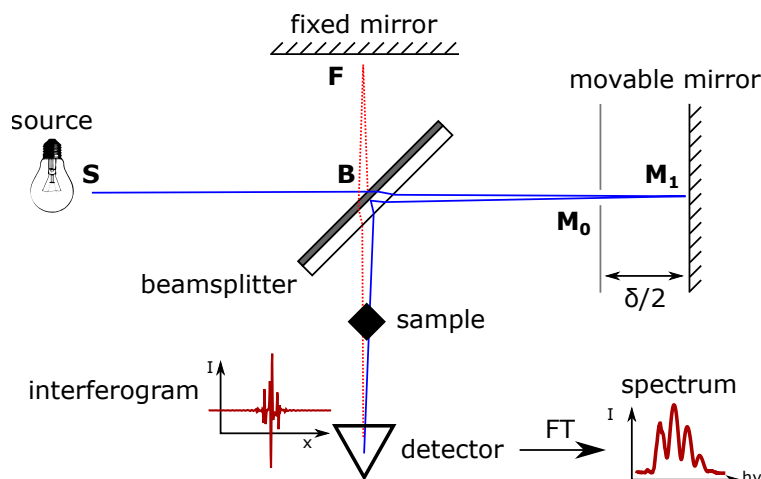
Figure 7: Principle of FTIR: a Fourier transform converts the interferogram into a spectrum. See the text.

When the beams from $\mathbf{F}$ and $\mathbf{M_1}$ recombine at the beamsplitter the path difference δ produces a phase difference $k\delta$, where $k = 2\pi/\lambda$ is the wavevector. For equal path lengths (zero phase difference) $100\,\%$ of the light is directed towards the sample where it is reflected, scattered, absorbed or transmitted before reaching the detector. For a phase difference of π, no light enters the sample compartment: all is reflected back towards the source. As the phase difference $k\delta$ changes with the motion of the mirror $\mathbf{M_1}$, an interference pattern (an interferogram) is generated at the detector: a plot of intensity as a function of mirror position. The final spectrum is the Fourier-transform of this interferogram.

When compared to dispersive (grating) spectrometers, FTIR instruments offer unmatched speed, accuracy and sensitivity[22]. Specific advantages are named after those who reported them. The *Connes*[24] advantage expresses the high wavenumber accuracy of FTIR instruments (exceeding $0.01\,\mathrm{cm^{-1}}$), which results from the use of helium-neon (HeNe) laser interferometry to position the moving mirror[25]. The *Jacquinot*[26] (or throughput) advantage emphasises the vastly greater flux admitted through the large circular apertures of FTIR instruments: more than 100 times that passed through the gratings of dispersive spectrometers[20]. The *Fellgett*[27] (or multiplex) advantage notes the rapid acquisition speed (above one scan/second[25]) which results because FTIR instruments simultaneously measure all frequencies. These latter advantages give FTIR spectrometers signal-to-noise ratios which exceed those of dispersive instruments by orders of magnitude[20]. Finally, FTIR spectrometers have better resolving power than dispersive instruments and offer unique spectral processing capabilities[21].

2.1.1 Fundamental principles: inteferograms and spectra

While FTIR instruments use broadband sources emitting a continuum of frequencies, intially let us consider a single monochromatic plane wave, which is incident on the interferometer in Figure 7. The electric field of the wave is

$$\mathbf{E}(\mathbf{x}, t; \mathbf{k}) = \mathbf{E_0}(\omega) \cos(\mathbf{k}\mathbf{x} - \omega t) \tag{2.1}$$

where $\mathbf{E_0}(\omega)$ and $\mathbf{k}$ denote the electric field amplitude and wavevector ($|\mathbf{k}| = 2\pi/\lambda = \omega/\mathrm{c}$). The beam splits evenly at the beamsplitter and after reflecting from the fixed and moving mirrors, the field after recombining and exiting the beamsplitter is

$$\begin{aligned}
\mathbf{E}(\mathbf{x}, t; \mathbf{k}, \delta) &= \mathbf{E_0}(\omega) \left[\cos(\mathbf{k}\mathbf{x} - \omega t) + \cos(\mathbf{k}\mathbf{x} - \omega t + k\delta)\right] \\
&= \mathbf{E_0}(\omega) \left[\cos\left(\mathbf{k}\mathbf{x} - \omega t + \frac{k\delta}{2} - \frac{k\delta}{2}\right) + \cos\left(\mathbf{k}\mathbf{x} - \omega t + \frac{k\delta}{2} + \frac{k\delta}{2}\right)\right] \\
&= 2\mathbf{E_0}(\omega) \cos\left(\mathbf{k}\mathbf{x} - \omega t + \frac{k\delta}{2}\right) \cos\left(\frac{k\delta}{2}\right)
\end{aligned} \tag{2.2}$$

where the relation $\cos(\alpha + \beta)$ is used in the last step. Recalling that the instantaneous power flux is proportional to the square of the electric field (from the Poynting vector $\mathbf{S} = \mathbf{E} \times \mathbf{H}$ and the relation $B = \mu H = E/c$; see Eqn. (3.16) on page 46). Squaring (2.2) gives an intensity

$I(\mathbf{x}, t; \mathbf{k}, \delta)$ at the detector (with no sample)

$$I(\mathbf{x}, t; \mathbf{k}, \delta) \propto 4\,\mathbf{E}_0^2(\omega) \cos^2\left(\mathbf{kx} - \omega t + \frac{k\delta}{2}\right) \cos^2\left(\frac{k\delta}{2}\right) \tag{2.3}$$

With the trigonometric relation $\cos^2 x = {}^1\!/\!2\big(\cos(2x) + 1\big)$, the time-average of $\cos^2(\omega t)$ is observed to be ${}^1\!/\!2$. Using this relation twice in (2.3) gives

$$\langle I(\tilde{\nu}, \delta)\rangle_t = \langle I(\mathbf{x}, t; \mathbf{k}, \delta)\rangle_t \propto \frac{4\,\mathbf{E}_0^2(\omega)}{2 \times 2}\big(1 + \cos(k\delta)\big) = \mathbf{E}_0^2(\omega)\big(1 + \cos(2\pi\tilde{\nu}\delta)\big) \tag{2.4}$$

where $\tilde{\nu}$ is the wavenumber ($\tilde{\nu} = {}^k\!/\!2\pi = {}^1\!/\!\lambda$). Practically, the squared amplitude $\mathbf{E}_0^2(\omega)$ of the source is better described by some function $B(\tilde{\nu})$ which includes the effects of beamsplitter efficiency and various other instrumental factors. Thus, for a continuum of frequencies incident from a broadband source, the detected intensity results by integrating (2.4) over wavenumber[23,28]

$$\langle I(\delta)\rangle_t = \int_0^\infty \langle I(\tilde{\nu}, \delta)\rangle_t \, \mathrm{d}\tilde{\nu} \propto \int_{-\infty}^\infty B(\tilde{\nu})\big(1 + \cos(2\pi\tilde{\nu}\delta)\big)\,\mathrm{d}\tilde{\nu} \tag{2.5}$$

where we define $B(\tilde{\nu}) = B(-\tilde{\nu})$. The limits for the integral in (2.5) have been chosen to allow application of a Fourier transform; this choice is permitted as the spectral amplitude is fully described by $B(\tilde{\nu})$. The right-hand side of (2.5) contains a constant term and, significantly, a cosine term which is a function of δ: the interferogram, $I_{\mathrm{f}}(\delta)$

$$\langle I(\delta)\rangle_t \propto \underbrace{\int_{-\infty}^\infty B(\tilde{\nu})\,\mathrm{d}\tilde{\nu}}_{\text{constant term}} + \underbrace{\int_{-\infty}^\infty B(\tilde{\nu})\cos(2\pi\tilde{\nu}\delta)\,\mathrm{d}\tilde{\nu}}_{\text{interferogram}} = I_{\mathrm{const}} + I_{\mathrm{f}}(\delta) \tag{2.6}$$

The interferogram has the form of a cosine Fourier-transform, i.e. a cosine basis with only real coefficients. Inversion of the cosine transform gives the incident spectral amplitude $B(\tilde{\nu})$

$$B(\tilde{\nu}) = \int_{-\infty}^\infty I_{\mathrm{f}}(\delta)\cos(2\pi\tilde{\nu}\delta)\,\mathrm{d}\delta = 2\int_0^\infty I_{\mathrm{f}}(\delta)\cos(2\pi\tilde{\nu}\delta)\,\mathrm{d}\delta \tag{2.7}$$

where in the right-hand side the symmetry of the cosine is used. Remarkably, equation (2.7) reveals that the incident spectrum $B(\tilde{\nu})$ can be recovered with complete accuracy if the interferogram is measured to infinite retardation δ. Practically, the extremes of retardation are constrained within some limits (e.g. $-\Delta \le \delta \le \Delta$), and so experimentally we determine

$$B'(\tilde{\nu}) = \int_{-\Delta}^\Delta I_{\mathrm{f}}(\delta)\cos(2\pi\tilde{\nu}\delta)\,\mathrm{d}\delta \tag{2.8}$$

Now, comparing (2.7) and (2.8), the restriction on the limits means that the incident spectrum is not determined to complete accuracy and $B'(\tilde{\nu}) \ne B(\tilde{\nu})$; however, because the magnitude of $I_{\mathrm{f}}(\delta)$ falls sharply with increasing $|\delta|$, we'll see that (2.8), which we'll call the computed spectrum, is a good approximation to the actual spectrum (2.7).

2.1.2 Instrumental line-shape, resolution and apodisation

This section shows how the computed spectrum $B'(\tilde{\nu})$ differs from the actual spectrum $B(\tilde{\nu})$. In discussing this, the important concepts of resolution and apodisation are explained. First,

rewrite (2.8) in a completely equivalently form

$$B'(\tilde{\nu}) = \int_{-\infty}^{\infty} I_{\mathrm{f}}(\delta)D(\delta)\cos(2\pi\tilde{\nu}\delta)\,\mathrm{d}\delta\,, \quad \text{with } D(\delta) = \begin{cases} 0, & |\delta| > \Delta \\ 1, & -\Delta \le \delta \le \Delta \end{cases} \quad (2.9)$$

where with this assignment $D(\delta)$ is called a boxcar function. Now, it is a standard result in Fourier analysis that the Fourier transform of the product of two functions $\mathcal{F}(gh)$ (where $\mathcal{F}(\cdot)$ denotes a Fourier transform) is the convolution of the Fourier transform of each function $\mathcal{F}(g) * \mathcal{F}(h)$. The convolution operation $(*)$ on the functions $g(x)$ and $h(x)$ is defined as

$$f(u) = g(x) * h(x) \equiv \int_{-\infty}^{\infty} g(x)h(u-x)\,\mathrm{d}x = \int_{-\infty}^{\infty} g(u-x)h(x)\,\mathrm{d}x \quad (2.10)$$

so that the computed spectrum $B'(\tilde{\nu})$ in (2.9) may be written as the convolution

$$B'(\tilde{\nu}) = \mathcal{F}\Big(I_{\mathrm{f}}(\delta)D(\delta)\Big) = \underbrace{\mathcal{F}\Big(I_{\mathrm{f}}(\delta)\Big)}_{B(\tilde{u})} * \underbrace{\mathcal{F}\Big(D(\delta)\Big)}_{f(\tilde{u})} = \int_{-\infty}^{\infty} B(\tilde{u})f(\tilde{\nu}-\tilde{u})\,\mathrm{d}\tilde{u} \quad (2.11)$$

As (2.7) showed that the Fourier transform of the interferogram $I_{\mathrm{f}}(\delta)$ is the actual spectrum $B(\tilde{u})$, equation (2.11) reveals that the computed spectrum $B'(\tilde{\nu})$ is the convolution of the actual spectrum $B(\tilde{u})$ and a function $f(\tilde{u})$, known as the instrument lineshape function (ILS). For the ILS, the Fourier transform of the boxcar function $D(\delta)$ as defined in (2.9) gives

$$f(\tilde{\nu}) = \int_{-\infty}^{\infty} D(\delta)\cos(2\pi\tilde{\nu}\delta)\,\mathrm{d}\delta = \int_{-\Delta}^{\Delta}\cos(2\pi\tilde{\nu}\delta)\,\mathrm{d}\delta = \left[\frac{\sin(2\pi\tilde{\nu}\delta)}{2\pi\tilde{\nu}}\right]_{-\Delta}^{\Delta} = 2\Delta\,\mathrm{sinc}(2\pi\tilde{\nu}\Delta)$$
$$(2.12)$$

With the ILS, the computed spectra $B'(\tilde{u})$ may be assessed for different source spectra $B(\tilde{u})$. For a monochromatic source at wavenumber $\tilde{\nu}_0$, set $B(\tilde{\nu}) = B_0 \cdot \delta(\tilde{\nu} - \tilde{\nu}_0)$, where $\delta(\cdot)$ is the Dirac delta function. Plugging $B(\tilde{\nu})$ and (2.12) into (2.11) then gives

$$B'(\tilde{\nu}) = \int_{-\infty}^{\infty} B_0\,\delta(\tilde{u}-\tilde{\nu}_0)\cdot 2\Delta\,\mathrm{sinc}\big(2\pi\Delta\{\tilde{\nu}-\tilde{u}\}\big)\,\mathrm{d}\tilde{u} = 2B_0\Delta\,\mathrm{sinc}\big(2\pi\Delta\{\tilde{\nu}-\tilde{\nu}_0\}\big) \quad (2.13)$$

Figure 8 shows the computed spectrum $B'(\tilde{\nu})$. The delta function, which has infinitesimal width in the actual spectrum $B(\tilde{\nu})$, broadens into a peak in the computed spectrum with a baseline zero-to-zero width of $1/\Delta$. Notice also the diminishing sidelobe series.
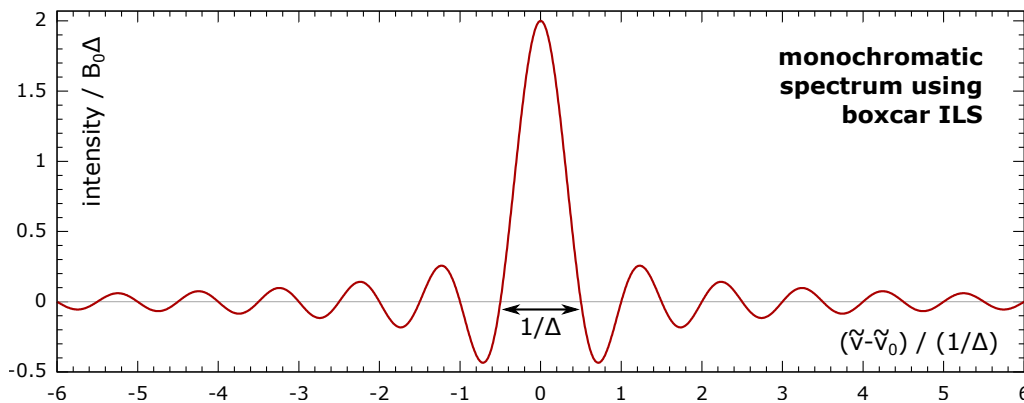
Figure 8: Monochromatic spectrum with boxcar ILS. The retardation limit Δ sets the zero-to-zero width.

20
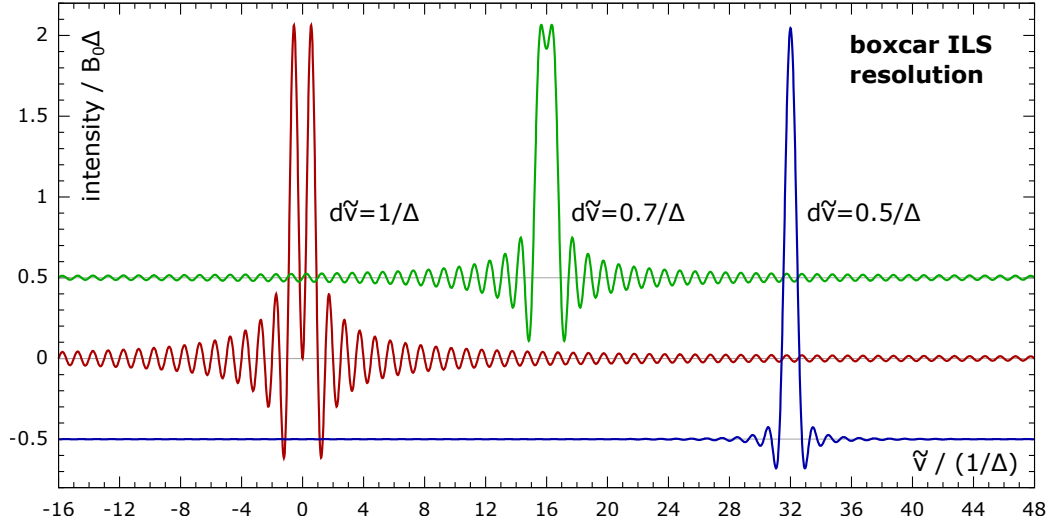
Figure 9: Resolution with boxcar lineshape. Two lines are well-resolved when separated by $1/\Delta$.

Figure 8 lends support to the claim made after equation (2.8) that the computed spectrum is a fair approximation to the real spectrum: the delta function in $B(\tilde{\nu})$ produced an unambiguous principal line in $B'(\tilde{\nu})$. For polychromatic spectra, $B(\tilde{\nu})$ may be decomposed into a sum of delta functions. Due to the sum rule (or linearity) of integration, each component in $B(\tilde{\nu})$ results in a term in the computed spectrum as equation (2.13). In particular, consider pairs of closely spaced lines; the resolution limit is determined by the zero-to-zero width of the principal peak in Figure 8. With the first zero at $\tilde{\nu} = 1/2\Delta$, two lines are evidently resolved to baseline if they are separated by $1/\Delta$. This problem is similar to that addressed by Rayleigh's criterion, where two point sources are considered resolved if the diffraction maximum of one is no closer than the first minimum of the other.

The envelope over doublets separated by $1/\Delta$, $0.7/\Delta$ and $0.5/\Delta$ is illustrated in Figure 9. Differentiating (2.13) finds the first minimum at $0.715/\Delta$ and the lines are barely resolved at such a spacing; lines separated by $0.5/\Delta$ are no longer resolved (where the maximum of one is set at the first zero of the other). Because the lines are clearly resolved when separated by $1/\Delta$, this is taken as the nominal resolution for FTIR spectroscopy. The first minima in Figure 8 has a magnitude which is 22 % of the principal. This makes resolving weak lines difficult when they are adjacent to stronger lines, even when at the nominal resolution. To weaken the sidelobes, the boxcar function is replaced with an apodisation function which weights the interferogram. For example, with a triangular apodisation function

$$D(\delta) = \begin{cases} 0, & |\delta| > \Delta \\ 1 - |\delta|/\Delta, & -\Delta \leq \delta \leq \Delta \end{cases} \qquad (2.14)$$

The triangular instrumental lineshape function results from the Fourier transform of (2.14)

$$f(\tilde{\nu}) = \int_{-\infty}^{\infty} D(\delta) \cos(2\pi\tilde{\nu}\delta)\, \mathrm{d}\delta = 2\Delta \operatorname{sinc}(2\pi\tilde{\nu}\Delta) - \frac{2}{\Delta} \int_{0}^{\Delta} \delta \cos(2\pi\tilde{\nu}\delta)\, \mathrm{d}\delta = \Delta \operatorname{sinc}^2(\pi\tilde{\nu}\Delta)$$

$$(2.15)$$

The computed spectrum which results from this lineshape is presented in Figure 10, together with the boxcar lineshape for comparison. With a triangular apodisation, the principal
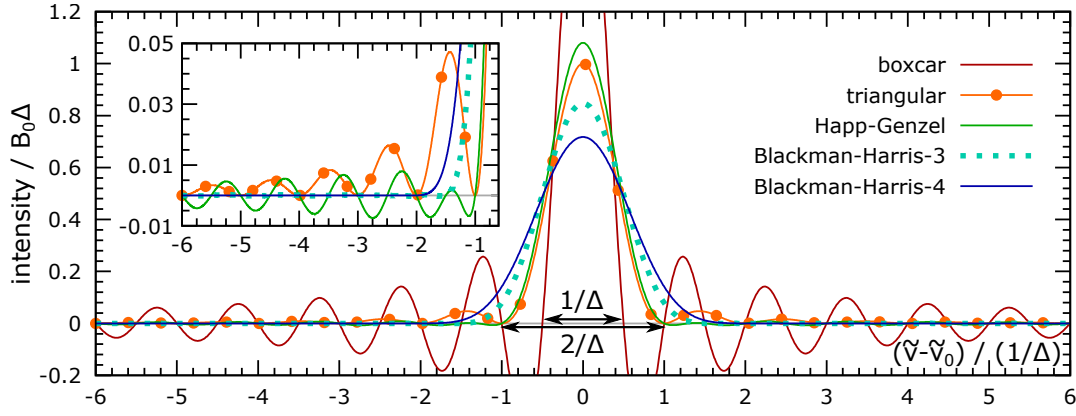
Figure 10: Various ILS functions with increasing sidelobe reduction: note the line width broadening.

maximum and zero-to-zero baseline width are respectively half and twice those of the boxcar lineshape. Although the sidelobe series is greatly reduced, the line width is significantly broader. Generally, apodisation function selection is a compromise between decreasing either linewidth or sidelobe amplitude; while the former improves resolution, the latter improves contrast between weak and strong lines. The Happ–Genzel function (seen in Figure 10) has similar resolution to a triangular ILS but with improved sidelobe reduction; for general work the Blackman–Harris 3-term apodisation function (seen in Figure 10 and used in this thesis) strikes an acceptable balance*. For particularly high-resolution work, the boxcar and Norton–Beer (weak) apodisation functions are favoured[31]. Because apodisation may be performed at any time post-measurement (as long as the inteferogram is available), such considerations need not cause delays at experiment time.

2.1.3 Resolution and signal-to-noise ratio

The signal to noise ratio (SNR) of a FTIR spectrometer is a function of the incident irradiance $\tilde{B}(\nu)$ (the power per unit area per unit wavenumber), the detector area A_D, the resolution $\Delta\tilde{\nu}$ and the detector efficiency $\eta(\tilde{\nu})$. The noise power N (inherent due to the amplifier electronics of the detector) and the measurement time t then lead to a signal-to-noise ratio[23]

$$\frac{S}{N} = \frac{\tilde{B}(\tilde{\nu})\, A_D\, \Delta\tilde{\nu}\, \eta(\tilde{\nu})\, \sqrt{t}}{N} \tag{2.18}$$

Thus achieving the same SNR after improving the resolution by a factor of two (halving $\Delta\tilde{\nu}$) requires a four-fold increase in measurement time.

---

*The Happ–Genzel and Blackman–Harris apodisation functions have the following form with $N + 1$ terms

$$D(\delta) = a_0 + \sum_{m=1}^{N} a_m \cos\left(m\pi\frac{\delta}{\Delta}\right) \quad \text{for} \quad |\delta| \leq \Delta, \quad \text{else} \quad D(\delta) = 0. \tag{2.16}$$

The coefficients for the Happ–Genzel and Blackman–Harris apodisation functions are as follows[21,29,30]

Apodisation function	a_0	a_1	a_2	a_3
Happ–Genzel	0.54	0.46	0	0
Blackman–Harris 3-term	0.42323	0.49755	0.07922	0
Blackman–Harris, 4-term	0.35875	0.48829	0.14128	0.01168

Beware of typos, e.g. the Blackman–Harris function in Griffiths[23]. Fourier transformation of (2.16) gives a lineshape

$$f(\tilde{\nu}) = a_0 \cdot 2\Delta\mathrm{sinc}(2\pi\Delta\tilde{\nu}) + \sum_{m=1}^{N} a_m\, (-1)^{m+1}\, \frac{4\tilde{\nu}\Delta^2 \sin(2\pi\Delta\tilde{\nu})}{\pi\,(m^2 - 4\Delta^2\tilde{\nu}^2)} \tag{2.17}$$

2.1.4 Interferogram sampling and phase correction

The Nyquist sampling criterion states that sinusoidal signals must be sampled at at least twice their frequency to be unambiguously determined[23]; FTIR spectrometers use interferometry of the HeNe laser extrema (and zero-crossings) to sample the interferogram accurately. The sampling of the inteferogram is thus limited by the wavelength of the HeNe laser.

Earlier in (2.7) the inteferogram was suggested to be a symmetric function with retardation either side of the centreburst. However, practically, various optical, electronic and sampling factors produce *chirped* or asymmetric interferograms; such effects might arise if the centre-burst fails to coincide with a sampling point, so that no point exists where all wavelengths are in phase. This is equivalent to a retardation transform $\delta \rightarrow \delta - \delta_0$ (where δ_0 is the actual position of the centre-burst), which introduces $\sin(2\pi\tilde{\nu}\delta_0)$ terms, most easily represented in (2.7) as a complex Fourier transform. The complex forward and inverse Fourier transforms and their corresponding discrete transforms (for N data samples) are

$$F(k) = \int_{-\infty}^{\infty} f(x)\,\mathrm{e}^{-i2\pi kx}\,\mathrm{d}x \quad \longrightarrow \quad F_n = \sum_{k=0}^{N-1} f_k\,\mathrm{e}^{-i2\pi nk/N} \tag{2.19}$$

$$f(x) = \int_{-\infty}^{\infty} F(k)\,\mathrm{e}^{i2\pi kx}\,\mathrm{d}k \quad \longrightarrow \quad f_k = \sum_{n=0}^{N-1} F_n\,\mathrm{e}^{i2\pi nk/N} \tag{2.20}$$

To correct such asymmetries, a phase correction procedure for double-sided interferograms (sampled on each side of the centreburst) due to Mertz[32] is used. The first step is to replace the cosine Fourier transform in (2.7) with the full complex transform

$$\tilde{B}(\tilde{\nu}) = \int_{-\infty}^{\infty} I(\delta)\,\mathrm{e}^{-i2\pi\delta\tilde{\nu}}\,\mathrm{d}\delta \tag{2.21}$$

The complex Fourier transform gives a complex spectrum $\tilde{B}(\tilde{\nu})$. If the complex spectrum $\tilde{B}(\tilde{\nu})$ differs by only a phase factor, then the modulus $|\tilde{B}(\tilde{\nu})|$ is the symmetric interferogram

$$\left|\tilde{B}(\tilde{\nu})\right| = \sqrt{\mathrm{Re}^2\{\tilde{B}(\tilde{\nu})\} + \mathrm{Im}^2\{\tilde{B}(\tilde{\nu})\}} = B(\tilde{\nu}) \tag{2.22}$$

Although the phase has been removed in (2.22), this relation is not a reliable correction due to noise vulnerabilities. Instead, the symmetric inteferogram may be written

$$\tilde{B}(\tilde{\nu}) = B(\tilde{\nu}) \cdot \mathrm{e}^{i\theta(\tilde{\nu})} \qquad \therefore \quad B(\tilde{\nu}) = \tilde{B}(\tilde{\nu}) \cdot \mathrm{e}^{-i\theta(\tilde{\nu})} \tag{2.23}$$

where $\theta(\tilde{\nu})$ is the phase angle of the complex number $\tilde{B}(\tilde{\nu})$, written as

$$\theta(\tilde{\nu}) = \tan^{-1} \frac{\mathrm{Im}\,\tilde{B}(\tilde{\nu})}{\mathrm{Re}\,\tilde{B}(\tilde{\nu})} \tag{2.24}$$

The second relation of (2.23) is pure real. Taking the real part of the right-hand side gives

$$B(\tilde{\nu}) = \left\{\mathrm{Re}\,\tilde{B}(\tilde{\nu})\right\}\cos\left\{\theta(\tilde{\nu})\right\} + \left\{\mathrm{Im}\,\tilde{B}(\tilde{\nu})\right\}\sin\left\{\theta(\tilde{\nu})\right\} \tag{2.25}$$

which is the Mertz phase correction. As $\theta(\tilde{\nu})$ is often slowly varying, in single-sided measurements, a small retardation $\delta < 0$ may be used and then interpolated to full resolution.

2.1.5 Synthesis: conversion of interferogram to spectrum

While operational details for FTIR spectroscopy are covered in great detail in many places, explicit descriptions of the conversion from interferogram to spectrum are somewhat hard to find. Whilst such an implementation is seemingly a trivial task, various aspects are easily overlooked and a concrete example proves instructive as a synthesis of the main ideas. A minimal MATLAB code (preferred here for its native functionality and concise implementation) is given below which converts a real double-sided (forward-only) interferogram to a spectrum which was consistent with that produced by the instrument.

```matlab
% setup required parameters, load double-sided interferogram
iCentreBurstRow=3811; % row one is first matrix row in Matlab (1-based)
fMaxWavenum=15799.17;
fResolutionPerCm=8;
mxIfgram=dlmread('interferogram.txt','\t',0,0); % columns: wavenum, intensity.

% remove any baseline and perform triangular apodisation
cvIfgram=mxIfgram(:,2)-mean(mxIfgram(:,2));
cvTriangularRamp=[ (0:1:(iCentreBurstRow-2))/(iCentreBurstRow-1), 1,( size(cvIfgram,1)- ...
    (iCentreBurstRow+1:1:size(cvIfgram,1)) )/(size(cvIfgram,1)-iCentreBurstRow) ]';
cvIfgram=cvIfgram.*cvTriangularRamp;

% rotate centreburst to zero retardation and pad with zeros to next power of two
iNumRowsToNextPowerOf2=2^(floor(log(size(cvIfgram,1))/log(2))+1);
cvIfgram=[cvIfgram(iCentreBurstRow:end,1); ...
    zeros(iNumRowsToNextPowerOf2-size(cvIfgram,1),1); cvIfgram(1:iCentreBurstRow-1,1)];

% perform Fourier transform and Mertz phase correction
cvSpectrum=fft(cvIfgram);
theta=atan2(imag(cvSpectrum),real(cvSpectrum)); % note: take care with inverse tangent
cvSpectrum=real(cvSpectrum).*cos(theta)+imag(cvSpectrum).*sin(theta);
cvSpectrum=cvSpectrum/fResolutionPerCm; % fix intensity: divide by resolution (see eqn X.XX)

% compute abscissa scale in wavenumber/cm. Note: the second-half of a Fourier transform
% is a mirror image of first-half, so we only need (iNumRowsToNextPowerOf2/2) rows
cvWavenums=( (0:1:(iNumRowsToNextPowerOf2/2-1))*fMaxWavenum/(iNumRowsToNextPowerOf2/2) )';
mxSpectrum=[cvWavenums,cvSpectrum(1:iNumRowsToNextPowerOf2/2,1)];

plot(mxSpectrum(:,1),mxSpectrum(:,2));
```

In the first lines the centre-burst position, resolution and maximum wavenumber are set explicitly and the inteferogram is loaded from the file `interferogram.txt` (with wavenumber and intensity columns) as exported from a Bruker VERTEX 70v spectrometer. Single measurements may comprise multiple scans, which are averaged before the interferogram is stored. While this code requires that the source file contains a single interferogram, FTIR instruments may measure inteferograms on forward and backward mirror movements and these may then be concatenated within the exported interferogram file: if so they should be phase-corrected separately. The remainder of the code is reasonably self-explanatory.

The measurement of a double-sided interferogram begins with the mirror set at maximum retardation on one side of the centreburst. Hence before applying the Fourier-transform, the interferogram must be rotated so that the centreburst is set at zero retardation. The interferogram is zero-padded to the first power of two; this gives a wavenumber separation equal to the resolution. Typically, zero-padding would proceed to the next power of two, which would give a wavenumber separation of half the resolution and allow observation of the dip in intensity between two closely spaced lines at the resolution limit[33]. The entire interferogram is used here for phase correction, yet smaller samples are often used.

2.1.6 Reduction to reflectivity and transmissivity

Upon acquisition, raw spectra describe the detector response to some unknown incident light flux. For meaningful spectra, these must be reduced to describe *only* the sample and to be free from instrumental artefacts. Ultimately, reflectivity $R(\tilde{\nu})$ and transmissivity $T(\tilde{\nu})$ spectra are sought for each sample, defined as the respective ratios of reflected and transmitted power. For the transmission beam path, transmissivity is determined as follows

$$T\left(\tilde{\nu}\right) = \frac{\text{power transmitted at } h\nu}{\text{power incident at } h\nu} = \frac{B'_{\text{sample}}\left(\tilde{\nu}\right)}{B'_{\text{aperture}}\left(\tilde{\nu}\right)} \tag{2.26}$$

while for the reflection beam path, reflectivity is determined as

$$R\left(\tilde{\nu}\right) = \frac{\text{power reflected at } h\nu}{\text{power incident at } h\nu} = \frac{B'_{\text{sample}}\left(\tilde{\nu}\right) - B'_{\text{aperture}}\left(\tilde{\nu}\right)}{B'_{\text{mirror}}\left(\tilde{\nu}\right) - B'_{\text{aperture}}\left(\tilde{\nu}\right)} \cdot R_{\text{mirror}}\left(\tilde{\nu}\right) \tag{2.27}$$

where B'_{sample}, B'_{aperture} and B'_{mirror} are respectively the computed spectra of the sample, the empty aperture and a gold mirror. In reflection, the empty aperture signal B'_{aperture} is usually negligible but must be assessed experimentally. The gold mirror is set identically in the sample position to act as a perfect reflection reference; because no mirror is a perfect reflector, a correction may be necessary to account for the actual reflectivity of the mirror $R_{\text{mirror}}(\tilde{\nu})$.

To determine $R_{\text{mirror}}(\tilde{\nu})$, a tabulated refractive index is used with the Fresnel coefficients: see equations (3.13), (3.15) and (3.17) in Section 3.3. The reflectivity derived from gold refractive index of Rakić et al.[34] is seen in Figure 11. The significance of the correction depends on the investigated regime: below $1.9\,\text{eV}$ the reflectivity is sufficiently good that no correction may be needed; above $1.9\,\text{eV}$ the reflectivity diminishes rapidly, and would introduce substantial errors if neglected, i.e. overestimated sample reflectivities.

Neglecting inelastic photon interactions, which convert high energy photons to low energy photons (e.g. non-linear crystals or photon down-conversion), a maximum value of unity is expected for the sum $R(\tilde{\nu}) + T(\tilde{\nu})$. Violations of this then suggest the presence of some instrumental issue, e.g. non-linear detector response, variability of source, inconsistent aperture position, etc. Such aspects are discussed in Section 3.12.
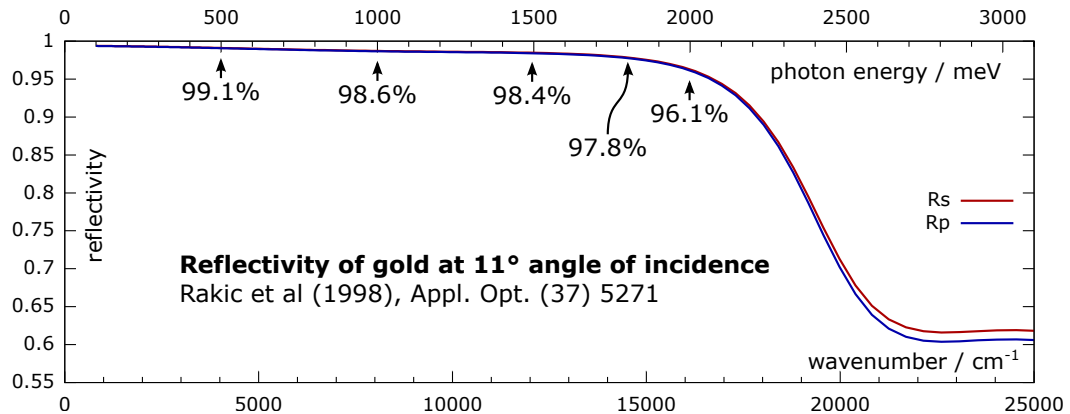
Figure 11: Reflectivity of gold via refractive index of Rakić et al.[34]. Note the drop above $1.9\,\text{eV}$.

2.1.7 Experimental configuration specifics

Figure 12 shows the configuration of the Bruker Vertex 70v used in this work[35]. Light is generated at one of two air-cooled sources which cover the near-infrared (NIR) and mid-infrared (MIR) spectral ranges: an ohmic-heated silicon-carbide *Globar* MIR source (350-$8000\,cm^{-1}$) (typically running at $\sim 1300\,K$) and a tungsten wire filament NIR source (4000-$15\,500\,cm^{-1}$). A moving mirror selects the required source. Immediately after the source sit optical aperture and filter wheels; at the interferometer, quartz or calcium fluoride beamsplitters are locked into position according to the required experimental spectral ranges. All experiments were conducted under a $2\,mbar$ vacuum (to minimise atmospheric absorption lines) using a combined reflection/transmission accessory with an $11°$ angle of incidence. This accessory utilised a moving mirror to acquire R and T spectra at experiment time without disturbing the sample. For work with liquid-helium, the sample was secured to the cold finger of a continuous-flow cryostat; the pressure-casing of the cryostat fitted within this accessory.

Both quantum and pyroelectric detectors were used in the work. The former type uses an absorber comprising a photodiode (a reverse-biased pn-junction) in photoconductive mode: giving a current proportional to irradiance (for photon energies above the semiconductor band gap). A silicon detector was used for work above $1.1\,eV$, while a liquid-nitrogen cooled, mercury cadmium telluride ($Hg_{1-x}Cd_xTe$ or MCT) detector provided coverage from $50\,meV$ through $1.5\,eV$. Because these compounds have different band gaps (CdTe has a $1.5\,eV$ gap, while HgTe is a semimetal with no bandgap) but the same space group and similar lattice parameters, any band gap $E_g < 1.5\,eV$ may be selected in an alloy of these materials.

Although MCT detectors have great sensitivity, these detectors also have a non-linear response: the detector response weakens with increasing intensity. If photometric accuracy is desired, e.g. to determine absorption spectra, then software non-linearity correction and either small apertures or neutral density filters should be used if feasible[23]. A deuterated (l-analine doped) triglycine sulfate (DLATGS) pyroelectric detector was also used for work above $30\,meV$. This detector-type comprises a ferroelectric light-absorbing material with a large, temperature-induced dielectric constant change; when used in capacitive circuits a strong voltage signal is determined with a linear response as a function of intensity[25].
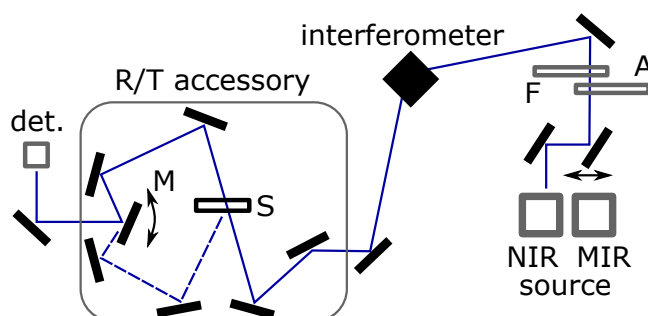
Figure 12: FTIR instrument configuration. Either the NIR or MIR source is used, as selected by the moving mirror. Light passes through aperture (A) and filter (F) wheels before traversing the interferometer. Once inside the sample compartment, a moving mirror (M) selects the transmission (solid line) or reflection (dashed) beampath. The sample is positioned at (S) via either a cryostat or standard accessory.
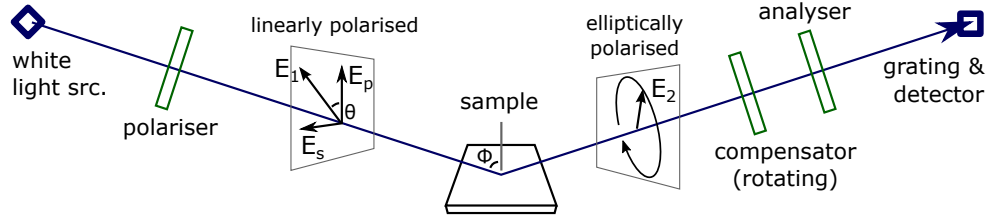
Figure 13: Schematic configuration for spectroscopic ellipsometry. Either the compensator or analyser may rotate. Wavelengths are separated by a grating at the detector. The compensator may be omitted.

2.2 Spectroscopic ellipsometry

Ellipsometry is a characterisation technique which **determines optical properties via changes in the polarisation state of light on reflection**[36,37]. Spectroscopic ellipsometry uses a broadband (white) light source rather than a monochromatic source. Typically ellipsometry is used to study surfaces, interfaces and thin films, one advantage over transmission spectroscopy being that no reference spectrum is required: results arise fully from geometrical changes in the ratio of the polarisations[36,37]; this may give ellipsometry certain advantages for rough samples. Because only reflectivity is used, results may be obtained for optically-thick media which would be unsuitable for transmission spectroscopy (see Section 3.14.3); however, if the surface reconstructs, has oxide or impurity layers or differs significantly from the bulk, the inherent surface sensitivity may bring difficulties not seen in transmission spectroscopy.

In principle, ellipsometry requires a collimated, polarised beam to be directed at a sample at oblique incidence (which serves to enhance differences between the p- and s-polarisations). Upon reflection, **the beam is generally elliptically polarised**[38]; subsequent analysis of the polarisation state then determines the reflection ratio ρ

$$\rho = \frac{r_p}{r_s} = \mathrm{e}^{i\Delta} \tan \Psi \qquad \text{with} \qquad \begin{array}{c} -180 \leq \Delta < 180° \\ 0 \leq \Psi < 90° \end{array} \qquad (2.28)$$

where r_p and r_s are the Fresnel reflection coefficients for p- and s-polarised light (see Section 3.4), while the modulus and argument of the complex number ρ are by convention[36] written as $\tan \Psi = |r_p|/|r_s|$ and $\Delta = \arg r_p - \arg r_s$. The angles Ψ and Δ are depicted in Figure 14; while Ψ is fully determined in all ellipsometers, in less capable instruments only the modulus of the angle Δ is determined.

A typical configuration for spectroscopic ellipsometry is shown in Figure 13. White light from the source is first linearly polarised (at azimuth $\theta \sim 45°$) before reflecting from the sample (at angle ϕ), where the (now elliptically polarised) light passes through a compensator and the final analyser. Either the compensator or analyser rotates to determine intensities in the plane perpendicular to the propagation direction (see again Figure 14). Wavelengths are finally separated at a grating just before the CCD detector. Ellipsometric instrumental configurations are often described[39] with the letters P, S, C and A, which respectively refer to sequential arrangements of polariser, sample, compensator and analyser (with a subscript R for rotating components), e.g. PSA$_\mathrm{R}$ describes a simple ellipsometer with a rotating analyser and no compensator: such a configuration determines only the modulus of Δ.

(a)

$\sqrt{R_s}=|r_s|$

t

a

b

Δ

Ψ

$\sqrt{R_p}=|r_p|$

(b)

$\Delta=\pm\pi/2$
$\Delta=\pm\pi/3$
$\Delta=\pm\pi/6$
$\Delta=0$

$\Delta=\pm\pi/2$
$\Delta=\pm 2\pi/3$
$\Delta=\pm 5\pi/6$
$\Delta=\pm\pi$

(c)
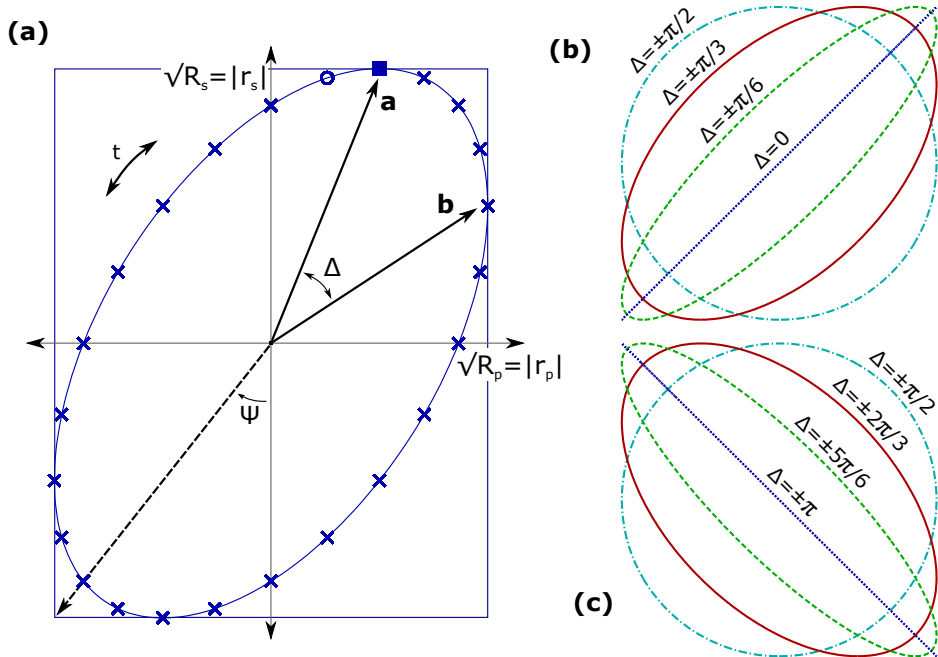
Figure 14: Elliptically polarised light as seen by a rotating analyser ellipsometer. The figure shows the plane perpendicular to the propagation direction (which is out of the page). The detected intensity changes as the analyser rotates, describing the blue ellipse. A check-mark appears for every $15°$ of phase Δ.

The signal seen by a rotating analyser ellipsometer (PSA$_R$) is shown in Figure 14; the plot uses two cosines, with $|r_p| = 0.6|r_s|$ and a phase shift $\arg r_p = \arg r_s + 60°$. The signal results from the addition of electric fields in the perpendicular p- and s-polarisation directions, these given by the square roots of the measured reflectivities R_p and R_s (see page 46). Geometric relations are shown for $\tan\Psi$, the ratio of amplitudes (the maximum values) of r_p and r_s, and for the phase difference Δ which sets the eccentricity of the ellipse.

Geometric discussions of the phase shift Δ are not easily found in the literature, yet these may be the simplest means of explaining the working principles of ellipsometry. Assuming that the $\mathbf{E}_p$ and $\mathbf{E}_s$ waves are in phase before reflection, then Δ is (by definition) the phase difference between some reference point on the reflected $\mathbf{E}_p$ and $\mathbf{E}_s$ waveforms. Arbitrarily choosing the maximum values of $\mathbf{E}_p$ and $\mathbf{E}_s$ as reference points, the angle Δ is then that between the vectors **a** and **b** in Figure 14a (although any reference point may be chosen, e.g. the first zeros after each maximum). A difficulty arises as the size of angular increments at the origin changes with angle, as seen if sectors are projected from the origin to the check-marks every $15°$ on the ellipse; however, here the check-marks show that $|\Delta| = 60°$, as required.

Rotating analyser ellipsometers (PSA$_R$) determine only the absolute value $|\Delta|$, as the signal is unchanged for traversals of the ellipse in either direction, i.e. for clockwise (right) or anti-clockwise (left) elliptically polarised light; in the former case, the $\mathbf{E}_s$ maximum leads $\mathbf{E}_p$, so $\Delta = \arg r_p - \arg r_s = -|\Delta|$; likewise for left- or anticlockwise polarisations, $\Delta = +|\Delta|$. The phase shift Δ is fully determined with the addition of a compensator (i.e. PSCA$_R$) which introduces an additional phase shift in one polarisation[40], e.g. with an initial Δ (without any compensator) of $+\pi/6$ or $-\pi/6$, insertion of a $\pi/2$ compensator (retarding say $\mathbf{E}_s$) changes Δ respectively to $2\pi/3$ or $\pi/3$, which are then distinct via Figures 14b and c.

Additionally, PSA$_R$ ellipsometers have poor accuracy at $\Delta = 0$ and $\Delta = 180°$; accuracy is optimal when the signal is close to circular polarisation[40]. Thus the PSCA$_R$ configuration both improves accuracy and fully determines Δ. While insertion of the compensator allows discrimination of left- and right-handed polarisations[†], the compensator does not necessarily need to be physically inserted or removed from the beam path; it is sufficient that two measurements are made with different compensator settings: this may simply involve a rotation of the compensator such that the p- and s-polarisations see different refractive indices (and so propagate with different phase velocities). Recently, rotating compensator ellipsometers (e.g. PSC$_R$A) have become available which use fixed analysers and eliminate the need for two measurements with different compensator settings, thus improving performance. Rotation of the compensator introduces a time-dependent phase shift $\delta\Delta = \frac{\pi}{2}\sin(\omega t)$, which continuously deforms the ellipse as seen in Figures 14b and c. Practically, the measured signals are decomposed by Fourier analysis into frequency components for further processing.

With the values of $\Psi(\hbar\omega)$ and $\Delta(\hbar\omega)$ determined across the spectral range, a suitable multi-layer optical model is selected and fitted to the experimental spectra via a least-squares method. The considerations follow in a similar manner to those discussed for the transfer-matrix approach in Section 3.11; optical models generally require homogeneous, coherent layers, with parallel boundaries and without inelastic photon interactions which would convert high-energy photons into low-energy photons (photon down-conversion). The reflectivity ratio $\rho(\hbar\omega)$ is then a function of the angle of incidence and the thicknesses and complex refractive index spectrum of each optical layer $\{d_i, \tilde{n}_i(\hbar\omega)\}$. Alternatively, a pseudo-dielectric constant may be written[‡] immediately by treating the sample as a semi-infinite medium (with no internal reflections) illuminated from the vacuum at an angle of incidence ϕ

$$\tilde{\epsilon} = \epsilon_1 + i\epsilon_2 \equiv \tilde{n}^2 = (n + i\kappa)^2 = \sin^2\phi \cdot \left[1 + \left(\frac{1+\rho}{1-\rho}\right)^2 \tan^2\phi\right] \qquad (2.29)$$

Refractive indices of interior layers or substrates may be determined through measurements of individual specimens. Fits are often improved by measuring samples at several different angles of incidence, with the highest sensitivities seen near the Brewster angle: where the reflectivity of p-polarised light vanishes in dielectric (non-absorbing) material. Fits for more complicated optical models may be greatly improved via the use of virtual media[39], such as the Bruggeman effective medium: intended for porous materials comprising voids of another medium (such as the vacuum)[42]. While dielectric functions may be fitted directly to ellipsometric spectra without physical constraints (e.g. via B-splines), often some physical dielectric model[39] is imposed (such as Lorentz oscillators or the models of Sellmeier, Cauchy, Tauc or Forouhi–Bloomer), the parameters of which are then fitted to the ellipsometric spectra. As an additional constraint, the real and imaginary parts of the dielectric function may be constrained to be consistent under the Kramers–Kronig relation[43,44].

[†] Take care with polarisation nomenclature. Here the convention of Born and Wolf is used: that right-handed elliptical polarisation describes an ellipse traversed in a clockwise direction when looking at the source; likewise left-handed polarisation gives an anti-clockwise traversal[41].

[‡] Take care with the sign of ρ, which depends on the form of the Fresnel coefficients; Eqn. (2.29) uses (3.15) and (3.13).

2.3 X-ray diffraction

X-ray diffraction (XRD) uses **the scattering of x-rays by electrons to study crystalline material**. The inherent long range order of crystals preferentially scatters x-rays in characteristic directions. In Figure 15(i), an x-ray beam incident at an angle α from the top left illuminates two atomic planes separated by an inter-plane spacing d. Consider the phase difference $\Delta\phi$ between the two rays scattered at an angle β by the top plane of atoms[45]

$$\Delta\phi = \frac{2\pi}{\lambda}\left(CD - AB\right) = \frac{2\pi}{\lambda}AD\left(\cos\alpha - \cos\beta\right) = 0, \qquad \text{for} \quad \alpha = \beta \qquad (2.30)$$

Thus coherent reflections only result from the condition $\alpha = \beta$. Using this and requiring coherent reflections from the optical path length EFG associated with the second plane[45]

$$\Delta s = EF + FG = 2d\,\sin\alpha = n\lambda \longrightarrow 2d\,\sin\theta = n\lambda \qquad (2.31)$$

where the right-hand side gives the familiar statement of Bragg's law for diffraction. Thus knowledge of the x-ray wavelength λ and measurement of the reflection angle θ determines the inter-plane spacing d. Figure 15(ii) depicts the general case of diffraction from an arbitrary bulk object with an origin at the point O. The phase difference $\Delta\phi$ associated with the path ABC due to the reflection of a ray with wavevector $\mathbf{k}$ at a position described by a vector $\mathbf{r}$ is

$$\Delta\phi = \frac{2\pi r}{\lambda}\left(\cos\beta + \cos\delta\right) = \left(\mathbf{k} - \mathbf{k}'\right)\cdot\mathbf{r} = -\Delta\mathbf{k}\cdot\mathbf{r} \qquad (2.32)$$

where $\Delta\mathbf{k} = \mathbf{k}' - \mathbf{k}$. Thus integrating the phase $e^{-i\Delta\mathbf{k}\cdot\mathbf{r}}$ over space with electron density $n(\mathbf{r})$

$$F_{\Delta k} = \int_{\text{space}} d^3\mathbf{r}\, n(\mathbf{r})\,e^{-i\Delta\mathbf{k}\cdot\mathbf{r}} \rightarrow \sum_{\mathbf{G}} \int_{\text{space}} d^3\mathbf{r}\, n_{\mathbf{G}}\,e^{i(\mathbf{G}-\Delta\mathbf{k})\cdot\mathbf{r}} \qquad (2.33)$$

where in the right-side, the Fourier expansion of $n(\mathbf{r})$ finds a significant result only for $\mathbf{G} = \Delta\mathbf{k}$. Inserting this condition into the middle term in (2.33) and using the periodicity of the lattice, $n(\mathbf{r}) = n(\mathbf{R}+\mathbf{s}) = n(\mathbf{s})$ (with $\mathbf{R}$ a lattice vector), gives an integral over the unit cell[46]

$$F_{\mathbf{G}} = N \int_{\text{cell}} d^3\mathbf{s}\, n(\mathbf{s})\,e^{-i\mathbf{G}\cdot\mathbf{s}} = N \sum_{j=1}^{\text{basis}} e^{-i\mathbf{G}\cdot\mathbf{s}_j} \int_{\text{cell}} d^3\mathbf{s}\, n_j(\mathbf{s}-\mathbf{s}_j)e^{-i\mathbf{G}\cdot(\mathbf{s}-\mathbf{s}_j)} \qquad (2.34)$$

where the electron density $n(\mathbf{s})$ is written as the sum of that due to each atom $n_j(\mathbf{s} - \mathbf{s_j})$ in
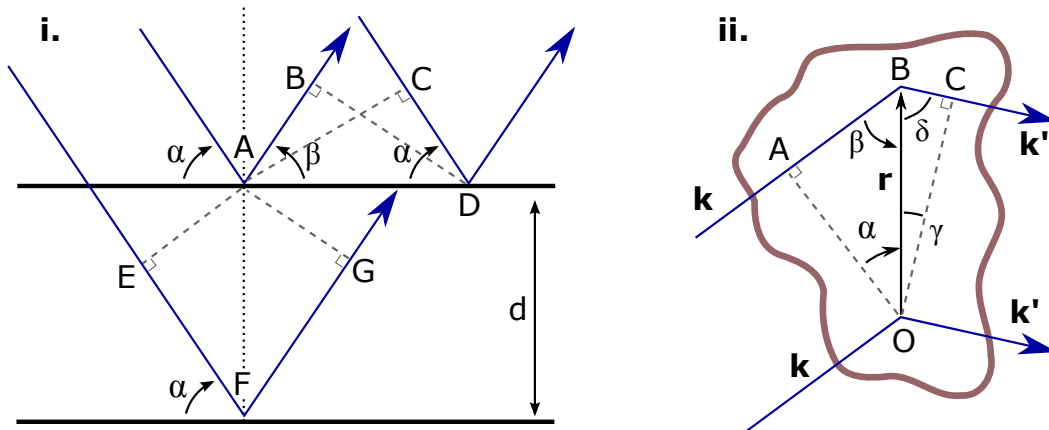
Figure 15: Constructions for x-ray diffraction: reflections from the same, parallel and arbitrary planes.

the basis with position vector $\mathbf{s}_j$, while N is the number of unit cells in the crystal. The integral on the right-side of (2.34) is the **atomic form factor** $f_j(\mathbf{G})$, with which (2.34) takes a simple form[46], with $\mathbf{s_j} = x_j\mathbf{a}_1 + y_j\mathbf{a}_2 + z_j\mathbf{a}_3$ and reciprocal lattice vector $\mathbf{G} = h\mathbf{b}_1 + k\mathbf{b}_2 + l\mathbf{b}_3$

$$F_{\mathbf{G}} = N \sum_{j=1}^{\text{basis}} f_j(\mathbf{G}) \, e^{-i\mathbf{G}\cdot\mathbf{s}_j} = N \sum_{j=1}^{\text{basis}} f_j(\mathbf{G}) \, \exp\left(-i2\pi\left\{hx_j + ky_j + lz_j\right\}\right) \qquad (2.35)$$

where the sum is the **structure factor** $S(\mathbf{G})$. Equation (2.35) gives an amplitude for a reflection associated with the reciprocal lattice vector $\mathbf{G}$. In conclusion, while **reflection positions depend only on the crystal lattice, reflection intensities depend only on the crystal basis**[45].

2.3.1 Fitting x-ray diffraction patterns

Equation (2.35) suggests an approach for fitting XRD patterns. Rather than separately fitting individual reflections in a pattern and attempting to establish the crystal structure from the fitted positions, alternatively **the entire pattern may be fitted directly from a model unit cell**. The Rietveld[47] method is one such full-pattern approach in which a model crystal structure is fitted to an experimental pattern by a least-squares method. With the reflection positions set by (2.31), the intensity $I(\theta)$ calculated for the diffraction pattern envelope is[48]

$$I(2\theta) = A \sum_j m_j \, L_j \, |S_j|^2 \, G(2\theta, \theta_{j0}, W) + B(2\theta) \qquad (2.36)$$

where A is an overall scale factor, j counts over distinct reflections, m_j is a multiplicity factor[45] [e.g. in a cubic crystal the reflections (100), $(\bar{1}00)$, (010), $(0\bar{1}0)$, (001), $(00\bar{1})$ are equivalent], L_j is a geometric (Lorentz-polarisation[45]) factor, and S_j and θ_{j0} are the structure factor and line position for reflection j. A single line-shape function $G(2\theta, \theta_{j0}, W)$, with width parameter $W(\theta)$, is used for all reflections in the pattern. Various line shape profiles including Gaussians and Lorentzians are used; one favoured profile is the pseudo-Voigt function[48], a linear combination of Gaussian and Lorentzian functions, so-called as it is intended to approximate a Voigt lineshape (a convolution of these functions). The width parameter $W(\theta)$ is often given an angular dependence[48]. A background function $B(2\theta)$ accounts for any background evolution: via polynomials or linear interpolation between discrete points.

The first task in fitting is to use the positions of pattern reflections to determine the crystal system and lattice parameters in a process called *indexing the pattern*[45]. For low-symmetry polycrystalline or powder samples, this process is non-trivial and indexing tools, such as GSAS2[49] and DICVOL06[50], assist with finding potential candidates. For single crystals, the space group may be determined after indexing via systematic absences of certain reflections; with polycrystalline material this may not be possible: but often the lattice parameters and space group are known from prior work. Once the crystal system and space group are known, a crystal basis may be fitted via the Rietveld method. The symmetry operations of the space group are employed to reduce the number of free-parameters in the fit[51]: a significant gain as fits may easily go astray (but the space group may detrimentally constrain the relaxation of certain atoms). Fitting typically initiates with a few parameters (e.g. scale factor, lattice parameters and line-width); further parameters are then carefully added incrementally.
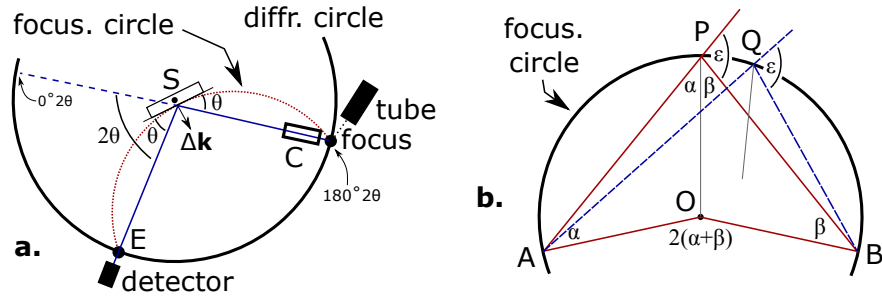
Figure 16: Bragg–Brentano configuration and inscribed angles, (**a**) the diffraction and focus circles are shown, along with the x-ray tube, collimator (C), sample (S), receiving slit (E) and detector. (**b**) the inscribed angle theorem of geometry claims that for any points A, B and P on a circle, then $\angle AOB = 2\angle APB$. The proof comprises three parts: only the first is required for XRD. Consider $\angle AOB = 2\pi - (\pi - 2\alpha) - (\pi - 2\beta) = 2(\alpha + \beta) = 2\angle APB = 2\angle AQB$. Thus, the same (hkl) planes in polycrystalline samples positioned on the focussing circle diffract x-rays by the same angle $2\theta = \epsilon$, and all of these rays are brought to a focus at the detector at B.

Rietveld fits may additionally model isotropic and anisotropic atomic temperature factors[48]; fractional occupancy of atomic sites; co-existing phases and preferential orientation; x-ray absorption and sample transparency[51], and diffractometer zero error or sample displacement; however, unphysical values may result from unsuitable models, and without complementary evidence a cautious approach may be most wise. Various quality-of-fit metrics are used[48,51], e.g.

$$R_{\mathrm{wp}} = \sqrt{\frac{\sum_{i=1}^{N}\left(y_{\mathrm{expr},i} - y_{\mathrm{calc},i}\right)^2/\sigma_i^2}{\sum_{i=1}^{N} y_{\mathrm{expr},i}^2/\sigma_i^2}} \qquad \text{with} \qquad R_{\mathrm{exp}} = \sqrt{\frac{N - P}{\sum_{i=1}^{N} y_{\mathrm{expr},i}^2/\sigma_i^2}} \qquad (2.37)$$

where $y_{\mathrm{expr},i}$, $y_{\mathrm{calc},i}$ and σ_i are the experimental and computed intensities, and the standard deviation for datapoint i of N, for a model with P parameters: such that ideally $R_{\mathrm{wp}} \to R_{\mathrm{exp}}$ and the reduced-χ^2 metric is then $R_{\mathrm{wp}}^2/R_{\mathrm{exp}}^2$. Where a Rietveld model fails to satisfactorily predict reflection intensities, full-pattern methods may still be utilised without predicting these (freely fitting intensities for each reflection). Examples of such methods have been described by Pawley[52,53], and Le Bail[54,55]: the former method is applied in Section 4.4.

2.3.2 Diffractometer geometry and components

One common diffraction configuration is the Bragg–Brentano arrangement[45] seen in Figure 16a, which positions the x-ray source and detector on a diffraction circle centred upon the sample. During measurement, the detector moves around the diffraction circle from $\sim 0°\,2\theta$ to higher angles; practically the detector cannot reach $180°\,2\theta$ as the x-ray tube is in the way. During this motion the sample holder rotates so that the normal to the sample surface bisects the angle between source and detector to maintain the $\theta_{\mathrm{incident}} = \theta_{\mathrm{reflected}}$ diffraction geometry. Focussing is optimised by aligning the focus of the x-ray tube, the detector receiving slit, and the tangent to the sample surface upon a second focussing circle[45]. As seen in Figure 16b, the inscribed angle theorem of geometry, which states that all angles inscribed inside a circle from two endpoints are identical (and half the angle from the centre to these points), then ensures that reflections from the same (hkl) planes in polycrystalline (or powder) samples set on the focussing circle are focussed back at the detector: giving enhanced intensities.

The collimator in Figure 16a typically comprises a set of parallel plates between two vertical slits (called a Soller slit), which respectively limits beam divergence in the directions perpendicular and parallel to the plane of the diagram: leading to sharper diffraction peaks. The purity of the beam may be improved with the addition of a monochromator, which typically rejects $K\beta$ radiation leaving only the $K\alpha_1$ and $K\alpha_2$ components (see Section 2.4). Monochromators comprise carefully prepared crystals (e.g. graphite) set at a specific angle to generate (and transmit) a diffraction reflection associated with only the desired frequency component. Monochromator crystals may be curved to better satisfy the focussing circle conditions of Figure 16b, and may be positioned either just after the x-ray tube, or after the receiving slit[45]. Modern instruments may integrate the tasks of monochromation and collimation into a single device[56]. Typically, the largest errors in XRD result from diffractometer zero errors and sample displacement away from the focussing circle; one benefit of full-pattern methods is their ability to account for such errors by using all the information in the pattern[57].

2.3.3 Data reduction for 2D detectors

Typically x-rays have been detected with photo-diode point-detectors; however, modern instruments intended for single-crystal work may employ 2D area detectors with $n \times n$ pixels. When examining polycrystalline material, these detectors record images of diffraction cones, and the number of photon counts within each ring must be integrated to generate a diffraction pattern for analysis, see Figure 17. The value in any single pixel then results from

$$\text{pixel value} = \text{gain} \times (\text{bias} + \text{dark electron count} + \text{Q.E.} \times \text{incident photon count}) \quad (2.38)$$

where the gain is the ratio of amplifier output to input, the system bias is evaluated by averaging a set of zero-time exposures (with closed shutter), the dark electron count is the number of electrons accumulated over an equal measurement time with closed shutter (due to thermal noise), while the actual number of electrons generated due to the incident light is given by the product of quantum efficiency (Q.E.) with the number of photons incident over the measurement duration. With a known Q.E., amplifier gain and a separate dark image (shutter-closed) measurement, equation (2.38) may be inverted to obtain the incident photon count for any pixel[58]. Circular multi-pixel bins, each associated with a diffraction cone with interior half-angle 2θ to $2\theta + \delta\theta$ and covering C_i pixels, may then be drawn upon each 2D image and the photon count N_i integrated for each bin; Poisson statistics then gives the standard deviation of bin i as $\sqrt{N_i}$. The conventional powder diffraction pattern then results after normalising the photon count and standard deviation as N_i/C_i and $(\sqrt{N_i})/C_i$ respectively.
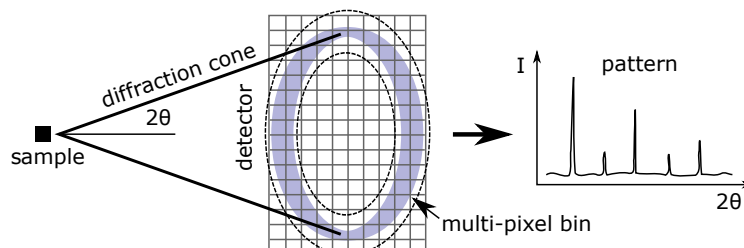
Figure 17: Powder pattern conversion for 2D detectors. See the text for details.

2.4 X-ray photoelectron spectroscopy

X-ray photoelectron spectroscopy (XPS) **studies occupied electron states via the photoelectric effect** under ultra-high vacuum, establishing the chemical composition of $\sim 10\,\mathrm{nm}$ thick surface layers (for elements above Li with atomic concentrations above $0.05\,\%$), the dependence of binding energies on environment (chemical shifts), band structure and valence band density of states[59]. Characteristic x-rays are produced on a Bremsstrahlung background after core electrons are liberated from a target anode by an incident electron beam; x-ray emission then occurs as electrons from higher levels relax to fill the core holes, see Figure 18a.

As sketched in Figure 18b, after illumination with x-ray photons of energy $h\nu$, electrons are ejected from isolated samples with kinetic energy $T_{X,i}$ from a state i with binding energy $E_{B,i}$ (the energy difference between the state E_i and the Fermi level E_F), from material with a work function Φ_{sample} (the energy difference between E_F and the local vacuum level $E_{vac,X}$). The photoelectric effect $T_{X,i} = h\nu - E_{B,i} - \Phi_{sample}$ thus holds for such samples. When the sample is placed within a spectrometer, charge may transfer between the sample and (metallic) instrument; the local vacuum levels then differ, and the sample Fermi level equalises with the instrumental E_F (which does not change due to a large density of states). If kinetic energies $T_{Y,i}$ of photoelectrons are measured within the spectrometer, and the instrumental work function Φ_{spec} is known, then the binding energy of the initial state is determined

$$E_{B,i} = h\nu - T_{Y,i} - \Phi_{spec} \tag{2.39}$$

The instrument work function Φ_{spec} may be found via analysis of the Fermi edge of a clean metallic sample (e.g. Au, Ag) by setting $E_{B,i} = 0$ in (2.39). Electron kinetic energies are evaluated with a hemispherical electron analyser, which employs a centripetal electric field to select electrons with some desired *pass energy*: which then traverse a circular path to the detector. Results include line-spectra (typically decomposed into pseudo-Voigt functions) and approximate valence-band density of states (as the VB is near fully occupied)[60]. Monochromators are often used in XPS for improved energy resolution (see Section 2.3.2).
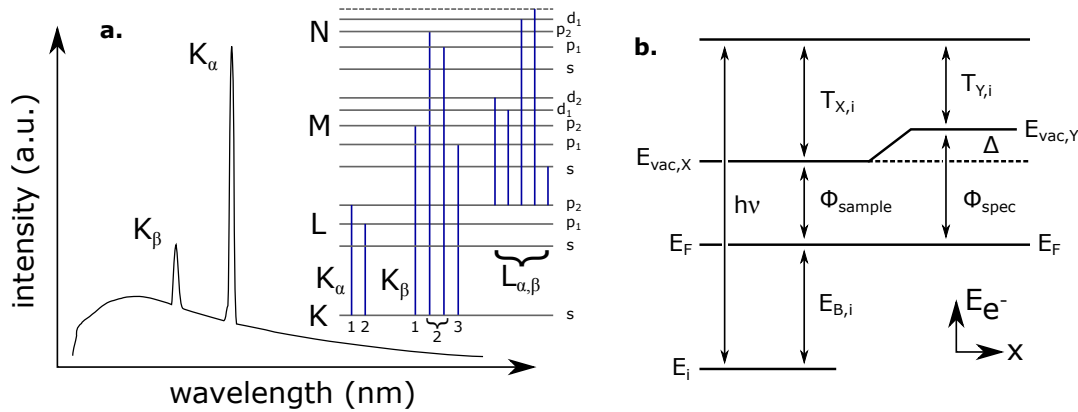
Figure 18: Characteristic x-rays and band structure for XPS. (**a**) x-rays result from interband relaxations; these are named for the principal quantum numbers of the final and initial states ($n = 1, 2, \cdot$ giving $K, L, \cdot$ and $\alpha, \beta, \cdot$) and the line intensity, e.g. $K_{\alpha,1}$ and $K_{\alpha,2}$, the former being the strongest line to the ground state. The spectra result from the selection rules for interband transitions: $\Delta j = \pm 1$, $\Delta m_j = \pm 1, 0$. (**b**) the band structure for a XPS spectrometer (labelled Y) with the sample (X) in position. See the text.

2.5 Scanning electron microscopy and energy dispersive spectroscopy

Scanning electron microscopy (SEM) uses electrons as probes to examine surface morphology and cross-sectional composition, to qualitatively discriminate between regions of different composition and (when supplemented with energy-dispersive spectroscopy) to perform quantitative elemental analysis[61]. The technique is frequently used in materials science and is suitable for conducting, vacuum-safe samples which are not damaged by the incident irradiation (non-conducting or organic samples may be gold-coated for analysis)[62].

The main components of an electron microscope are depicted schematically in Figure 19. An **electron gun** accelerates electrons to 1-100 keV, these are confined and focussed by **condensing and objective lenses** and **raster-scanned** across the sample surface by **magnetic deflection coils**[61]. Electrons liberated from the specimen are detected, and an image forms as the beam scans across the sample, whose position is precisely controlled (in translation and rotation) by a **motorised stage**. Spatial resolution is determined by the spot size at the specimen surface and the beam energy E_B. Spot size may range from 1 nm to 5 μm as beam current increases from 1 pA through 10 nA. The beam energy sets the penetration depth R (typically a few microns), estimated[61] as $R \propto A V^{1.7} / (\rho Z^{0.9})$, with atomic and mass numbers Z and A, applied voltage V and mass density ρ.

The detected signal is due to various phenomena in the sub-surface interaction volume. **Secondary electrons**, due to inelastic electron collisions, have energies < 50 eV and are particularly sensitive to topography[62,63]. **Backscattered electrons**, due to elastic collisions with atoms, emerge with energies up to E_B (peaked at $0.5 E_B$) and give compositional contrast due to atomic number sensitivity[62]. **X-rays** and other photons result, both due to characteristic radiation as electrons from higher levels relax to fill core holes, and due to a Bremsstrahlung continuum as electrons loose energy in nuclear Coulomb fields. Finally **Auger electrons**, ejected with characteristic energies as electrons relax to fill core holes, give composition information (with 5 % accuracy) on a 1 nm surface layer[62].

In energy dispersive spectroscopy (EDS or EDX), characteristic x-rays are measured by a supplementary detector (with typical resolution 3 eV), allowing quantitative elemental analysis for elements with atomic number $Z > 4$ (sensitive to concentrations of 0.1 % with accuracies of 1 %)[62]. As x-ray production is enhanced with larger interaction volumes, EDS typically uses intermediate beam energies (~ 20 keV) so is not a surface sensitive technique[61,64].
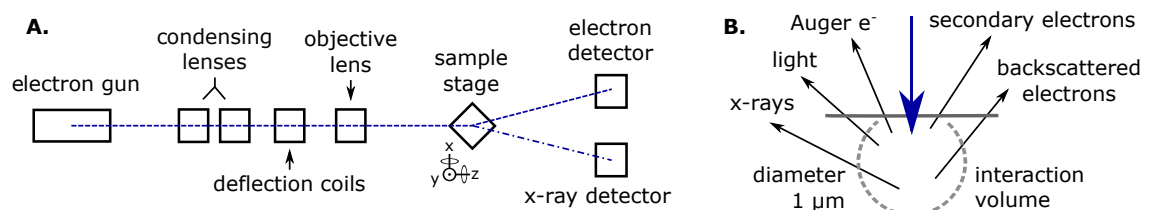
Figure 19: Schematic layout for scanning electron microscope (with EDS) and electron interactions.
A. electrons focussed by lenses and scanned across the sample surface by magnetic deflection coils liberate electrons and x-rays for detection. **B** shows the incident beam, the interaction region and various interactions.

2.6 Density functional theory

Density-functional theory (DFT) seeks to determine the properties of many-electron systems from first-principles, including: lattice parameters, bond lengths and angles; charge densities; cohesive & ionisation energies and affinities; ionic forces, bulk and elastic moduli, vibrational modes and molecular dynamics. The works of Martin[65] and others[66,67] have been invaluable in this section. Consider the Schrödinger equation for a neutral system of N electrons

$$\mathcal{H}\,\Psi\{\mathbf{r}_i; \mathbf{r}_I\} = (T_\mathrm{e} + V_\mathrm{ee} + V_\mathrm{en} + V_\mathrm{nn} + T_\mathrm{n})\,\Psi\{\mathbf{r}_i; \mathbf{r}_I\} \tag{2.40}$$

where I and i respectively count over nuclei and electrons, whose kinetic energies and Coulomb interactions are identified in the Hamiltonian. Expanding this Hamiltonian gives

$$-\frac{\hbar^2}{2m_\mathrm{e}}\sum_i \nabla_i^2 + \frac{e^2}{4\pi\epsilon_0}\left(\sum_{i\neq j}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{i,I}\frac{Z_I}{|\mathbf{r}_i - \mathbf{r}_I|} + \sum_{I\neq J}\frac{Z_I Z_J}{|\mathbf{r}_I - \mathbf{r}_J|}\right) - \frac{\hbar^2}{2}\sum_I \frac{\nabla_I^2}{m_\mathrm{I}} \tag{2.41}$$

In the Born–Oppenheimer (or adiabatic) approximation, the nuclear kinetic energy term T_n is considered negligible in comparison to the other terms: as the nuclear mass is ~ 2000 times heavier than the electron mass. Hartree[68] approached the problem by writing $\Psi\{\mathbf{r}_i\}$ as a product of single electron wavefunctions, $\psi_1(\mathbf{r}_1)\psi_2(\mathbf{r}_2)\cdots\psi_N(\mathbf{r}_N)$; Fock[69] then wrote $\Psi\{\mathbf{r}_i\}$ as an anti-symmetric Slater determinant, $\det\big[\psi_1(\mathbf{r}_1), \psi_1(\mathbf{r}_2), \cdots, \psi_1(\mathbf{r}_N); \cdots \psi_N(\mathbf{r}_N)\big]$, which respected the Pauli exclusion principle and improved energies through the introduction of an exchange potential[70]. Even with such approaches, the information content of the wavefunction $\Psi\{\mathbf{r}_i\}$ grows exponentially[§], limiting applications to systems below ~ 10 atoms[71,72].

The many-body problem was reformulated in terms of an idealised electron density by Thomas[73], Fermi[74] and Dirac[75], giving a function in just three coordinates; subsequently, Hohenberg and Kohn[76] found theorems which suggested a solution approach $\big[$after replacing the terms $V_\mathrm{en} + V_\mathrm{nn} + T_\mathrm{n}$ in the Hamiltonian of (2.40) with an external potential $V_\mathrm{ext}(\mathbf{r})\big]$, and which proved that **ground state density is a good variable in any many-particle system** :

1. for any system of interacting particles in some external potential $V_\mathrm{ext}(\mathbf{r})$, **the ground state particle density $n_0(\mathbf{r})$ results from a unique potential $V_\mathrm{ext}(\mathbf{r})$, up to a constant.** Thus there is a one-to-one relation[¶] between the ground state density n_0, wavefunction $\Psi_0\{\mathbf{r}_i\}$ and external potential $V_\mathrm{ext}(\mathbf{r})$. This non-trivial result is clarified in Figure 20.

2. **a universal functional[‖] $E[n(\mathbf{r})] = F[n(\mathbf{r})] + \int \mathrm{d}\mathbf{r}\, V_\mathrm{ext}(\mathbf{r})\, n(\mathbf{r})$ exists for the total energy,** written in any density $n(\mathbf{r})$ for any external potential $V_\mathrm{ext}(\mathbf{r})$. **The ground state energy and density $n_0(\mathbf{r})$ result at the global minimum of $E[n(\mathbf{r})]$.** *Universal* implies that one functional contains all the physics needed to treat particles of that class, e.g. electrons.

Thus the first statement enables the second statement; however, the work gives no guidance on selection of the correct functional $F[n]$: which is evidently highly complicated[65].

[§] Plotting the wavefunction $\psi_1(\mathbf{r}_1)\cdots\psi_N(\mathbf{r}_N)$ over a $10 \times 10 \times 10$ spacial grid requires $(10^3)^N$ complex numbers.

[¶] If sufficient computational machinery is found to implement the mappings of Figure 20, the many-body Hamiltonian is then fully determined (up to a constant) by the ground state density $n_0(\mathbf{r})$; assuming a fixed external potential, solution of the Hamiltonian gives wavefunctions and energies for all states in the system (ground and excited).

[‖] While a function maps a scalar to a scalar, e.g. $f(x) = x^2$, a functional maps a function to a scalar, e.g. $g[f] = \int_a^b \mathrm{d}x\, f^2(x)$.
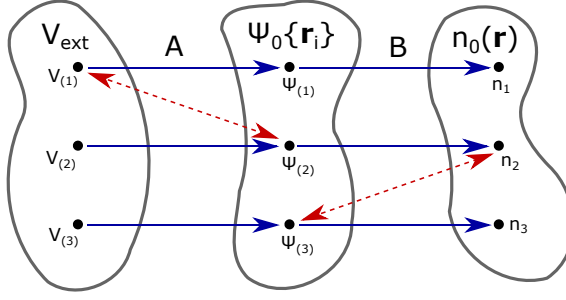
Figure 20: Conceivable mappings between external potentials, ground state wavefunctions and densities. The sets contain: every possible external potential $V_{ext}(\mathbf{r})$ differing by more than a constant and leading to a non-degenerate ground state; all resultant ground states $\Psi_0\{\mathbf{r}_i\}$ and all densities $n_0(\mathbf{r})$. The forward mappings (blue, solid arrows) ensure that each element has at least one mapping (surjectivity); Hohenberg and Kohn showed that no further mappings exist (injectivity), eliminating the diagonal mappings (red, dashed arrows). The forward A mapping is removed by construction (i.e. non-degeneracy). The remaining mappings lead to contradictions if two $V_{ext}(\mathbf{r})$ or $\Psi(\mathbf{r}_i)$ respectively share the same $\Psi(\mathbf{r}_i)$ or $n_0(\mathbf{r})$. See Engel[67].

Subsequently, Kohn and Sham[77] proposed that the problem could be addressed by replacing the interacting N-electron problem with an auxiliary system of N fictitious, independent spin-half particles, with the same ground state density $n_0(\mathbf{r})$ as the many-body system (assuming the feasibility of this). The auxiliary system has a kinetic energy functional

$$T_{KS}[n] \rightarrow -\frac{\hbar^2}{2m} \sum_{i,\sigma} \int \mathrm{d}^3\mathbf{r} \, \langle \psi_i^\sigma(\mathbf{r})|\nabla^2|\psi_i^\sigma(\mathbf{r})\rangle \tag{2.42}$$

with eigenfunctions $\psi_i^\sigma(\mathbf{r})$, where i and σ count over spacial and spin eigenstates; the density is then $n(\mathbf{r}) = \sum_{i,\sigma} |\psi_i^\sigma(\mathbf{r})|^2$. In the ground state, only the N states of lowest energy are occupied. The universal Hohenberg and Kohn energy functional becomes

$$E_{KS}[n] = T_{KS}[n] + \frac{1}{2}\frac{e^2}{4\pi\epsilon_0} \int \mathrm{d}^3\mathbf{r}\,\mathrm{d}^3\mathbf{r}' \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + \int \mathrm{d}^3\mathbf{r}\, V_{ext}(\mathbf{r})\, n(\mathbf{r}) + E_{II} + E_{xc}[n] \tag{2.43}$$

where the first integral term is the Hartree energy $E_{Hartree}[n]$, and all difficult many-body effects are contained in the exchange-correlation functional $E_{xc}[n]$: named for corrections associated with antisymmetry under exchange (Pauli exclusion), and for many-body screening which weakens Coulomb interactions between pairs of electrons. A Lagrange multiplier $\epsilon \langle \psi_i^\sigma | \psi_i^\sigma \rangle - 1) = 0$ subtracted from (2.43) ensures orthonormality; minimising (2.43) with this multiplier using functional differention[**] and the chain rule gives the stationary point[65,66]

$$\frac{\delta E_{KS}[n]}{\delta \psi_i^{\sigma *}(\mathbf{r})} = \frac{\delta}{\delta \psi_i^{\sigma *}(\mathbf{r})}\left(T_{KS}[n] - \epsilon \langle \psi_i^\sigma | \psi_i^\sigma \rangle\right)$$
$$+ \frac{\delta}{\delta n(\mathbf{r},\sigma)}\left(E_{Hartree}[n] + \int \mathrm{d}^3\mathbf{r}\, V_{ext}(\mathbf{r})\, n(\mathbf{r}) + E_{xc}[n]\right)\frac{\delta n(\mathbf{r},\sigma)}{\delta \psi_i^{\sigma *}(\mathbf{r})} = 0 \tag{2.44}$$

Rearranging gives the Kohn–Sham equations

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + \frac{e^2}{4\pi\epsilon_0}\int \mathrm{d}^3\mathbf{r}' \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + V_{ext}(\mathbf{r}) + \frac{\delta E_{xc}[n]}{\delta n_i(\mathbf{r})}\right]\psi_i^\sigma(\mathbf{r}) = \epsilon_i^\sigma \psi_i^\sigma(\mathbf{r}) \tag{2.45}$$

[**] Variation of the functional $F[g(x)] = \int \mathrm{d}x\, g(x)h(x)$ gives $\delta F[g(x)] = F[g+\delta g] - F[g] = \int \mathrm{d}x\, \frac{\delta F[g]}{\delta g}\delta g$; so $\frac{\delta F[g]}{\delta g} = h(x)$.

where the integral term is the Hartree potential $V_{\text{Hartree}}(\mathbf{r})$. At this point, equation (2.45) is still an exact solution of the many-body system: all the complexities are hidden in the exchange-correlation term in (2.45), sometimes written $V_{\text{xc}}(\mathbf{r})$. Subtracting the Kohn–Sham functional from the Hohenberg–Kohn functional gives insight into the origin of this term

$$E_{\text{xc}}[n] = \langle T_{\text{e}} \rangle - T_{\text{KS}}[n] + \langle V_{\text{ee}} \rangle - V_{\text{Hartree}}[n] \tag{2.46}$$

Thus $E_{\text{xc}}[n]$ accounts for any differences between the kinetic energies of the many-body and Kohn–Sham systems, and for any differences between the many-body Coulomb interaction and the Hartree energy. One of the strengths of the Kohn–Sham approach is that by explicitly accounting for the long-range Hartree and independent particle kinetic energies, the $V_{\text{xc}}(\mathbf{r})$ term is typically only required to make small, localised energy corrections[65]. If the *universal* exchange-correlation functional was known exactly, then the real many-body problem could be solved exactly through the solution of the Kohn–Sham equations. With approximations for E_{xc}, we may expect to do less well, yet the local-density approximation (LDA), generalised gradient approximation (GGA) and various hybrid functionals have proved to be surprisingly successful.

For the typical densities seen in solids, the spatial extent of exchange and correlation effects is similar to that seen in the homogeneous electron gas. In the local density approximation (LDA), the exchange-correlation energy is approximated by the (known) exchange-correlation energy of such a gas with the same density

$$E_{\text{xc}}^{\text{LDA}}[n^\uparrow, n^\downarrow] = \int \mathrm{d}^3\mathbf{r}\, n(\mathbf{r})\, \epsilon_{\text{xc}}^{\text{LDA}}\big(n^\uparrow(\mathbf{r}), n^\downarrow(\mathbf{r})\big) \tag{2.47}$$

where $\epsilon_{\text{xc}}^{\text{LDA}}(n^\uparrow, n^\downarrow)$ is the homogeneous electron gas exchange-correlation energy density (energy per electron) parametrised by the spin-up and spin-down electron densities. To improve on LDA, various generalised-gradient approximations (GGA) use gradients of the density[78]

$$E_{\text{xc}}^{\text{GGA}}[n^\uparrow, n^\downarrow] = \int \mathrm{d}^3\mathbf{r}\, n(\mathbf{r})\, \epsilon_{\text{xc}}\big(n^\uparrow(\mathbf{r}), n^\downarrow(\mathbf{r}), \nabla n^\uparrow(\mathbf{r}), \nabla n^\downarrow(\mathbf{r})\big) \tag{2.48}$$

where $\epsilon_{\text{xc}}(\cdots)$ is some exchange-correlation energy density derived from the homogeneous electron gas, but incorporating an electron density gradient. Semiconductor band gaps are always underestimated by the LDA and GGA functionals, due to the treatment of the exchange interaction. Hybrid functionals such as HSE06[79] typically do better by replacing some fraction of the GGA exchange energy with that evaluated via Hartree–Fock or other means[65].

Strong correlation effects emerge in systems on the boundary between those with localised and delocalised electrons, and may explain high-temperature superconductors, metal-insulator transitions[80] and various other phenomena[65]. Effects are most evident in the later $3d$ and $4d$ transition metals, the anomalous[††] $4f$ lanthanides (rare earths) and the $5f$ actinides, and represent cases where many-body effects are dominant and the approximations used in current functionals break down[65].

[††] These materials exhibit peculiar transport and magnetic behaviour and bulk properties inconsistent with the (usually) stable and localised $4f$ states, possibly due to interband couplings[81].

2.6.1 Solving the Kohn–Sham equations

On attempting to solve the Kohn–Sham equations (2.45), the Hamiltonian is found to include Hartree and exchange-correlation terms which depend on the electron density $n(\mathbf{r})$ which derives from the same wavefunctions we wish to find! To resolve this conundrum, an initial guess is made $\left[\text{for either } n_0(\mathbf{r}) \text{ or } \Psi_i(\mathbf{r})\right]$ and the Kohn–Sham equations are then solved iteratively until a self-consistent solution is achieved. In summary, our procedure is to

1. guess an initial ground state density $n_0(\mathbf{r})$.
2. construct the Hamiltonian and solve the Kohn–Sham eqns (2.45): giving $\psi_i^\sigma(\mathbf{r})$ and ϵ_i^σ.
3. calculate the new ground state (GS) density $n_0'(\mathbf{r}) = \sum_{i,\sigma}^{\text{GS}} |\psi_i^\sigma(\mathbf{r})|^2$.
4. if self-consistent, i.e. if $n_0(\mathbf{r}) \approx n_0'(\mathbf{r})$ to acceptable accuracy, then finish.
5. otherwise, assign $n_0'(\mathbf{r})$ to $n_0(\mathbf{r})$ and continue from step 2.

Solution gives the ground state density, total energy and ionic forces $\big($where from the many-body wavefunction one has $\mathbf{F}_I = -\langle\Psi|\partial\mathcal{H}/\partial\mathbf{r}_I|\Psi\rangle\big)$, while estimating phonon frequencies, electronic band structures and the density of states[82]. While no physical meaning is claimed for the Kohn–Sham eigenfunctions and eigenvalues[‡‡], band structures of these are often remarkably similar to those derived via other methods[83]; for this reason, Kohn–Sham band structures are frequently presented in the literature and carefully studied by workers.

Due to the periodic nature of crystalline material, wavefunctions may be expanded in Bloch sums of plane waves

$$\Psi_{n,\mathbf{k}}(\mathbf{r}) = u_{n,\mathbf{k}}(\mathbf{r})\, e^{i\mathbf{k}\mathbf{r}} = \sum_{\mathbf{G}} c_{n,\mathbf{k},\mathbf{G}}\, e^{i(\mathbf{k}+\mathbf{G})\mathbf{r}} \tag{2.49}$$

where $\mathbf{G}$ is a reciprocal lattice vector, $c_{n,\mathbf{k},\mathbf{G}}$ is a constant, and the eigenstates previously labelled generally by i are now explicitly labelled with a band index n and wavevector $\mathbf{k}$. Ideally, the sum in (2.49) would include very many plane waves; practically, the expansion is truncated at some carefully selected cut-off energy $E_{\text{cut}} = (\hbar^2/2m)\,|\mathbf{k} + \mathbf{G}|^2$.

While valence electrons are nicely handled by such expansions, *very many* plane waves are needed to represent core electron wavefunctions: which fluctuate rapidly to maintain orthogonality on approach to the nucleus as the Coulomb potential $V_{\text{nucl}}(r) = -Ze^2/4\pi\epsilon r$ diverges to negative infinity. Rather than using real nuclear Coulomb potentials, the problem is greatly simplified though the use of pseudopotentials: which modify the ionic Coulomb potential (of the nucleus and core electrons) such that it converges smoothly to some finite value at $r = 0$. Because material properties are predominantly set by their valence electrons, pseudopotentials lead to a workable number of plane waves (*soft* pseudopotentials using fewer planewaves) without significantly affecting results. Core electrons are then not explicitly included in calculations as a consequence of their implicit inclusion in the pseudopotential.

[‡‡] Excepting the highest occupied eigenvalue which, with an exact exchange-correlation functional, is the negative of the first ionisation energy.

Many quantities are conveniently evaluated as integrals in reciprocal space over the Brillouin zone, e.g. evaluation of the ground state (GS) density requires consideration of all states within the Brillouin zone volume Ω_{BZ}

$$n_0(\mathbf{r}) = \frac{1}{\Omega_{BZ}} \sum_{n,\sigma}^{GS} \int_\Omega d^3\mathbf{k} \left|\psi_{n,\mathbf{k}}^\sigma(\mathbf{r})\right|^2 \longrightarrow \sum_{n,\sigma}^{GS} \sum_{\mathbf{k}}^{IBZ} w_{\mathbf{k}} \left|\psi_{n,\mathbf{k}}^\sigma(\mathbf{r})\right|^2 \qquad (2.50)$$

where in the right-hand side, the integral has been approximated by a weighted sum over a mesh of special points in the irreducible Brillouin zone (IBZ), which is the smallest fraction of the full Brillouin zone which is sufficient (under symmetry operations) to reproduce all physics of the material[65]. The Monkhorst and Pack[84] grid is the most popular choice for placing special points; this method uniformly distributes points on a grid about $\mathbf{k} = 0$. Improved performance may result with the use of offset grids[85].

2.6.2 Practical considerations for calculations

After selection of an exchange-correlation functional, the accuracy of calculations is set by the value of E_{cut} (which sets the completeness of plane wave basis), the density of the selected k-point grid and the convergence criteria for the self-consistent field loop. Where pseudopotentials are used, the suitability of these may also need to be confirmed.

Calculations typically start with a selection of candidate grids appropriate for the intended Brillouin zone volume. Studies are then performed with the candidate grids, to assess whether the total energy converges sufficiently, and if not, to choose a grid with an higher k-point density. Finally, a total energy convergence study is performed by varying E_{cut}, concluding when further increases in E_{cut} produce only negligible changes in the total energy of the system. For example, acceptability criteria (set by the operator) may determine that convergence is achieved when further increases in E_{cut} lead to consistant total energies, differing by less than 1×10^{-2} meV.

Structural optimisation studies relax the ionic positions while minimising forces or total energy. For significant structural changes the entire cycle is repeated to verify the validity of the selected E_{cut} and k-point grid density (for extreme displacements the space group may change during relaxation). Electronic band structure calculations should trace all symmetry lines in the irreducible Brilluoin zone[86]; band structures are often seen which omit important symmetry lines (and potentially important physics). The results of all calculations should be compared with literature values to identify potential discrepancies.

Chapter 3

Determination of optical parameters in absorbing multilayers

3.1 Optical spectroscopy in photovoltaics and problem statement

Optical characterisation methods are particularly useful as these reveal dielectric properties and discrete and continuous absorption modes while typically being non-damaging and relatively simple to perform. In the particular case of photovoltaics, research primarily seeks materials with fundamental absorption onsets near the optimum photon energy of $1.4\,\mathrm{eV}$, with strong absorption above this threshold. Furthermore, discrete or continuous absorption features below the fundamental absorption threshold are undesirable: as these potentially act as intermediate states with deteriorative non-radiative recombination properties for photovoltaics. Refractive index spectra are sought for photovoltaic light-trapping and anti-reflection coatings, and for index matching between absorption- and window-layers.

Despite the utility of optical spectroscopy for these various applications, spectral interpretation is non-trivial, and the task is further complicated as sample material often comprises multiple, inhomogeneous layers: either intentionally (for mechanical or operational needs) or non-intentionally. The key problem of optical spectroscopy is then to choose a reliable means of inverting the measured spectra to allow accurate claims about the specimen. While aspects of this task are addressed in the literature, nowhere is the problem fully addressed from a practical perspective appropriate for photovoltaic work; this is the aim of this chapter.

The next sections start by considering physical phenomena and reviewing the solutions for electromagnetic waves in absorbing media and at dielectric boundaries. Additionally, dielectric processes for discrete and continuous absorption are considered (such as interband and excitonic absorption). This is a sufficient foundation to consider the problem of light propagating through a single material layer. The transfer-matrix method for multiple layers is then introduced. Armed with these tools, we can tackle the inversion of the measured reflection and transmission spectra to obtain refractive index and absorption spectra. The task is in fact a multi-valued inverse problem and in the most general case, no good solution may exist. Common approaches to the problem used in the literature are assessed, and we finish with a comparative study of the different approaches for important photovoltaic materials.
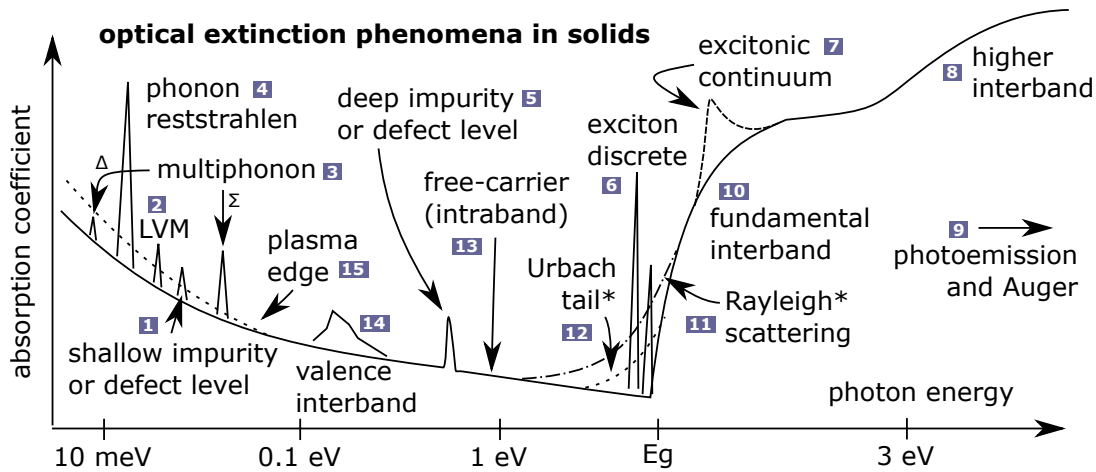
Figure 21: Optical extinction phenomena in semiconductors, where *extinction* refers to any process which removes photons from the system. Various phenomena are shown at the typical energy scales seen in semiconductors. The ordinate scale is not accurate. See the text for further details.

3.2 Optical extinction and physical phenomena

When light is incident upon a solid, photons are reflected, transmitted or absorbed: with the precise behaviour determined by the material's complex dielectric function, $\tilde{\epsilon}_r(\hbar\omega)$. Absorbed photons create excited states which either decay radiatively via photoluminescence (in one or more transitions), or non-radiatively via phonon interactions (also called thermalisation). Depending upon the crystal structure, the dielectric properties may be isotropic or anisotropic (e.g. for the cubic or orthorhombic crystal systems); in the latter case, the dielectric properties take on a tensor form[87].

Because photoluminescence sends photons in all directions, the detected photoluminescent flux is typically negligible in comparison to the reflected or transmitted flux; this generally ensures against observing an increased number of low energy photons (from photoluminescent photon down-conversion). Various microscopic phenomena (shown schematically in Figure 21) contribute to the dielectric function, each with varying degrees of relevance to the experimentalist according to the material quality or research goal. The various phenomena depicted in Figure 21 are briefly described below:

The **fundamental absorption edge** is typically the strongest feature in the absorption spectrum of semiconductors, resulting from the absorption of photons with sufficient energies $\hbar\omega \geq E_g$ to excite electrons from the valence band across the band gap E_g into the conduction band. The absorption edge may include contributions of different magnitudes from direct and indirect transitions, as discussed in detail in Section 3.6. Beyond the absorption edge, strong features are seen associated with **higher interband transitions** to conduction bands at higher energies. At much higher energies, x-ray photons have sufficient energy to liberate strongly bound valence electrons (in **x-ray photoemission**), and **Auger electrons** may be liberated subsequently as atoms relax to the ground state.

Absorption edges are often adorned with exponentially decaying **Urbach tails**, due to inhomogeneities or thermal disorder which perturb the band edges (see Section 3.8). Alter-

natively the tail may comprise a continuum of optically-active, shallow defect or impurity states proximal to the band edges[88]. At low temperatures in high-quality material, electrons and holes form bound-states called **excitons** (see Section 3.9), resulting in discrete lines below E_g which converge into a continuum producing enhanced absorption strength near E_g.

In **Rayleigh scattering**, photons are elastically scattered by inhomogeneities much smaller than their wavelength λ. Rayleigh scattering is $\sim 10^6$ times stronger than Raman scattering[89] (the inelastic scattering of a photon and optical phonon), and attenuates light[87] as $\exp(-\alpha x)$ with $\alpha = C/\lambda^4$, so becomes more important with increasing photon energy (explaining why the sky is blue), and may complicate the determination of absorption spectra.

Absorption features found in the middle of the band gap are generally undesirable for photovoltaics. At finite temperature in most semiconductors, some charge carriers are always present in the conduction and valence bands due to ionised dopants, impurities or defect states. In **free carrier absorption**, these carriers are excited into higher states within the same band. Such processes need a phonon to conserve momentum (going as $\alpha \propto \lambda^2$) so become important at low energy[44]. Similarly in p-type material, a weak absorption ($\sim 100\,\mathrm{cm}^{-1}$) is seen as holes in the valence band make **interband transitions** to lower valence bands. **Deep impurity or defect** levels may also present absorption features in the middle of the gap.

In conducting materials, no static electric field may exist as charge carriers always reconfigure to cancel any applied field; however, high-frequency electromagnetic waves (oscillating electric fields) may exist, as there is insufficient time for reconfiguration to occur. The **plasma frequency** defines a point above which light may propagate[87]: below which absorption is finite and reflectivity approaches 100 %.

At low energies (1-100 meV) in polar crystals, light may be absorbed through the excitation of vibrational modes associated with transverse optical phonons, which leave characteristic **reststrahlen** features in reflectivity spectra. More complicated but weaker **multi-phonon** processes, involving the absorption of a photon (possibly with additional phonons) with the emission of one or more phonons (and their harmonics), are feasible and also enable lattice absorption in non-polar materials such as silicon[44]. At the same energy scale, absorption may occur due to electronic excitations at **shallow impurities** and **defects**[44], and at **local vibrational modes**: due to vibrational modes associated with lattice impurities[90].

With the exception of the plasma edge, multi-phonon processes and local vibrational modes, all of the remaining phenomena depicted in Figure 21 play some direct or indirect role in the work described in this thesis. To study these processes further, it is useful to revise the solutions of Maxwell's equations for waves in absorbing media. These equations relate foundational optical concepts: such as the wavevector, dielectric function, refractive index, extinction and absorption coefficients, and the electric field amplitude. Then, after recalling the Fresnel coefficients and Poynting relation, we have sufficient tools to predict the propagation of energy in optical transmission and reflection experiments, and to attempt to *invert* experimental spectra to acquire material parameters: such as absorption spectra.

3.3 Electromagnetic waves in homogeneous media

Consider Maxwell's equations which embody the laws of Gauss, Faraday and Ampère[91,92]

$$\nabla \cdot \mathbf{D} = \rho(\mathbf{r}) \qquad \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \qquad \nabla \times \mathbf{H} = \mathbf{j} + \frac{\partial \mathbf{D}}{\partial t} \qquad (3.1)$$

with the charge density $\rho(\mathbf{r})$, magnetic fields $\mathbf{B}$ and $\mathbf{H}$ and the electric field $\mathbf{E}$, displacement $\mathbf{D}$ and current density $\mathbf{j}$. A vector relation may be used to simply $\nabla \times \nabla \times \mathbf{E}$

$$\nabla \times \nabla \times \mathbf{E} = \nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} = -\frac{\partial}{\partial t}(\nabla \times \mathbf{B}) \qquad (3.2)$$

In homogeneous material $\mathbf{D} = \epsilon_0 \epsilon_r \mathbf{E}$ and $\mathbf{B} = \mu_0 \mu_r \mathbf{H}$, where $\mu_0 \mu_r$ and $\epsilon_0 \epsilon_r$ are the electric permittivity and magnetic permeability. Rewriting $\nabla \times \mathbf{H}$ with these relations and $\mathbf{j} = \sigma \mathbf{E}$

$$\nabla \times \mathbf{B} = \mu_0 \mu_r \mathbf{j} + \mu_0 \mu_r \frac{\partial \mathbf{D}}{\partial t} = \mu_0 \mu_r \sigma \mathbf{E} + \mu_0 \mu_r \epsilon_0 \epsilon_r \frac{\partial \mathbf{E}}{\partial t} \qquad (3.3)$$

with σ the conductivity. Using (3.3) in (3.2) with no free charges $\nabla \cdot \mathbf{E} = 0$ gives a wave equation[41]

$$\nabla^2 \mathbf{E} = \mu_0 \mu_r \sigma \frac{\partial \mathbf{E}}{\partial t} + \mu_0 \mu_r \epsilon_0 \epsilon_r \frac{\partial^2 \mathbf{E}}{\partial t^2} \qquad (3.4)$$

A solution $\mathbf{E}(\mathbf{x}, t) = \mathbf{E}_0 \, e^{i(\tilde{\mathbf{k}}\mathbf{x} - \omega t)}$ with a complex wavevector $\tilde{\mathbf{k}} = \mathbf{k}_1 + i\mathbf{k}_2$ gives a wave $\mathbf{E}(\mathbf{x}, t) = \mathbf{E}_0 \, e^{i(\mathbf{k}_1 \mathbf{x} - \omega t)} e^{-\mathbf{k}_2 \mathbf{x}}$ with a spatially attenuated amplitude*. Using this in (3.4) gives

$$\tilde{\mathbf{k}}^2 = i\omega\mu_0\mu_r\sigma + \omega^2\mu_0\mu_r\epsilon_0\epsilon_r = \omega^2\epsilon_0\mu_0\mu_r\left(\frac{i\sigma}{\omega\epsilon_0} + \epsilon_r\right) \qquad (3.5)$$

Now $\tilde{\mathbf{k}}^2 = (k_1^2 - k_2^2) + i2\mathbf{k}_1 \cdot \mathbf{k}_2$. Using this in (3.5) and equating real and imaginary parts gives

$$k_1^2 = k_2^2 + \omega^2\mu_0\mu_r\epsilon_0\epsilon_r \qquad\qquad 2\mathbf{k}_1 \cdot \mathbf{k}_2 = \omega\mu_0\mu_r\sigma \qquad (3.6)$$

Assuming $\mathbf{k}_1 \cdot \mathbf{k}_2 = k_1 k_2$, eliminating k_2 in (3.6) and solving the quadratic for k_1^2 gives

$$k_1 = \omega\sqrt{\frac{\mu\epsilon}{2}}\sqrt{1 + \sqrt{1 + \frac{\sigma^2}{\omega^2\epsilon^2}}} \qquad\qquad k_2 = \sigma\sqrt{\frac{\mu}{2\epsilon}}\left[1 + \sqrt{1 + \frac{\sigma^2}{\omega^2\epsilon^2}}\right]^{-0.5} \qquad (3.7)$$

where we have relabelled $\mu = \mu_o\mu_r$ and $\epsilon = \epsilon_0\epsilon_r$. For poor conductors $\sigma^2/\omega^2\epsilon^2 \to 0$, so $\tilde{k} \approx \omega\sqrt{\mu\epsilon} + i\sigma\sqrt{\mu/4\epsilon}$, while for good conductors $\sigma^2/\omega^2\epsilon^2 \to \infty$, giving $\tilde{k} \approx (1+i)\sqrt{\omega\mu\sigma/2}$. The bracketed term in (3.5) suggests a complex dielectric constant $\tilde{\epsilon}_r$ including ohmic losses

$$\left|\tilde{\mathbf{k}}\right| = \omega\sqrt{\epsilon_0\mu_0\mu_r\tilde{\epsilon}_r} = \omega\sqrt{\epsilon_0\mu_0\mu_r\left(\frac{i\sigma}{\omega\epsilon_0} + \epsilon_r\right)} \qquad (3.8)$$

The wave's phase velocity v is then $dx/dt = \omega/\tilde{k}$, which for the vacuum gives $c = 1/\sqrt{\epsilon_0\mu_0}$. For absorbing media, this leads to a complex refractive index $\tilde{n} = n + i\kappa$

$$\tilde{n} = \frac{c}{v} = \sqrt{\mu_r\tilde{\epsilon}_r} = \frac{cB}{E} = c\,\mu_0\,\mu_r\,\frac{H}{E} \qquad (3.9)$$

---

* Some care is needed in the sign choice for the imaginary part of the refractive index, dielectric constant and wavevector (usually only one of which need be considered). The choice must correspond to the sign used in the spatial part of the plane wave phase if ohmic contributions are to produce attenuation, e.g. a wave $\exp(i\{\omega t - \mathbf{k} \cdot \mathbf{x}\})$ requires $\tilde{n} = n - i\kappa$.
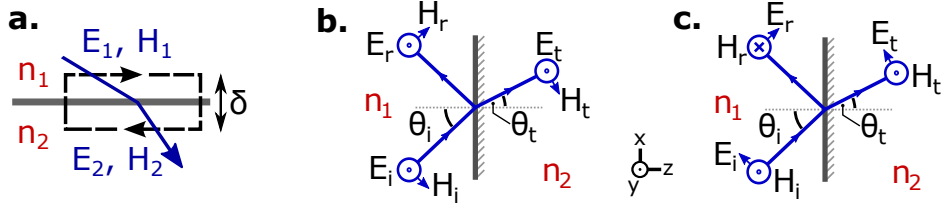
Figure 22: Electric E and magnetic H fields at dielectric boundaries. The media have refractive indices n_1 and n_2. The symbols $\otimes$ and $\odot$ denote vectors pointing respectively into and out of the page.

3.4 Dielectric boundaries and the Fresnel coefficients

Optical experiments use electric and magnetic fields to study materials and their interfaces. Maxwell's equations describe the behaviour of the electric $\mathbf{E}$ and magnetic $\mathbf{H}$ fields at dielectric boundaries. Consider Figure 22a, which shows an electric field crossing such a boundary. A rectangular circuit L of height δ and length ℓ (with area S) traverses the boundary. Integrating Maxwell's equation $\nabla \times \mathbf{E}$ over the area S and using the Stokes' theorem gives

$$\int_{S(L)} \nabla \times \mathbf{E}\, d\mathbf{s} = \int_{S(L)} -\frac{\partial \mathbf{B}}{\partial t}\, d\mathbf{s} = \oint_{L(S)} \mathbf{E} \cdot d\mathbf{l} \xrightarrow{\delta \to 0} (E_{1,p} - E_{2,p})\,\ell = 0 \qquad \therefore \Delta E_p = 0 \quad (3.10)$$

where in the right-hand side the height δ is reduced to the vanishing limit and $E_{1,p}$ and $E_{2,p}$ denote the components of the fields parallel to the boundary. Equation (3.10) requires the continuity of the $\mathbf{E}$ field parallel to the boundary; a similar relation exists for the $\mathbf{H}$ field.

Now consider Figure 22b which shows electromagnetic waves incident upon a boundary between media with refractive indices n_1 and n_2. The electric field is polarised perpendicular to the plane of incidence; the $\mathbf{E}$ and $\mathbf{H}$ fields parallel to the boundary are continuous

$$E_i + E_r = E_t \qquad \text{and} \qquad H_r \cos \theta_i - H_i \cos \theta_i = -H_t \cos \theta_t \qquad (3.11)$$

Recalling the optical impedance $Z = E/H = \mu E/B = \sqrt{\mu_0 \mu_r / \epsilon_0 \epsilon_r} = \mu_0 \mu_r c / n$ gives

$$\frac{n_1}{\mu_{r,1}} \left(E_r \cos \theta_i - E_i \cos \theta_i \right) = -\frac{n_2}{\mu_{r,2}} E_t \cos \theta_t \qquad (3.12)$$

which with (3.11) gives the Fresnel (s-polarised or TE) reflection and transmission coefficients

$$r_s = \frac{E_r}{E_i} = \frac{n_1 \cos \theta_1 - n_2 \cos \theta_2}{n_1 \cos \theta_1 + n_2 \cos \theta_2} \qquad\qquad t_s = \frac{E_t}{E_i} = \frac{2 n_1 \cos \theta_1}{n_1 \cos \theta_1 + n_2 \cos \theta_2} \qquad (3.13)$$

where non-magnetic media are assumed, $\mu_{r,1} = \mu_{r,2}$. Now consider Figure 22c where the electric field is polarised parallel to the plane of incidence. Using the continuity of the $\mathbf{E}$ and $\mathbf{H}$ fields parallel to the interface and again the relation $H = nE/(\mu_0 \mu_r c)$ gives

$$\frac{n_1}{\mu_{r,1}} \left(E_i - E_r \right) = \frac{n_2}{\mu_{r,2}} E_t \qquad \text{and} \qquad E_i \cos \theta_i + E_r \cos \theta_i = E_t \cos \theta_t \qquad (3.14)$$

The Fresnel (p-polarised or TM) reflection and transmission coefficients result from (3.14)

$$r_p = \frac{E_r}{E_i} = \frac{n_1 \cos \theta_2 - n_2 \cos \theta_1}{n_1 \cos \theta_2 + n_2 \cos \theta_1} \qquad\qquad t_p = \frac{E_t}{E_i} = \frac{2 n_1 \cos \theta_1}{n_1 \cos \theta_2 + n_2 \cos \theta_1} \qquad (3.15)$$

Equations (3.13) and (3.15) relate the $\mathbf{E}$ field amplitudes of light reflected from or transmitted through a boundary to **the physical properties of the media comprising the boundary**.

The Poynting relation $\mathbf{S} = \mathbf{E} \times \mathbf{H}$ expresses the irradiance (the power per unit area) of an electromagnetic wave[41,91,92]. Carefully using the real parts of $\mathbf{E}$ and $\mathbf{H}$ in this relation with the solution $\mathbf{E}(\mathbf{x}, t)$ below (3.4), then converting $\mathbf{H}$ to $\mathbf{E}$ via (3.9) and taking the time average[†], allows the irradiance to be written entirely in terms of the electric field amplitude E_0

$$\left| \langle \mathbf{S} \rangle \right| = \left| \langle \operatorname{Re} \mathbf{E} \times \operatorname{Re} \mathbf{H} \rangle \right| = \frac{1}{2} \left| \operatorname{Re} \left(\mathbf{E}_0 \times \mathbf{H}_0^* \right) \right| = \frac{E_0^2 \operatorname{Re} \tilde{n}^*}{2 \mu_0 \mu_r c} \tag{3.16}$$

where the amplitudes $\mathbf{E}_0$ and $\mathbf{H}_0$ are orthogonal. The reflectivity R and transmissivity T, respectively **the fractions of power reflected from and transmitted through an interface**, then result directly from the Poynting relation (3.16) and the Fresnel relations (3.13) and (3.15)

$$R_s = |r_s|^2; \quad T_s = \frac{\mu_1 \operatorname{Re}\left(\tilde{n}_2 \cos\theta_2 \right)}{\mu_2 \, n_1 \, \cos\theta_1} |t_s|^2 \quad \text{and} \quad R_p = |r_p|^2; \quad T_p = \frac{\mu_1 \operatorname{Re}\left(\tilde{n}_2^* \cos\theta_2 \right)}{\mu_2 \, n_1 \, \cos\theta_1} |t_p|^2 \tag{3.17}$$

where the incident medium is assumed absorption-free and the pre-factors in the transmissivities appear as the incident and transmitted waves reside in different media, while for the reflectivities, both waves propagate in the same medium and so these factors cancel identically. Equation (3.17) conceals various subtleties[‡]; the cosine terms appear as the Poynting relation describes an energy flux and the cross-sectional areas of the beams before and after the interface differ according to $\mathbf{S} \cdot \hat{\mathbf{z}}$ (note also the often-missed complex conjugate $\tilde{n}_2^*$).

Frequently the aim of research is to determine the absorption spectrum $\alpha(\hbar\omega)$; the absorption coefficient α is the reciprocal of the distance a wave must travel before it's intensity diminishes to $1/\mathrm{e}$ of its initial value, which is seen in the Beer–Lambert relation

$$I = I_0 \exp(-\alpha x) \tag{3.18}$$

The absorption coefficient and complex refractive index $\tilde{n} = n + i\kappa$ (with κ the extinction coefficient) are simply related by $\kappa = \alpha/2k$ (with k the wavevector), seen for the solution below (3.4) if $\tilde{\mathbf{k}} \to \tilde{n}k$. An important relation in optics is the Snell law, relating the angles of the incident θ_1 and transmitted θ_2 waves to the refractive indices n_1 and n_2 of the media

$$\tilde{n}_1 \sin\theta_1 = \tilde{n}_2 \sin\theta_2 \tag{3.19}$$

If the left side of (3.19) is real and the transmitted medium is absorbing, then $\tilde{n}_2$ and θ_2 must be complex if the relation is to hold. This is true in absorbing media; rather than a plane wave, an inhomogeneous wave propagates in the medium: so-called because the planes of constant phase and constant amplitude are inclined relative to each other (the latter generally being parallel to the boundary). Various works discuss inhomogeneous waves further[93,95].

[†]The Poynting relation in complex fields $\langle \operatorname{Re} \mathbf{E} \times \operatorname{Re} \mathbf{H} \rangle = \frac{1}{2} \operatorname{Re}\left(\mathbf{E}_0 \times \mathbf{H}_0^* \right)$ is shown on page 33 of Born and Wolf[41].

[‡] For the s- and p-polarisations of Fig. 22, we have $\{E_x = 0, H_x = -H\cos\theta\}$ and $\{H_x = 0, E_x = E\cos\theta\}$ so that

$$\hat{\mathbf{z}} \cdot \operatorname{Re}\mathbf{E} \times \operatorname{Re}\mathbf{H} = \hat{\mathbf{z}} \cdot \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ \operatorname{Re}E_x & \operatorname{Re}E_y & \operatorname{Re}E_z \\ \operatorname{Re}H_x & \operatorname{Re}H_y & \operatorname{Re}H_z \end{vmatrix} = \operatorname{Re}E_x \operatorname{Re}H_y - \operatorname{Re}E_y \operatorname{Re}H_x = \begin{cases} \text{s-polarised}: \ \operatorname{Re}(E)\operatorname{Re}(H\cos\theta) \\ \text{p-polarised}: \ \operatorname{Re}(E\cos\theta)\operatorname{Re}(H) \end{cases}$$

from which (3.17) results after inserting these into a time-average as (3.16) and dividing by the incident irradiance. Because the quantity $\mathbf{S} \cdot \hat{\mathbf{z}}$ is equivalently the Poynting vector component *normal* to the interface, the same results are perhaps found more elegantly after rewriting the Fresnel equations in only field components *parallel* to the interface: then called the thin film equations[93]. The incident and reflected waves are not independent when the incident media is absorbing[93,94,95].

3.5 Microscopic treatment of the dielectric function

In this section the microscopic theory for the dielectric function is briefly presented. Such an analysis sheds light on various aspects of optical refraction and absorption phenomena: such as vibrational and phonon modes and plasma resonances. The generality of this analysis covers electrons, phonons and semiconductor phenomena in one treatment.

Consider a system of bound charged particles, such as valence electrons bound to their nuclei or positive and negative ions bound together within a crystal lattice. The oscillating electric field of an incident electromagnetic plane wave induces the particles to oscillate, as described by the equation of motion[87]

$$m\frac{\partial^2 \mathbf{x}}{\partial t^2} + m\gamma\frac{\partial \mathbf{x}}{\partial t} + \kappa\mathbf{x} = -qE_0 e^{-i\omega t} \tag{3.20}$$

where m, q and $\mathbf{x}$ are respectively the particle mass, charge and position vector at time t; E_0 and ω are the amplitude and angular frequency of the incident plane wave; while κ and γ are the restoring force and damping constants. This differential equation has a solution $\mathbf{x}(t) = \mathbf{x}_0 e^{-i\omega t}$. Using this in (3.20) gives

$$m\left(-\omega^2\mathbf{x_0}e^{-i\omega t}\right) + m\gamma\left(-i\omega\mathbf{x_0}e^{-i\omega t}\right) + \kappa\,\mathbf{x_0}e^{-i\omega t} = -qE_0 e^{-i\omega t} \tag{3.21}$$

Using the relation $\omega_0^2 = \kappa/m$ with the natural angular frequency ω_0 and rearranging gives

$$\mathbf{x_0} = \frac{qE_0}{m} \cdot \frac{1}{\omega^2 - \omega_0^2 + i\omega\gamma} \tag{3.22}$$

This gives for the electric polarisation[87]

$$\mathbf{P}(t) = -Nq \cdot \mathbf{x}(t) = -\frac{Nq^2\mathbf{E_0}}{m} \cdot \frac{e^{-i\omega t}}{\omega^2 - \omega_0^2 + i\omega\gamma} \tag{3.23}$$

Now, a material's bulk polarisation has contributions from various microscopic processes, and it is helpful to write these as separate contributions to the electric displacement

$$\mathbf{D} = \epsilon_0\mathbf{E} + \mathbf{P}_1 + \mathbf{P}_2 + \cdots = \epsilon_0(1 + \chi_1 + \chi_2 + \cdots)\mathbf{E} = \epsilon_0\epsilon_r\mathbf{E}(t) \tag{3.24}$$

where χ_i are the electric susceptibilities. Often just one specific phenomena is of interest and all remaining processes are hidden within a catch-all susceptibility χ_{media}, so that

$$\mathbf{D}(t) = \epsilon_0(1 + \chi_{\text{media}} + \chi_{\text{resonant}})\mathbf{E}(t) = \epsilon_0\left(1 + \chi_{\text{media}} - \frac{Nq^2}{m\epsilon_0} \cdot \frac{1}{\omega^2 - \omega_0^2 + i\omega\gamma}\right)\mathbf{E}(t) \tag{3.25}$$

Thus the relative electric permittivity or dielectric function $\epsilon_r(\omega)$ is

$$\epsilon_r(\omega) = 1 + \chi_{\text{media}} - \frac{Nq^2}{m\epsilon_0} \cdot \frac{1}{\omega^2 - \omega_0^2 + i\omega\gamma} = 1 + \chi_{\text{media}} - \frac{\omega_p^2}{\omega^2 - \omega_0^2 + i\omega\gamma} \tag{3.26}$$

where the plasma frequency w_p is defined by $\omega_p^2 = {}^{Nq^2}/m\epsilon_0$. The dielectric function is sometimes written in terms of the static $\epsilon_{\text{st}} \equiv \epsilon(0)$ and high frequency $\epsilon_\infty \equiv \epsilon(\infty)$ limits, which
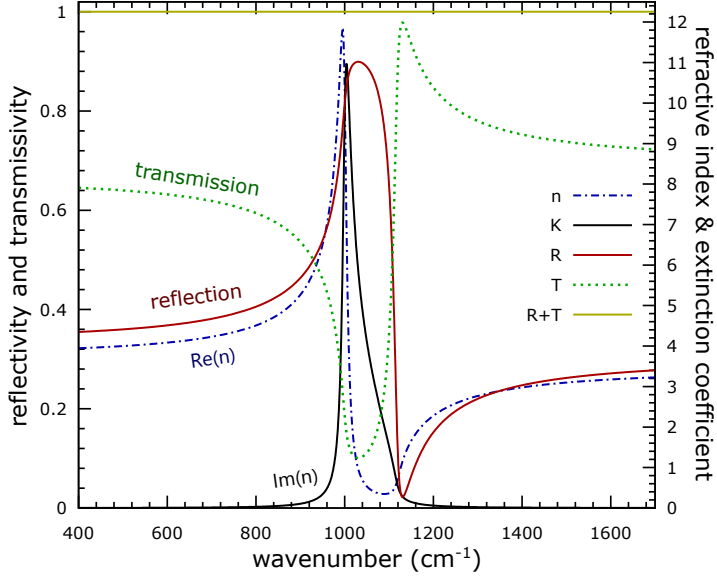
Figure 23: Reflection R and transmission T at boundary with absorbing medium with dielectric parameters: $w_0 = 1000\,\text{cm}^{-1}$, $\gamma = 15\,\text{cm}^{-1}$, $\epsilon_{\text{st}} = 15$ and $\epsilon_\infty = 12$ (see the text).

encapsulate the above considerations and may be more suitable for measurement

$$\epsilon_r(\omega) = \epsilon_\infty - (\epsilon_{\text{st}} - \epsilon_\infty)\frac{\omega_0^2}{\omega^2 - \omega_0^2 + i\omega\gamma} \tag{3.27}$$

or alternatively, separating the real and imaginary parts

$$\epsilon_r(\omega) = \epsilon_\infty - (\epsilon_{\text{st}} - \epsilon_\infty)\frac{\omega_0^2(\omega^2 - \omega_0^2 - i\omega\gamma)}{(\omega^2 - \omega_0^2)^2 + \omega^2\gamma^2} \tag{3.28}$$

Figure 23 uses (3.27), (3.17), (3.19) and (3.9) to plot the reflectivity R, transmissivity T and refractive index at a boundary between the vacuum and an absorbing medium at normal incidence. As the boundary is infinitesimal, no energy is absorbed: so the sum $R + T = 1$. Figure 24 additionally uses (3.93) to plot R and T for a similar $2\,\mu\text{m}$ incoherent single layer. Note that the maximum of R and minimum of T now sit at different wavenumbers.
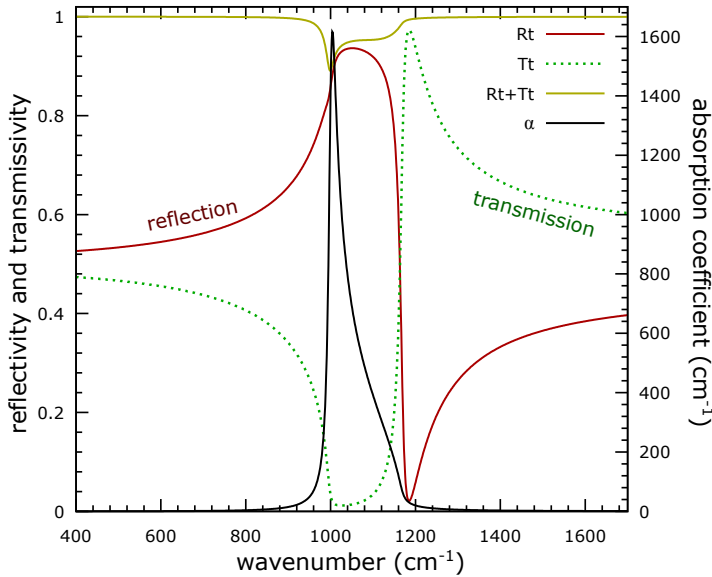
Figure 24: Reflection R, transmission T and absorption α at an incoherent single layer with $2\,\mu\text{m}$ thickness and dielectric parameters: $w_0 = 1000\,\text{cm}^{-1}$, $\gamma = 15\,\text{cm}^{-1}$, $\epsilon_{\text{st}} = 15$ and $\epsilon_\infty = 11$ (see the text).

3.6 Interband optical absorption in semiconductors

This section discusses interband optical absorption in semiconductors. While many texts include non-trivial treatments of the topic, e.g. Yu & Cardona[44], Hamaguchi[96], Bassani[97] and Basu[43], each has some peculiarity, typographical error, indulgence or omission which complicates the exposition. The intention here is to present the main ideas plainly, in compact form without unnecessary diversions. Before starting, it is beneficial to revisit time-dependent perturbation theory: as this work is the foundation for all subsequent results; we broadly follow Eisberg & Resnick[14], Zettili[98] and Sakurai & Napolitano[99]. Next the joint density of states and Van Hove singularities are introduced before direct absorption is discussed in some detail. We conclude with a look at indirect absorption.

3.6.1 Time-dependent perturbation theory

Time-dependent perturbation theory offers approximate solutions to problems which are close to systems with known solutions[98]. Consider the Hamiltonian $\mathcal{H}(x,t)$

$$\mathcal{H}(x,t) = \mathcal{H}_0(x) + V(x,t) \tag{3.29}$$

where the time-dependent perturbation $V(x,t)$ is small in comparison to the unperturbed Hamiltonian $\mathcal{H}_0(x)$, which has exactly-known time-independent eigenstates $|\phi_m(x)\rangle$

$$\mathcal{H}_0(x)|\psi_m(x,t)\rangle = i\hbar\frac{\partial}{\partial t}|\psi_m(x,t)\rangle \qquad \text{with} \qquad |\psi_m(x,t)\rangle = \mathrm{e}^{-iE_m t/\hbar}|\phi_m(x)\rangle \tag{3.30}$$

The explicit parameter dependence of the stationary states $|\psi_m\rangle$ will be omitted subsequently. The general solution $|\Psi(x,t)\rangle$ to (3.29) may be expanded in a basis over the $|\psi_m\rangle$

$$|\Psi(x,t)\rangle = \sum_m c_m(t)|\psi_m\rangle \tag{3.31}$$

Inserting $|\Psi(x,t)\rangle$ and (3.29) in the time-dependent Schrödinger equation, $\mathcal{H}|\Psi\rangle = i\hbar\frac{\partial}{\partial t}|\Psi\rangle$

$$\sum_m c_m(t)\left(\left\{\mathcal{H}_0(x) - i\hbar\frac{\partial}{\partial t}\right\} + V(x,t)\right)|\psi_m\rangle = i\hbar\sum_m \frac{\mathrm{d}c_m}{\mathrm{d}t}|\psi_m\rangle \tag{3.32}$$

where the term in braces $\{\cdot\}$ is just the time-independent Schrödinger equation (3.30) and so vanishes. Pre-multiplying by $\langle\psi_f|$ and using orthogonality $\langle\psi_f|\psi_m\rangle = \delta_{fm}$ gives

$$\sum_m c_m(t)\langle\psi_f|V(x,t)|\psi_m\rangle = i\hbar\sum_m \frac{\mathrm{d}c_m}{\mathrm{d}t}\langle\psi_f|\psi_m\rangle = i\hbar\frac{\mathrm{d}c_f}{\mathrm{d}t} \tag{3.33}$$

With the system prepared in the state $|\psi_i\rangle$ and the perturbation $V(x,t)$ switched on at time $t = 0$, the coefficient $c_f(t)$ is now needed at time $\tau > 0$, which gives **the probability of a transition to a state** $|\psi_f\rangle$. The following approximations hold to first-order[14]

$$c_m(0) = \begin{cases} 1, & m = i \\ 0, & m \neq i \end{cases} \qquad \text{and} \qquad c_m(t) \text{ is } \begin{cases} \simeq 1, & m = i \\ \ll 1, & m \neq i \end{cases} \tag{3.34}$$

These conditions (3.34) make the sum in the left-hand side of (3.33) entirely dominated by the coefficient $c_i(t)$. Integrating with respect to time from $t = 0$ to τ and using (3.34) gives

$$\left[c_f(t)\right]_0^\tau = c_f(\tau) = -\frac{i}{\hbar} \int_0^\tau \underbrace{\langle \phi_f | V(x,t) | \phi_i \rangle}_{V_{fi}(t)} e^{i\omega_{fi}t}\, dt \tag{3.35}$$

where the eigenstates of the unperturbed Hamiltonian (3.30) are used to explicity write the energy separation between the initial and final states, $\hbar\omega_{fi} = E_f - E_i$ (and $V(x,t)$ is assumed to have no operator $\partial/\partial t$). The transition probability $P_{if}(t)$ at time t to first order is

$$P_{if}(t) = |c_f(t)|^2 = \left| -\frac{i}{\hbar} \int_0^t V_{fi}(t')\, e^{i\omega_{fi}t'}\, dt' \right|^2 \tag{3.36}$$

For a **time-independent perturbation** $V(x,t) = V(x)$ the transition probability becomes

$$P_{if}(t) = \frac{|V_{fi}|^2}{\hbar^2} \left| \int_0^t e^{i\omega_{fi}t'} dt' \right|^2 = \frac{|V_{fi}|^2}{\hbar^2} \left| \frac{e^{i\omega_{fi}t}-1}{i\omega_{fi}} \right|^2 = \frac{4\,|V_{fi}|^2}{(\hbar\omega_{fi})^2} \sin^2 \frac{\omega_{fi}t}{2} \tag{3.37}$$

The sine term in (3.37) causes the transition probability to oscillate as a function of time. We may rewrite (3.37) by recalling some properties of the delta function[98]

$$\delta(y) = \lim_{t\to\infty} \frac{\sin^2(yt)}{\pi y^2 t} \qquad \text{and} \qquad \delta(ax) = \frac{1}{|a|}\delta(x) \tag{3.38}$$

For large $t \gg 2\pi/\omega$, sequentially using the relations of (3.38) in (3.37) gives a probability

$$P_{if}(t) = \frac{\pi t}{\hbar^2} |V_{fi}|^2\, \delta\left(\frac{\omega_{fi}}{2}\right) = \frac{2\pi t}{\hbar} |V_{fi}|^2\, \delta(\hbar\omega_{fi}) \tag{3.39}$$

so that the **transition probability per unit time**, the transition rate W_{if}, is

$$W_{if} = \frac{dP_{if}}{dt} = \frac{2\pi}{\hbar} |V_{fi}|^2\, \delta(\hbar\omega_{fi}) = \frac{2\pi}{\hbar} |V_{fi}|^2\, \delta(E_f - E_i) \tag{3.40}$$

Note that (3.38) shows that the dimensions of the Dirac delta function are the reciprocal of its argument, so these last two equations are valid via dimensional analysis. With a **continuum of final states** at energies E_f, the transition rate (3.40) is conveniently written[99]

$$W_{if} = \int \frac{dP_{if}}{dt} \rho(E_f)\, dE_f = \frac{2\pi}{\hbar} \overline{|V_{fi}|^2} \int \delta(E_f - E_i)\, \rho(E_f)\, dE_f = \frac{2\pi}{\hbar} \overline{|V_{fi}|^2}\, \rho(E_f)\Big|_{E_f \simeq E_i}$$

where $\rho(E_f) = \frac{dN}{dE}$ is the density of final states and $\overline{|V_{fi}|^2}$ denotes an energy average over the interval around E_i. This result is the well-known **Fermi Golden rule**. Let's now consider a **harmonic perturbation** $V(x,t)$

$$V(x,t) = U(x)e^{i\omega t} + U^\dagger(x)e^{-i\omega t} \tag{3.41}$$

where $U(x)$ is a time-independent operator; writing the perturbation in this form (the sum of a term and its complex conjugate) guarantees that $V(x,t)$ is a real quantity. We take care to discriminate between the frequency ω of the driving harmonic perturbation and that of

ω_{fi}, previously associated with the energy difference between the initial and final states. Using the harmonic perturbation of (3.41) in the transition probability (3.36) and following our previous steps through (3.40) gives for the **transition rate**

$$W_{if} = \frac{2\pi}{\hbar} \left[|U_{fi}|^2 \delta(E_f - E_i + \hbar\omega) f_f (1 - f_i) + |U_{fi}^\dagger|^2 \delta(E_f - E_i - \hbar\omega) f_i (1 - f_f) \right] \quad (3.42)$$

where the additional Fermi functions $f_i \equiv f(E_i)$ and $f_f \equiv f(E_f)$ account for the occupancies of the initial and final states[43]. This equation suggests that transitions are most probable for

$$E_f = \begin{cases} E_i - \hbar\omega & \text{(stimulated emission)} \\ E_i + \hbar\omega & \text{(absorption)} \end{cases} \quad (3.43)$$

We are mainly interested in absorption phenomena, given by the second term in (3.42).

3.6.2 Joint density of states

The familiar density of states $g(E)$ counts the number of states (per unit volume) in some energy interval from E to $E + dE$ for a *single* band; the *joint* density of states applies when both initial and final states lie within continuous bands[87] with an energy separation $E_{cv} = E_c - E_v$. The number of states δN in an infinitesimal k-space volume element δV_k is then

$$\delta N = \frac{\delta V_k}{(2\pi)^3} = \frac{\delta S_k}{(2\pi)^3} \cdot |\nabla_{\mathbf{k}}(E_{cv})|^{-1} \cdot dE_{cv} \quad (3.44)$$

where δS_k is the infinitesimal k-space area element for the surface S_k with constant energy E_{cv}. Dividing by dE_{cv} and integrating over the surface $S_k(E_{cv})$ gives the density of states

$$g(E_{cv}) = \frac{1}{4\pi^3} \int_{S_k(E_{cv})} \frac{dS_k}{|\nabla_{\mathbf{k}}(E_{cv})|} \quad (3.45)$$

where a factor of two has been included for spin-degeneracy. This expression is valid for any band. Singularities exist in the integrand of (3.45) where the bands are parallel and $\nabla_{\mathbf{k}}(E_{cv}) = 0$. By expanding E_{cv} in a Taylor series to second order in $\mathbf{k}$ about such a point $\mathbf{k_0}$,

$$E_{cv}(\mathbf{k}) = E_{cv}(\mathbf{k_0}) + (\mathbf{k} - \mathbf{k_0}) \cdot \underbrace{\nabla_{\mathbf{k}} E_{cv}(\mathbf{k_0})}_{=0} + \frac{1}{2} \left((\mathbf{k} - \mathbf{k_0}) \cdot \nabla_{\mathbf{k}} \right)^2 E_{cv}(\mathbf{k_0}) + O(\mathbf{k^3}) \quad (3.46)$$

Van Hove classified the M_0, M_1, M_2 and M_3 critical points according to the respective number of negative components of the vector $(\partial_{kx}^2, \partial_{ky}^2, \partial_{kz}^2) E_{cv}$; the M_0 and M_3 points are respectively energy minima and maxima, whilst M_1 and M_2 denote saddle points. To evaluate an analytic expression for the joint density of states, consider an M_0 critical point with parabolic and spherically symmetric valence and conduction bands described by respective effective masses m_v^* and m_c^* (noting $m_v^* = -m_h^*$). The joint (direct) transition energy is

$$E_{cv} = (E_g + E_c) - E_v = E_g + \frac{(\hbar k)^2}{2m_e^*} + \frac{(\hbar k)^2}{2m_h^*} = E_g + \frac{(\hbar k)^2}{2\mu} \quad (3.47)$$

with the reduced mass $\mu = (m_e^* m_h^*)/(m_e^* + m_h^*)$. To evaluate the joint density of states (3.45) is used while taking advantage of the spherical symmetry; the area element dS_k in spherical

polar coordinates is $k^2 \sin\theta \, d\phi \, d\theta$. With $\nabla_{\mathbf{k}}(E_{cv}) = \hbar^2 k/\mu$, this gives in (3.45) with (3.47)

$$g(E_{cv}) = \frac{k\mu}{4\pi^3\hbar^2} \int_0^\pi \sin\theta \, d\theta \int_0^{2\pi} d\phi = \frac{\sqrt{2\mu^3}}{\pi^2\hbar^3} \sqrt{E_{cv} - E_g} \tag{3.48}$$

3.6.3   Absorption in direct-gap semiconductors

The Hamiltonian for a particle with charge q and mass m in an electromagnetic field is[44,100]

$$\mathcal{H} = \frac{1}{2m}(\mathbf{p} + q\mathbf{A})^2 + V(\mathbf{r}) = \left[\frac{\mathbf{p}^2}{2m} + V(\mathbf{r})\right] + \frac{q}{m}\mathbf{p}\cdot\mathbf{A} + \frac{q^2\mathbf{A}^2}{2m} = \mathcal{H}_0 + \mathcal{H}_{er} + O(\mathbf{A}^2) \tag{3.49}$$

where $\mathbf{A}$ is the vector potential and $\mathcal{H}_{er} = \frac{q}{2m}(\mathbf{p}\cdot\mathbf{A} + \mathbf{A}\cdot\mathbf{p}) = \frac{q}{m}\mathbf{p}\cdot\mathbf{A}$ is the electron-radiation interaction Hamiltonian written in the Coulomb gauge[44]. For small $\mathbf{A}$, the term $\mathcal{H}_{er}$ is a small perturbation. Consider an electromagnetic wave polarised parallel to the unit vector $\hat{\mathbf{a}}$

$$\mathbf{A}(\mathbf{r}, t) = \hat{\mathbf{a}}\frac{A_0}{2}\left(e^{i(\mathbf{q}\cdot\mathbf{r}-\omega t)} + e^{-i(\mathbf{q}\cdot\mathbf{r}-\omega t)}\right) \tag{3.50}$$

Upon inserting this into the electron-radiation Hamiltonian, the pre-factor of $e^{-i\omega t}$ is then the $U^\dagger(x)$ of (3.41). The absorption transition rate then follows from the absorption term in (3.42)

$$\Gamma_{if} = \frac{2\pi}{\hbar}\frac{e^2|A_0|^2}{4m^2}\left|\left\langle\psi_f\left|e^{i\mathbf{q}\cdot\mathbf{r}}\,\hat{\mathbf{a}}\cdot\mathbf{p}\right|\psi_i\right\rangle\right|^2\delta(E_f - E_i - \hbar\omega)\cdot\underbrace{f(E_i)[1 - f(E_f)]}_{[f(E_i)-f(E_f)]} \tag{3.51}$$

A similar expression[101] for emission results with the first term in (3.42). Noting the symmetry of the matrix element $|p_{fi}| = |p_{if}|$ and the evenness of the delta function $\delta(a) = \delta(-a)$, **the net absorption rate** is then the difference $\Gamma_{if} - \Gamma_{fi}$, giving the simplified right-side of (3.51). The Bloch functions for electrons in the valence and conduction bands are

$$|v\mathbf{k}_v\rangle = \frac{1}{\sqrt{V}}u_{v,\mathbf{k}_v}(\mathbf{r})\,e^{i\mathbf{k}_v\cdot\mathbf{r}} \qquad \text{and} \qquad |c\mathbf{k}_c\rangle = \frac{1}{\sqrt{V}}u_{c,\mathbf{k}_c}(\mathbf{r})\,e^{i\mathbf{k}_c\cdot\mathbf{r}} \tag{3.52}$$

where V is the crystal (normalisation) volume. Using (3.52) in (3.51) gives the matrix element

$$\left\langle\psi_f\left|e^{i\mathbf{q}\cdot\mathbf{r}}\,\hat{\mathbf{a}}\cdot\mathbf{p}\right|\psi_i\right\rangle = \frac{1}{V}\int_V d^3\mathbf{r}\,u_{c,\mathbf{k}_c}^\dagger(\mathbf{r})\,e^{-i\mathbf{k}_c\cdot\mathbf{r}}\,e^{i\mathbf{q}\cdot\mathbf{r}}\,\hat{\mathbf{a}}\cdot\mathbf{p}\,u_{v,\mathbf{k}_v}(\mathbf{r})\,e^{i\mathbf{k}_v\cdot\mathbf{r}} \tag{3.53}$$

Now as $\mathbf{p} = -i\hbar\nabla$, the effect of the momentum operator in (3.53) is

$$\mathbf{p}\,u_{v,\mathbf{k}_v}(\mathbf{r})\,e^{i\mathbf{k}_v\cdot\mathbf{r}} = \left[\mathbf{p}\,u_{v,\mathbf{k}_v}(\mathbf{r})\right]e^{i\mathbf{k}_v\cdot\mathbf{r}} + u_{v,\mathbf{k}_v}(\mathbf{r})\,\hbar\mathbf{k}_v\,e^{i\mathbf{k}_v\cdot\mathbf{r}} \tag{3.54}$$

Rewriting the position vector $\mathbf{r}$ as $\mathbf{R} + \mathbf{r}'$, where $\mathbf{R}$ a primitive lattice vector and $\mathbf{r}'$ is a unit cell vector, then the integral of (3.53) can be written as one over the unit cell (of volume Ω)

$$\text{Eqn. (3.53)} = \frac{1}{V}\sum_{\mathbf{R}}e^{i(\mathbf{k}_v-\mathbf{k}_c+\mathbf{q})\cdot\mathbf{R}}\int_\Omega d^3\mathbf{r}'\,e^{i(\mathbf{k}_v-\mathbf{k}_c+\mathbf{q})\cdot\mathbf{r}'}\,u_{c,\mathbf{k}_c}^\dagger(\mathbf{r}')\,\hat{\mathbf{a}}\cdot(\mathbf{p} + \hbar\mathbf{k}_v)\,u_{v,\mathbf{k}_v}(\mathbf{r}') \tag{3.55}$$

where the periodicity $u_{n,\mathbf{k}}(\mathbf{R}+\mathbf{r}') = u_{n,\mathbf{k}}(\mathbf{r}')$ of the Bloch cellular functions and (3.54) are used. Now, the leading exponential sum can be shown to reduce to a delta function $\delta(\mathbf{w}\cdot\mathbf{R})$, with the result that (3.55) is zero unless $\mathbf{w} = (\mathbf{k}_v - \mathbf{k}_c + \mathbf{q})$ is a reciprocal lattice vector $\mathbf{G}$.   For the dominant term $\mathbf{G} = \mathbf{0}$, the exponential inside the integral of (3.55) goes to unity, and the term $u_{c,\mathbf{k}_c}^\dagger(\mathbf{r})\,u_{v,\mathbf{k}_v}(\mathbf{r})$ vanishes due to the orthogonality of the Bloch functions.

Photon wavevectors are essentially negligible in comparison to those of electrons: the wavevectors of electrons at the Brillouin zone edge ($k = [\pi/5] \times 10^8 \text{ cm}^{-1}$ for lattice parameters of 5 Å) are some $1000\times$ greater than those of photons at 1 µm ($q = 2\pi \times 10^4 \text{ cm}^{-1}$). With $\mathbf{q} \sim \mathbf{0}$, the condition $\mathbf{k}_\text{v} - \mathbf{k}_\text{c} + \mathbf{q} = \mathbf{0}$ essentially requires *direct* transitions with $\mathbf{k}_\text{v} = \mathbf{k}_\text{c}$. Approximating the function $u_{\text{c},\mathbf{k}_\text{c}}$ in (3.55) by the first term[§] in the Taylor expansion of $u_{\text{c},\mathbf{k}_\text{v}+\mathbf{q}}$ then gives [44]

$$\left\langle \psi_\text{f} \left| e^{i\mathbf{q}\cdot\mathbf{r}} \, \hat{\mathbf{a}} \cdot \mathbf{p} \right| \psi_\text{i} \right\rangle \approx \hat{\mathbf{a}} \cdot \mathbf{p}_{fi} = \left\langle \psi_\text{f} \left| \hat{\mathbf{a}} \cdot \mathbf{p} \right| \psi_\text{i} \right\rangle = \frac{1}{\Omega} \int_\Omega \mathrm{d}^3\mathbf{r}' \, u_{\text{c},\mathbf{k}}^\dagger(\mathbf{r}') \, \hat{\mathbf{a}} \cdot \mathbf{p} \, u_{\text{v},\mathbf{k}}(\mathbf{r}') \tag{3.56}$$

where k_v is relabelled simply as k, and the crystal volume outside the integral in (3.53) is replaced by that of the unit cell $\Omega = V/N$, due to the sum over the N unit cells in (3.55). Setting (3.56) in (3.51), the transition rate *per unit volume* at zero Kelvin results from consideration of all wavevectors $\mathbf{k}$ within this volume (see Section 3.7 for work above zero Kelvin)

$$\Gamma_\text{if}(\hbar\omega) = \frac{\pi}{\hbar} \frac{e^2 E_0^2}{2m^2\omega^2} \sum_\mathbf{k} \left| \hat{\mathbf{a}} \cdot \mathbf{p}_\text{fi} \right|^2 \delta\big(E_\text{f}(\mathbf{k}) - E_\text{i}(\mathbf{k}) - \hbar\omega\big) \tag{3.57}$$

where the relations $\nabla \times \mathbf{E} = -\partial\mathbf{B}/\partial t$ and $\mathbf{B} = \nabla \times \mathbf{A}$ have been used, giving $\mathbf{E} = -\partial\mathbf{A}/\partial t$ and an electric field amplitude $E_0 = \omega A_0$ from (3.50). Next, the transition rate (3.57) is connected to the Beer–Lambert absorption coefficient, which describes the attenuation of the incident optical intensity $I_0(\hbar\omega)$ (a power) as a function of propagation length x

$$I(x) = I_0 \, e^{-\alpha x} \quad \therefore \quad \frac{\mathrm{d}I}{\mathrm{d}x} = -\alpha \, I(x) \quad \therefore \quad \alpha(\hbar\omega) = -\frac{1}{I}\frac{\mathrm{d}I}{\mathrm{d}x} \approx -\frac{\delta I/I}{\delta x} = \frac{\hbar\omega \cdot \Gamma_\text{if}(\hbar\omega)}{\langle S(\hbar\omega)\rangle} \tag{3.58}$$

where in the rightmost expression, the fraction of absorbed intensity ($-\delta I/I$) at position δx is written as the power absorbed per unit area ($\delta x \cdot \hbar\omega \cdot \Gamma_\text{if}$) over the power incident per unit area: given via the time-averaged Poynting relation (3.16)

$$\left| \langle \mathbf{S} \rangle \right| = \left| \langle \text{Re}\,\mathbf{E} \times \text{Re}\,\mathbf{H} \rangle \right| = \frac{1}{2} \left| \text{Re}\,(\mathbf{E}_0 \times \mathbf{H}_0^*) \right| = \frac{E_0^2 \, \text{Re}\,\tilde{n}^*}{2\mu_0\mu_r c} \tag{3.59}$$

Setting the transition rate (3.57) and Poynting relation (3.59) in (3.58) gives an absorption coefficient

$$\alpha(\hbar\omega) = \frac{\pi\mu_r e^2}{\omega\epsilon_0 \, cm^2 \text{Re}\,\tilde{n}} \sum_\mathbf{k} \left| \hat{\mathbf{a}} \cdot \mathbf{p}_\text{fi} \right|^2 \delta\big(E_\text{f}(\mathbf{k}) - E_\text{i}(\mathbf{k}) - \hbar\omega\big) \tag{3.60}$$

The joint density of states $g(E_\text{cv})$ offers a convenient means of evaluating the sum over $\mathbf{k}$

$$\alpha(\hbar\omega) = \frac{\pi\mu_r e^2}{\omega\epsilon_0 \, cm^2 \text{Re}\,\tilde{n}} \left| \hat{\mathbf{a}} \cdot \mathbf{p}_\text{fi} \right|^2 \int g(E_\text{cv})\delta\big(E_\text{cv} - \hbar\omega\big) \, \mathrm{d}E_\text{cv} \tag{3.61}$$

where the approximation is used that $\left| \hat{\mathbf{a}} \cdot \mathbf{p}_\text{fi} \right|$ is not strongly $\mathbf{k}$-dependent. Thus, for spherical parabolic bands at a M_0 critical point (3.48) the direct absorption coefficient is [96]

$$\alpha(\hbar\omega) = \frac{\mu_r e^2 \left| \hat{\mathbf{a}} \cdot \mathbf{p}_\text{fi} \right|^2}{\pi\omega\epsilon_0 \, cm^2\hbar^3 \, \text{Re}\,\tilde{n}} \sqrt{2\mu^3 \, (\hbar\omega - E_g)} \tag{3.62}$$

[§]Subsequent terms in the Taylor expansion lead to electric quadrupole and magnetic dipole transitions

3.6.4 Absorption in indirect gap semiconductors

Because photons carry negligible momentum, absorption at indirect onsets (taking electrons from wavevector $\mathbf{k}_v$ to $\mathbf{k}_c$) requires the assistance of a phonon to conserve momentum[102]

$$\hbar\mathbf{k}_v \pm \hbar\mathbf{q} = \hbar\mathbf{k}_c \tag{3.63}$$

where the $\pm$ sign allows for absorption or emission of a phonon with wavevector $\mathbf{q}$, the probability of the former process given by the Bose–Einstein phonon occupation number[44], $n(E,T) = \{\exp(E/k_BT) - 1\}^{-1}$; similarly, because phonon emission increases the phonon occupation number, the probability goes as $\{n(E,T) + 1\} = \{1 - \exp(-E/k_BT)\}^{-1}$. Absorption is then the sum of the temperature-dependent phonon absorption and emission edges[102]

$$\alpha(\hbar\omega, T) = B\left[\frac{\{R(\hbar\omega - E_g + E_{ph})\}^2}{\exp(E_{ph}/k_BT) - 1} + \frac{\{R(\hbar\omega - E_g - E_{ph})\}^2}{1 - \exp(-E_{ph}/k_BT)}\right] \tag{3.64}$$

where E_g, E_{ph}, k_B and B are respectively the indirect band gap, averaged phonon energy, Boltzmann constant and a material-specific constant, whilst the ramp functions ($R(x)=x$ for $x > 0$, else $R(x)=0$) ensure conservation of energy. The first and second terms respectively represent phonon absorption and emission; due to the negligible phonon occupation number, only the emission term contributes at low temperature. As discussed by Hamaguchi[96] and Yu & Cardona[44], the form of the numerators results from consideration of the electron-photon and electron-lattice Hamiltonian matrix elements in second-order perturbation theory, giving a transition rate w_{if}

$$w_{if} = \frac{2\pi}{\hbar}\left|\sum_m \frac{\langle f|\mathcal{H}|m\rangle\langle m|\mathcal{H}|i\rangle}{E_i - E_m}\right|^2 \delta(E_i - E_f) \tag{3.65}$$

where the sum runs over all intermediate electron states $|m\rangle$. This expression describes consecutive electron-photon and electron-phonon interactions (and the reverse): as seen by paths 1 and 2 in Figure 25. Due to the phonon energy in the ramp function arguments in (3.64), the photon energy of the indirect onset will be somewhat displaced from the position of the associated indirect gap. Note that the phonons involved in (3.63) are those with significant wavevectors (not the zone-centre phonons seen in low-energy optical work).
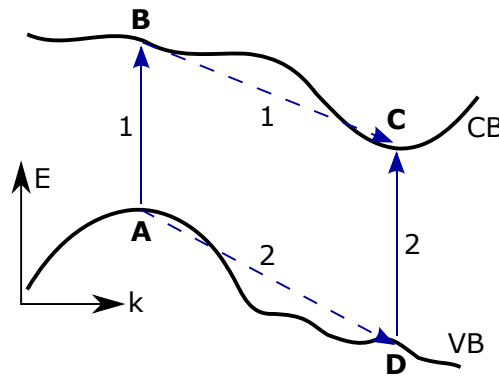
Figure 25: Indirect absorption. Electrons at **A** in the valence band may absorb a photon and transition to **B**, then transition to **C** by absorption or emission of a phonon (path 1). Transitions may also occur in the opposite order: leading to path 2. Alternatively, path 2 may be considered as an electron (photon) transition from **D** to **C**, followed by an electron (phonon) transition from **A** to **D**.
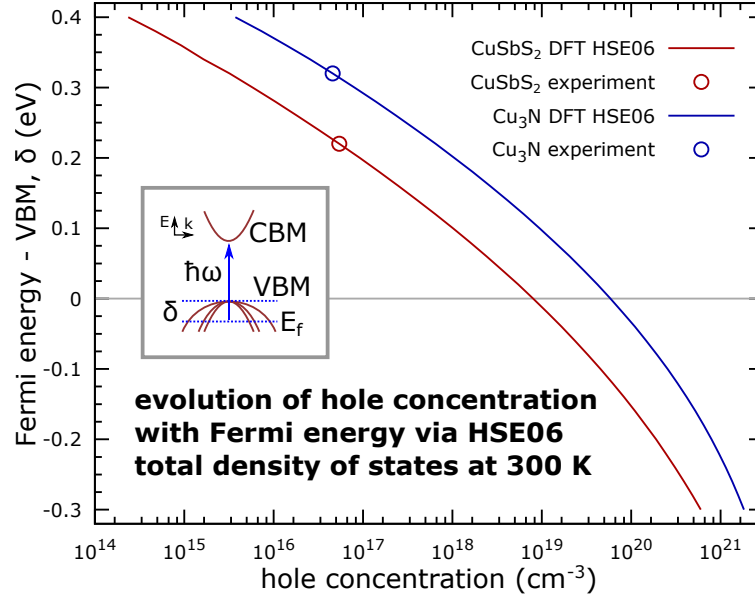
Figure 26: The Burstein–Moss shift. Carrier concentrations are evaluated at different Fermi energies via (3.66) with the HSE06 total density of states for the p-type semiconductors $CuSbS_2$ and Cu_3N at 300 K. Inset shows the blueshift $|\delta|$ of the optical gap $\hbar\omega$ as the Fermi level E_F enters the valence band.

3.7 Band filling considerations: the Burstein–Moss shift

Optical studies of heavily-doped or degenerate semiconductors may be subject to a band-filling phenomenon known as the Burstein–Moss shift[103,104], first observed in the degenerate n-type semiconductor indium antimonide but feasible in either n- or p-type material. The effect is illustrated in Figure 26 for some p-type semiconductors to be discussed later. At high carrier concentrations (10^{19} to 10^{20} cm^{-3}), the Fermi level E_F passes into the valence band, thus increasing the optical band gap $\hbar\omega$ beyond the fundamental gap E_g by the magnitude of the difference $|\delta| = |E_F - E_{VBM}|$. In InSb this is due to a particularly light conduction band effective mass[105]: which reduces the conduction band density of states and leads to a conduction band which is more readily filled by donor electrons¶ (i.e. band-filling).

Practically, it is useful to estimate the significance of any band-filling. A simple estimation compares the experimental carrier concentration with those calculated for n- or p-type material, using the density of states $g(E)$ and a Fermi function, $f(E) = \{\exp([E - E_F]/k_BT) + 1\}^{-1}$

$$n(E_F, T) = \int_{E_{CBM}}^{\infty} g(E)f(E)\,\mathrm{d}E \qquad p(E_F, T) = \int_{-\infty}^{E_{VBM}} g(E)[1 - f(E)]\,\mathrm{d}E \qquad (3.66)$$

where the hole occupancy is written as $1 - f(E)$. Selecting an acceptable model for the density of states $g(E)$ is non-trivial, particularly for p-type material as the valence band typically comprises multiple degenerate levels with different dispersions. Density functional theory gives an estimate of the total density of states $g(E)$, with which the integrals of (3.66) may be evaluated numerically, as seen in Figure 26 for a range of Fermi energies. These calculations find that the experimental carrier concentrations (discussed on pages 88 and 116) are 1000 or 100 times too small for band-filling effects to be significant in Cu_3N or $CuSbS_2$.

¶ If $E = \hbar^2 k^2/2m^*$, a small effective mass $m^* = \hbar^2/\mathrm{d}^2 E/\mathrm{d}k^2$ makes $dE/dk = \hbar^2 k/2m^*$ large, suggesting a small density of states $g(E) \sim |dE/dk|^{-1}$ by (3.45). Conversely, p-type InSb has negligible band-filling due to a heavier m_h^*. Generally the Burstein–Moss shift competes with conduction band *lowering* due to perturbations from impurity states[106].
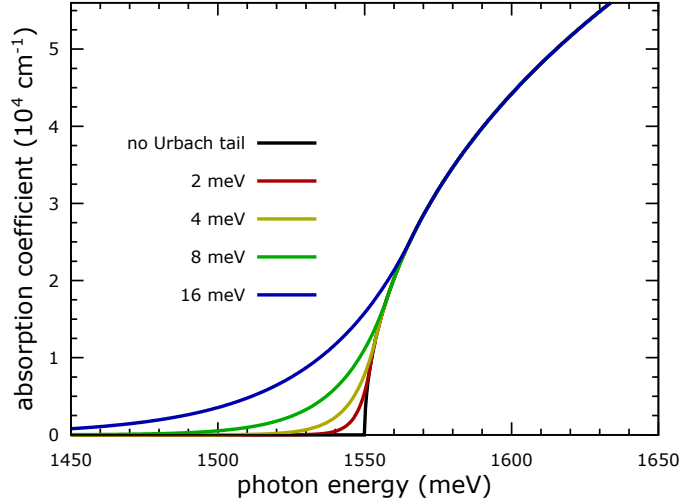
Figure 27: Urbach tails and direct onsets. The tails shown join the direct edge at up to $16\,\mathrm{meV}$ above E_g.

3.8 Direct absorption edges and the Urbach tail

Consider the absorption coefficient $\alpha(\hbar\omega)$ at a direct onset, as given by equation (3.62)

$$\alpha(\hbar\omega) = \frac{A}{\hbar\omega} \cdot \sqrt{\hbar\omega - E_\mathrm{g}} \tag{3.67}$$

where E_g and A are respectively the band gap and a material constant. Without high-quality material, absorption onsets in the form of (3.67) are rarely seen. Instead, static and dynamic effects (i.e. disorder due to impurities, defects and phonons[107,108]) lead to an exponentially decaying Urbach tail[109] which joins the direct absorption edge smoothly and continuously at some energy $E_\mathrm{g} + E_\mathrm{U}$. The absorption due to the Urbach tail is then

$$\alpha_\mathrm{U}(\hbar\omega) = B \exp\left(\gamma \left\{\hbar\omega - E_\mathrm{g} - E_\mathrm{U}\right\}\right) \tag{3.68}$$

where B and γ are additional material constants. To attach the Urbach tail smoothly the gradients of (3.67) and (3.68) are needed with respect to photon energy. Differentiating gives

$$\frac{\mathrm{d}\alpha}{\mathrm{d}(\hbar\omega)} = -\frac{A}{(\hbar\omega)^2}\sqrt{\hbar\omega - Eg} + \frac{A}{2\hbar\omega}\frac{1}{\sqrt{\hbar\omega - Eg}} \quad \text{and} \quad \frac{\mathrm{d}\alpha_\mathrm{U}}{\mathrm{d}(\hbar\omega)} = \gamma\,\alpha_\mathrm{U} \tag{3.69}$$

Equating (3.67) and (3.68) for continuity at the intersection energy $\hbar\omega = E_\mathrm{g} + E_\mathrm{U}$ gives

$$B = \frac{A}{E_\mathrm{g} + E_\mathrm{U}}\sqrt{E_\mathrm{U}} \tag{3.70}$$

Equating the equations (3.69) for smoothness at $\hbar\omega = E_\mathrm{g} + E_\mathrm{U}$ and using (3.70) gives

$$\gamma A\sqrt{E_\mathrm{U}} = -\frac{A}{E_\mathrm{g} + E_\mathrm{U}}\sqrt{E_\mathrm{U}} + \frac{A}{2}\frac{1}{\sqrt{E_\mathrm{U}}} \quad \longrightarrow \quad \gamma = \frac{1}{2E_\mathrm{U}} - \frac{1}{E_\mathrm{g} + E_\mathrm{U}} \tag{3.71}$$

Thus the tail may be written in a new form, with only one free parameter E_U (see Figure 27).

$$\alpha_\mathrm{U}(\hbar\omega) = \frac{A\sqrt{E_\mathrm{U}}}{E_\mathrm{g} + E_\mathrm{U}} \exp\left\{\frac{(\hbar\omega - E_\mathrm{g} - E_\mathrm{U})(E_\mathrm{g} - E_\mathrm{U})}{2E_\mathrm{U}(E_\mathrm{g} + E_\mathrm{U})}\right\} \quad \text{for} \quad \hbar\omega \le (E_\mathrm{g} + E_\mathrm{U}) \tag{3.72}$$

Various relations are seen in the literature, with up to three parameters[43,110,111]. Typically $\alpha(\hbar\omega) = A\exp(B\{\hbar\omega - C\})$, where C is often (but not always[112]) the zero Kelvin band gap[113], while B sometimes has a temperature dependence[43,110], $B = D/k_\mathrm{B}T$ (as in Urbach's original work[109]). Tails typically broaden with increasing temperature[113] (see Tropf et al.[110]).

3.9   Excitons (and shallow impurities)

The absorption processes discussed thus far neglect the Coulomb interaction between the hole and electron. Upon considering this, bound states of electrons and holes (called excitons) become viable and significantly change the absorption spectra. Excitons are typically described using one of two different treatments: Frenkel excitons have wavefunctions written in atomic basis functions and describe tightly-bound states usually found in molecular systems; Wannier–Mott excitons (discussed below) use the effective mass approximation and are suited to the weakly-bound states often seen in inorganic semiconductors.

The excitonic Schrödinger equation is written with the addition of a Coulomb term [44,96,114,115]

$$\left(-\frac{\hbar^2}{2m_e^*}\nabla_{\mathbf{r}_e}^2 - \frac{\hbar^2}{2m_h^*}\nabla_{\mathbf{r}_h}^2 - \frac{e^2}{4\pi\epsilon_0\epsilon_r|\mathbf{r}_e - \mathbf{r}_h|} \right)\Psi(\mathbf{r}_e, \mathbf{r}_h) = E\,\Psi(\mathbf{r}_e, \mathbf{r}_h) \qquad (3.73)$$

where respectively $\mathbf{r}_e$, $\mathbf{r}_h$ and m_e^*, m_h^* are the electron and hole position vectors and effective masses. The centre-of-mass position $\mathbf{R}$ and relative displacement $\mathbf{r}$ are

$$\mathbf{R} = \frac{m_e^*\,\mathbf{r}_e + m_h^*\,\mathbf{r}_h}{m_e^* + m_h^*} \qquad \text{and} \qquad \mathbf{r} = \mathbf{r}_e - \mathbf{r}_h \qquad (3.74)$$

Writing the Hamiltonian (3.73) in relative and centre-of-mass coordinates via (3.74) gives

$$\left(-\frac{\hbar^2}{2\mu}\nabla_{\mathbf{r}}^2 - \frac{e^2}{4\pi\epsilon_0\epsilon_r|\mathbf{r}|} - \frac{\hbar^2}{2M}\nabla_{\mathbf{R}}^2 \right)\Psi(\mathbf{r}, \mathbf{R}) = E\,\Psi(\mathbf{r}, \mathbf{R}) \qquad (3.75)$$

with the translational mass $M = m_e^* + m_h^*$ and reduced mass $\mu = 1/(1/m_e^* + 1/m_h^*)$. Equation (3.75) has solutions in the form of the product of a plane wave and hydrogenic function $\psi(\mathbf{r})$

$$\Psi(\mathbf{r}, \mathbf{R}) = \psi(\mathbf{r}) \cdot e^{i\mathbf{K}\cdot\mathbf{R}} \qquad (3.76)$$

with the exciton wavevector $\mathbf{K} = \mathbf{k}_e + \mathbf{k}_h = \mathbf{k}_e - \mathbf{k}_v$ equal to the sum of those of the electron and hole. The energy of the orbital motion arises identically to the hydrogenic Bohr series

$$E_n = E_g - \frac{\mu\,e^4}{2\left(4\pi\epsilon_0\epsilon_r n\hbar\right)^2} = E_g - \frac{\mu}{m_e\epsilon_r^2}\cdot\frac{R_H}{n^2} \qquad \text{where} \quad n = 1, 2, \cdots \qquad (3.77)$$

where E_g is the associated gap, R_H=13.6 eV and R_x are respectively the Rydberg energy for hydrogen and for the exciton, n is the principal quantum number and ϵ_r is the real part of the dielectric constant evaluated at E_n. The total energy is then $E = E_n + \hbar^2 K^2/2M$, but conservation of momentum and the negligible photon wavevector again lead to vertical (direct) optical transitions, so $\mathbf{K} \approx 0$. Excitons are expected at critical points where the bands are parallel ($\nabla_k E_c = \nabla_k E_v$), where electron and hole have similar group velocities [116,117] ($\mathrm{d}\omega/\mathrm{d}k = 1/\hbar \cdot \mathrm{d}E/\mathrm{d}k$). A similar hydrogenic analysis holds for shallow impurity-centres [44].

Excitonic absorption spectra consist of discrete lines and a continuum for large n [44,115]. Line spectra are fitted with Lorentzians and Gaussians with respective 10 and 100 meV widths in homogeneous and inhomogeneous material [118,119]. The relation for orbital energies (3.77) varies according to selection rules, anisotropy or dimensionality; three-dimensional excitons may have discrete line-widths which follow a $T^{1.5}$ temperature dependence, and line-positions which evolve linearly at low temperature [120].

3.10 Propagation of light in finite media: ray-tracing analysis

Consider the propagation of coherent light through a system of three optical media separated by parallel interfaces, see Figure 28. The amplitudes of waves reflected and transmitted at each interface are given by the Fresnel coefficients r_{pq} and t_{pq} as functions of the refractive indices (n_0, n_1 and n_2), with p denoting the incident medium. The amplitudes of outbound waves in the incident and final medium include contributions from internally reflected waves.
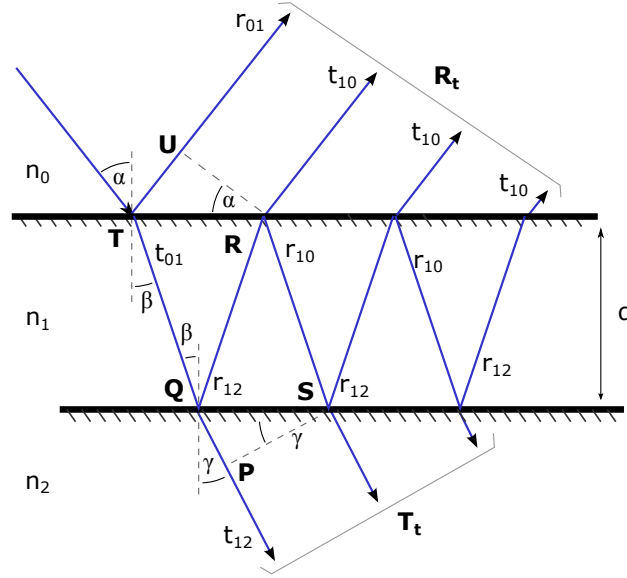
Figure 28: Propagation of light through a single layer, showing internal reflections within the layer.

The **total Fresnel reflection (amplitude) coefficient** r_{tot} is written by considering a ray which is transmitted into the second medium and which subsequently undergoes repeated reflections (see Figure 28). The expression for r_{tot} is a sum of terms, one for each ray contributing to the outbound wave, each in the Fresnel coefficients and phase factors $\mathrm{e}^{i\phi}$, with ϕ the phase difference between adjacent rays in the incident medium

$$r_{\text{tot}} = r_{01} + t_{01} \underbrace{\mathrm{e}^{i\phi} r_{12}}_{\beta} t_{10} + t_{01} \underbrace{\mathrm{e}^{i\phi} r_{12}}_{\beta} r_{10} \underbrace{\mathrm{e}^{i\phi} r_{12}}_{\beta} t_{10} + \cdots$$

$$= r_{01} + t_{01}\,\beta\,t_{10} \left[1 + \beta\,r_{10} + (\beta\,r_{10})^2 + \cdots \right] \longrightarrow r_{01} + \frac{t_{01}t_{10}r_{12}\mathrm{e}^{i\phi}}{1 - r_{10}r_{12}\mathrm{e}^{i\phi}} \qquad (3.78)$$

where r_{mn} and t_{mn} are the Fresnel reflection and transmission coefficients for light incident in medium m upon the boundary with medium n. In the right-hand side of (3.78), the geometric series in (βr_{10}) is written as the infinite sum: $1 + x + x^2 + \cdots = 1/(1 - x)$. To evaluate the phase ϕ, consider the difference in optical path lengths between the waves reflected from point T and transmitted through point R with reference to Figure 28,

$$\Delta s = (TQ + QR)\,n_1 - UT\,n_0 \qquad (3.79)$$

A few relations result from the geometry, with d the thickness of the second medium

$$TQ = QR = \frac{d}{\cos\beta}\,, \qquad\qquad UT = RT\,\sin\alpha\,, \qquad\qquad RT = 2d\,\tan\beta \qquad (3.80)$$
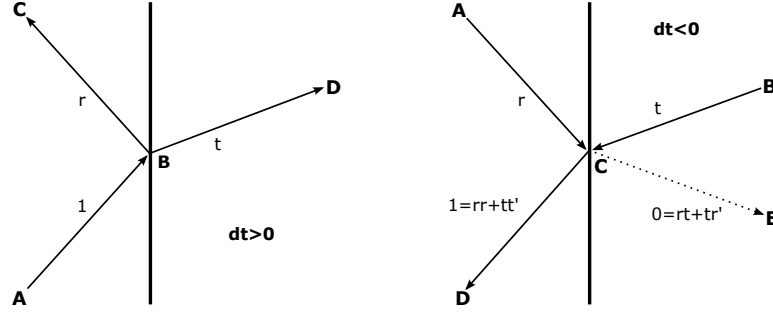
Figure 29: The Stokes relations. The ray AB with unit amplitude divides into rays BC and BD with respective amplitudes r and t. In reversed time, these rays must combine to give ray CD (and not CE).

Inserting (3.80) into (3.79) and using Snell's law ($n_0 \sin \alpha = n_1 \sin \beta$) gives the path difference

$$\Delta s = \frac{2n_1 d}{\cos \beta} \left(1 - \sin^2 \beta\right) = 2n_1 d \cos \beta \tag{3.81}$$

The **phase difference for each adjacent pair of waves** results with the incident wavevector k

$$\phi = 2 k n_1 d \cos \beta \tag{3.82}$$

which may be complex due to n_1. Similarly, the transmitted wave path difference Δs is

$$\Delta s = (QR + RS) n_1 - QP n_2 \qquad QP = (2d \tan \beta) \sin \gamma \qquad QR = RS = \frac{d}{\cos \beta} \tag{3.83}$$

Hence the phase difference of adjacent waves in transmission results with $n_1 \sin \beta = n_2 \sin \gamma$

$$\phi = k \, \Delta s = \frac{2kn_1 d}{\cos \beta} \left(1 - \sin^2 \beta\right) = 2k n_1 d \cos \beta \tag{3.84}$$

The expression for the total reflection (3.78) is further simplified with the Stokes relations, which are derived with time-reversal symmetry. Consider light incident on an interface between two optical media, see Figure 29. The incident ray AB of unit amplitude is reflected and transmitted according to the Fresnel coefficients r and t. Requiring the same physics under reversed time $\delta t < 0$, the incident rays AC and BC with amplitudes r and t must sum to give the outbound ray CD of unit amplitude, and not produce a ray CE, giving the relations

$$1 = rr + tt' \longrightarrow tt' = 1 - r^2 \qquad \text{and} \qquad tr' + rt = 0 \longrightarrow r = -r' \tag{3.85}$$

where r' and t' are the Fresnel reflection and transmission coefficients written for rays incident from the opposite side of the interface. Alternatively, the relations (3.85) may be derived from the algebra of the Fresnel coefficients. Using (3.85) to simplify (3.78) gives

$$r_{\text{tot}} = \frac{r_{01} \left(1 - e^{i\phi} r_{12} r_{10}\right) + t_{01} e^{i\phi} r_{12} t_{10}}{1 - e^{i\phi} r_{12} r_{10}} = \frac{r_{01} + e^{i\phi} r_{12}}{1 - e^{i\phi} r_{12} r_{10}} \tag{3.86}$$

where $r_{10} = -r_{01}$ and $t_{01} t_{10} = 1 - r_{01}^2$ via the Stokes relations (3.85). The relevant quantity in experiments is the **power reflection coefficient** R_{tot}, which is proportional to $|r_{\text{tot}}|^2$

$$R_{\text{tot}} = |r_{\text{tot}}|^2 = \frac{|r_{01}|^2 + r_{01} r_{12}^* e^{-i\phi} + r_{12} r_{01}^* e^{i\phi} + |r_{12}|^2 e^{-2 \operatorname{Im} \phi}}{1 - r_{12}^* r_{10}^* e^{-i\phi} - r_{12} r_{10} e^{i\phi} + |r_{12} r_{10}|^2 e^{-2 \operatorname{Im} \phi}} \tag{3.87}$$

In a similar manner to (3.78), **the transmission amplitude coefficient** t_{tot} from Figure 28 is

$$t_{\text{tot}} = t_{01}t_{12}e^{i\frac{\phi}{2}}\left(1 + r_{12}r_{10}e^{i\phi} + r_{12}^2 r_{10}^2 e^{2i\phi} + \cdots\right) \longrightarrow \frac{t_{01}t_{12}e^{i\frac{\phi}{2}}}{1 - r_{12}r_{10}e^{i\phi}} \tag{3.88}$$

where the phase $e^{i\frac{\phi}{2}}$ appears as transmitted rays make an odd number of passes through the layer. The **power transmission coefficient** T_{tot} then results from the squared modulus

$$T_{\text{tot}} = |t_{\text{tot}}|^2 = \frac{|t_{01}t_{12}|^2 \, e^{-\text{Im}\,\phi}}{1 - r_{12}r_{10}e^{i\phi} - r_{12}^* r_{10}^* e^{-i\phi} + |r_{12}r_{10}|^2 \, e^{-2\,\text{Im}\,\phi}} \tag{3.89}$$

3.10.1 Coherent and incoherent single layers and common approximations

In the most elementary work, the material under study is surrounded on each side by vacuum, in which case the refractive indices $n_0 = n_2$ and the Fresnel coefficients $r_{12} = r_{10}$. Equations (3.87) and (3.89) are simplified with

$$T_{\text{tot}} = |t_{\text{tot}}|^2 = \frac{|t_{01}t_{12}|^2 \, e^{-2\,\text{Im}\,\phi}}{1 - r_{12}r_{10}e^{2i\phi} - r_{12}^* r_{10}^* e^{-2i\phi} + |r_{12}r_{10}|^2 \, e^{-4\,\text{Im}\,\phi}} \tag{3.90}$$

$$R_{\text{tot}} = \frac{2R_{01}\left(1 - \cos\phi\right)}{1 - 2R_{01}\cos\phi + R_{01}^2} = \frac{4\,R_{01}\,\sin^2(\frac{\phi}{2})}{\left(1 - R_{01}\right)^2 + 4\,R_{01}\,\sin^2(\frac{\phi}{2})} \tag{3.91}$$

If the system is incoherent, then phase effects are averaged out and may be ignored. Instead, the reflection and transmission power coefficients are more conveniently deduced by considering an infinite series of internal reflections written in terms of the absorption coefficient α and the power reflection coefficient $R_0 = |r_{01}|^2 = |r_{10}|^2$ (giving a transmitted power $T_0 = 1 - R_0$)

$$R_{\text{tot}} = R_0 + R_0\left(1 - R_0\right)^2 e^{-2\alpha d}\left(1 + R_0^2 e^{-2\alpha d} + \cdots\right) = R_0 + \frac{R_0(1 - R_0)^2 e^{-2\alpha d}}{1 - R_0^2 \, e^{-2\alpha d}} \tag{3.92}$$

$$T_{\text{tot}} = \left(1 - R_0\right)^2 e^{-\alpha d}\left(1 + R_0^2 e^{-2\alpha d} + \cdots\right) = \frac{(1 - R_0)^2 \, e^{-\alpha d}}{1 - R_0^2 \, e^{-2\alpha d}} \tag{3.93}$$

These equations are often seen in the literature where they are inverted to solve for absorption spectra $\alpha(\hbar\omega)$; however, a difficulty arises due to the unknown quantity R_0. Assuming $R_{\text{tot}} = R_0$ leads to an error in $\alpha(\hbar\omega)$ which is particularly significant at weak absorption strength, $\alpha(\hbar\omega) < 10^4 \, \text{cm}^{-1}$.

Frequently, under the approximation of strong absorption, an expression is written for the total transmission as (3.93) but ignoring internal reflections, giving a transmission $T_{\text{tot}} = \left(1 - R_0\right)^2 e^{-\alpha d}$ and an absorption coefficient

$$\alpha = \frac{1}{d}\log\left(\frac{(1 - R_0)^2}{T_{\text{tot}}}\right) \tag{3.94}$$

which again contains the unknown quantity R_0. See the further discussions in Section 3.13.3.
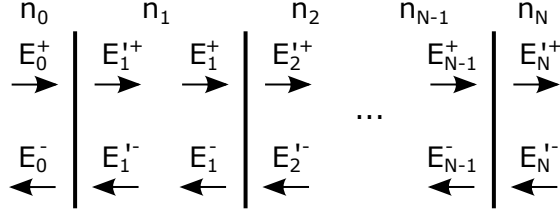
Figure 30: Transfer Matrix method. Waves with electric field amplitudes $E_0^{\pm}$ propagate in the forward and reverse directions in each medium with a complex refractive index $\tilde{n}$. Unprimed and primed magnitudes respectively denote amplitudes immediately before and after each interface.

3.11 The transfer-matrix method for multi-layers

The ray-tracing method described in Section 3.10 quickly becomes unwieldy for optical systems with multiple layers. Alternatively, rather than considering internal reflections within each layer, the problem is elegantly treated using an approach due to Abelès[121] known as the transfer-matrix method[93,122,123,124], which writes the fields within each layer as linear combinations of those of the forward and reverse-bound waves (see Figure 30).

Consider a boundary between media a and b with Fresnel reflection and transmission (amplitude) coefficients r_{ab} and t_{ab} for forward propagation. Waves in each medium have electric field amplitudes $E_a^{\pm}$ and $E_b^{\pm}$, where the superscripts $+$ and $-$ respectively label forward and reverse propagating waves. The waves reflected and transmitted from the boundary are then

$$E_b^+ = t_{ab}\,E_a^+ + r_{ba}\,E_b^- \qquad \text{and} \qquad E_a^- = r_{ab}\,E_a^+ + t_{ba}\,E_b^- \qquad (3.95)$$

Solving (3.95) for the electric field amplitudes E_a^+ and E_a^- in terms of E_b^+ and E_b^- gives

$$\begin{pmatrix} E_a^+ \\ E_a^- \end{pmatrix} = \frac{1}{t_{ab}} \begin{bmatrix} 1 & -r_{ba} \\ r_{ab} & t_{ab}t_{ba} - r_{ab}r_{ba} \end{bmatrix} \begin{pmatrix} E_b^+ \\ E_b^- \end{pmatrix} \longrightarrow \frac{1}{t_{ab}} \begin{bmatrix} 1 & r_{ab} \\ r_{ab} & 1 \end{bmatrix} \begin{pmatrix} E_b'^+ \\ E_b'^- \end{pmatrix} \qquad (3.96)$$

where the Stokes relations (3.85) simplify the right-hand side, and the unprimed $E_a^{\pm}$ and primed $E_b'^{\pm}$ electric fields denote the respective amplitudes immediately before and after the boundary. Under propagation the phases and amplitudes of the fields change as follows

$$\begin{pmatrix} E_a'^+ \\ E_a'^- \end{pmatrix} = \begin{bmatrix} \mathrm{e}^{i\phi} & 0 \\ 0 & \mathrm{e}^{-i\phi} \end{bmatrix} \begin{pmatrix} E_a^+ \\ E_a^- \end{pmatrix} = P_a \begin{pmatrix} E_a^+ \\ E_a^- \end{pmatrix} \qquad (3.97)$$

As discussed in the footnote on page 44, a plane wave phase written as $\exp\{i(k\tilde{n}x - wt)\}$ requires a refractive index $\tilde{n} = n + i\kappa$ and a phase term $\phi = -k\tilde{n}_a d_a \cos\theta_a$ in the propagation matrix P_a in (3.97). The interface I_{ab} and propagation P_a matrices allow the computation of reflection and transmission behaviour for any multilayer system. For **the coherent single-layer system** discussed in Section 3.10 the transfer matrix $T_{02} = I_{01}P_1I_{12}$ is written

$$\begin{pmatrix} E_0^+ \\ E_0^- \end{pmatrix} = T_{02} \begin{pmatrix} E_2^+ \\ E_2^- \end{pmatrix} = \frac{1}{t_{01}t_{12}} \begin{bmatrix} 1 & r_{01} \\ r_{01} & 1 \end{bmatrix} \begin{bmatrix} \mathrm{e}^{i\phi_1} & 0 \\ 0 & \mathrm{e}^{-i\phi_1} \end{bmatrix} \begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix} \begin{pmatrix} E_2^+ \\ E_2^- \end{pmatrix} \qquad (3.98)$$

where the reflection and transmission coefficients at each interface are r_{01}, t_{01} and r_{12}, t_{12}.

The reflection and transmission coefficients as laboriously computed in Section 3.10 for a coherent single layer may then be immediately written via the matrix elements

$$r_{02} = \left. \frac{E_0^-}{E_0^+} \right|_{E_2^- = 0} = \frac{\{T_{02}\}_{21}}{\{T_{02}\}_{11}} \qquad \text{and} \qquad t_{02} = \left. \frac{E_2^+}{E_0^+} \right|_{E_2^- = 0} = \frac{1}{\{T_{02}\}_{11}} \qquad (3.99)$$

which give identical results as equations (3.86) and (3.88) noting the different phase factor (3.84) used in those equations. Likewise, more complicated relations are trivially written. For N layers, the transfer matrix is

$$T_{0N} = I_{01} \prod_{i=1}^{N} P_i I_{i,i+1} \qquad (3.100)$$

where the product operator works left-to-right with increasing i, preserving the order of the matrices, such that **a coherent two layer film** has a transfer matrix $T_{03} = I_{01} P_1 I_{12} P_2 I_{13}$

$$T_{03} = \frac{1}{t_{01} t_{12} t_{23}} \begin{bmatrix} 1 & r_{01} \\ r_{01} & 1 \end{bmatrix} \begin{bmatrix} e^{i\phi_1} & 0 \\ 0 & e^{-i\phi_1} \end{bmatrix} \begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix} \begin{bmatrix} e^{i\phi_2} & 0 \\ 0 & e^{-i\phi_2} \end{bmatrix} \begin{bmatrix} 1 & r_{23} \\ r_{23} & 1 \end{bmatrix} \qquad (3.101)$$

Giving reflection and transmission amplitude coefficients as for (3.99), e.g. $t_{03} = 1/\{T_{03}\}_{11}$

$$t_{03} = \frac{t_{01} t_{12} t_{13}\, e^{-i(\phi_1 + \phi_2)}}{1 + r_{12} r_{23}\, e^{-2i\phi_2} + r_{01} r_{12}\, e^{-2i\phi_1} + r_{01} r_{23}\, e^{-2i(\phi_1 + \phi_2)}} \qquad (3.102)$$

$$r_{03} = \frac{r_{01}\left(1 + r_{12} r_{23}\, e^{-2i\phi_2}\right) + r_{12}\, e^{-2i\phi_1} + r_{23}\, e^{-2i(\phi_1 + \phi_2)}}{1 + r_{12} r_{23}\, e^{-2i\phi_2} + r_{01} r_{12}\, e^{-2i\phi_1} + r_{01} r_{23}\, e^{-2i(\phi_1 + \phi_2)}} \qquad (3.103)$$

Equations (3.103) are identical to those found using more complicated ray-tracing methods (see Crook's relations for a double-layer[125], also written on page 63 of Heavens[126]).

As discussed by Katsidis[127], the transfer-matrix method may be utilised to write equations for entirely or partially incoherent optical multilayers by matching the intensities at the coherent-incoherent boundary. A trivial **single incoherent layer** is then

$$T_{02}^{\text{int}} = \frac{1}{|t_{01} t_{12}|^2} \begin{bmatrix} 1 & -|r_{10}|^2 \\ |r_{01}|^2 & \left(|t_{01} t_{10}|^2 - |r_{01} r_{10}|^2\right) \end{bmatrix} \begin{bmatrix} |e^{i\phi_1}|^2 & 0 \\ 0 & |e^{-i\phi_1}|^2 \end{bmatrix} \begin{bmatrix} 1 & -|r_{21}|^2 \\ |r_{12}|^2 & \left(|t_{12} t_{21}|^2 - |r_{12} r_{21}|^2\right) \end{bmatrix} \qquad (3.104)$$

which with matrix element relations analogous to (3.99) gives equations (3.92) and (3.93), e.g. the transmissivity $T_{02} = 1/\{T_{02}^{\text{int}}\}_{11}$. Notice in (3.104) that the interface intensity matrices are written without the simplifications used in the right-side of (3.96). Thin-films may then be modelled as **two-layer systems comprising a coherent film with incoherent substrate**

$$T_{03}^{\text{int}} = \frac{1}{|t_{02} t_{23}|^2} \begin{bmatrix} 1 & -|r_{20}|^2 \\ |r_{20}|^2 & \left(|t_{02} t_{20}|^2 - |r_{02} r_{20}|^2\right) \end{bmatrix} \begin{bmatrix} |e^{i\phi_2}|^2 & 0 \\ 0 & |e^{-i\phi_2}|^2 \end{bmatrix} \begin{bmatrix} 1 & -|r_{32}|^2 \\ |r_{23}|^2 & \left(|t_{23} t_{32}|^2 - |r_{23} r_{32}|^2\right) \end{bmatrix} \qquad (3.105)$$

where r_{02} and t_{02} are the films amplitude coefficients from (3.99) or (3.86) and (3.88), and again the reflectivity and transmissivity result respectively from the matrix elements $\{T_{03}^{\text{int}}\}_{21}/\{T_{03}^{\text{int}}\}_{11}$ and $1/\{T_{03}^{\text{int}}\}_{11}$. With such tools, arbitrarily complex models may be constructed to evaluate the optical response of multi-layered systems.

3.12 Simulated R and T spectra from selected optical models

The transfer matrix method equips workers with the tools needed to compute reflectivity and transmissivity spectra for any arbitrary multi-layer system. This is beneficial as experimental spectra show great scope for complexity; before any results may be extracted, the specimen is first represented by an appropriate optical model. It is worthwhile gaining an appreciation of how different models broadly influence the measured spectra. Four typical models are considered here: coherent and incoherent single-layers (labelled $1c$ and $1i$), and the same upon incoherent substrates ($2c$ and $2i$). Generally photovoltaic absorbers, which strongly absorb ($\alpha > 10^4\,\mathrm{cm}^{-1}$) photons with energies above the band gap, must be thin ($\sim 1\,\mu\mathrm{m}$) if much of the absorption edge is to be visible: so transparent substrates are typically needed and multi-layers are important.

The films and substrates considered have respective thicknesses of $2\,\mu\mathrm{m}$ and $1\,\mathrm{mm}$. The films have a semiconductor-like dielectric function: with a refractive index $n = 3$ and a fundamental absorption onset (and Urbach tail) associated with a $1.5\,\mathrm{eV}$ direct gap. A discrete $1.2\,\mathrm{eV}$ absorption feature demonstrates the effect of a possible impurity level. The substrates are glass-like with $n = 1.5$ and non-absorbing. To compute the resulting spectra, a transfer-matrix method implementation (see Section 3.11) was written in MATLAB to reliably perform all the matrix algebra (this engine is utilised subsequently to invert spectra for determination of optical parameters: see Section 3.14). Verification of this code included two-layer test spectra exactly matching Figures 2 and 3 of Swanepoel[128] and Minkov[129], and agreement with the explicit algebra of Crook[125,126] (for oblique, two-layered, coherent, absorbing systems) to the tenth decimal place. Figure 31 shows the R and T spectra associated with each model.
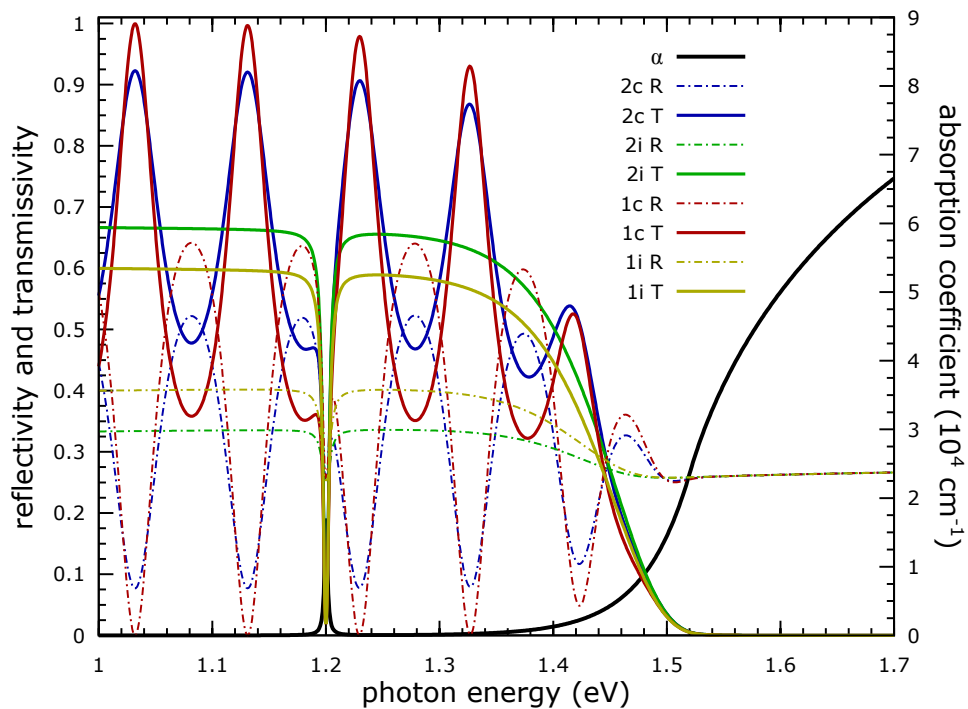
Figure 31: Reflectivity and transmissivity spectra from various optical models at normal incidence. The sample is a coherent or incoherent $2\,\mu\mathrm{m}$ film (labelled 1c and 1i), and the same upon a $1\,\mathrm{mm}$ non-absorbing substrate (2c and 2i), with refractive indices $n_2 = 1.5$. See the text.

The primary purpose of studying the spectra of Figure 31 is to identify spectral characteristics which may allow researchers to select the most appropriate optical models for the optical system under investigation. Various aspects are apparent from the figure:

- **significant interference oscillations appear** in the reflectivity and transmissivity spectra for both coherent models, due to the preservation of phase relations for internal reflections. No oscillations are seen for incoherent models (by definition).

- **oscillations offer additional information**, e.g. with a refractive index n, the simple relation $2k_i n d = 2m_i \pi$ crudely estimates the film thickness d, with k_i and m_i a wavevector and increasing integer for consecutive maxima; however, oscillations also complicate the absorption spectra and may be affected by further complexities (see Section 3.12.2).

- **oscillations diminish with increasing absorption strength**; ultimately the signals from all models converge as the absorption strength exceeds $10^4 \, \mathrm{cm}^{-1}$. This suggests the importance of model selection in the medium absorption regime below $10^4 \, \mathrm{cm}^{-1}$.

- the reflectivity R and transmissivity T oscillations fluctuate about the value of the associated incoherent signal and follow the direction of this signal. Note that the incoherent signal does not intercept the oscillations at the half-maximum value and that the oscillations are not sinusoidal.

- as expected **the sum $R + T$ never exceeds unity**, but may be close to unity well away from strong absorption features (where the R and T signals should be mirror images); $R + T$ sometimes exceeds unity in experimental work: often a sign of an inconsistency.

- however, the sum $R + T$ (and the signals R and T) are **significantly affected even at medium absorption strength**. For all models in Figure 31, the sum $R + T \approx 80\,\%$ at $\hbar\omega = 1.4 \, \mathrm{eV}$: even though the absorption coefficient $\alpha(\hbar\omega)$ here is only $10^3 \, \mathrm{cm}^{-1}$.

- under medium absorption **a substrate significantly changes the R and T amplitudes**. A substrate with lower refractive index reduces the index contrast at the film–substrate interface, reducing the intensity of internal reflections and amplitude of interference oscillations.

- similarly, **the transmission is enhanced and the reflection is reduced** by a substrate of lower index (at medium absorption), as seen in the one- and two-layer incoherent spectra (again due to the reduced contrast for reflections at the film-substrate interface).

- discrete absorption resonances leave characteristic features in both the R and T spectra.

These aspects will be seen to have significance in the next section: where the experimental R and T spectra are inverted to obtain refractive index spectra. We will see that various errors occur due to approximations made during this work.

3.12.1 Dielectric phenomena: R and T typically have opposite gradients

Experimental studies of dielectric properties often utilise external influences such as temperature, pressure, or applied electric or magnetic fields: variation of which reveals further information about the specimen. In support of such work, it is useful to establish some general rules for the behaviour of the reflectivity R and transmissivity T under broad changes in dielectric function (or equivalently refractive index). Our goal is to decide whether changes observed in experimental reflectivity or transmissivity spectra originate specifically from dielectric or other phenomena; in such experiments, any behaviour not explained by dielectric processes likely indicates an undesirable contribution or systematic error.

The easiest way to examine how R and T evolve with dielectric phenomena is simply to plot these after choosing a suitable refractive index function and optical model from those discussed in Sections 3.10 and 3.11. In Figure 32 the reflectivity and transmissivity are shown for a material with weak and strong absorption regimes (see the caption for details). It is clear that under weak absorption, R and T generally have gradients $dR/d(\hbar\omega)$ and $dT/d(\hbar\omega)$ of opposite sign. This is less obvious in the strong absorption region, where the change in the internal reflection contribution (related to the transmissivity gradient), dominates over that from the refractive index change; however, in the strong absorption limit (not plotted in Figure 32) the reflectivity again increases with increasing photon energy. Such results may be seen analytically by differentiation, but the graphical method is much simpler. Thus a rule emerges that **the reflectivity and transmissivity gradients generally have opposite sign**, where exceptions are made for regions of extreme change: such as absorption edges.
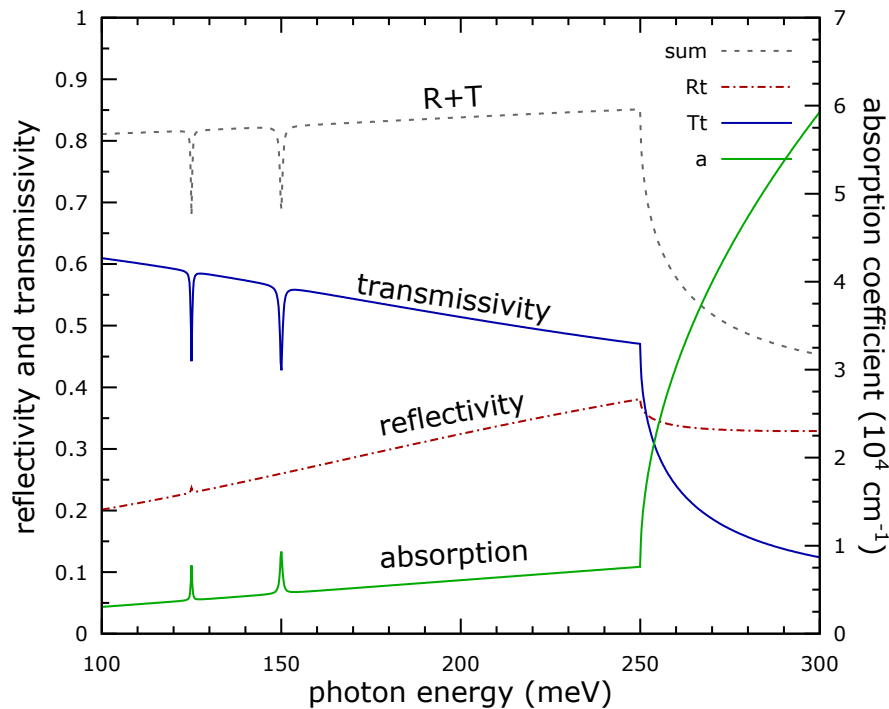
Figure 32: Behaviour of the reflectivity and transmissivity under dielectric changes. The optical model is a coherent $1\,\mu m$ thin film upon a coherent $1\,mm$ substrate. The film has a refractive index $\tilde{n} = 3.1 - 0.03i$ (with increasing real part) and low and high absorption regions which feature discrete Lorentzian absorption features at 125 and $150\,meV$ and a strong direct absorption onset at $250\,meV$.

3.12.2 Impact of thickness variations on spectra

Film thickness is a key parameter in the evaluation of absorption spectra due to the interdependence of the thickness and absorption coefficient. Typically uniform thickness is assumed, yet without high-quality synthesis methods (and particularly for polycrystalline material) it's perhaps more realistic to use an appropriate statistical distribution. Thickness is often determined by (cross-sectional) scanning electron microscopy, profilometry or spectroscopic methods, e.g. ellipsometry, or envelope[128,129,130] or photometric[131,132,133] analysis.

Interference oscillations in spectra are often taken as evidence of good material quality (and so uniform thickness); conversely, oscillations would not be possible in samples with random thickness variations. Between these extremes, coherent samples with thicknesses following various functional forms (e.g. wedges, saw-teeth or sinusoids) lead to oscillations with suppressed amplitude[130,134]. Thickness variations lead to significant inaccuracies if not accounted for: as suggested by the different transmission edges in Figure 33a.

Polycrystalline material is potentially modelled as an array of coherent crystallites (i.e. each coherently reflecting or transmitting light) with a Gaussian thickness distribution about some mean d_0, and a detected intensity given by a sum over all crystallites. Care must be taken normalising such a Gaussian for thin films: as the low thickness tail of the distribution may terminate before the distribution is negligible. The results of such an analysis in MAT-LAB for a coherent, $2\,\mu m$ single-layer, and for the same with Gaussian-distributed crystallite thicknesses (with FWHM of $4\,\mu m$ and $16\,\mu m$) are shown in Figure 33b (see also the inset). Surprisingly, **identical spectra result from each model** despite the different spreads in thickness. Although adding two spectra from different crystallites causes a significant spectral change (as seen in Figure 33a), any differences cancel under a continuous Gaussian distribution, suggesting that real-world spectra may not be significantly affected by thickness variations.
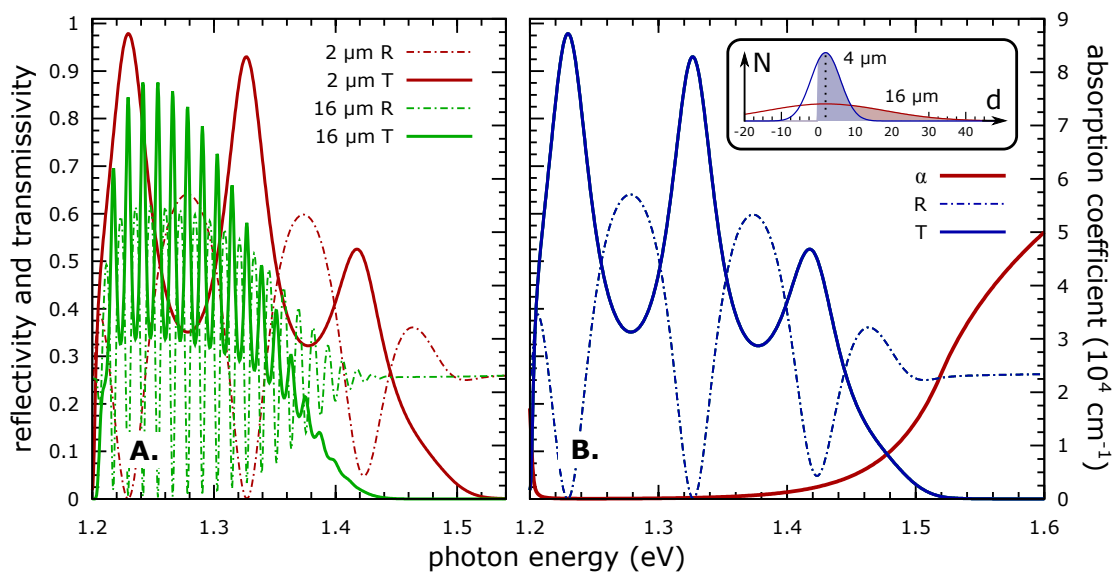
Figure 33: Impact of thickness variations on reflectivity and transmissivity spectra. **A.** shows spectra for perfect 2 and $16\,\mu m$ thick coherent single layers, using the refractive index of Figure 31. **B.** shows identical R and T spectra for $2\,\mu m$ thick films without and with Gaussian thickness variations of 4 and $16\,\mu m$ (FWHM).

3.13 Determination of optical constants from $R(\hbar\omega)$ and $T(\hbar\omega)$ spectra

This section considers various means of inverting the experimental reflectivity $R(\hbar\omega)$ and transmissivity $T(\hbar\omega)$ for accurate refractive index spectra $\tilde{n}(\hbar\omega) = n(\hbar\omega) + i\kappa(\hbar\omega)$. One might reasonably expect that, after determination of two experimental quantities (the reflectivity and transmissivity), the optical model should be solvable for two unknowns (the refractive index n and extinction coefficient κ).

However, the task turns out to be substantially more complex than it first appears, and in the most general cases, no good solutions for the experimental spectrum may exist. The difficulty arises because the optical constants n and κ appear both inside and outside complex exponential phase factors, making the inversion task a particularly ill-conditioned inverse problem with many local (and with no global) solutions[131,135]. Crudely, the problem is analogous to seeking a single solution θ for $\sin(\theta)$=1. Ultimately, even a perfectly determined R and T may be consistent with several different $\tilde{n}$ at some specific wavelength[136].

3.13.1 Established approaches for optical constant determination

Methods for experimental determination of optical constants and film thickness are discussed extensively in the literature. The published methods span the range between accuracy and experimental practicality; to this end, many approaches attempt solutions from limited information (such as those utilising only transmission spectra[132]) or depend on assumptions, such as non-absorbing substrates or normal incidence. As discussed by Heavens[123], methods fall broadly into three categories

- **Photometric methods which compute solutions from reflectivity or transmissivity**. All of these methods typically attempt to invert the optical equations (see Section 3.11 or Heavens[126]). Notable examples include the work of Nilsson[137], who found a method using forward and reverse R and T measurements; Tomlin[138], who posed the problem in an algebraically simpler form; Vriens et al.[139] who discussed the problem of accounting for absorbing substrates; Hjortsberg[140] who claimed enhanced accuracy with both R and T and modified half-metallised substrates; and Ruíz-Perez et al.[130] and Aqili & Maqsood[141] who respectively used reflection and transmission methods for inhomogeneous samples with variable thickness. The tendency in the literature is to use substantially more elementary approaches: these are discussed in Section 3.13.3.

  Practically, R and T determination is complicated by experimental difficulties with reflectivity measurements[135] (for which a reference reflectivity is required) or due to necessary inconsistencies between the R and T channels. One solution is the transmission-only approach of Chambouleyron et al.[132], who restated the problem by **requiring solutions to satisfy certain physical constraints**, e.g. a refractive index n which always increases with photon energy; these explicit physical constraints were subsequently relaxed by Birgin et al.[131], whose numerical code has been released[133]. Alternatively, the Kramers–Kronig relations may be used with R for thick homogeneous specimens[44,46].

When faced with many solutions to a problem one approach is to **consider every result in the solution space**. Such ideas were possibly first considered by Phillips[142] but were more clearly stated by Paulick[136], who outlined the main issues with automated inversion methods: the existence of multiple, non-unique solutions $\tilde{n}$ for the measured R and T, and coinciding (or very closely separated) solutions $\tilde{n}$: which make it difficult to converge to the correct solution even with accurate measurements. Paulick's method involves the construction of a mesh $\{n, \kappa\}$ upon which graded contours of reflectivity and transmissivity are plotted; solutions then correspond to all intercepts of the relevant R and T contours. A derived method is used in Section 3.14 (see also Figures 34-41).

- **Envelope methods which utilise the positions of interference maxima or minima.** Such methods were first highlighted by Manifacier et al.[143] for an infinitely-thick substrate in transmission, and then subsequently revised for a finite substrate[128] and thickness variations[134] by Swanepoel, for whom the method is now named. A similar method in reflectivity was found by Minkov[129]; while Chiao et al. discusses a composite envelope and photometric method[144]; these approaches are not applicable to incoherent systems.

  Generally, each of these methods sets an applicability region, which is typically in the medium absorption regime, $\alpha \leq 10^3 \, \text{cm}^{-1}$. Results are then computed only at the positions of interference extrema; a dielectric model fitted to these positions provides results away from the extrema or into regions of strong absorption. Accuracy depends upon the validity of the model and the number of extrema used, e.g. results published[145] for Cu_3N suggest $n = 2$ and $\alpha = 2 \times 10^4 \, \text{cm}^{-1}$ at 1.8 eV, somewhat different from the $n = 3$ and $\alpha = 5 \times 10^4 \, \text{cm}^{-1}$ determined by FTIR and ellipsometry in Chapter 4.

- **Polarisation methods which look for a polarisation change (typically in reflectivity).** of which the technique of ellipsometry is a primary example. While ellipsometry is a powerful technique due to the use of an internal standard (polarisation change of light)[38], problems may arise due to enhanced surface sensitivity[135]. Often ellipsometry and transmission spectra are mutually complementary, as seen in Section 3.14.

3.13.2 Solution space for selected optical models

The previous section noted how determination of the optical constants (the refractive index $n(\hbar\omega)$ and absorption coefficient $\alpha(\hbar\omega)$) from the reflectivity $R(\hbar\omega)$ and transmissivity $T(\hbar\omega)$ spectra is generally a difficult inverse problem with many local, non-global solutions. This is clearly seen by plotting contours for the optical models of Figure 31 in $\{n, \alpha\}$ space. In an approach similar to that used by Paulick[136], R and T are evaluated (using the relevant model from page 63) for a specific photon energy over a mesh spanning a domain in $\{n, \alpha\}$ space. Subtracting the experimental R and T (from Figure 31) from the computed R and T (from the relevant mesh) gives δR and δT residuals across the mesh, respectively represented by the thin and bold contours in Figures 34–41. Solutions exist at the intersections $\delta R = \delta T = 0$.

Figure 34: Three solutions ◇ for a coherent single film at 1.5 eV. Real solution marked ∗.
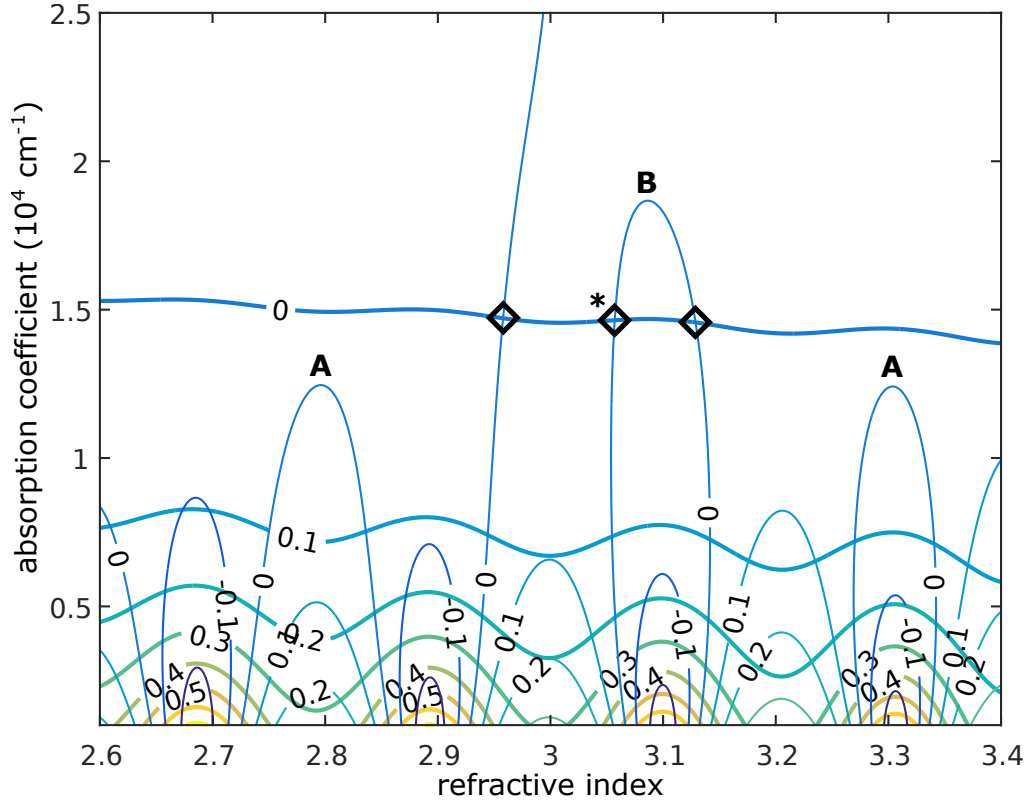
Figure 35: Four solutions ◇ for a coherent single film at 1.4 eV. Real solution marked ∗.
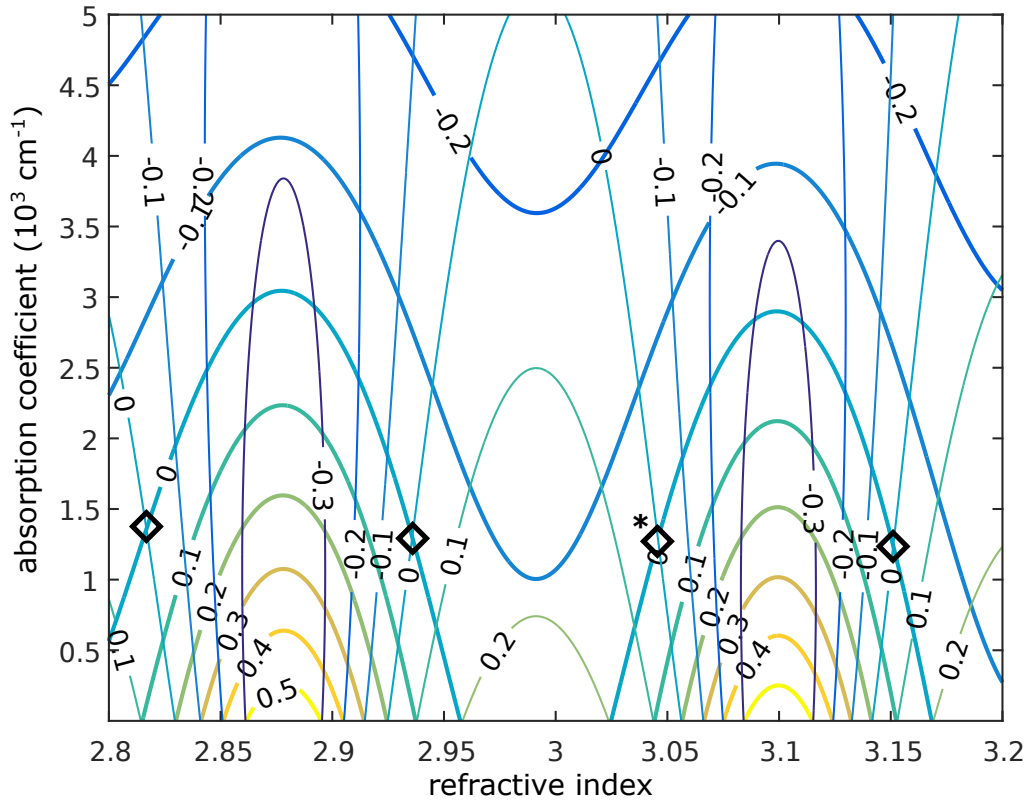
While coherent films have unique solutions well above the 1.5 eV absorption onset (where internal reflections are negligible), generally, multiple solutions exist and selecting the correct solution is non-trivial: even when inverting exact simulated data. Figure 35 shows the difficulties with multiple solutions at low energies. The proximity of different solutions illustrates

Figure 36: One solution ⬦ for an incoherent single film at $1.5\,\mathrm{eV}$.
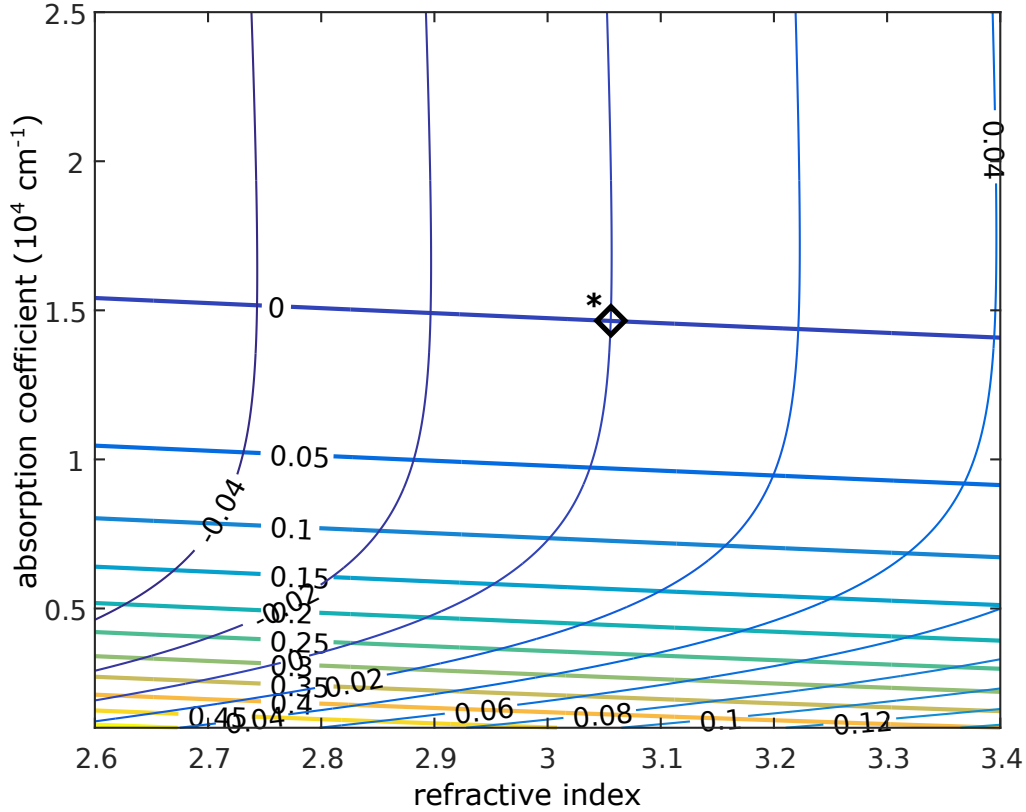
Figure 37: One solution ⬦ for an incoherent single film at $1.4\,\mathrm{eV}$.
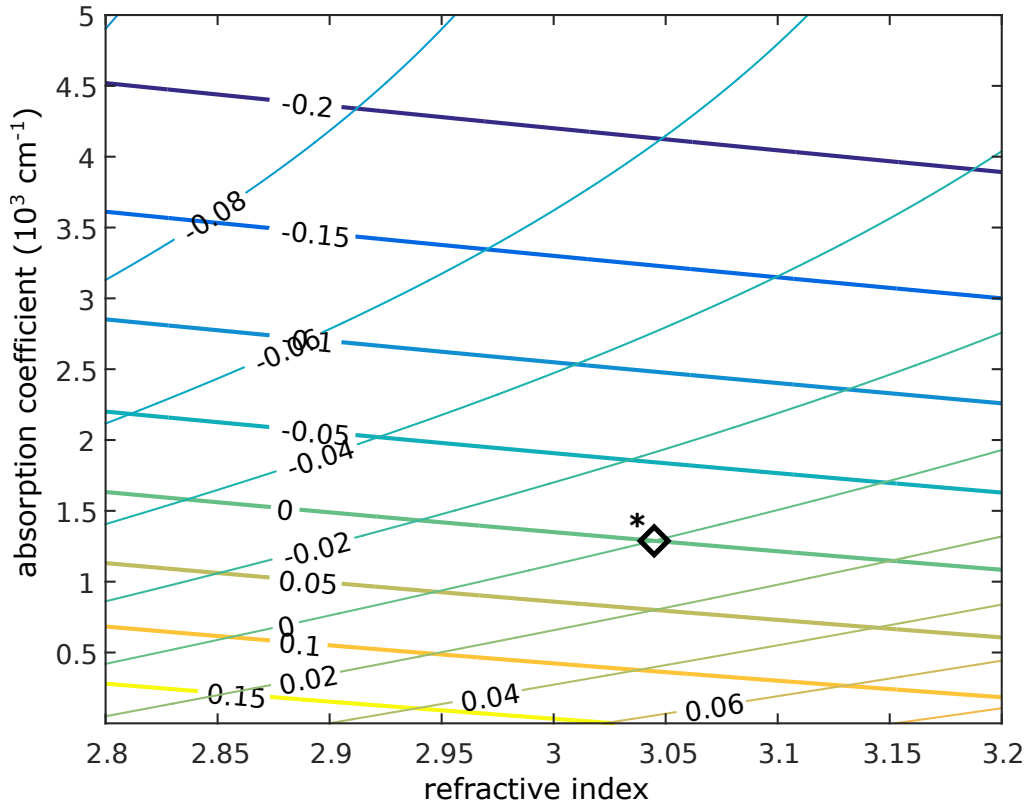
the scope for errors, particularly with inaccuracies in R and T; in such conditions the near-coincident points marked **A** in Figure 34, could introduce invalid solutions. In contrast, incoherent single films show unique solutions in the space searched, see Figures 36 and 37.

Figure 38: Three solutions ◇ for a coherent film on an incoherent substrate at 1.5 eV. **Real solution marked** ∗.
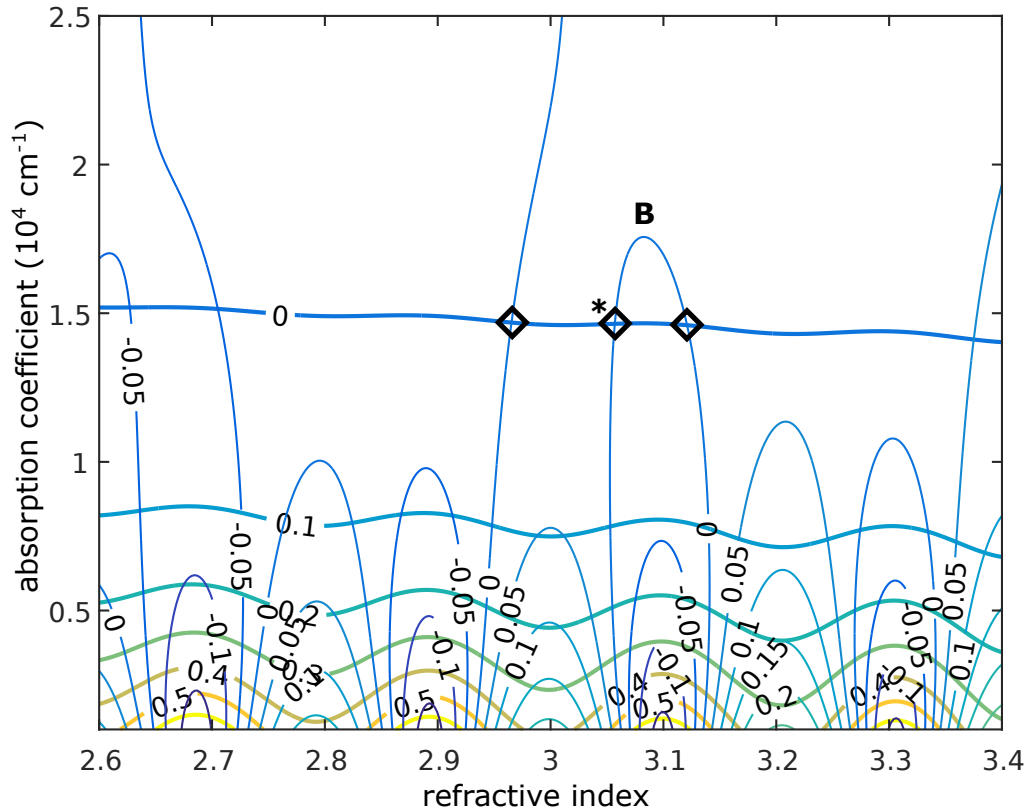
Figure 39: Four solutions ◇ for a coherent film on an incoherent substrate at 1.4 eV. **Real solution marked** ∗.
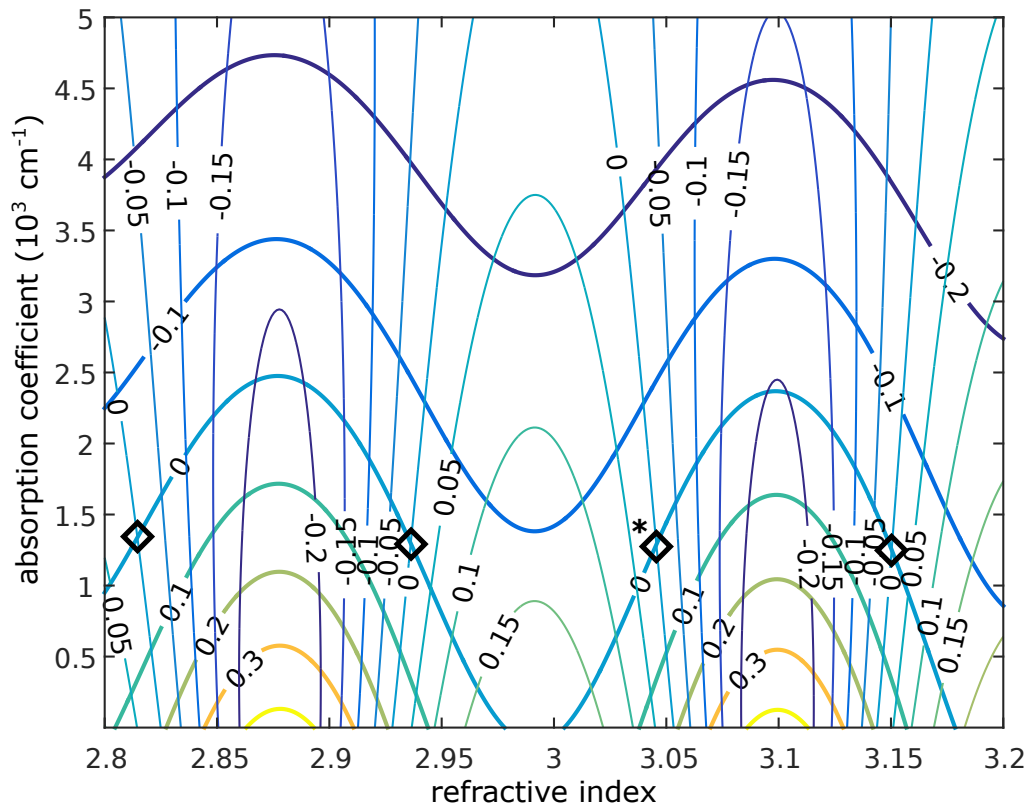
Multiple solutions are again seen for coherent films on incoherent substrates. The point marked **B** in Figure 38 illustrates a problematic condition which occurs when solutions lie at reflectivity extrema: two nearby solutions may diverge with evolving photon energy.

Figure 40: One solution ◇ for an incoherent film on incoherent substrate at $1.5\,\text{eV}$.
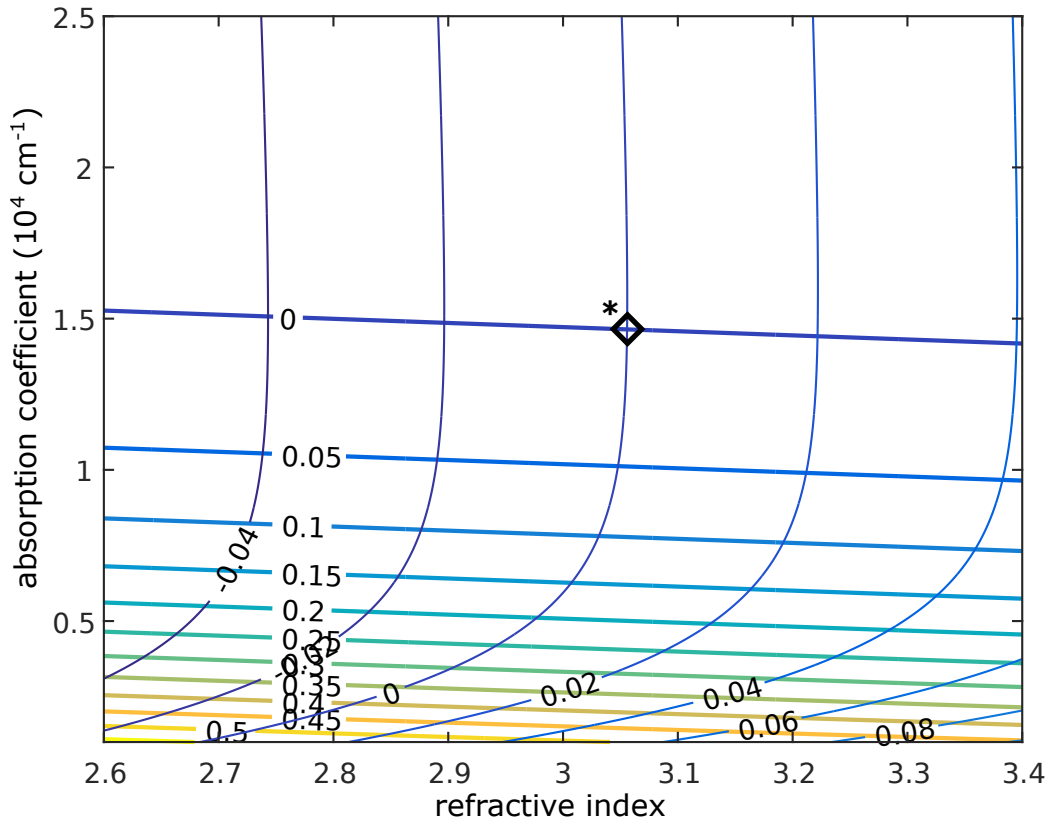
Figure 41: One solution ◇ for an incoherent film on incoherent substrate at $1.4\,\text{eV}$.
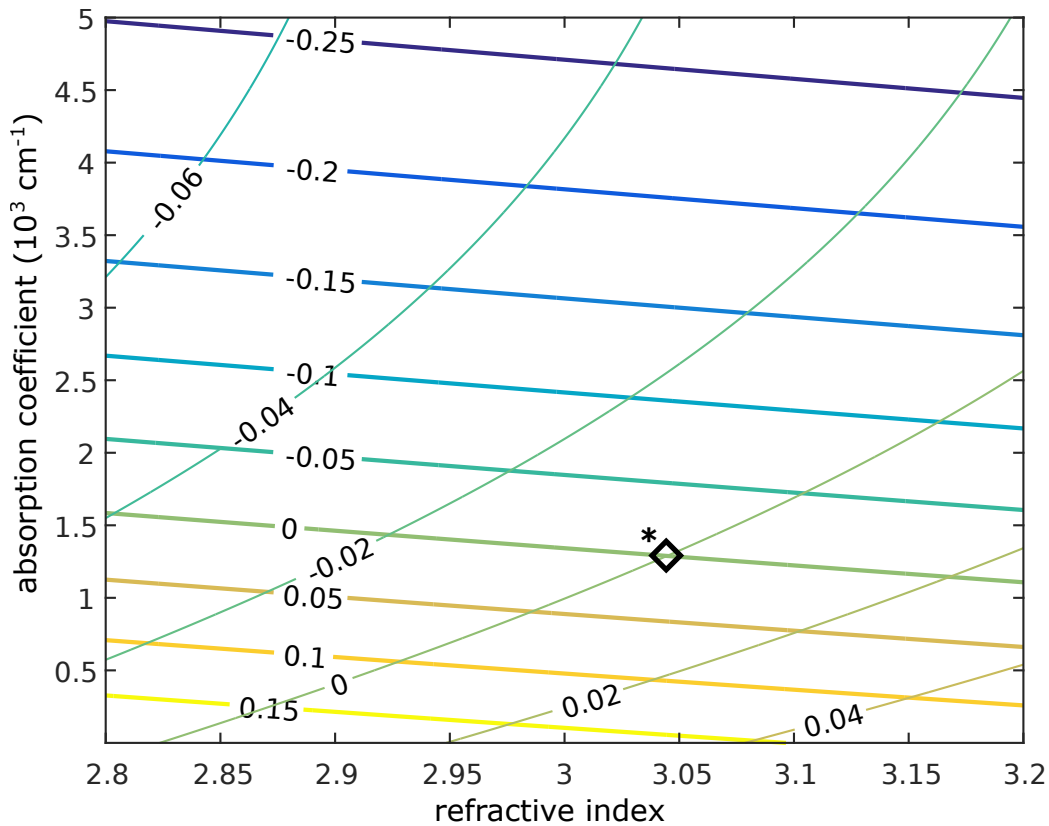
Incoherent films on incoherent substrates again have unique solutions in the domain shown; however, experimental spectra frequently show interference oscillations.

3.13.3 Analysis of common relations for absorption spectra

Whilst optical spectroscopy is a convenient and non-damaging approach for material characterisation, Section 3.13 discusses how inversion of the measured reflectivity and transmissivity (for the refractive index and absorption coefficient) is non-trivial and potentially a multivalued inverse problem. In practical work, the complexities discussed in Section 3.13 are often ignored by researchers and instead various simpler relations are used (usually derived from single layer models). While there is some merit in the use of simple relations which are both easily applied and which make few assumptions about the specimen, some aspects exist which are not always appreciated.

One typical relation is written for the strong absorption regime where internal reflections are negligible; the absorption spectra $\alpha(\hbar\omega)$ for a single optical layer of thickness d, with semi-infinite reflectivity R_0 (at the front and rear interfaces), is written by considering transmission through both interfaces and attenuation via propagation within the medium

$$T_{\text{tot}} = (1 - R_0)\,e^{-\alpha d}\,(1 - R_0) \quad \longrightarrow \quad \alpha(\hbar\omega) = \frac{1}{d}\log\left(\frac{\{1 - R_0(\hbar\omega)\}^2}{T_{\text{tot}}(\hbar\omega)}\right) \tag{3.106}$$

where T_{tot} is the resultant transmissivity for the system (as measured experimentally). Under strong absorption, the experimental and semi-infinite reflectivities are equal, $R_{\text{tot}} = R_0$; this essentially requires that the product $(\alpha d) \geq 4$. Conversely, contributions from internal reflections become increasingly important below this limit. Because workers are often interested in accurately determining the photon energy at the fundamental absorption onset, where the absorption is necessarily weak, **the relation may fail exactly at the place where it is required to give sensible results**. In such cases the transmissivity of equation (3.93), which considers incoherent internal reflections, may be solved as a quadratic in $e^{-\alpha d}$ giving

$$\alpha = -\frac{1}{d}\log\left(\frac{-(1 - R_0)^2 + \sqrt{(1 - R_0)^4 + 4T_{\text{tot}}^2 R_0^2}}{2T_{\text{tot}} R_0^2}\right) \tag{3.107}$$

where the sign of the quadratic is chosen to give a positive argument to the logarithm, and the expression is written in the (unmeasured) semi-infinite reflectivity R_0. Generally, accuracy is always improved over (3.106) by using (3.107) with the assumption that the experimental-determined reflectivity R_{tot} (which includes internal reflections) is a good approximation to R_0. A scheme is described in Section 3.13.4 which improves on this approximation.

Frequently, where no reflectivity spectra are available, equation (3.106) is used with $R_0 = 0$. This rarely leads to sensible results as significant reflectivities exceeding 20% are seen regularly. To evaluate the behaviour of the various relations in the medium absorption regime ($< 1 \times 10^4\,\text{cm}^{-1}$), reflectivity R and transmissivity T spectra are calculated (and made more realistic with an additional 1% signal noise) for an incoherent $1\,\mu\text{m}$ single-layer using relations (3.92) and (3.93). The R and T spectra are then solved for absorption spectra $\alpha(\hbar\omega)$, using equations (3.106) (with and without reflection set to zero) and (3.107). An additional trace uses a corrected semi-infinite reflectivity as discussed in Section 3.13.4.
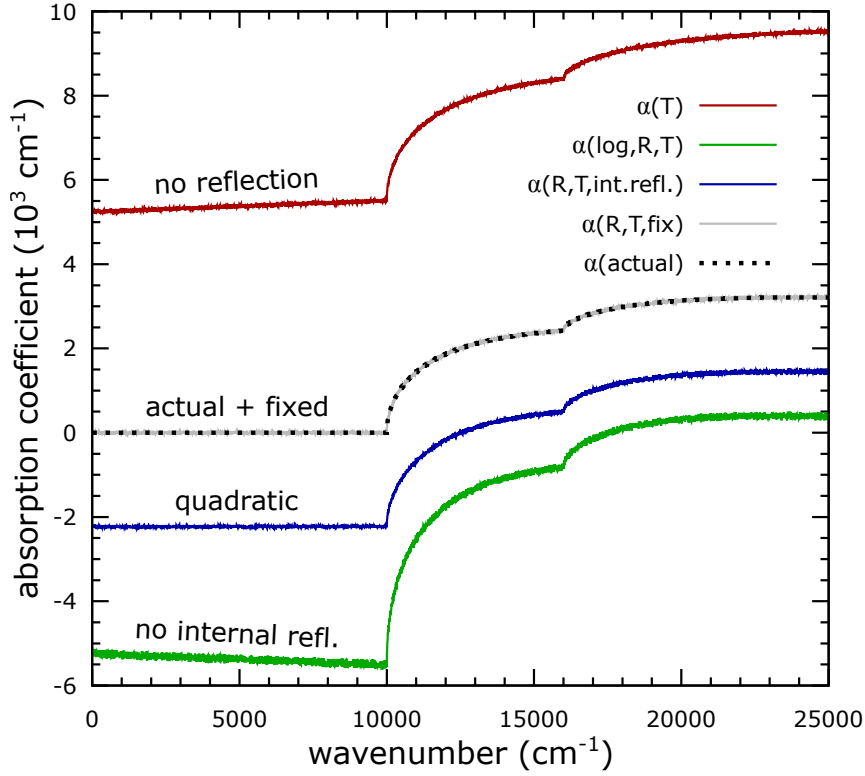
Figure 42: Comparison of accuracies of methods used to compute absorption spectra. Equations (3.92) and (3.93) are used to simulate reflectivity R and transmissivity T spectra (with 1 % signal noise) for an incoherent 1 µm single-layer at 11° angle-of-incidence with refractive index $\tilde{n} = 3 + 0.05i$, from which two direct absorption onsets develop at 10 000 and 16 000 cm^{-1}. Equations (3.106), (3.107) and the method of Section 3.13.4 are then used to compute absorption spectra from R and T.

The resulting absorption spectra are shown in Figure 42. Ignoring reflections produces a significant error across the spectrum; neglecting internal reflections is no better, and the claimed absorption remains troublesome even when these are considered using the quadratic solution of (3.107): with negative absorption resulting below 15 000 cm^{-1}. The corrected semi-infinite reflectivity (discussed in Section 3.13.4) accurately describes the real absorption spectrum, with the fraction $\alpha_{\mathrm{fixed}}/\alpha_{\mathrm{actual}} = 1$ with 1 % standard deviation in the medium absorption regime (above 10 000 cm^{-1}). Accuracy reduces significantly in the weak absorption regime (here below $\tilde{\nu} = 10\,000\,\mathrm{cm}^{-1}$) where the absorption coefficient makes a negligible contribution and R and T are instead dominated by signal noise.

The results of Figure 42 suggest that, where reflection spectra are not available (3.106) may be very inaccurate in the medium absorption regime; similarly, where reflectivity spectra are available (3.107) should always be used. The significance of these findings depends upon how the absorption spectra are used, e.g. accurate absorption edge positions are still feasible if the specimen possesses a sharp absorption edge with strong absorption. While the results for the corrected semi-infinite reflectivity are encouraging, this correction is derived from the same optical model used to generate the spectra, so good results may be expected. The correction is still preferred for other optical models, but does less-well: when a glass substrate is introduced the accuracy falls to some 85 % of the real absorption.

Interestingly for incoherent models, the refractive index is particularly resilient to noise

across the entire spectrum, with fraction $n_{\text{fixed}}/n_{\text{real}} = 1.0168(21)$. Solving the square modulus $R = |r|^2$ of the Fresnel equations (3.13) or (3.15) at normal incidence for the real part of the refractive index $\tilde{n} = n + i\kappa$ (with vacuum as the incident medium), gives

$$n = \frac{1}{R-1}\left\{-(R+1) - \sqrt{(R+1)^2 - (R-1)^2(\kappa^2+1)}\right\} \qquad (3.108)$$

where the sign is chosen to give a positive index. This expression only gives reliable results after correcting the measured reflectivity R as discussed in Section 3.13.4.

To more clearly establish the reliability of equations (3.106) and (3.107), contours of these relations may be plotted over a complex refractive index space. Consider a typical $1\,\mu\text{m}$ incoherent semiconducting film upon a thick glass substrate (with refractive index $n = 1.5$) illuminated by photons with energy $1.4\,\text{eV}$ from vacuum at an $11°$ angle of incidence. Evaluation of the reflectivity and transmissivity over this space using equations (3.92) and (3.93) gives the claimed absorption coefficients α_1 via (3.106) and (3.107). The ratio of the claimed absorption and real absorption coefficients α_1/α then gives the solution topography.
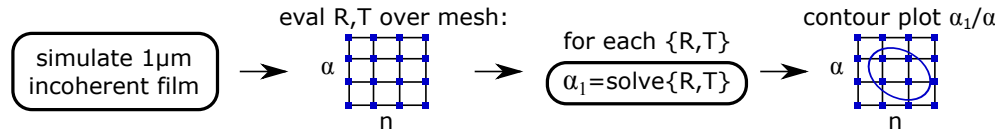
Ideally, if (3.106) and (3.107) found absorption spectra accurately, these plots would show a flat surface near α_1/α=1. Figures 43, 44 and 45 show the resulting contour plots: respectively ignoring reflection, with external reflection (but not internal reflections) and with internal reflections. The contours reveal **significant errors in the claimed absorption coefficients for media with large refractive indices under weak absorption**. While accuracy improves consecutively after considering external and internal reflections, at $\{n, \alpha\} = \{3.2, 10^4\,\text{cm}^{-1}\}$ only $80\,\%$ accuracy is achieved even after considering internal reflections. The discrepancy results from the assumption of strong absorption (discussed on page 73): that the reflectivity R of the film-vacuum interface is equivalent to the total reflectivity of the optical system R_{tot}. Section 3.13.4 discusses an approach which attempts to correct this issue.

Because the optical path length within the film increases as the film's refractive index increases, so internal reflectivity corrections (and the questionable assumption of strong absorption) become more significant with increasing refractive index. One outcome, for typical semiconductors with refractive index $n = 3$, is that (3.106) and (3.107) suggest negative absorption for respective (actual) absorption strengths below 3×10^3 and $2 \times 10^3\,\text{cm}^{-1}$. Although Figures 43, 44 and 45 are computed for $1.4\,\text{eV}$, barely any difference is evident at other photon energies. Slightly more accurate results are achieved at normal incidence; correspondingly, much worse results occur with increasing angle of incidence. The $11°$ angle of incidence selected here is typical of that seen in instruments which measure reflectivity. Attempting to determine absorption spectra **using only transmissivity leads to errors which reach up to a factor of at least** 50 in the examined domain. Accuracy and consistency of the various relations improves under strong absorption conditions ($\alpha > 10^4\,\text{cm}^{-1}$).

Figure 43: Ratio of absorption (computed with only transmission) to real absorption, see Eqn. (3.106)

Figure 44: Ratio of absorption (computed without internal reflections) to real absorption, see Eqn. (3.106)

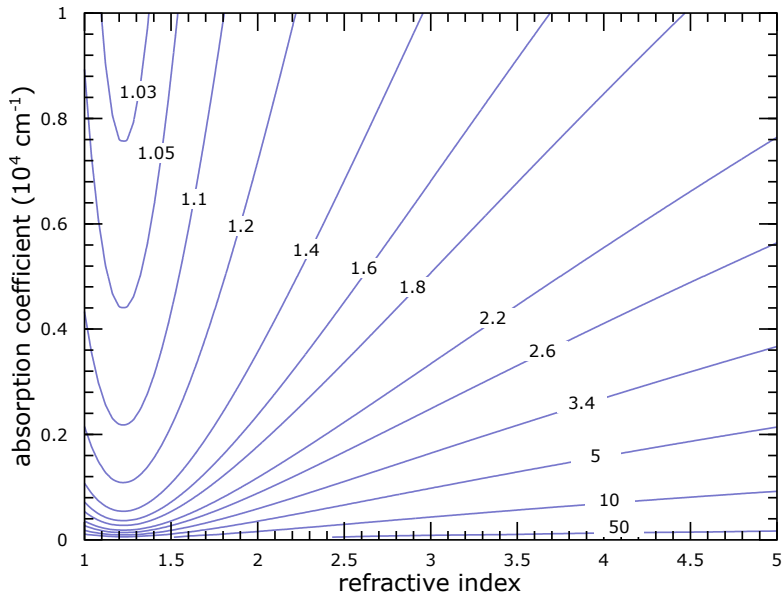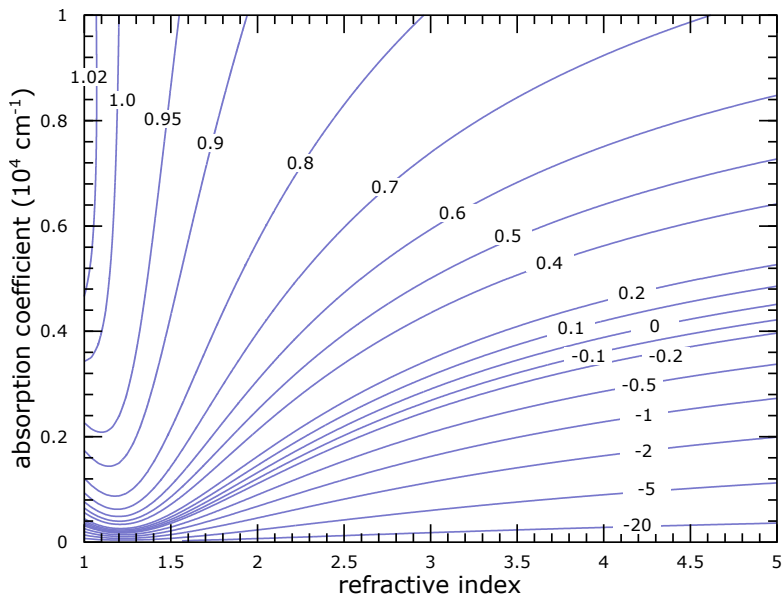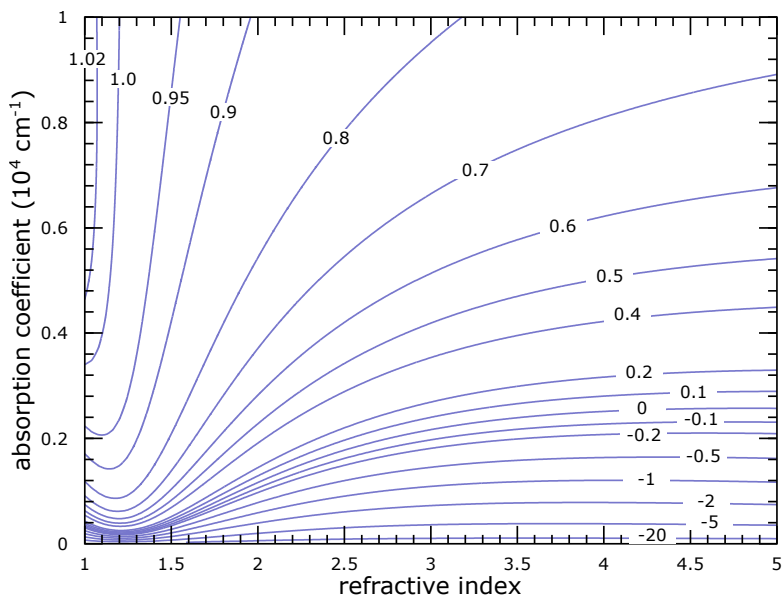**Figure 45: Ratio of absorption (computed with internal reflections) to real absorption**, see Eqn. (3.107)

3.13.4 Absorption correction via a self-consistent intrinsic reflectivity

Algebraic relations for absorption spectra $\alpha(\hbar\omega)$, such as those discussed in Section 3.13.3, are typically written in the intrinsic reflectivity $R(\hbar\omega)$ which is the reflectivity seen by light at a planar boundary between the vacuum and a semi-infinite expanse of the material under investigation; equivalently this is the reflectivity $R_0(\hbar\omega)$ in equations (3.92) and (3.93)

$$R_{\text{tot}} = R_0 \left[1 + \frac{(1 - R_0)^2 \, e^{-2\alpha d}}{1 - R_0^2 \, e^{-2\alpha d}} \right] \quad \text{and} \quad T_{\text{tot}} = \frac{(1 - R_0)^2 \, e^{-\alpha d}}{1 - R_0^2 \, e^{-2\alpha d}} \tag{3.109}$$

Experiments do not directly measure the reflectivity $R_0(\hbar\omega)$; instead, the total reflectivity R_{tot} is seen, which includes contributions from internal reflections at each face. Typically, workers assume strong absorption so that $R_0 \approx R_{\text{tot}}$, yet as discussed in Section 3.13.3, this can lead to errors: particularly for absorption strengths below $10^4 \, \text{cm}^{-1}$. In incoherent media (without interference oscillations), R_{tot} should always satisfy $R_{\text{tot}} \geq R_0$, due to contributions from internal reflections. When R_{tot} differs significantly from R_0, equations (3.109) are not satisfied under the assumption $R_0 \approx R_{\text{tot}}$. Our approach then is to solve (3.109) self-consistently, starting from the guess $R_0 = R_{\text{tot}}$. Some rearrangement of (3.109) gives

$$\frac{1}{T_{\text{tot}}} \left(\frac{R_{\text{tot}}}{R_0} - 1 \right) - \exp(-\alpha d) = 0 \tag{3.110}$$

While (as seen) the transmissivity (3.109) is quadratic in the absorption coefficient, giving

$$\alpha = -\frac{1}{d} \log \left[\frac{-(1 - R_0)^2 + \sqrt{(1 - R_0)^4 + 4 \, T_{\text{tot}}^2 \, R_0^2}}{2 \, T_{\text{tot}} \, R_0^2} \right] \tag{3.111}$$

With a known film thickness d, equations (3.110) and (3.111) are easily solved for the absorption coefficient $\alpha(\hbar\omega)$ and intrinsic reflectivity $R_0(\hbar\omega)$ using a non-linear least squares method (such as the Levenberg–Marquardt implementations in Python or MATLAB: see below). In each iteration, our guess for R_0 is used first in (3.111) and the resulting α creates a residual in the left-hand side of (3.110) which is the minimised quantity.

```matlab
function Main() % TODO: set sFileNameWavenumRT, iNumHeaderRows, iColumnR, iColumnT, iColumnWavenum, fThicknessM
   global mxExprWavenumRT iThisRow fThicknessM fAlphaPerM;
   mxExprWavenumRT=dlmread(sFileNameWavenumRT,'\t',iNumHeaderRows,0);
   mxAbsorptionPerMetre=[];
   for iThisRow=1:size(mxExprWavenumRT,1)
      fIntrinsicR=[mxExprWavenumRT(iThisRow,iColumnR)];
      fIntrinsicR=fsolve(@FitIntrinsicReflectivity,fIntrinsicR);
      mxAbsorptionPerMetre=[mxAbsorptionPerMetre; mxExprWavenumRT(iThisRow, iColumnWavenum), fAlphaPerM ];
   end % finished: now save [mxAbsorptionPerMetre]
end
function y=FitIntrinsicReflectivity(fIntrinsicR) % TODO: implement SolveAbsorption()
   global mxExprWavenumRT iThisRow fThicknessM fAlphaPerM;
   fAlphaPerM=SolveAbsorption(fIntrinsicR,mxExprWavenumRT(iThisRow,iColumnT),fThicknessM);
   y=(mxExprWavenumRT(iThisRow,iColumnR)/fIntrinsicR-1)/mxExprWavenumRT(iThisRow,iColumnT)-exp(-fAlphaPerM*fThicknessM);
end
```

With the assumption of a specific substrate refractive index, the method may be extended for double-layers. As seen in Section 3.13.3 the correction gives impressive results for incoherent single layers. In such systems, using the intrinsic reflectivity R_0 in equation (3.108) gives a refractive index which is particularly stable against noise.

3.14 Comparison of approaches for photovoltaic materials

This section evaluates approaches for refractive index and absorption spectra determination (as discussed) for the thin-film photovoltaic absorbers cadmium telluride (CdTe) and gallium arsenide (GaAs). A gallium antimonide (GaSb) upon GaAs multi-layer is also evaluated. In addition to the methods of Section 3.13.3, the MATLAB transfer-matrix code (page 63) was extended to allow the optical constants of coherent and incoherent, single and double multi-layers (respectively labelled $1c$, $1i$, $2ci$, $2ii$) to be fitted via a non-linear least squares method to experimental reflectivity R and transmissivity T spectra. Because distinct solutions are feasible under strong absorption (see page 69), an initial point $\{n(\hbar\omega), \alpha(\hbar\omega)\}$ is selected at an energy well above the fundamental absorption onset, and fitting then proceeds consecutively towards lower energies. After the first few solutions, the energy derivatives of the trajectory through $\{n, \alpha\}$ space are used to predict the positions of subsequent initial points.

3.14.1 Cadmium telluride on glass (by magnetron sputtering)

Cadmium telluride is one of the leading compounds for commercial thin-film photovoltaics. With cell efficiencies exceeding $20\,\%$, CdTe modules are competitive with polycrystalline silicon. To illustrate aspects encountered in spectroscopy of real-world materials, FTIR reflection and transmission spectra were taken of a $2\,\mu m$ CdTe film deposited by magnetron sputtering on soda-lime glass; absorption spectra computed by various alternative methods are presented in Figure 46. Because the R and T spectra showed interference oscillations, the models chosen included coherent and incoherent films upon incoherent substrates, incoherent single-layers as well as the correction of Section 3.13.4 and the relations (3.106) and (3.107).
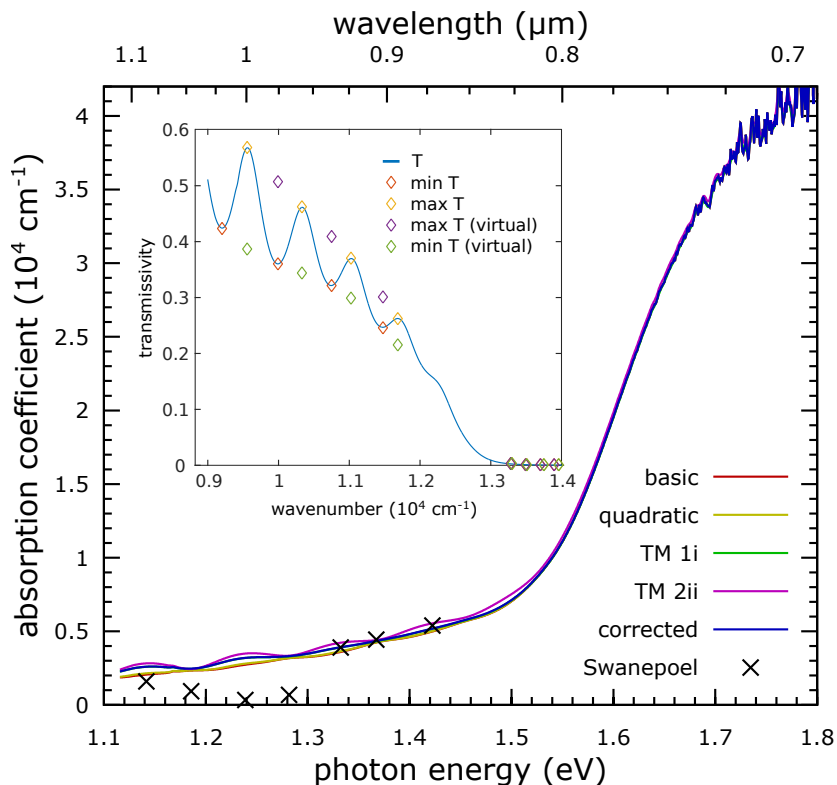
Figure 46: CdTe absorption spectra, via the **basic** and **quadratic** relations (3.106) and (3.107), incoherent films via the transfer matrix (TM) method, the Swanepoel method, and the correction of Section 3.13.4.

Surprisingly, in contrast to Figure 42 the different approaches lead to essentially no significant differences between the absorption spectra; however, the reflectivities ($R \sim 14\,\%$) were considered somewhat smaller than seen typically, and above $1.5\,\text{eV}$ the real part of the claimed refractive index was just 2.2, whereas elsewhere this is found to be nearer 3 for both zinc-blende and wurzite phases[146,147,148]. Considering possible issues, a Swanepoel[128] implementation written in MATLAB was applied: as this envelope method computes absorption spectra entirely from the transmissivity (and additionally provides an independent thickness estimate). After locating extrema via an automated method (see the inset in Fig. 46), this analysis indicated a $1.75\,\mu\text{m}$ film thickness and $n = 2.6$ for the highest three extrema chosen (in the medium absorption regime near $1.4\,\text{eV}$), suggesting an inconsistency.

For a successful analysis the measured signals must be reliable and the selected optical model should be relevant for the specimen. In particular, it is possible that the bulk reflection and transmission behaviour is poorly described by the specular reflection and Snell refraction angles if the sample surfaces are especially rough (feasible for certain synthesis methods, e.g. consider light deflected by the angled sides of a prism): in which case adequate signal-referencing may be impossible. Under such conditions, further complications arise due to the different acceptance angles of the R and T instrumental beam-paths. Additional systematic errors may arise from unidentified experimental factors or inappropriate optical models, e.g. significant changes may occur if the surface reconstructs or if another compound is present (such as oxides[149,150] which develop on CdTe).

With the luxury of prior-published data for the material under study, one simple and valuable means of assessing the assumed optical model is to simulate the expected R and T spectra using the methods of Section 3.11. Figure 47 shows such simulated spectra (using the ellipsometric work of Treharne et al.[146]) alongside the experimental FTIR spectra.
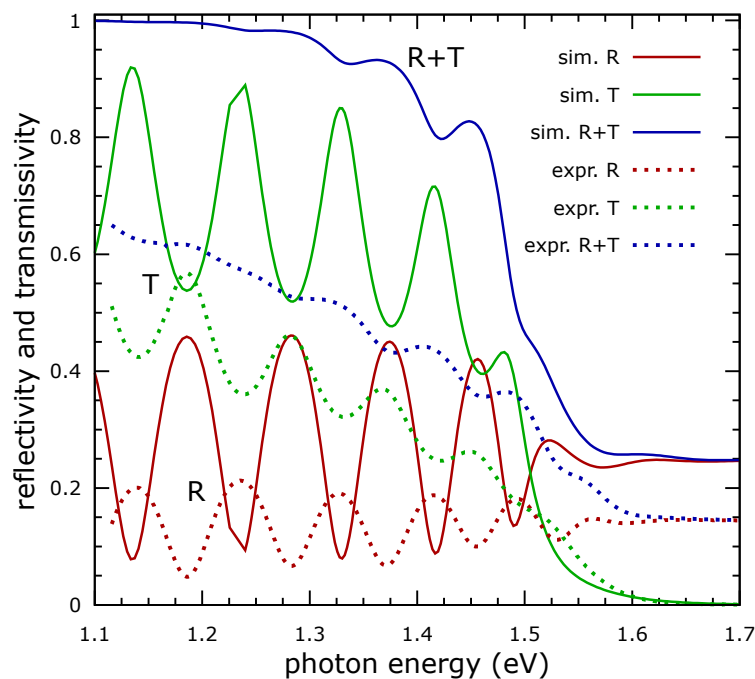
Figure 47: Simulated and experimental CdTe R and T spectra, suggesting $\sim 35\,\%$ of light is undetected.

The comparison suggests that both the experimental R and T signals are significantly under-estimated. While an underestimated sample thickness could explain a reduced magnitude for T, this is not the case for R, leading to the conclusion that either all the published CdTe work is in error (and a lower 2.2 index is in fact sensible), or some systematic factor is present (such as uncollected diffusely reflected light). The latter hypothesis is favoured here: both as a rough surface seems feasible given the sputtered nature of the material, and specifically in light of the rounded and angular CdTe crystallites reported in the sputtered synthesis of Ghorannevis et al.[151] and Kwon[152] (see Figure 4.3.3 on Kwon's page 82). No traces from coherent models are shown in Figure 32 as no continuous solutions existed below $1.6\,\mathrm{eV}$ (supporting the hypothesis). Conventional spectrometers, which require the conditions of geometrical optics, are vulnerable to errors when measuring rough samples: as the diffuse R and T spectra cannot be reliably attained. Instead, instruments which measure energy over a hemisphere (such as those equipped with integrating spheres) or methods which utilise internal references (such as spectroscopic ellipsometry) are necessary.

3.14.2 Gallium antimonide upon gallium arsenide (by molecular-beam epitaxy)

With ideal optical absorption properties, gallium arsenide is the premier material for high-performance thin-film and multi-junction photovoltaic applications. To assess approaches for the determination of optical constants in highly crystalline systems, FTIR reflection and transmission spectra were taken of a $2.6\,\mu\mathrm{m}$ thick gallium antimonide film, grown by molecular beam epitaxy (MBE) upon a $0.5\,\mathrm{mm}$ thick gallium arsenide substrate, and separately of a semi-insulating $0.5\,\mathrm{mm}$ thick GaAs substrate. Both sets of spectra showed $R + T \approx 1$ for sub-gap photon energies, indicating conditions of geometrical optics.
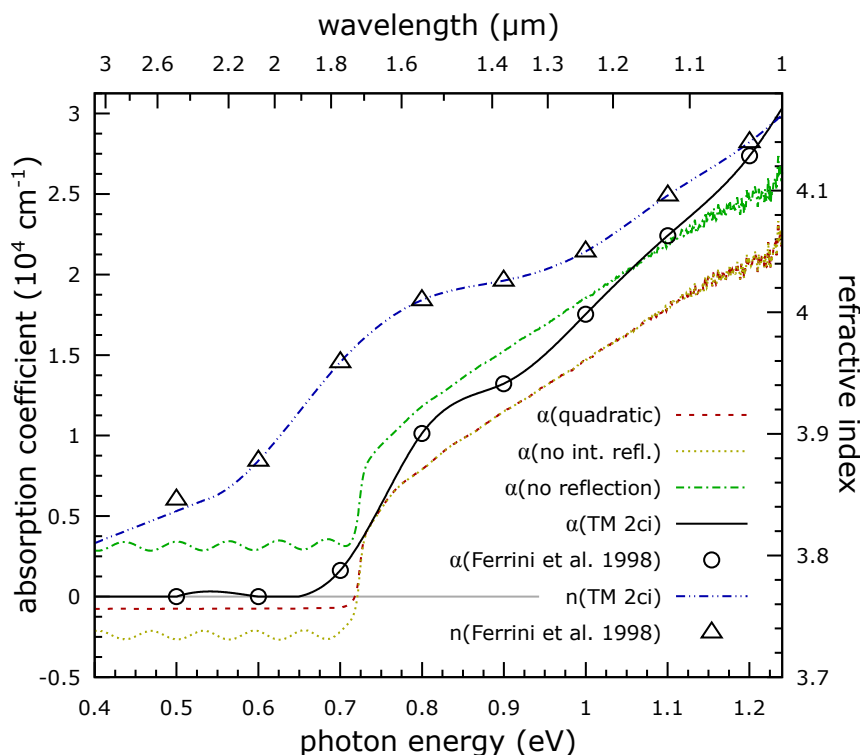
Figure 48: GaSb absorption and refractive index via the quadratic relation (3.107), eqn. (3.106) with R and with R=0, and separately via the transfer matrix (TM) method, consistent with Ferrini et al.[153]

As the GaSb/GaAs spectra showed interference oscillations, a coherent film upon incoherent substrate model was used for evaluation of absorption and refractive index spectra: these and the results of relations (3.106) and (3.107) are shown in Figure 48.

As was found in Figure 42, ignoring external or internal reflections leads to significant errors in the medium absorption regime below $10^4 \, cm^{-1}$, and negative absorption results for relations (3.106) and (3.107) below $0.72 \, eV$; interestingly, because these simple relations neglect the substrate (which reduces the optical path length in GaSb), the absorption strength is increasingly underestimated with increasing photon energy: by chance this happens to make the no-reflection approach initially seem reasonable by a double cancellation of errors. As seen in Figure 48, the refractive index and absorption spectra evaluated by the transfer matrix method here agree rather impressively with the ellipsometric spectra of Ferrini et al.[153], leading to some confidence in the approach for high quality systems.

3.14.3 Gallium arsenide (Czochralski wafer)

The GaAs absorption and refractive index spectra, with the specimen treated as a single incoherent model, are shown in Figure 49, along with the results of relations (3.106) and (3.107). The same discrepancies as noted previously are evident in the elementary relations; however, the thickness of this sample precludes examination of much of the absorption edge (the sample being ~ 500 times too thick). Nevertheless, the sub-gap refractive index computed by the transfer-matrix method is reasonably close to that found by Skauli et al.[154], who used a non-linear optics technique. It is interesting to examine the spectra which result from examination of this sample by spectroscopic ellipsometry (see Figure 50), as ellipsometric spectra are not necessarily dependent upon transmissivity data.
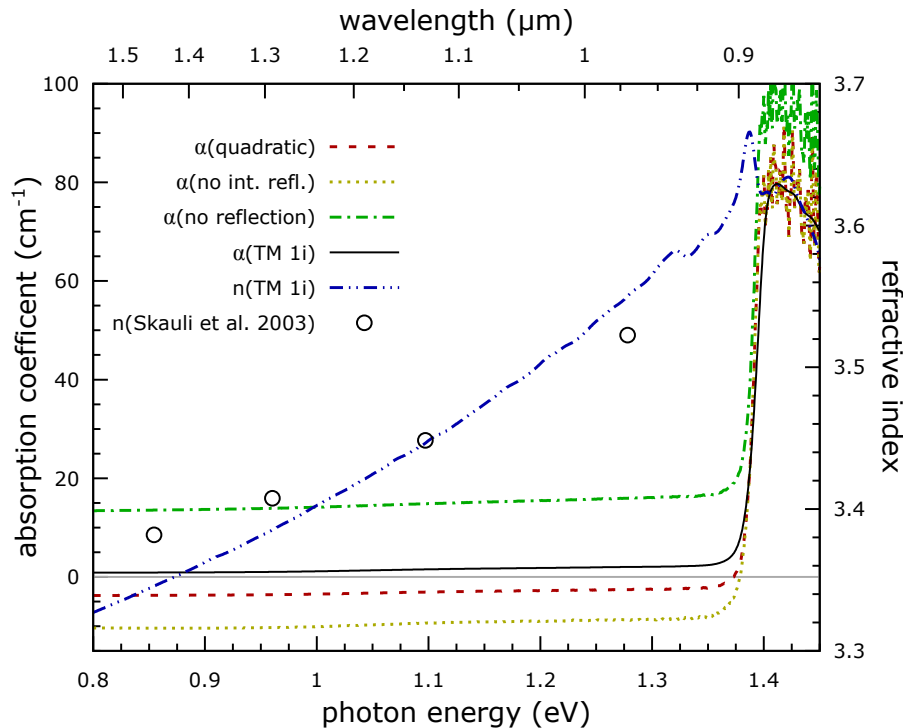
Figure 49: GaAs absorption coefficient and refractive index via (3.106) and (3.107) and single-layer TM method. The index is close to Skauli et al.[154]. The sample is too thick to fully observe the absorption edge.
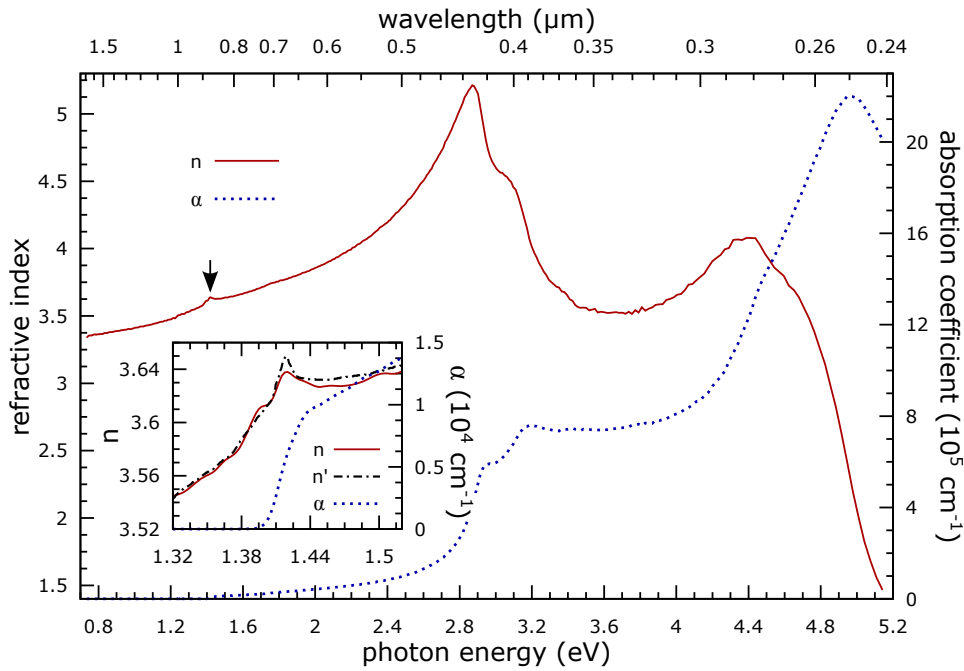
Figure 50: GaAs ellipsometric absorption coefficient and refractive index. Inset expands region near 1.42 eV. Spectra taken via Woollam M-2000-UI at 300 K and 65, 70 and 75° with a 0.01 eV B-spline fit.

The ellipsometric refractive index and absorption spectra seen in Figure 50 match those found previously [155,156] rather impressively: with significant spectral features consistent over a wide wavelength range. The inset expands the region in the proximity of the GaAs band gap near 1.42 eV and shows good agreement with the refractive index of Casey et al. [155], who also reports absorption onsets initiating prior to 1.42 eV. Interestingly, despite the impressive efficiencies of GaAs cells the initial absorption strength is not especially strong when compared to emerging photovoltaic technologies. The absorption edge weakens [157] upon reaching $10^4 \, \mathrm{cm}^{-1}$, then increases relatively slowly [155], only reaching $2 \times 10^4 \, \mathrm{cm}^{-1}$ near 1.7 eV. As seen, the ellipsometer is a fantastic tool which provides supporting and complementary evidence for optical spectra determined by other means.

3.15 Conclusions

This chapter has reviewed various practical aspects for the determination of optical constants (from reflectivity $R(\hbar\omega)$ and transmissivity $T(\hbar\omega)$ spectra) within a photovoltaic context. In the prior pages we have seen: how R and T should generally have gradients of opposite signs with respect to photon energy (or wavelength); that the sum $R + T$ should not normally exceed unity; that for thin-films ($d = 1 \, \mu m$), one should be wary of small sums $R + T < 0.5$: which may suggest diffuse reflection or transmission; that solving for $\{n, \alpha\}$ may be a nontrivial inverse problem; that R and T simulations (e.g. via the transfer-matrix method) are practically useful for spectral assessment; that spectral envelope methods may provide additional information; the strengths and weaknesses of different approaches: the transfer-matrix method may be very successful for fitting good quality specimens (but is not foolproof), and that where possible, spectra should be corroborated with a complementary method, such as ellipsometry. The following chapters utilise various ideas developed in this chapter.

Chapter 4

Temperature-dependent structural and optical properties of copper nitride Cu$_3$N

4.1 Introduction

Copper nitride Cu$_3$N is a relatively unexplored metastable semiconductor which has been studied for potential applications in write-once optical data storage[158,159,160], resistive random access memories[161], hybrid organic-inorganic solar cells[162], magnetic nanostructures and spin barrier tunnel junctions[163], and for rendering conductive dots and lines by mask-less laser or electron beam writing[164,165,166]. Some of these applications exploit the low decomposition temperature of $\sim 300\,°C$, above which Cu$_3$N decomposes to metallic copper and N$_2$. Lately Cu$_3$N has attracted interest as a candidate non-toxic, earth-abundant absorber for thin-film photovoltaics[161] (TF-PV). Favourable characteristics for PV applications include a near-optimal band gap for AM1.5 solar illumination; strong above-onset absorption; material which is dopable both p- and n-type: suggesting potential realisation of pn-homojunctions; defect tolerance, and a surface which may be passivated by a native oxide; as well as good material stability and established polycrystalline growth routes[167].

Defect tolerance is highly desirable for the commercial viability of emerging photovoltaic technologies; this is demonstrated in the impressive efficiencies achieved by Cu(In,Ga)Se$_2$ (CIGS) solar cells, which now exceed 20 % despite the use of polycrystalline, defective and non-stoichiometric material[168,169]. In Cu$_3$N, defect tolerance arises from a conduction band minimum and valence band maximum which derive from respective bonding (lower energy) and anti-bonding (higher energy) molecular levels, contrary to the typical arrangement seen in semiconductors[167]. With this arrangement, defect states are set within the bands or at shallow levels near the band edges, with potentially reduced non-radiative recombination and improved cell efficiencies. Such a band configuration is also seen in CuTaN$_2$, which shares the N-Cu-N bond chains seen in Cu$_3$N[161]. The recent findings, based on Rutherford backscattering spectroscopy[170,171], ellipsometry and theoretical work[167], that the Cu$_3$N (001) surface may be protected by a $\sim 6\,nm$ native Cu$_2$O oxide layer with reduced energy and zero density of states, are particularly attractive to solar cell designers as additional passivation steps are often used to trap photo-generated carriers and minimise surface recombination.

Cu_3N crystallises in a rather open, cubic anti-ReO_3 structure (space group $Pm\bar{3}m$, number 221), as first determined by Juza and Hahn[172], comprising a cubic network of vertex-connected NCu_6 octahedra, see Figure 51. The structure is rather similar to the anti-perovskite structure ABX_3 (where A and B are anions and X is a cation) without the A anion in the $\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$ body-centre position (Wyckoff b-position). The space vacated leaves substantial room to accommodate structural distortions or for hosting impurities. Indeed, the similarity with the perovskites suggests that the rich display of perovskite structural distortions may present in the Cu_3N system. This possibility is studied in Section 4.5.

Some spread is apparent in various physical properties reported in the literature for Cu_3N, as illustrated by the lattice parameter and band gap which range from $3.807(4)$[172] through $3.89\,\text{Å}$[171] and 0.25 to $1.9\,\text{eV}$[173] respectively. Such variability is thought to result from differences in nitrogen content, which may vary at growth-time due to nitrogen liberation[173], and is implicated in affecting the stoichiometry[174], band gap[175], native acceptor concentration[176], conductivity[177], preferential orientation[174] and decomposition temperature[173]. Nitrogen may also migrate towards the surface, creating two phases with distinct nitrogen concentrations, lattice parameters and physical properties[171]. In the worst cases nitrogen emission may introduce voids and bubbles in films seen under atomic force microscopy[171], or copper inclusions leading to metallic conductivity[178]. Stoichiometric, copper- or nitrogen-rich films may all be prepared, with excess copper or nitrogen atoms likely incorporated in the body-centre position or at grain boundaries[171,179].
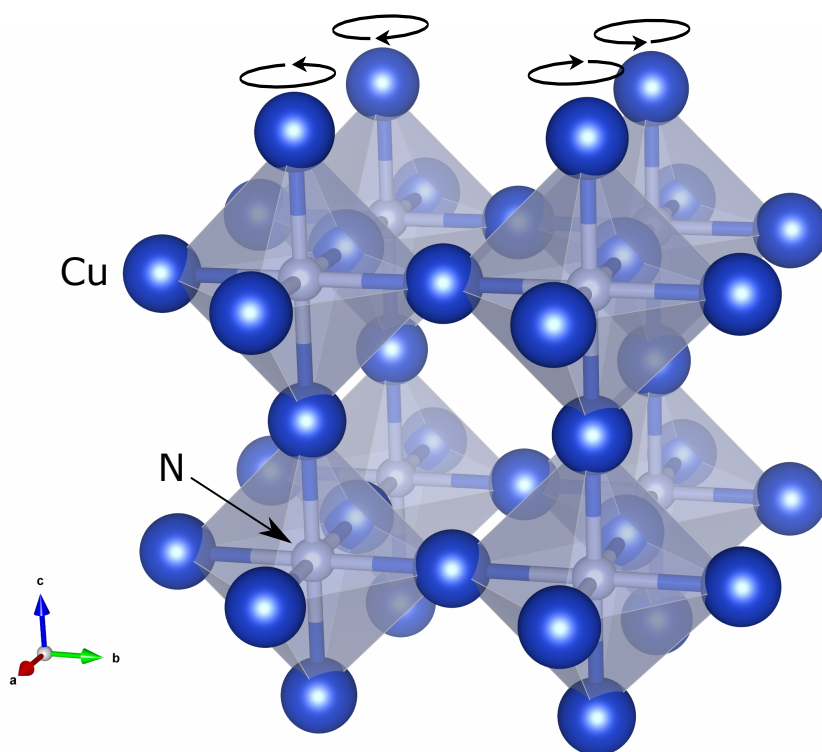
Figure 51: Anti-ReO_3 structure of Cu_3N. The unit cell comprises one nitrogen (silver) and three copper (blue) atoms; the crystal structure is then a cubic network of vertex-connected NCu_6 octahedra. The central void provides space for accommodation of guest species, leading to the Cu_4N and Cu_3N_2 phases. The rotations at the top depict potential perovskite rigid unit modes, as discussed further in Section 4.5.

Given the above, one could take the view that stoichiometric control in Cu_3N is a significant headache; however, because excess copper and nitrogen atoms incorporate respectively as intrinsic donors and acceptors[167,177,179], the above variability may actually be considered an asset. Indeed, the nature of the structure suggests significant opportunities for synthesis with inert or functionalising components in the body-centre position. Copper-rich material[167,173,177,180] is n-type with resistivities from 10^{-6} to $10\,\Omega\,m$, mobilities of 0.1 to $19.5\,cm^2/(Vs)$ and carrier concentrations of 10^{17} to $10^{20}\,cm^{-3}$. Nitrogen-rich, p-type material[167,179,181,182] shows resistivities of 10^{-5} to $11\,\Omega\,m$, mobilities of 1 to $10\,cm^2/(Vs)$ and carrier concentrations of 10^{16} to $10^{17}\,cm^{-3}$. Resistivities typically decrease by an order of magnitude with increasing dopant concentration[173,179,181], with band gaps reducing by some $300\,meV$ (it is not yet clear whether band gap reduction arises from perturbations from dopant levels, or some other factor). Typical resistivities of 0.01 to $0.3\,\Omega\,m$ are seen in stoichiometric material[178,183]. For comparison, doped[184,185] and high resistance[186] GaAs show respective resistivities of 6×10^{-4} to $0.03\,\Omega\,m$ and $10^5\,\Omega\,m$; while metals such as copper[187] have resistivities of $1.67 \times 10^{-8}\,\Omega\,m$. The observation of Maruyama and Morishita[177], that Cu_3N is a conductor when found with lattice parameters greater than $3.868\,\text{Å}$, should be taken in light of these considerations.

Under copper- or nitrogen-rich growth conditions, physical properties may have significant contributions from the nearby cubic anti-perovskite $Pm\bar{3}m$ Cu_4N and Cu_3N_2 phases[188], which are identical but for a respective copper or nitrogen atom sited in the $\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$ body-centre position. In x-ray diffraction, Cu_3N, Cu_4N and Cu_3N_2 each show reflections in identical positions: the Cu_3N and Cu_3N_2 patterns differ only in the relative intensity of the (100) reflection, whilst certain Cu_4N lines show very weak intensities, e.g. (100), (110), (210) and (211). Where preferential orientation is also present, it may be difficult to even confirm the presence of Cu_4N or Cu_3N_2. As stoichiometry is sometimes determined solely from the XRD lattice parameter using Vegard's law type arguments[171,176], i.e. $a = (1 - x)a_{Cu_3N} + x\,a_{Cu_4N}$, this explains some of the variability seen in the literature. Preferential orientation is often observed in x-ray diffraction of Cu_3N films, with (100) and (111) orientations rather common[161,189]. Orientation is strongly dependent on substrate temperature: which controls the surface diffusion of molecules during synthesis[190]. (111)- and (100)-oriented films are seen at respectively lower and higher substrate temperatures[190]. Because (111) planes comprise only nitrogen atoms, while $(m00)$ planes feature predominantly copper atoms, liberation of these atoms from the surface leaves (111) and $(m00)$ oriented films which are respectively Cu- and N-rich[173,174,183].

Because Cu_3N is a metastable compound it would not exist in thermodynamic equilibrium. While the compound is more stable than free copper or nitrogen atoms, it is less stable than metallic copper or molecular N_2, as seen in the following reactions[161]

$$3\,Cu + N \rightarrow Cu_3N + 14.4\,eV \tag{4.1}$$

$$Cu_3N \rightarrow 3\,Cu(\text{metal}) + 0.5\,N_2 + 0.8\,eV \tag{4.2}$$
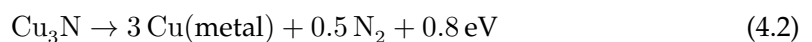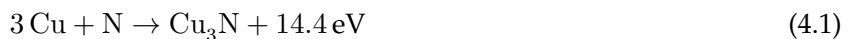
Thus a small decomposition barrier enables the existence of Cu_3N. The precise temperature at which decomposition occurs varies in the literature (as discussed above) from 100 to $470\,°C$, with an average of $280 \pm 80\,°C$, with the lower temperatures seen on annealing under vacuum[180,191,192]. Perhaps due to the metastability of the Cu_3N system, the phase diagram is somewhat unexplored. Searches for pressure- and temperature-dependent structural phase transitions[193] found a reversible transition to a metallic tetragonal phase designated HP-Cu_3N (space group I4/mmm, number 139) at $5\,GPa$. Despite the potential for temperature-dependent reconfiguration of the Cu_3N structure, no temperature-dependent structural phase transitions have yet been reported. Section 4.5 discusses possible perovskite-style rigid unit mode distortions in Cu_3N.

Reactive RF magnetron sputtering is by far the favoured method[161,167,183,190,194] for Cu_3N synthesis, followed by reactive DC magnetron sputtering[179,182,187,195]. Additional synthesis methods include DC triode sputtering[171,176], ion assisted deposition[158], DC plasma nitridation of copper substrates, chemical vapour deposition[189], chemical ammonolysis[172,196,197], reactive pulsed laser deposition[198], atomic layer deposition[199], molecular beam epitaxy[170] and use of a RF plasma jet chemical reactor[200]. Thin films of Cu_3N then have a mahogany or orange-brown colour[201], with thicker layers appearing dark green. After decomposition, the material takes on a red colour[191]. Post synthesis, Cu_3N films seem to be rather stable, surviving unchanged in conditions of $60\,°C$ and $95\,\%$ humidity for 15 months[158], and with no structural (x-ray diffraction) or conductivity changes seen after 12 months[161,200].

4.2 Research objectives and context

This chapter reports the results of temperature-dependent x-ray diffraction (XRD) and Fourier transform infrared spectroscopy (FTIR) studies of polycrystalline Cu_3N films deposited on glass substrates (Corning Eagle XG, an alkaline earth boro-aluminosilicate glass[202]) by reactive magnetron sputtering, comprising Ar and N ion irradiation of Cu metal targets. Further details of the deposition are discussed in Zakutayev et al.[167]. For extensive reviews of magnetron sputtering, see Kelly and Arnell[203] or Bräuer et al.[204].

Prior to this work, XRD of microcrystalline Cu_3N material[193] at atmospheric pressure between 20 and $300\,K$ suggested a linear thermal expansion coefficient of $6.4(3) \times 10^{-6}\,K^{-1}$; in contrast, XRD of Cu_3N single crystals at $130\,K$ and $294\,K$ found no significant changes in the $3.819(1)\,Å$ lattice parameter[197], but suggested highly anisotropic copper displacements and a minor distortion from the anti-ReO_3 structure: with the N-Cu-N bond angle relaxing from 173° at $294\,K$ to 176° at $130\,K$. This effect was explained as due to decreased mobility of the rigid NCu_6 octahedra at the lower temperature; however, while accepting copper anisotropy and some disorder, static octahedral distortions were rejected in higher precision neutron diffraction studies[196]. Elsewhere, extended x-ray absorption fine structure (EXAFS) between 10 and $300\,K$ set an upper limit of $0.005\,Å$ on any elongation of the Cu-N and Cu-Cu bonds[205], consistent with the $0.0037\,Å$ elongation expected from linear expansion[193]. These works leave some questions on the nature of thermal expansion in Cu_3N.

4.3 Experimental methods

Temperature-dependent x-ray diffraction (XRD) was performed at five temperatures between 100 and $280\,$K using a Supernova x-ray diffractometer (Rigaku Oxford Diffraction) with monochromated molybdenum anode ($\lambda_{K\alpha1}$=0.709 317 15 Å and $\lambda_{K\alpha2}$=0.713 607 Å)[206] and Eos CCD detector, and separately at three temperatures between 4.2 and $100\,$K using the XMaS laboratory source (ESRF, Grenoble, France) with a point detector (avalanche photodiode) and monochromated copper anode ($\lambda_{K\alpha1}$=1.540 592 90 Å and $\lambda_{K\alpha2}$=1.544 427 40 Å)[206]. For the work with Mo $K\alpha$, several micrograms of Cu_3N crystallites were removed from the film via scapel and fixed to the staging fibre with high vacuum grease (Dow Corning). With the sample temperature set by a nitrogen Cryojet (Oxford Instruments), phi scans ($0 \leq \phi < 360°$) were collected with a sample-detector distance of $120\,$mm, dark exposures of $120\,$s and 1x1 pixel binning; the resultant 2D images were integrated into $6 < 2\theta \leq 57°$ powder patterns by *CrysAlis Pro* 38.41. Before investigating Cu_3N with this instrument, the vacuum grease was scanned at 100 and $200\,$K to verify that this adhesive contributed no additional reflections. The work with Cu $K\alpha$ covered $20 \leq 2\theta < 71.5°$ and studied the film as-deposited with a $2\,$mm entrance slit and a closed-cycle helium cryostat (Institut Laue-Langevin) with Joule-Thomson jet and beryllium windows. After an initial broad scan of $2\,$min per $0.035°$ 2θ step at $4.2\,$K, a series of narrow scans targeted only the most significant reflections at higher resolution to achieve satisfactory signal-to-noise, given the instrument time available.

Infrared transmission and specular reflection spectroscopy was performed at $11°$ angle of incidence for 17 temperatures between 4.2 and $300\,$K, and photon energies of 0.37 to $3.1\,$eV (3.3 to $0.4\,$μm) using a Bruker Vertex 70v Fourier-transform infrared (FTIR) spectrometer equipped with a combined reflection-transmission accessory and an Oxford Instruments CFV2 continuous-flow helium cryostat. This spectral range was chosen primarily to evaluate any evolution in the fundamental absorption onset, with prior theoretical and experimental
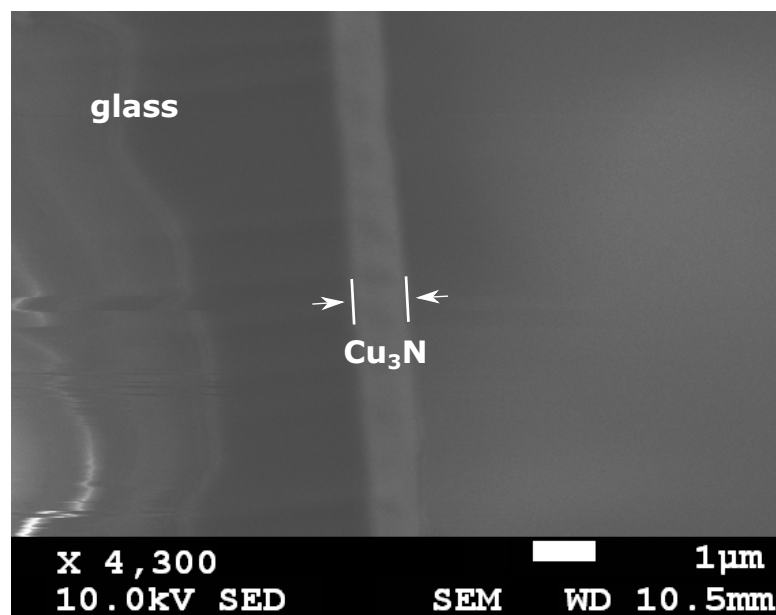
Figure 52: Cross-sectional scanning electron microscopy (SEM): suggesting a Cu_3N film thickness of $1\,$μm.
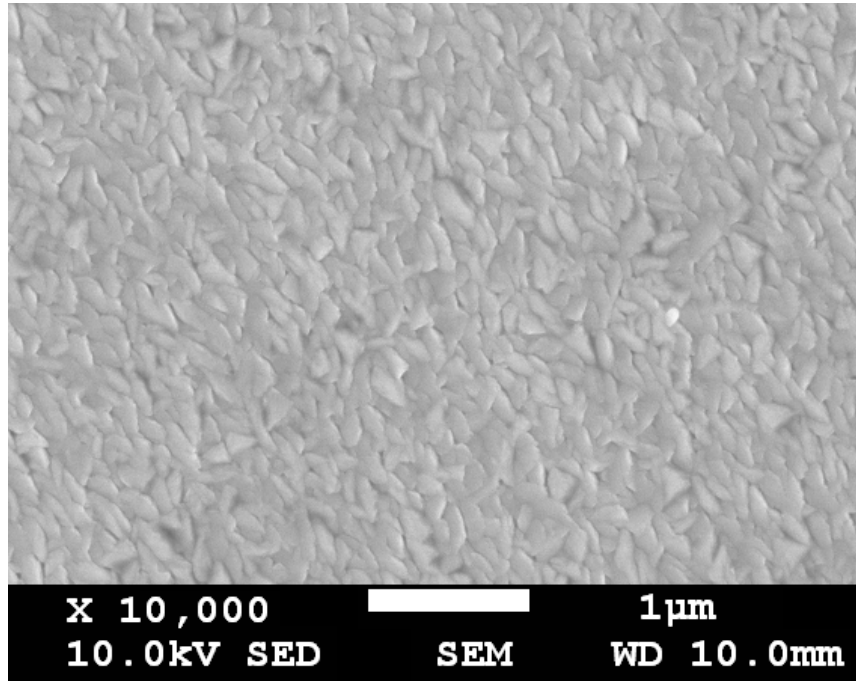
Figure 53: Surface morphology. A 200 nm grain size is suggested by scanning electron microscopy (SEM).

reports suggesting indirect and direct absorption onsets between 0.6 through 2.2 eV, and initial examinations which showed a long transmission edge starting below 1 eV and spanning some 800 meV. Two instrumental configurations were necessary to capture the entire absorption edge from baseline, each with different detectors; a silicon detector was used above 1.1 eV, while a liquid-nitrogen cooled mercury cadmium telluride (MCT) detector was selected for lower photon energies. The conversion from reflection and transmission to absorption spectra is discussed in the following section. To complement the FTIR absorption spectra, the Cu_3N dielectric function was evaluated separately with a Woollam M2000UI (rotating compensator) spectroscopic ellipsometer between 0.7 to 5.1 eV (245 to 1690 nm) at 300 K using incident angles of 65°, 70° and 75°. A film-substrate model comprising surface roughness was fitted to the measured reflection, transmission and depolarisation data in *CompleteEASE* v5.08.

Post-growth Hall measurements in van der Pauw geometry with indium contacts and an Accent HL5500 system found p-type material with a hole concentration of 10^{16} to 10^{17} cm^{-3}, indicating negligible band-filling given the discussion in Section 3.7. X-ray photoemission spectra of the as-deposited films supported theoretical work on the valence band density of states. These spectra were acquired in ultrahigh vacuum with a SPECS monochromatic Al $K\alpha$ source ($\hbar\omega$=1486.6 eV) and PSP Vacuum Technology electron analyser: with 100 scans, 0.2 s dwell time and 10 eV pass energy. Scanning electron microscopy (SEM) using a JEOL JSM 7001F with a 10 keV beam energy revealed polycrystalline films with good substrate coverage and crystallites of typical dimension 100 nm×300 nm, see Fig. 53; cross-sectional SEM indicated a film thickness of 0.98(9) µm, see Fig. 52. Energy dispersive spectroscopy (EDS) using an Oxford Inca X-Act detector found no impurity lines but indicated a somewhat copper-rich film ($\sim$ 80 atomic %), possibly due to reduced nitrogen sensitivity under EDS.

4.4 Structural investigations by x-ray diffraction

4.4.1 Analysis of patterns, indexing and structural refinement

Typical x-ray diffraction patterns are shown in Figures 54 and 55. Reflections associated with the well-known cubic $Pm3m$ anti-ReO$_3$ phase appear in all patterns with little temperature dependence between 4.2 and $280\,$K; however, in initial work with Mo $K\alpha$ at 100, 150 and $200\,$K several additional reflections appeared alongside those associated with the $Pm3m$ phase: with the strongest at 9.73, 11.36, 16.79 and $24.61°\ 2\theta$, see Fig 55. In repeat work with Mo $K\alpha$ to investigate this further, these reflections only developed at $200\,$K. The set of new and $Pm3m$ reflections were not convincingly indexed to a single phase within *GSAS-II*[49].

The new reflections neither matched potential spectral components (Mo $K\alpha_2$, Mo $K\beta$, W Lβ)[206] nor impurity phases: including metallic copper, copper(I) oxide (Cu$_2$O), copper(II) oxide (CuO), copper nitride (Cu$_4$N), α-quartz, or the hexagonal (Ih) or cubic (Ic) phases of water ice. Neither was the borosilicate substrate implicated as the Cu$_3$N crystallites were detached from the substrate prior to starting work. The strong temperature-dependence of the new reflections suggested a distinct phase. With only four reflection positions assigned with confidence, an orthorhombic cell $(4.175, 3.583, 2.983)$ Å was tentatively fitted. Due to the accuracy required to assess lattice expansion in the $Pm3m$ phase, combined with poor
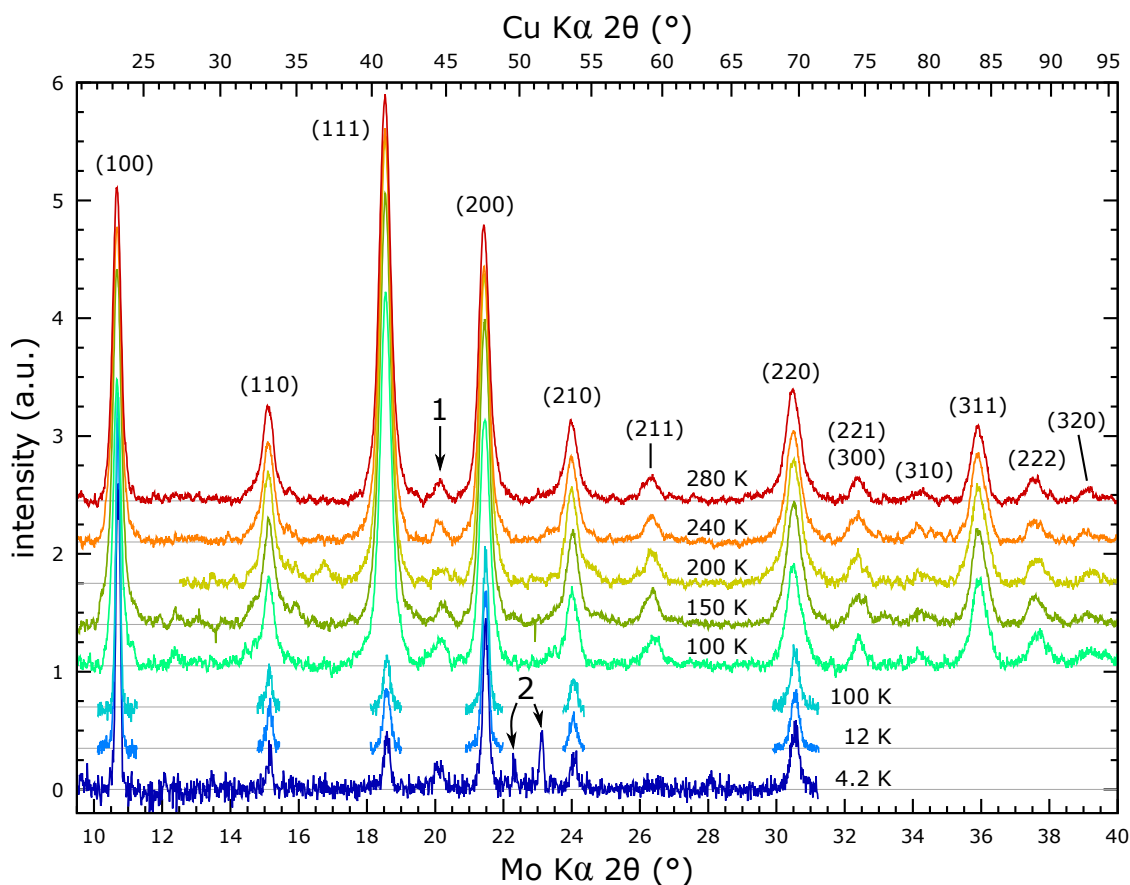
Figure 54: Temperature-dependent x-ray diffraction of Cu$_3$N. Mo and Cu anodes used respectively between 100 to $280\,$K and 4.2 through $100\,$K. Temperature dependent background fitted with splines and subtracted. Reflection possibly associated with Cu near $20°\ 2\theta$ Mo $K\alpha$ (labelled 1). Preferential $(m00)$ orientation and beryllium lines (labelled 2) seen at (49 through $52°\ 2\theta$) at $4.2\,$K.
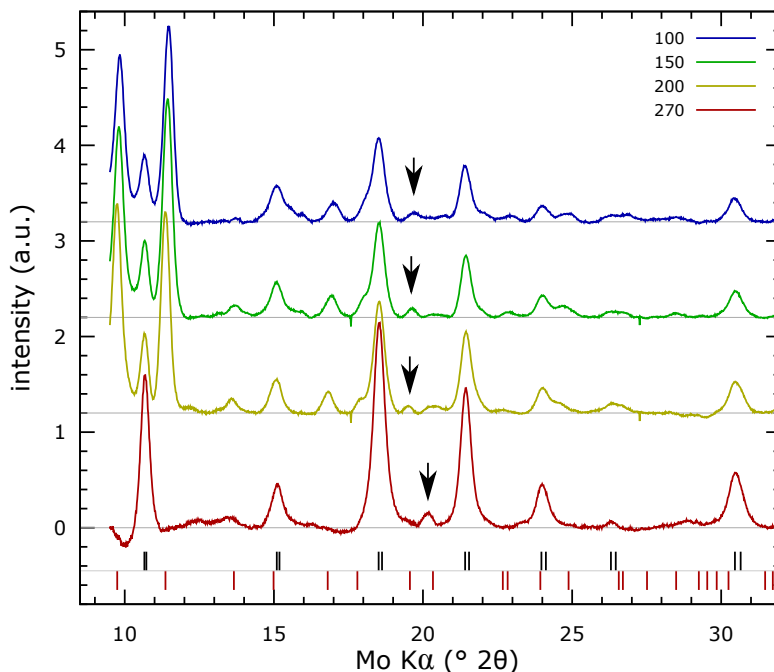
Figure 55: Intruding reflections in Cu₃N XRD pattern. A set of reflections (later established as due to crystalline adhesive grease) are seen in Mo $K\alpha$ with strong temperature-dependence between 100 and $270\,K$. The scale at the bottom gives the reflection positions for the cubic $Pm3m$ (upper) and tentative orthorhombic (lower) phases. The arrow indicates an additional unidentified reflection (see the discussion).

signal-to-noise and limited confidence in the assigned orthorhombic phase, no patterns were fitted where orthorhombic reflections were in close proximity to those from $Pm3m$. With no low-temperature Cu_3N phase transitions reported in the literature, the possibility was considered that the unknown reflections resulted from growing crystallinity in the high-vacuum grease used as an adhesive (use of grease is established practice in low-temperature XRD[207]). The experiments of Figure 55 used hour-long scans with progressively descending temperature: unknown reflections were seen below $270\,K$, yet these were not observed in a $100\,K$ grease-only measurement at the end-of-day. In the experiment of Figure 54, no reflections were seen in a grease-only measurement at $200\,K$ at the start-of-day; measures proceeded with one hour intervals in the sequence $150, 100, 200, 280$ and $240\,K$, with unknown reflections seen *only* at $200\,K$: suggesting a grease freezing point near $200\,K$ and supporting the idea that the grease may remain amorphous upon quenching below this temperature.

A further experiment sought time-dependent crystallisation in the adhesive grease at $200\,K$; crystallisation was successfully found within one hour at $200\,K$, see Figure 56. This pattern explains ~ 10 intruding reflections in the Cu_3N patterns. The evidence indicates a grease freezing point near $200\,K$ and that crystallisation is avoided by quenching (rapid cooling). Indexing Fig. 56 suggests the following candidate space groups and lattice parameters

Space group	M20	a, b, c (Å)	α, β, γ (°)	Vol. (Å³)
$C2/m$	44.9	$4.60, 10.23, 6.02$	$90, 83.97, 90$	282.0
$Cmmm$	18.3	$8.38, 11.95, 2.58$	$90, 90, 90$	258.8

where more reliable fits have higher M20 figures-of-merit[208]. During work, the culpability
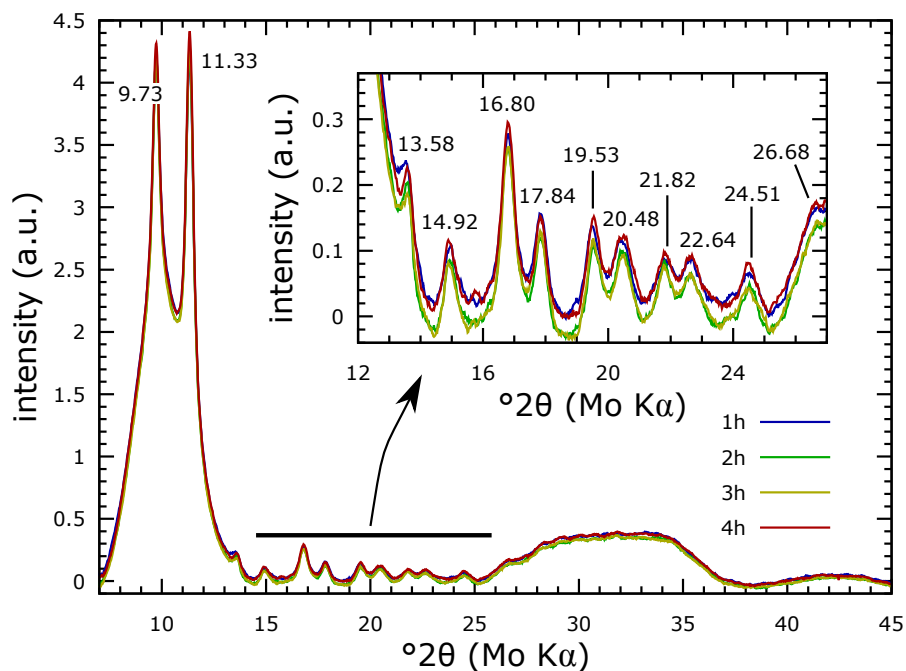
Figure 56: Crystallinity in Dow–Corning high-vacuum grease (used as sample adhesive). Crystallisation develops at 200 K within one hour, but may be avoided by quenching. XRD pattern (Mo $K\alpha$).

of the grease was confirmed rather belatedly, this led to various research directions seeking Cu_3N structural phase transitions (as discussed subsequently). No grease-associated reflections are seen in the Cu $K\alpha$ pattern at 4.2 K, but weak reflections from the beryllium cryostat domes are apparent (Fig 54). These reflections are shifted from their Bragg positions because the domes are displaced from the diffractometer axis. The strongest beryllium reflections (100), (002) and (101) are expected[209,210] respectively at 45.8, 50.9 and 52.8° 2θ, whilst the observed reflections lie at 44.5, 49.6 and 51.5° (Cu $K\alpha$). These angles do not change[211] significantly below 300 K. The expected and observed angles are then consistent within 0.04° 2θ via the relation[212] for sample displacement s, $\Delta2\theta(\mathrm{rad}) = 2s/R \cdot \cos(1\theta)$, with the ratio $2s/R = -0.0255$ and diffractometer radius R.

The development of unidentified reflections in the patterns at lower temperatures motivated consideration of potential distortions or structural phase transitions away from the well-known $Pm3m$ anti-ReO_3 structure, which could possibly present as a function of time or temperature. To evaluate this, the Cu_3N structure was treated as unknown and the 100 K pattern was indexed after fitting the background with a 13-coefficient Chebyschev polynomial in *GSAS-II*; three solutions with minimal volume had convincing M20 figures-of-merit[208]

Space group	M20	Lattice parameter (Å)	Vol. (Å^3)
$Pm3m$	65.3	3.816 57	55.59
$P4/mmm$	132.7	3.818 60, 3.806 07	55.50
$Pmmm$	367.2	3.803 07, 3.818 39, 3.814 13	55.39

The M20 statistic improves, perhaps expectedly, with increased parameter freedom. The patterns generated by the $Pm3m$ and $P4/mmm$ structures are sufficiently close to cubic $Pm3m$ that a favoured lattice is not discernible by eye. Rietveld refinement[47] of structures

in the $P4/mmm$ and $Pmmm$ space groups in *FullProf 2k*[51] (v5.70) failed to find preferred solutions away from $Pm3m$. For example, $P4/mmm$ was tried with N at (0,0,0) and Cu in the Wyckoff (b, f) or (d, f) positions, similar to $Pm3m$ Cu$_3$N respectively either compressed along the c-axis or with one Cu displaced to the body-centre position. Each attempt showed deteriorating R_{wp} and χ^2 statistics.

Rietveld refinement was initially preferred over extrapolation methods for structural evaluation: because full-pattern approaches exploit greater information to select the most appropriate systematic corrections[57]. Refinement of all patterns suggested $(m00)$ preferential orientation, evident in the mismatched Mo $K\alpha$ pattern (200) residuals in Fig 57. Preferential orientation in the Mo $K\alpha$ patterns likely results from incomplete separation of the Cu$_3$N crystallites after removal from the substrate. Perhaps expectedly (due to the borosilicate substrate), the Cu $K\alpha$ patterns showed significant $(m00)$ orientations, which were not fitted acceptably using the models in *FullProf 2k*. The Rietveld fit to the 4.2 K pattern, the dotted line in Fig. 57 (offset vertically for clarity), shows significant discrepancies in the (100) and (111) reflection amplitudes. Considering that proceeding with a Rietveld approach without an acceptable structural model would drive errors elsewhere in the fit, particularly in lattice parameters
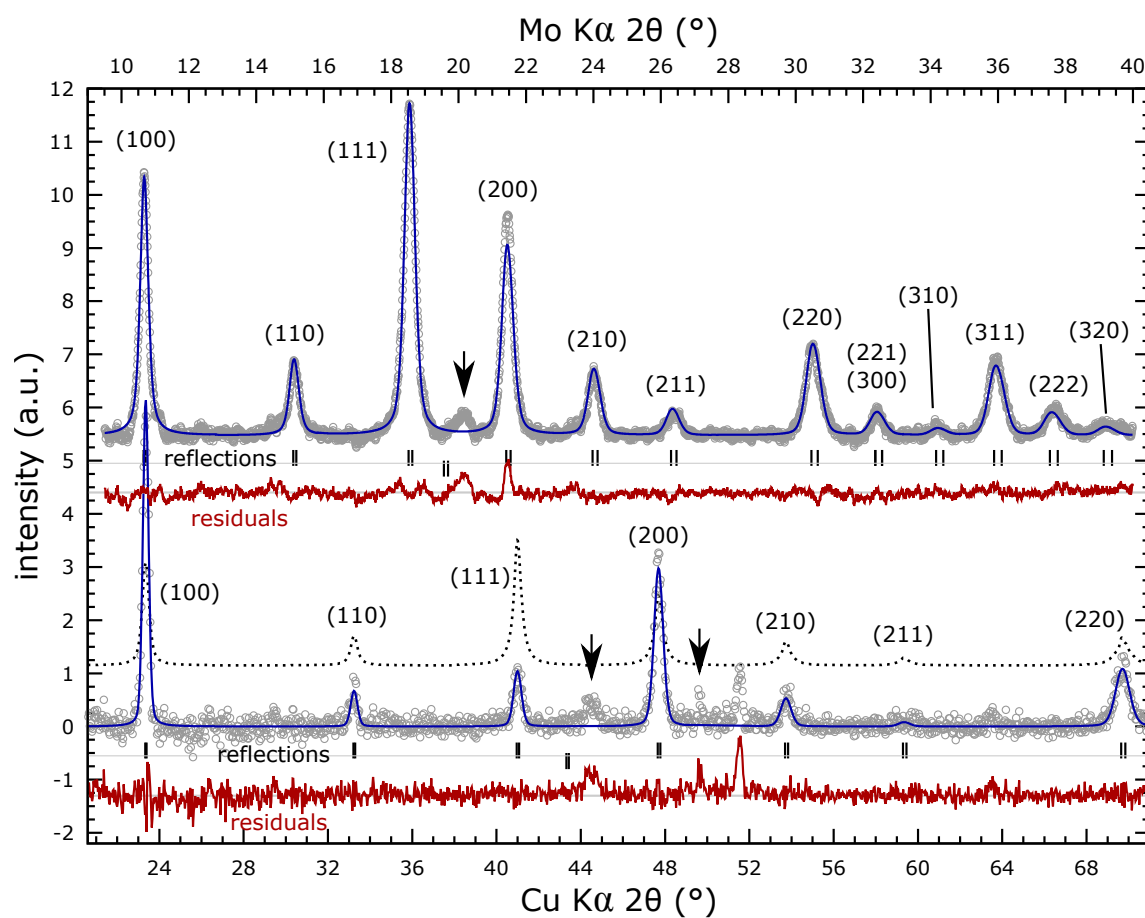
Figure 57: Strutural refinement of Cu$_3$N XRD. Rietveld refinement leads to good fit to Mo $K\alpha$ pattern at 100 K, but gives poor fit (dotted line, displaced vertically for clarity) to Cu $K\alpha$ 4.2 K pattern due to preferential $(m00)$ orientation. Pawley fit successfully fits the 4.2 K pattern after excluding lines associated with beryllium cryostat windows.

and their uncertainties, a Pawley-like[52,53] full-pattern fit was instead implemented in MAT-LAB with the reflection intensities fitted as free parameters. This code utilised the Levenberg–Marquardt non-linear least-squares algorithm and composed the pattern from a single pseudo-Voigt lineshape with conventional $\{U, V, W, \eta, X\}$ linewidth parameters[48], while incorporating corrections for sample displacement and zero error. Prior to each fit, the continuous x-ray background, which evolved somewhat as a function of temperature, was fitted with splines and subtracted. All spectra where fitted using the Pawley-like code to a cubic lattice with good statistics, mean reduced χ^2 of 1.2(1), mean degrees of freedom 3060 (Mo $K\alpha$), 883 (Cu $K\alpha$).

The thermal expansion determined by the Pawley fitting for Cu_3N is plotted in Fig. 58. The lattice parameter as-measured contracts respectively from $3.823\,05(24)$ to $3.8181(3)\,\text{Å}$ between 280 and 100 K; below 100 K the magnitude of thermal expansion reduces significantly, with a $3.8173(6)\,\text{Å}$ parameter seen at $4.2\,\text{K}$. Indeed, the thermal expansion below $150\,\text{K}$ is sufficiently small that the 4.2 and 150 K lattice parameters agree under a Gaussian 3-sigma consistency test

$$\left| a(4.2\,\text{K}) - a(150\,\text{K}) \right| < 3 \times \sqrt{\Delta a^2(4.2\,\text{K}) + \Delta a^2(150\,\text{K})} \qquad (4.3)$$

The consistency in the $100\,\text{K}$ lattice parameter from instruments in different laboratories lends some confidence to the results. Despite finding negligible expansion below $100\,\text{K}$, the results are insufficiently precise to confirm or reject zero or negative thermal expansion here.

The lattice parameter uncertainties resultant from the Pawley fitting, e.g. in Fig. 58, represent the standard error on the mean, as derived from the uncertainties ($\sqrt{N}$, due to Poisson statistics) on the initial 2θ-binned intensities (for Cu $K\alpha$), or from the uncertainties computed by the *CrysAlis Pro* 2D-to-powder pattern integration algorithm (for Mo $K\alpha$). These uncertainties estimate statistical variations in the data: for the correct model, the uncertainty should accurately describe the noise level. This is so for the Cu $K\alpha$ signal: in flat (constant signal level) regions, noise fluctuations are well estimated by the uncertainty magnitude; however, the Mo $K\alpha$ patterns showed excessively large uncertainties in all regions, some 2.5 times larger than the noise standard deviation. The excessive magnitudes seen in the Mo $K\alpha$ error bars thus propagate directly into excessive uncertainties on the computed lattice parameters, resulting in excessively small reduced-χ^2 fit statistics (below 0.2). To correct for this possible instrumental issue, the Mo $K\alpha$ error bars were each reduced by a factor of 2.5, giving mean reduced-χ^2 statistics of 1.2. Because only a scalar correction was performed, the convergence point of the least-squares algorithm was unchanged: the same lattice parameters were determined, with only the magnitudes of their uncertainties adjusted.

4.4.2 Fitting the lattice expansion with a Debye–Einstein–Grüneisen model

Before evaluation of the thermal (linear) expansion coefficient, the set of lattice parameters should be fitted to a suitable model (e.g. a polynomial). Because the expansion model then represents the entire dataset, the expansion coefficient gains resilience to statistical
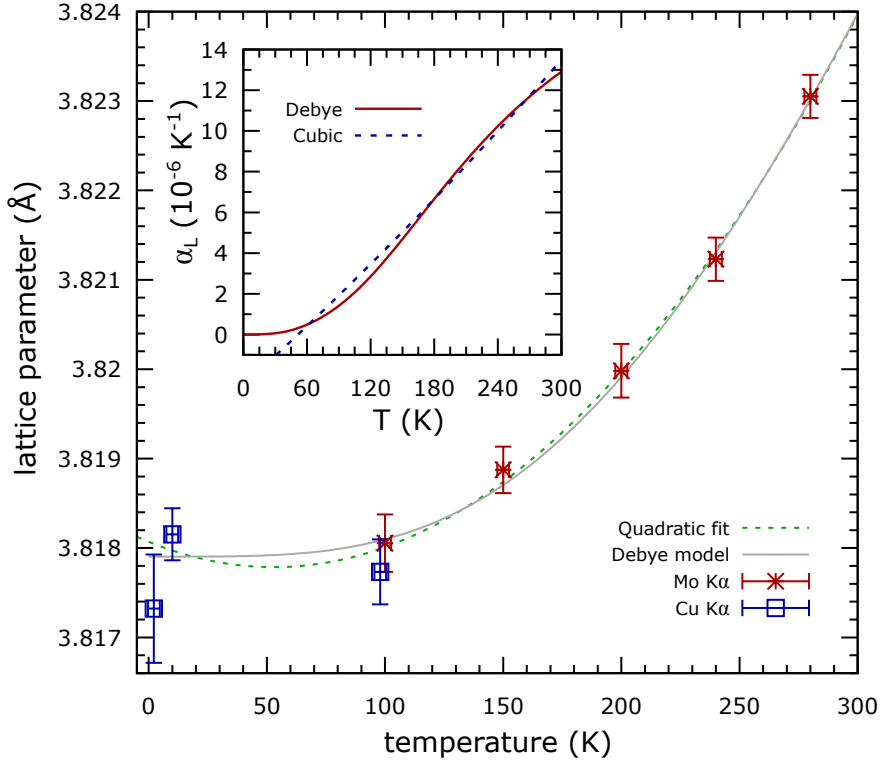
Figure 58: Thermal expansion of Cu₃N. The lattice parameters resultant of a Pawley-like full-pattern fit to a cubic lattice are shown; the thermal linear expansion coefficient resultant from fits to polynomials or to a quasi-harmonic Debye-Einstein model suggest vanishing thermal expansion below 100 K.

extrema in individual lattice parameters, and the thermal expansion coefficient may additionally be estimated away from the experimental datapoints. Here the Debye–Grüneisen model of Sayetat et al[213] is loosely followed. Aside from being a slightly less arbitrary treatment of thermal expansion (when compared to a polynomial), this quasi-harmonic model gives estimates of the Debye temperature (see Appendix B), the zero Kelvin lattice parameter and the average Grüneisen parameter. The approach is 'quasi-harmonic' as no thermal expansion would be seen with the Debye model alone, i.e. with only harmonic vibrations. Instead, the anharmonicity of the lattice vibrations is approximated by combining the Debye model of the lattice specific heat with the semi-empirical Grüneisen relation, which relates a material's vibrational modes and its volume. Starting with the Grüneisen relation between the thermal (volume) expansion coefficient α_V and the lattice specific heat under constant volume C_V,

$$\alpha_V = \gamma \, C_V K / V \tag{4.4}$$

where K, V and γ are respectively the isothermal compressibility, the crystal volume and the average Grüneisen parameter; multiplying by volume V on both sides gives

$$\frac{\mathrm{d}V}{\mathrm{d}T} = \gamma \, C_V K \tag{4.5}$$

Integrating with respect to temperature gives a system in four unknowns

$$V(T) = V(0) + \gamma \, K \int_0^T C_V \left(\theta_D \right) \, \mathrm{d}T' \tag{4.6}$$

In the right-hand side of (4.6) the zero Kelvin lattice parameter [or equivalently the volume $V(0)$], Grüneisen parameter γ, isothermal compressibility K and Debye temperature θ_D are then fitted by least-squares regression to the volumes $V(T)$ associated with the lattice parameters resulting from the Pawley analysis. The thermal expansion in Cu_3N as determined by (4.6), along with the thermal (linear) expansion coefficient as given by (4.4) and the relation between the linear α_L and volume α_V expansion coefficients ($\alpha_V = 3\,\alpha_L$), are plotted in Figure 58. Alternatively, fitting the temperature dependence to a quadratic suggested (with uncertainties indicated in brackets)

$$a(\text{Å}) = 1.01(11) \times 10^{-7}T^2 - 1.1(3) \times 10^{-5}T + 3.818\,07(22) \qquad (4.7)$$

The implementation used differed somewhat from that of Sayetat et al[213], who fitted the product of the isothermal compressibility and the Grüneisen parameter as a single parameter; instead, the Cu_3N bulk modulus B (the inverse of the compressibility, $B = 1/K$) was set to the predicted[214,215,216,217] (and typical) value of $110\,\text{GPa}$, then solved for the Grüneisen parameter directly. This minor change immediately reveals unacceptable Grüneisen parameters (a typographic error possibly exists in the Sayetat et al. paper leading to an error of ~ 30 orders of magnitude in their claimed Grüneisen parameter).

While the fit statistics for the Sayetat et al. model are acceptable, the computed Debye temperature of $930(230)\,\text{K}$ is somewhat larger than the mean (and standard deviation) values $390(50)\,\text{K}$ predicted[215,218,219] for Cu_3N. The Debye model is essentially conceived for the acoustic phonon regime, i.e. for rather low temperatures or for monatomic crystals with one atom per unit cell. While the Debye specific heat should be an accurate model[46] for temperatures well below $0.1\,\theta_D$, most of the points in Fig. 58 lie above this limit even with an overestimated Debye temperature. In Cu_3N, with four atoms per unit cell, optical modes may contribute significantly at medium temperatures, as perhaps supported by the anisotropic thermal displacement parameters determined in Rietveld refinements (discussed shortly).

In light of this, the lattice specific heat C_V in (4.6) was revised to comprise the sum of the Debye and Einstein specific heats, with the Einstein term weighted by a factor $(r - 1)$, where r the number of atoms in the crystal basis[70]. Analysis found that the increased parameter freedom which resulted from separate Einstein θ_E and Debye θ_D temperatures lead to results which were not physically acceptable. Instead, the Einstein temperature was fixed to be 1.5 times larger than the Debye temperature: a reasonable magnitude seen in typical semiconductor phonon dispersion relations. After this change, the Debye temperature converged to $550(150)\,\text{K}$, with a zero Kelvin lattice parameter of $3.817\,88(19)\,\text{Å}$ and an average Grüneisen parameter of $2.3(9)$.

While the lattice parameters are determined most reliably by the above Pawley analysis, a Rietveld refinement of the Mo $K\alpha$ patterns (for which preferential orientation was less of a concern) offers potential insights. Such a refinement in *FullProf 2k* determined lattice parameters in agreement with those of the Pawley analysis, with fractional differences well

below 5×10^{-5}. Refinement of the $200\,\text{K}$ Mo $K\alpha$ pattern was complicated by the proximity of grease reflections: excluding the most significant Cu_3N (100) reflection and the leading grease reflections gave a $200\,\text{K}$ lattice parameter of $3.8180(9)\,\text{Å}$, possibly perturbed by residual grease contributions and further supporting the Pawley approach.

Thermal displacement parameters resulting from the Rietveld analysis indicated respective isotropic and anisotropic displacements for the nitrogen and copper atoms, with copper displacements perpendicular to the N-Cu-N axis growing in magnitude above $150\,\text{K}$. For all fits the background-subtracted, weighted figure of merit R_{wp},

$$R_{wp} = \frac{\sum_i w_i \left(y_{\text{expr},i} - y_{\text{calc},i}\right)^2}{\sum_i w_i \left(y_{\text{expr},i}\right)^2} \tag{4.8}$$

(where $y_{\text{expr},i}$, w_i and $y_{\text{calc},i}$ are respectively the experimental intensity, variance and calculated intensity for each bin i in the subset which contribute Bragg reflections), lay between 11–14 % with a reduced χ^2 between 0.186 and 0.257 (using uncorrected uncertainties) and a statistic $4.12 < R_F < 6.85$ %. Small improvements in the fit were seen with phase admixtures of Cu_3N with either Cu_4N or Cu_3N_2: with R_{wp} statistics of 13.7 and 14.3 for respective Cu_4N and Cu_3N_2 occupancies of $6(1)$ % and $23(4)$ %; however, these did not allow discrimination of a favoured composition, the improved statistics likely resulting from increased parameter freedom in the fit.

4.4.3 Discussion: x-ray diffraction

The Cu_3N lattice parameters determined in this work (below $280\,\text{K}$) are consistent with those of Zachwieja et al[197], who studied single crystals, but are inconsistent with those for the microcrystalline material of Wosylus et al[193] and for the powders of Juza and Hahn[172]. The thermal expansion observed here between 4.2 and $280\,\text{K}$ is consistent with the $0.005\,\text{Å}$ reported in the extended x-ray absorption fine structure (EXAFS) study of Kuzmin et al[205] between 10 and $300\,\text{K}$. However, the suggestion[193] of a $6.4(3) \times 10^{-6}\,\text{K}$ linear thermal expansion coefficient perhaps underestimates and misrepresents the thermal expansion in Cu_3N. With increasing temperature the thermal linear expansion coefficient increases from $1 \times 10^{-6}\,\text{K}^{-1}$ (at $100\,\text{K}$) to $1.3 \times 10^{-5}\,\text{K}^{-1}$ (at $300\,\text{K}$). So perhaps $6.4(3) \times 10^{-6}\,\text{K}$ should be viewed as an average expansion in this range. Yet because the rate of thermal expansion changes with temperature, it is recommended to utilise equations fitted to the lattice expansion, such as (4.7). Very little thermal expansion is observed below $100\,\text{K}$: the use of 2D detectors or synchrotron radiation sources seems advisable for further investigations into thermal expansion in Cu_3N in this region.

Various limitations of the Sayetat et al. approach are discussed in the original paper. While the method has been successfully applied to polyatomic systems[213], evaluation work with this code found a significantly underestimated silicon Debye temperature of $\sim 80\,\text{K}$ (including Einstein terms and using data from Yim et al[220]) compared with the accepted value of

$\sim 650\,\mathrm{K}$. With improved precision and more data points, results might be expected to improve, and explicit values for the Einstein temperature and isothermal compressibility would perhaps be unnecessary; however, given that the approach employs an average Grüneisen parameter (when separate Grüneisen parameters should properly be written for each branch and wavevector) and neglects electronic and magnetic contributions and any temperature variability in the isothermal compressibility, it is suggested that the Cu_3N Debye temperature and Grüneisen parameter as reported above should be used only in qualitative work.

In all patterns, a small intensity excess is found amid the Cu_3N (111) and (200) reflections (at $20.15°\ 2\theta$ Mo $K\alpha$ or $44.5°\ 2\theta$ Cu $K\alpha$). This feature (see Fig 55) does not match the reflections of Cu, Cu_2O, CuO or H_2O, but is in the vicinity of the (111) reflection of metallic copper (at $19.6°\ 2\theta$ Mo $K\alpha$ or $43.3°\ 2\theta$ Cu $K\alpha$), yet is perhaps too distant for a copper association to be held with much confidence. While copper (111) reflections are frequently seen in diffraction patterns from Cu_3N[158,173,198], no such feature was reported in a survey of over thirty Cu_3N papers reporting x-ray diffraction. While the feature could result from strained copper nano-crystallites under an effective $3\,\%$ lattice parameter reduction, such contractions would differ to thermal expansion and shape effects seen previously in copper nano-crystallites[221]. Explorative structural work with simulated XRD patterns of Cu_3N super-cells, e.g. employing FCC sub-lattices offset by half a lattice constant and comprising Cu, N, O_2, N_2 or Ar atoms, failed to develop evidence for such a reflection, but tentatively suggested that the feature could be related to a proposed oxygen-terminated surface layer[167,170,171]. Further work including XRD with improved signal-to-noise is necessary to further assess the feature.

This work found that the Dow–Corning high-vacuum grease used as an adhesive was found to crystallise when held for one hour at $200\,\mathrm{K}$, but that the grease remains amorphous if rapidly quenched below this temperature. Indexing suggested space groups $C2/m$ or $Cmmm$. The silicone grease used comprises a siloxane metalloid compound (one possessing Si-O-Si linked groups)[222,223], and in Cu-$K\alpha$ radiation the strongest reflections from the grease are at 21.3 and $24.8°\ 2\theta$. In low-temperature XRD studies of polysiloxanes[224], discrete lines appear and the amorphous background decreases as the degree of crystallisation increases with time (hours). In particular see the table and the growing reflections near $20°$ in Figure 1 of Moskalenko et al[224]. While possibly not a good match to the Dow–Corning grease, Figure 5 of Mehta et al.[225] shows XRD of a Si-O-Si linked compound (titanasilsesquioxane) with two strong reflections near $22°$ similar to those seen in this chapter.

Our XRD results for the thermal expansion of Cu_3N should be considered alongside those for the related anti-ReO_3 compound NaN_3, for which powder and single crystal samples have been studied by x-ray and neutron diffraction[226]. This work finds a similar evolution of the thermal linear expansion coefficient, which decreases from 6 to $4.5\,\mathrm{K}^{-1}$ as temperature decreases from 393 to $20\,\mathrm{K}$. While this work reports no flattening of the thermal expansion below $100\,\mathrm{K}$, some inconsistencies are suggested in the powder diffraction data (see Fig. 3 in the paper) and the work demonstrates some difficulties inherent in such measurements[226].
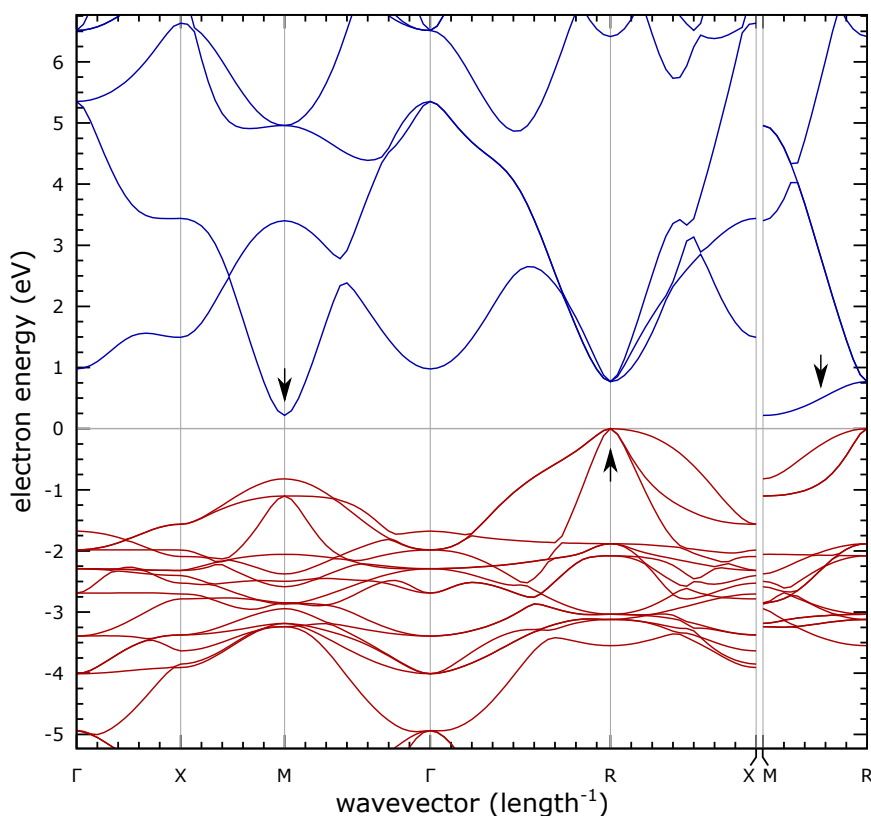
Figure 59: DFT band structure: The GGA conduction band minimum and valence band maximum are found respectively at the M and R points, with the smallest direct onset between M and R.

4.5 First-principles investigations

Prior theoretical work on Cu_3N has included investigations of electronic[167,214], structural[216] and optical[218] properties; mechanical properties such as shear and strain[214], Poisson's ratio and bulk moduli[215]; thermal expansion, heat capacity and Grüeinsen's relation[215]; studies of interstitials[227,228], nearby phases[229] and chemical enthalpies[217]; systematic family trends[219]; phonon dispersion[230]; defects[231,232] and surface properties[167].

Notable highlights include a GW quasi-particle band structure[167] with respective 1.0 and 1.6 eV indirect and direct band gaps close to those seen experimentally; calculation and experimental observation of the Γ-point optical phonon modes[230]; some early works[216] and others with a large breadth of physical studies[215]; and recent experimental and theoretical work[231] which suggests that bipolar doping results from copper vacancies and interstitials. Given the volume of work above, the calculations in this section focus upon evaluating various electronic and structural hypotheses in support of experimental work.

4.5.1 Details of calculations

Density functional theory (DFT) calculations were performed in the ABINIT[233,234] code, with the eigenfunctions expanded in a planewave basis with a ~ 1 keV cut-off energy, the ionic potentials approximated by Troullier–Martins pseudopotentials[235] and the exchange–correlation potential treated with the generalised gradient approximation (GGA) functional of Perdew, Burke and Ernzerhof[78] (PBE96). Independent calculations were performed by Chris Savory

(UCL, UK) in the VASP[236,237] code using the screened HSE06 hybrid density functional[79,238], a plane-wave basis and the projector augmented wave (PAW) method[239,240]. Convergence studies on the GGA planewave cut-off energy and appropriate k-point grid density achieved total energy convergence to better than $0.01\,\%$; while the self-consistent field cycle converged upon energy differences below $0.03\,\mathrm{meV}$. Our default grid density used 55 k-points in the irreducible Brillouin zone; results of structural relaxation studies did not change significantly after increasing the grid density to 204 k-points. The density of states was evaluated via the tetrahedron method with 506 k-points in the irreducible Brillouin zone.

Relaxation of the Cu_3N unit cell from the $3.807\,\text{Å}$ lattice parameter of Juza et al[172] converged on a parameter of $3.877\,\text{Å}$, somewhat greater than seen in other first-principles calculations (eight other DFT studies[215] showed a mean and standard deviation lattice parameter of $3.814(19)\,\text{Å}$), but consistent with the $3.868\,\text{Å}$ result of Yu et al[230], who similarly employed a GGA PBE96 functional. Indeed, if a $0.5\,\%$ error is assumed in reported DFT lattice parameters, then nearly all prior results are consistent with this value, with the few remaining exceptions arising from functionals using the local density approximation (LDA); LDA functionals have a general tendency to over-bind crystals (having larger cohesive energies, bigger bulk moduli[230] and underestimated lattice parameters), while GGA functionals tend to under-bind: with slightly weak cohesive energies and overestimated lattice parameters[216].

4.5.2 Electronic properties and density of states

Evaluation of the band structure along the reciprocal space path (Γ-X-M-Γ-R-X,M-R), which traces all edges of the irreducible primitive cubic Brillouin zone[86], shows a band dispersion and density of states remarkably consistent with that seen elsewhere[167,214,215,216], despite differing computational approaches. The valence band maximum (VBM) and conduction band minimum (CBM) are found respectively at the $R\left(\frac{1}{2}\frac{1}{2}\frac{1}{2}\right)$ and $M\left(\frac{1}{2}\frac{1}{2}0\right)$ symmetry points, with the smallest direct gap at $\left(\frac{1}{2}\frac{1}{2}\frac{1}{3}\right)$ between R and M, see Figs. 59 and 60. This VBM and CBM configuration is reported across all prior literature; however, because the literature often omits important symmetry lines, particularly the line between M and R, the smallest direct gap may occasionally be determined erroneously at the R-point when undiscovered positions may exist with smaller gaps. Noting this, the claim (after exploring all symmetry lines) that the smallest direct gap lies between R and M has not been reported previously; however, such a transition is not seen in Zakutayev et al[167], due to slight differences in the curvature of the lowest conduction band. Calculations indicated respective 189 and 720 meV
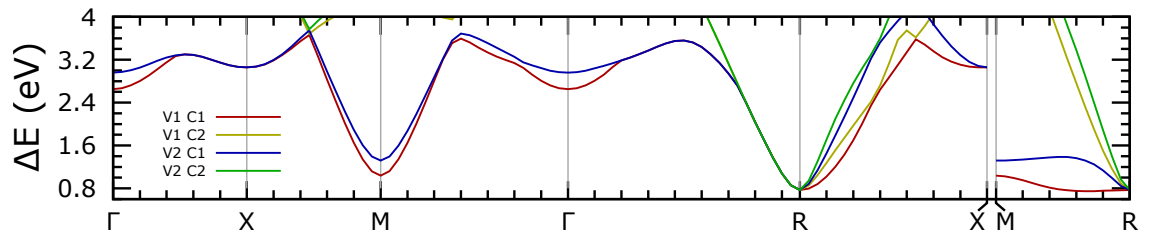
Figure 60: Extrema in the Cu_3N band structure: the GGA energy differences between each of the top two conduction and valence bands (respectively labelled $V2$, $V1$, $C1$ and $C2$).
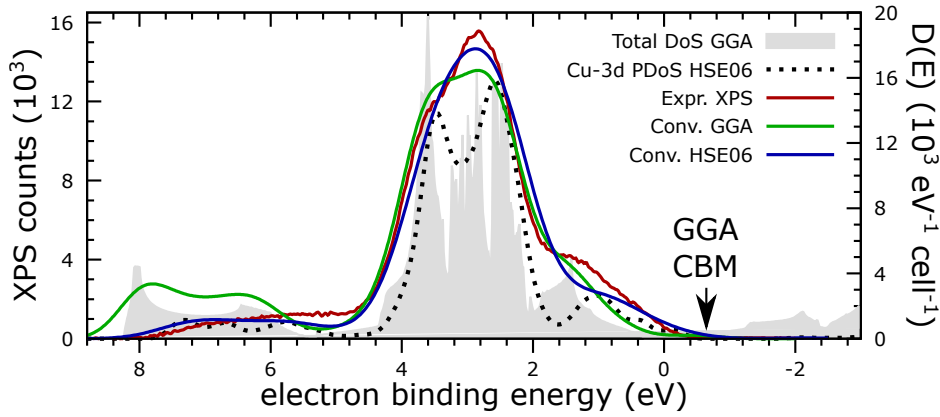
Figure 61: DFT density of states: The total GGA and partial HSE06 Cu-3d density of states are compared with the same broadened by convolution with Gaussian functions, and fitted to XPS spectra (Al-$K\alpha$) of the as-deposited Cu$_3$N films. The fit improves with HSE06, but an excessive $3.5\,\mathrm{eV}$ FWHM remains.

indirect and direct gaps, figures which are highly representative of the literature: where 10 studies showed respective mean (and standard deviation) indirect and direct gaps of $0.28(10)$ and $0.82(8)\,\mathrm{eV}$. Semiconductor band gaps are typically significantly underestimated by the LDA and GGA exchange-correlation functionals; more accurate calculations in the GW quasi-particle approximation[167] (perhaps the leading Cu$_3$N band structure yet published) show similar dispersion with respective indirect- and direct band gaps of 1.0 and $1.6\,\mathrm{eV}$, consistent with the gaps seen experimentally. Many similarities emerge on comparing the band dispersion of Figure 59 against that in this GW study, particularly in the topmost valence bands which differ by some $0.1\,\mathrm{eV}$ (differences of some $0.3\,\mathrm{eV}$ are seen between the conduction bands). Differences between energy separations in the GGA and GW band structures are qualitatively proportional to the magnitude of the energy separation. The HSE06 calculations suggest indirect and direct gaps of 1.0 and $1.9\,\mathrm{eV}$ and support the GGA dispersion.

The computed total GGA density of states seen in Fig. 61 likewise shows good agreement with previous work[214,215,216,218]. The valence band density of states may be studied experimentally via x-ray photoelectron spectroscopy (XPS), after consideration of relevant photo-emission cross-sections (assuming appropriate angle-integrated sampling and ignoring final-state effects)[241]. While the Cu$_3$N valence band derives from states with N-2p and Cu-3d character[167,216], the Cu-3d photo-ionisation cross-section[242] at $1486.6\,\mathrm{eV}$ is some 167 times larger than that for N-2p; using only these dominant Cu-3d states then, the partial HSE06 density of states was convolved with a Gaussian line-shape (to account for instrumental and lifetime broadening) and fitted to Shirley background-subtracted valence band XPS spectra of the as-deposited Cu$_3$N films, see Fig. 61. For comparison, the GGA total density of states was convolved similarly. A Fermi function removed unoccupied conduction-band states. Reasonable agreement is seen between the computed and experimental spectra, with the HSE06-derived spectrum fitting the XPS data more closely despite XPS analysis of the Cu-2p lines suggesting a possible oxide layer[243]: for which significantly reduced density of states are proposed[167], and which may account for some of the differences seen in Fig. 61.

4.5.3 Study of band gap dependence on lattice parameter

Accepting inaccuracies in the GGA band gap but considering that the band dispersion and lattice parameter dependence may be quantitatively reliable, the evolution of the bandgap was investigated as a function of lattice parameter. The lattice was rigidly compressed from 3.89 to $3.797\,\text{Å}$, during which the direct and indirect gaps decreased respectively by 0.247 and $0.237\,\text{eV}$, see Fig. 62. The direct and indirect band gap dependence on lattice parameter (a, in Angstrom) were fitted by the following quadratics

$$E_g^{\text{direct}}(\text{meV}) = -2.120a^2 + 18.84a - 40.29 \tag{4.9}$$

$$E_g^{\text{indirect}}(\text{meV}) = -1.261a^2 + 12.35a - 28.54 \tag{4.10}$$

To correct for inaccuracies due to the GGA functional, these gaps where scaled respectively by the factors 2.089 and 4.237 (noting the proportionality mentioned previously), which bring the gaps into coincidence with those of Zakutayev et al[167] at the $280\,\text{K}$ lattice parameter. The band gap evolution as a function of temperature is then estimated in Figure 62 by combining the fitted band gap dependence on lattice parameter with the Debye or quadratic fits to the thermal lattice expansion (see equation (4.7) on page 95), neglecting at this stage the contribution of the electron-phonon interaction to the band gap.
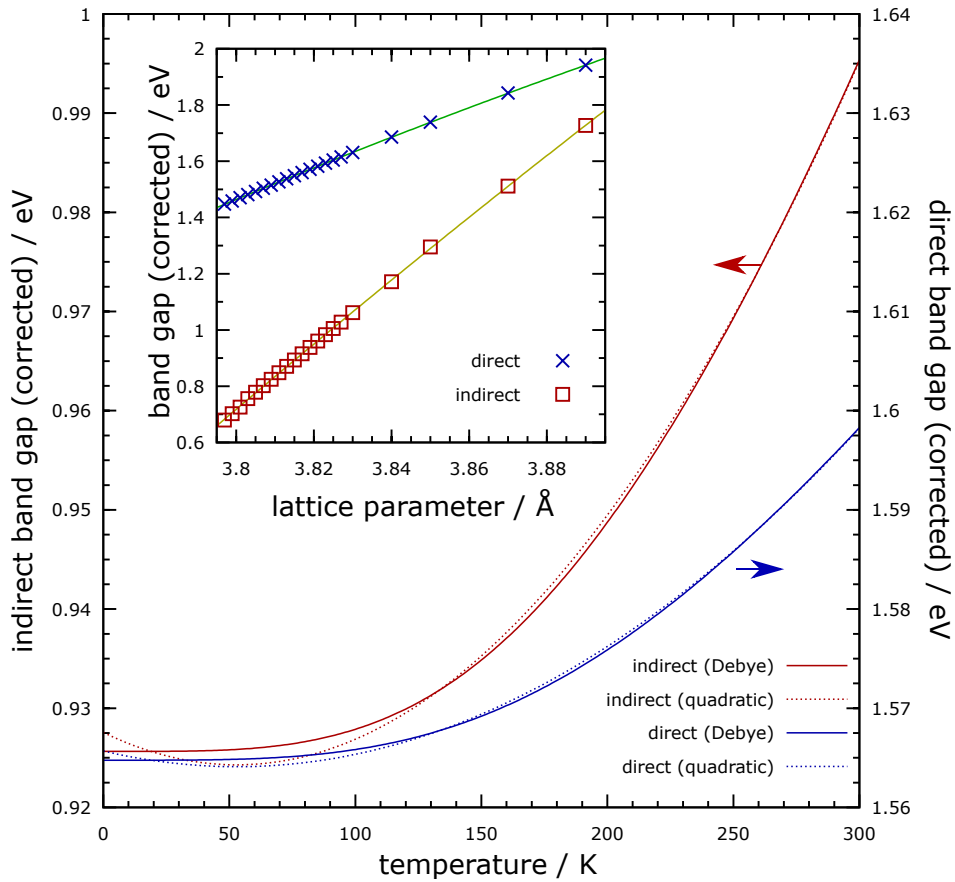
Figure 62: Temperature dependence of the Cu$_3$N band gaps: shows the computed temperature-dependence of the direct and indirect Cu$_3$N band gaps, from DFT calculations and temperature-dependent XRD work; the inset shows the band gap dependence on lattice parameter.
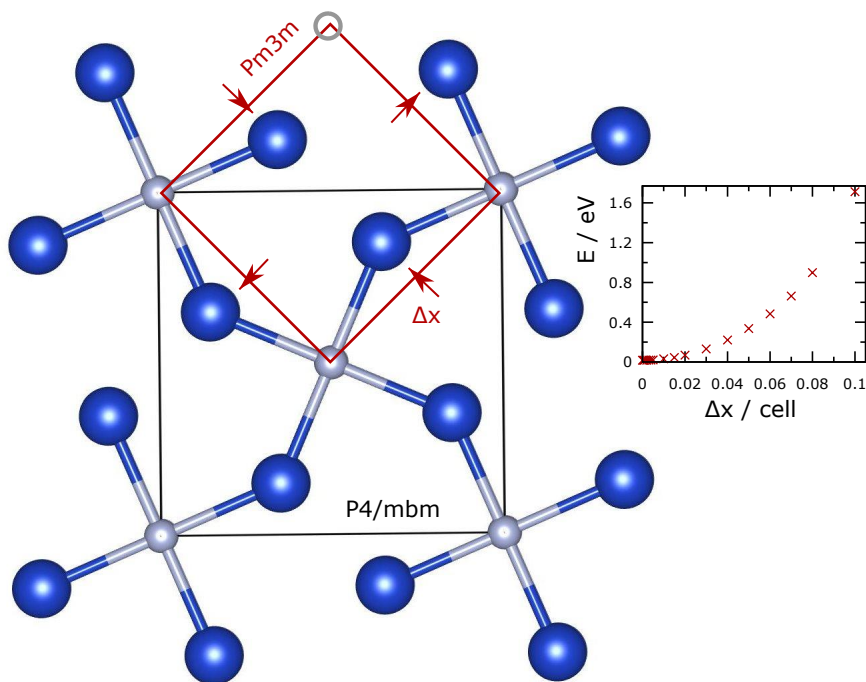
Figure 63: Candidate structural distortion modes in Cu_3N: The figure depicts a potential Cu_3N rigid-unit mode, parametrised by the relative displacement of Cu atoms in (100) planes. Such distortions lead to a tetragonal structure (P4/mbm) with side $\sqrt{2}$ larger than that of the cubic anti-ReO_3 ($Pm3m$) sub-cell. The inset plots the results of the total energy study: a monotonically increasing energy for increasing distortion.

4.5.4 Search for structural distortions including perovskite-like rigid-unit modes

Before the systematic effect of growing crystallinity in the adhesive grease during x-ray diffraction was confirmed, the possibility of various structural phase-transitions were considered (which may yet be of interest). The first hypothesis postulated a total energy local minimum for a potential co-existing Cu_3N orthorhombic phase, using the structural parameters fitted to the reflections in the XRD of Figure 55. The lattice was again relaxed starting from this structure, yet the cell returned entirely back to the $3.877\,\text{Å}$ cubic anti-ReO_3 structure: suggesting that the genesis of any orthorhombic Cu_3N phase would need to be more complicated, and providing some confidence in the determination of the $Pm3m$ structure.

Perovskite structures show a rich display of distortion phenomena including rigid-unit modes[226], i.e. rotations and oscillations of rigid octahedral structural units. A total energy study was performed to investigate such distortions in Cu_3N, parametrised by the rotation angle of the rigid NCu_6 octahedra, see Fig. 63. In this work, neighbouring octahedra were rotated in anti-sense with a constant cubic $Pm3m$ lattice parameter: such that all Cu-N bond lengths were identical and copper atoms in (100) planes were displaced perpendicular to the N-Cu-N axes. No total energy local minima were found for such distortions: the total energy increased exponentially for relative displacements $0.05 \leq \Delta x \leq 10\,\%$ of the $Pm3m$ unit cell, see the inset in Fig. 63. In prior theoretical, x-ray and neutron diffraction studies on the structurally similar anti-ReO_3 compound Na_3N, the failure to observe temperature-dependent distortions such as rigid-unit modes was explained as due to the absence of stereochemically active, lone electron pairs on the bridging sodium atoms[226] (these also absent in Cu_3N).

4.6 Band gap temperature dependence by FTIR spectroscopy

4.6.1 Evaluation and reduction of reflectivity and transmissivity spectra

The reflectivity $R(\hbar\omega)$ and transmissivity $T(\hbar\omega)$ as determined by the silicon and MCT detectors were first joined in the overlapping spectral region near $1.15\,\mathrm{eV}$. These spectra showed some evidence of systematic instrumental effects with $R + T > 1$ near $600\,\mathrm{meV}$, and further processing was necessary to join the spectra.

While MCT detectors offer great sensitivity, they are susceptible to non-linear (decreasing) response with increasing light intensity[23]. Reference spectra, which feature the greatest intensities, will be underestimated under such conditions, while sample reflectivities and transmissivities are correspondingly overestimated. Despite the use of software non-linearity correction, the MCT reflectivities and transmissivities showed excesses suggestive of residual non-linearity. To correct for this, the reflectivity and transmissivity signals in the overlapping spectral regime of the silicon and MCT detectors were used: the reflectivity and transmission were taken as free from non-linearities when the MCT signals coincide with those from the silicon detector (assumed to be non-linearity free) near $1.15\,\mathrm{eV}$. Because non-linearity errors contribute a low frequency (or approximately constant) excess to a measured spectrum, a linear correction $A \times I(\hbar\omega) + B$ is applied to each pair of MCT reflectivity and transmissivity spectra, where the constant factors A and B (typical magnitudes $A = 78\,\%$
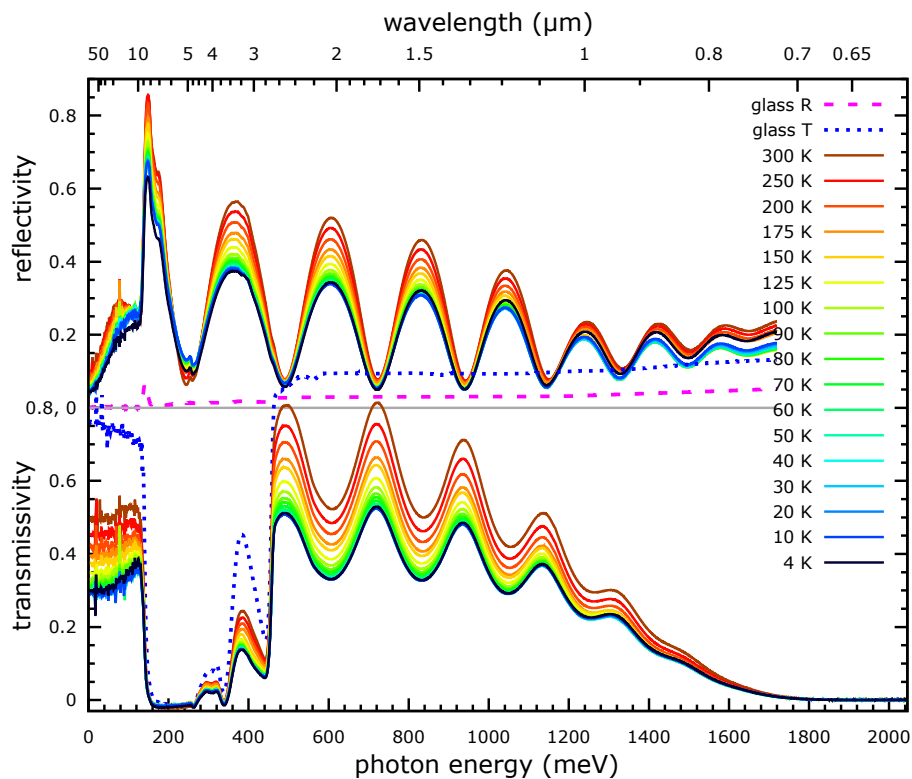
Figure 64: Temperature-dependent reflectivity and transmissivity: Composite FTIR spectra using NIR source, CaF$_2$ & Quartz beamsplitter and MCT & Si detector. The dotted line denotes the reflectivity and transmissivity of the borosilicate substrate. The MCT spectra are corrected for non-linearity using the levels of the silicon detector; this figure suggests that each of the reflection and transmission spectra increase in intensity with temperature; this phenomena is discussed further in the text.

and $B = -3\%$) ensure that the MCT traces coincide with those from the silicon detector with similar oscillation amplitudes and curvature. While such considerations are necessary, the MCT spectra serve mainly to determine the baseline position, and do not significantly impact on absorption edge evaluation above $1\,\mathrm{eV}$. The composite, temperature-dependent spectra after non-linearity correction are presented in Figure 64. This figure shows a series of interference fringes due to internal reflections from the film-substrate interface. The dotted traces in the figure denote the reflectivity and transmissivity of the borosilicate substrate; various quartz-like features imposed by the substrate are evident. these include the reflectivity feature at $140\,\mathrm{meV}$, the opaque region in the transmission between 160 and $260\,\mathrm{meV}$ and the absorption features at 200 and $340\,\mathrm{meV}$.

One interesting phenomena suggested in Figure 64 is the strong temperature-dependence of the magnitudes of reflectivity and transmissivity, which both increase monotonically with temperature. Upon conversion to absorption spectra, this behaviour leads to monotonically increasing absorption strength with *decreasing* temperature and strong absorption well below the semiconductor band edge ($5 \times 10^3\,\mathrm{cm}^{-1}$ at $500\,\mathrm{meV}$ and $4.2\,\mathrm{K}$). Possible explanations for such phenomena using dielectric function processes (e.g. absorption coefficients with carrier population statistics, inter-valence band transitions, free carrier absorption) run into difficulties primarily as one must explain how the reflectivity and transmissivity may acquire temperature gradients of the same sign, as well as sub-gap phenomena typically showing small absorption magnitudes ($100\,\mathrm{cm}^{-1}$), particularly at low temperature. Typically, if the
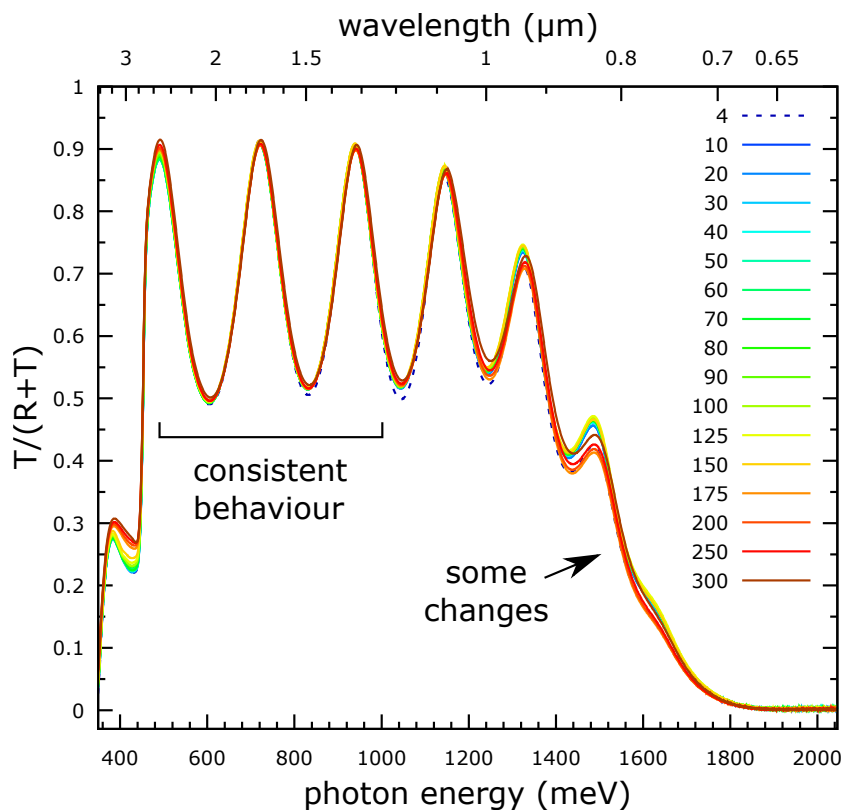
Figure 65: Evidence for time-dependent systematic effects. The plot shows the ratio of transmissivity to the sum of reflectivity and transmissivty, $T/(R + T)$, for each temperature. The fraction neglects absorption, but the coincidence below $1.2\,\mathrm{eV}$ suggests a systematic error in the spectra of Figure 64.
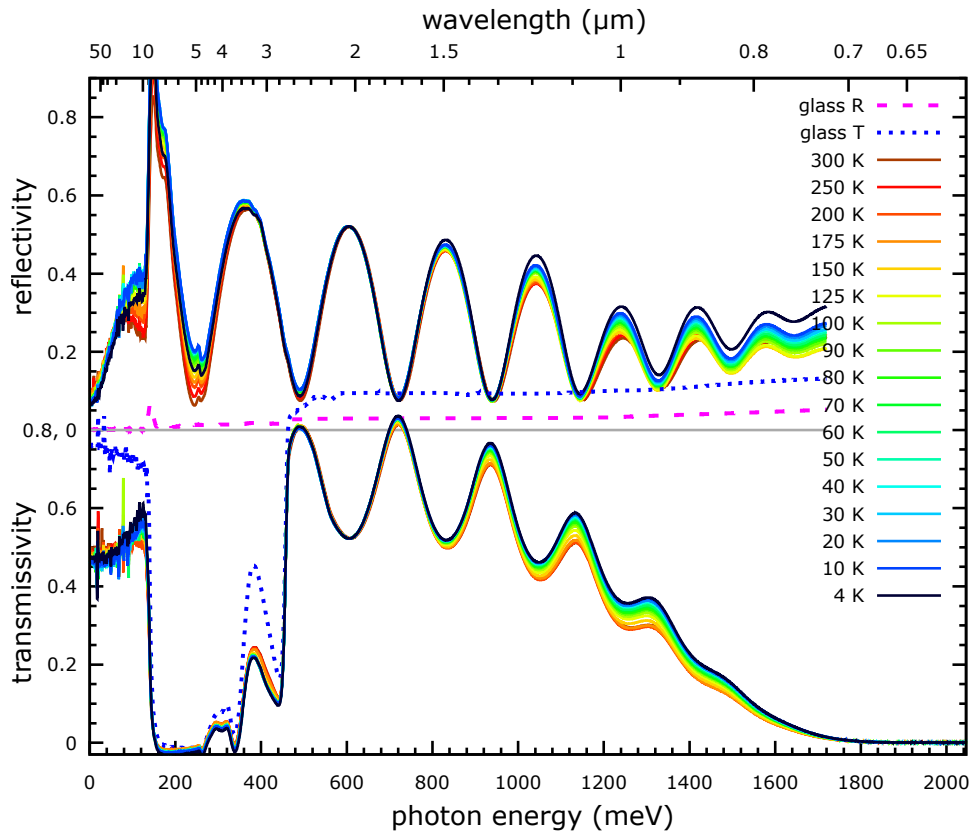
Figure 66: Temperature-dependent reflectivity and transmissivity. The spectra are shifted to coincide with the 300 K trace at 607.9 meV in the transparent region. This results in the transmission edge redshifting with increasing temperature, suggesting enhanced absorption.

transmissivity is changed by some dielectric phenomenon, then the reflectivity shall move in the counter direction. After studies of instrumental factors, along with further considerations (see the discussion in Section 3.12), this effect was considered as most likely a systematic instrumental effect due to changes to the light incident on the sample.

To illustrate this, Figure 65 shows a plot of the transmissivity over the sum of reflectivity and transmissivity, i.e. $T/(R+T)$. While R and T are simple fractional quantities (constant spectra for the material with some temperature-dependence), the measured signal comprises the product of these with the incident spectrum, which may change as a function of time during a long experiment. Where little absorption is present, as expected for semiconductors below the band edge, the fraction $T/(R+T)$ serves to divide out any time-dependent changes in the incident spectrum; here the incident spectrum is assumed to be identical for R and T spectra taken at some temperature (taken at nearly the same time), but is allowed to change over time so that, e.g. the spectra at 4.2 and $100\,\mathrm{K}$ could differ in some unknown manner.

As seen in Fig. 65, there are hardly any differences between the traces in the low-absorption region below $1.2\,\mathrm{eV}$, strongly suggesting that, in the absence of systematic effects, these spectra would be coincident in Figure 64. Above $1.2\,\mathrm{eV}$, some differences are evident due to the effects of strong absorption (which is neglected in Fig. 65). Strong absorption effects are most evident in the high temperature traces due to redshift of the band gap. To correct for these systematic effects, all traces were scaled to intercept the 300 K trace at $607.9\,\mathrm{meV}$ in the
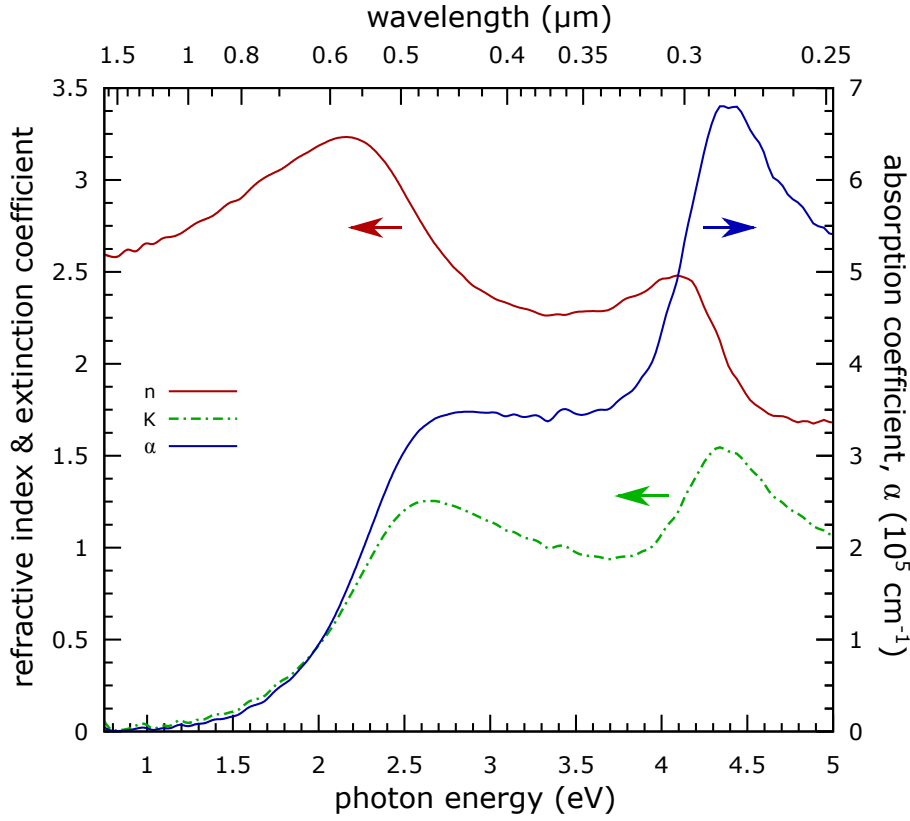
Figure 67: Ellipsometric absorption and refractive index spectra. The plot depicts the complex refractive index $\tilde{n} = n - i\kappa$ and absorption coefficient as fitted to spectroscopic ellipsometry (SE) at $300\,\text{K}$. No dip at $2.4\,\text{eV}$ is seen in the SE absorption spectrum.

transparent quartz-like region, with the $300\,\text{K}$ trace selected as a nominal signal as this was closest in time and temperature to the reference spectrum, and hence is expected to be least affected by systematic effects. The corrected spectra are plotted in Figure 66. Examining this figure, one finds that the subgap features are nearly unchanged while the absorption edge moves to lower photon energies with increasing temperature, suggesting that the approach is justified.

4.6.2 Absorption spectra determination

The measured reflectivity $R_t(\hbar\omega)$ and transmissivity $T_t(\hbar\omega)$ were then reduced to absorption spectra $\alpha(\hbar\omega)$ using equations which describe the reflection and transmission of light directed at normal incidence at an incoherent single optical layer of thickness d with parallel faces

$$R_t = R\left[1 + \frac{(1-R)^2 \exp(-2\alpha d)}{1 - R^2 \exp(-2\alpha d)}\right] \qquad \text{and} \qquad T_t = \frac{(1-R)^2 \exp(-\alpha d)}{1 - R^2 \exp(-2\alpha d)} \qquad (4.11)$$

where $R(\hbar\omega)$ is the (unmeasured) intrinsic reflectivity at a semi-infinite boundary between the vacuum and the Cu_3N film (in contrast, the total reflectivity R_t includes the contributions from internal reflections at each face). The intrinsic reflectivity R must be determined before proceeding; this is achieved using the procedure discussed in Section 3.13.4.

Absorption spectra evaluated from equations (3.110) and (3.111) with a $1\,\mu\text{m}$ film thick-

ness (as indicated by SEM and ellipsometry) are presented in Figure 68. Corresponding absorption and refractive index spectra resulting from spectroscopic ellipsometry are presented in Fig. 67. Whilst the ellipsometric model suggested a $1.0\,\mu m$ film thickness (with mean-squared error of 11.8), this figure was quoted with a $4\,nm$ uncertainty, which was considered to significantly overestimate the precision of the thickness determination. The magnitude of the ellipsometric mean-squared error perhaps suggests possible further complications in the ellipsometric fit. Given reports of Cu oxide layers protecting the Cu_3N film surface, nanometre-thick layers of these have been used in prior ellipsometric work[167]: yet no significant improvement was found here for models incorporating such layers.

The ellipsometric and FTIR absorption spectra are reasonably consistent below $1.8\,eV$, but a small underestimation ($\sim 20\,\%$) becomes apparent in the FTIR absorption spectra with increasing photon energy. The corrected absorption computed from (3.110) and (3.111) relies on the measured spectra being only slightly affected by the substrate. This approach, while practical, is legitimate only after careful consideration of the consequences. For an absorbing film upon a transparent substrate of lower refractive index, the reduced refractive index contrast (in comparison to a boundary with the vacuum) reduces the reflectivity of the rear interface; as a result, less light is trapped (and absorbed) in the film by internal reflections. Thus neglecting such a substrate leads to an overestimated optical path length, and hence an underestimated absorption coefficient. The error is rather constant across the corrected spectra: $\sim 85\,\%$ for a $1\,\mu m$ film with refractive index $n=3$ on glass.
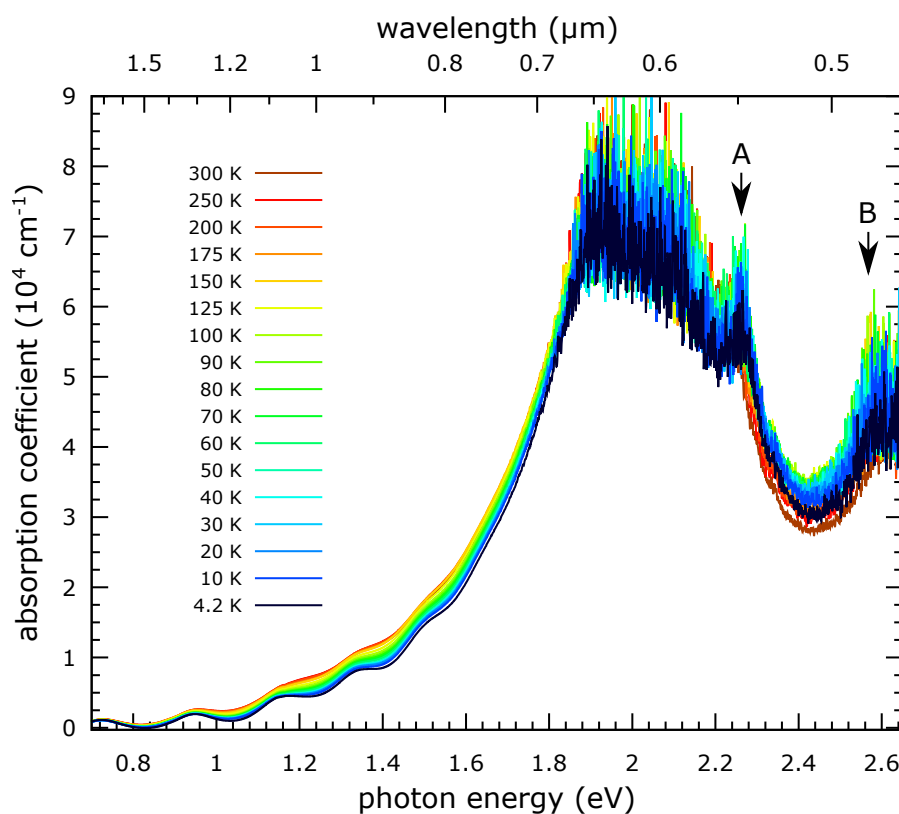
Figure 68: FTIR absorption spectra. shows temperature-dependent FTIR absorption spectra between 4.2 and 300 K, over which the absorption edge shifts by some $40\,meV$. Above $1.8\,eV$ two absorption features, labelled A and B, are distinct from the noise which results from feeble transmission.

Visual inspection of the FTIR spectra of Fig. 68 suggests a strong absorption onset near $1.5\,\mathrm{eV}$, possibly located above a weaker onset at $1.0\,\mathrm{eV}$. Such a configuration would be consistent with the respective positions of the calculated[167] direct and indirect band gaps in $\mathrm{Cu_3N}$. In the ellipsometric spectra of Fig. 67, the absorption strength grows strongly to a local maximum of $3.5 \times 10^5\,\mathrm{cm^{-1}}$ at $2.7\,\mathrm{eV}$, maintaining this strength until increasing again at $3.7\,\mathrm{eV}$. The strong absorption from $1.5\,\mathrm{eV}$ supports the potential for $\mathrm{Cu_3N}$ as a candidate photovoltaic absorber. The edges of Fig. 68 show rather weak temperature dependence, with the band gap decreasing by some $40\,\mathrm{meV}$ as the temperature is increased from 4.2 to $300\,\mathrm{K}$. Some other features are notable in the FTIR spectra of Fig. 68: residual oscillations due to internal reflections are seen below $1.5\,\mathrm{eV}$ as the incoherent model of equations (4.11) neglects interference effects; the noise seen approximately between 1.8 and $2.2\,\mathrm{eV}$ results from a feeble transmission signal due to strong absorption in the sample.

The fundamental absorption edge may result from various transitions from the first valence band. Features at 2.26 and $2.56\,\mathrm{eV}$, labellel A and B in Fig. 68, are possibly due to direct transitions between the second valence and first conduction bands, respectively at M and between M and R. A drop in absorption strength between A and B is not corroborated by the ellipsometric absorption spectrum of Fig. 67 (despite some 30 ellipsometric data points in this region). In explaining this difference, note that while the FTIR spectra follow reasonably directly from the experimental data through equations (4.11), the SE spectra depend upon the quality of fitting of the optical model, and could be vulnerable to certain discrepancies.

4.6.3 Fitting the absorption onset

The temperature-dependent absorption spectra $\alpha(\hbar\omega)$ of Fig. 68 are sufficiently featureless that it is not possible to confidently discern separate regions for fitting combined direct and indirect onsets. Given that the direct onset should in any case be orders of magnitude stronger, the temperature-dependent spectra are fitted with equations describing these

$$\alpha_{\mathrm{d}} \cdot \hbar\omega = A\sqrt{R\left(\hbar\omega - E_{\mathrm{g}}^{\mathrm{d}}\right)} \tag{4.12}$$

where $R(\cdot)$ is the ramp function, $R(x)=x$ for $x \geq 0$ and $R(x)=0$ for $x < 0$. Each direct onset includes an exponential Urbach tail which intrudes into the band gap

$$\alpha_{\mathrm{U}}(\hbar\omega) = B \exp\left(\gamma\left\{\hbar\omega - E_{\mathrm{g}}^{\mathrm{d}} - E_{\mathrm{U}}\right\}\right) \tag{4.13}$$

where E_{U} is a material-dependent parameter which sets the point where the Urbach tail (4.13) smoothly and continuously joins the absorption edge of (4.12), and B and γ are functions only of A [the same constant used in (4.12)], $E_{\mathrm{g}}^{\mathrm{d}}$ and E_{U}, as discussed in Section 3.8. Occasionally, E_{U} is given a temperature-dependence. Such that the absorption associated with a direct onset is

$$\alpha_{\mathrm{direct}}(\hbar\omega) = \begin{cases} \alpha_{\mathrm{U}}(\hbar\omega), & \hbar\omega \leq \left(E_{\mathrm{g}}^{\mathrm{d}}+E_{\mathrm{U}}\right) \\ \alpha_{\mathrm{d}}(\hbar\omega), & \text{otherwise} \end{cases} \tag{4.14}$$

The results of fitting equation (4.14) using the Levenberg-Marquardt algorithm in MATLAB are presented in Fig. 70, with sample fits and residuals presented in Fig. 69. In this fitting, all spectra were fitted between 492 and 1798 meV with a baseline evaluated for each spectrum using the average absorption coefficient in the transparent region between 492 and 723 meV. Constraints were set to ensure the solution converged to physically reasonable values. The direct gap and Urbach energy E_U where constrained to lie respectively between 1.4 and 1.8 eV and below 0.5 eV, while the absorption strength was constrained to be positive.

A maximum redshift of 32 meV is seen across the entire temperature range, giving a band gap temperature coefficient dE_g/dT of just $-0.11\,\mathrm{meV\,K^{-1}}$, rather smaller than seen typically in semiconductors[244]; in contrast, a comparison of twelve common semiconductors[245] including Si, Ge, GaAs, GaP and CdTe showed mean (and standard deviation) band gap temperature coefficients of $-0.36(7)\,\mathrm{meV\,K^{-1}}$. The Urbach intercept E_U increases with temperature from 106 to 118 meV, see the inset in Fig. 70, while the absorption strength A shows a scatter below 2 % around $9 \times 10^8\,\mathrm{eV^{0.5}cm^{-1}}$.

The band gap evolution seen in Fig. 70 differs somewhat from the behaviour seen in typical semiconductors, where enhanced and vanishing temperature-dependence are respectively seen at high and low temperatures[244,246]. For photon energies above 1.3 eV in Figs. 64 and 66, the reflectivity traces at 4.2 K and above 175 K are somewhat displaced (in the same direction) away from the bulk of the dataset. This effect, possibly an unidentified systematic contribution, propagates into the calculated absorption coefficients and leads to a reduced absorption magnitude (and higher estimated band gap) at each of these temperatures.
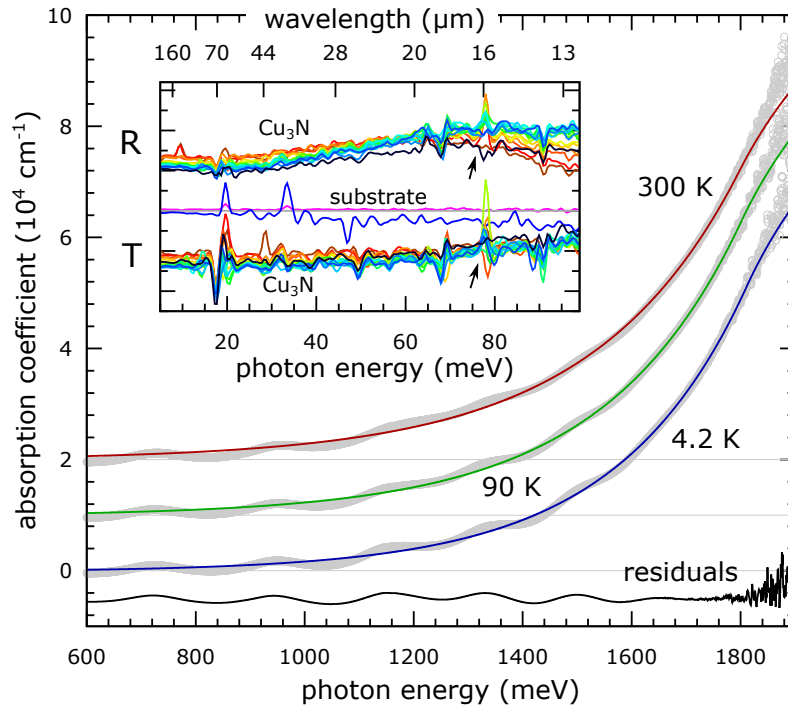
Figure 69: temperature dependence of Cu$_3$N band gap and sample fitted absorption onsets. Shows selected fitted absorption edges and residuals for the 4.2 K trace. The inset shows a feature in the reflectivity and transmissivity at 79 meV associated with the three-fold degenerate Γ_{15} IR-active phonon mode.
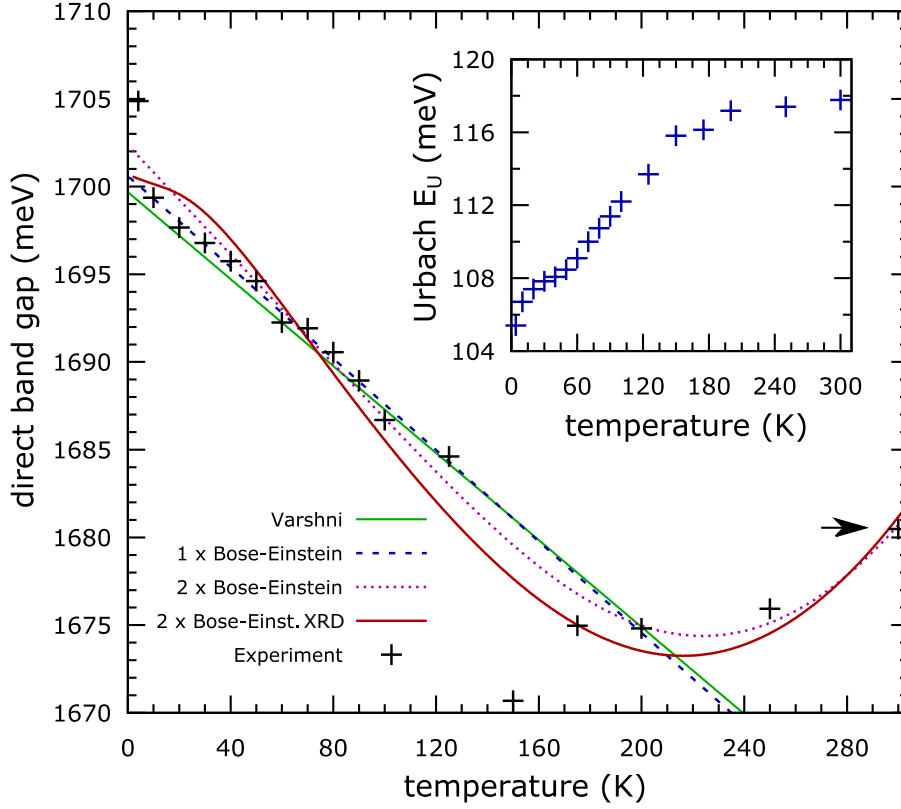
Figure 70: Temperature dependence of Cu₃N band gap, as determined by fits to the absorption edge, considered as a direct gap. The evolution is fitted using the Varshi relation, single and double Bose–Einstein oscillators, and using a model which combines the double Bose–Einstein oscillator with the band gap dependence expected from thermal lattice expansion (labelled Bose-Einst. XRD). The point marked with an arrow is not fitted via the Varshni relation or single Bose–Einstein oscillator. The inset shows the evolution of the Urbach energy E_U.

4.6.4 Fitting temperature-dependence with electron–phonon interaction

In light of the discussion in Appendix C, the temperature-dependence of the optically-determined band gaps of Fig. 70 is examined by fitting with the empirical Varshni relation[247]

$$E_g^{\text{Varshni}} = E_0 - \frac{\alpha T^2}{T + \beta} \tag{4.15}$$

and with models comprising one and two Bose-Einstein oscillators[246,248,249]

$$E_g^{BE} = E_0 + \sum_i \alpha_i \left[1 + \frac{2}{\exp\left(\frac{E_i}{k_B T}\right) - 1} \right] \tag{4.16}$$

The results of the least-squares fits are shown as the solid and broken lines in Fig. 70. The Varshni fit suggests respective E_0, α and β parameters of $1.70\,\text{eV}$, $0.10\,\text{meV\,K}^{-1}$ and $0.014\,\text{K}$; such values are well within the typical range seen for Varshni parameters: as seen by respective mean (and standard deviation) α and β values of $0.6(5)\,\text{meV\,K}^{-1}$ and $400(500)\,\text{K}$ for fits to 37 common semiconductors[247,250]. The single Bose–Einstein oscillator fit finds a zero Kelvin gap of $1.70\,\text{eV}$ with respective average phonon and electron-phonon interaction energies of 0.095 and $-0.061\,\text{meV}$. The quality of the Varshni and single Bose-Einstein oscillator fits significantly deteriorates when the experimental $300\,\text{K}$ datapoint is included; this

point was considered to possibly be affected by systematic factors (as discussed) and so was omitted for these fits. The physical model might be expected to improve with two oscillators as these cater for the different energy scales expected for acoustic and optical phonons (as used in the Debye-Einstein model for fitting the lattice expansion on page 95). The double Bose–Einstein oscillator fit suggests an E_0 of $1.63\,\text{eV}$, and respective average phonon and interaction energies $\{E_i, \alpha_i\}$ of $\{0.020, -0.017\}$ and $\{50, 65\}\,\text{meV}$.

The small average phonon energies fitted above should perhaps be viewed with suspicion as the acoustic phonon 3D density of states generally vanishes at low energy, as seen in the Debye density of states (B.11) of Appendix B (with a typical sound velocity of $1\,\text{km s}^{-1}$). Additionally, prior Cu_3N phonon density of states calculations[230] find maxima near 9 and $84\,\text{meV}$, with negligible states below $5\,\text{meV}$. These calculations (and experimental work) find two three-fold degenerate Γ_{15} zone centre IR-active modes, with energies of 12 and $81\,\text{meV}$. The inset of Figure 69 shows a temperature-dependent feature at $78\,\text{meV}$ in the reflectivity and transmissivity spectra which is distinct from the substrate spectra. This feature is tentatively associated with the upper Γ_{15} zone centre mode.

An interesting fit to Fig. 70 results by writing separate band gap contributions from the electron-phonon interaction and the lattice expansion: with two Bose–Einstein oscillators representing the former and the XRD lattice expansion and DFT band gap evolution estimating the latter (see page 101). The direct gap at a temperature T is then

$$E_g^{\text{direct}}(T) = E_g^{\text{XRD,DFT}}(T) + E_g^{BE}(T) \tag{4.17}$$

the E_0 parameter of (4.16) raises the $E_g^{\text{XRD,DFT}}$ gap by $181\,\text{meV}$, while the respective average phonon and interaction energies $\{E_i, \alpha_i\}$ of $\{98.5, -40.2\}$ and $\{8.73, -10.2\}\,\text{meV}$ ensure a good fit, see Figure 70. The lower optical phonon energy is thus much improved over that of the plain dual Bose–Einstein oscillator.

Each of the Bose–Einstein oscillator models suggests some counterplay between the signs of the zero Kelvin gap correction E_0 and the electron–phonon interaction α_i terms, which permit some freedom to fit the experimental data. Appendix C explains how different signs for these terms may be physically reasonable, particularly given phonon energies of different magnitudes. Both dual Bose-Einstein oscillator fits include the $300\,\text{K}$ datapoint and show improved fits above $150\,\text{K}$. Omitting the $4.2\,\text{K}$ datapoint for the dual Bose-Einstein oscillator fit with lattice expansion both improves the fit statistics and results in a band gap evolution more typical of that seen in semiconductors; as discussed, the relaxing evolution at high temperature is atypical, and could result from systematic effects in the last few datapoints.

4.7 Conclusions and future work

This chapter has discussed investigations into the optical and structural properties of copper nitride between 4.2 and $300\,\text{K}$. A Debye–Einstein–Grüneisen model suggested a zero Kelvin lattice parameter of $3.817\,88(19)\,\text{Å}$, with a Debye temperature of $550(150)\,\text{K}$ and average

Gruňeisen parameter of 2.3(9). The Cu_3N lattice parameters determined in this work below 280 K are consistent with those of Zachwieja et al[197]. The thermal linear lattice expansion decreases monotonically from $1.3 \times 10^{-5}\,K^{-1}$ at 300 K; below 100 K the expansion diminishes significantly and falls below $1.5 \times 10^{-6}\,K^{-1}$. While several reflections presented at 200 K due to growing crystallinity in the adhesive grease, no temperature-dependent structural phase transitions from the 300 K cubic $Pm3m$ structure of Cu_3N were observed.

Our result for the thermal lattice expansion is of similar order to that seen in the structurally similar anti-ReO$_3$ compound Na_3N, which shows[226] a linear expansion coefficient of $4.6 \times 10^{-6}\,K$ between 20 and 293 K. In light of the uncertainties achieved here and the vanishing thermal expansion below 100 K, further low-temperature studies of thermal expansion in Cu_3N would be well-advised to use either single crystals, or high-precision instruments (i.e. measuring to high 2θ) with good signal-to-noise (e.g. 2D detectors or synchrotron sources).

Density functional theory (DFT) calculations suggested a band structure very similar to that seen in prior literature: with a direct gap between the R and M points and a density of states qualitatively supported by experimental valence-band x-ray photoelectron spectroscopy. A study of the band gap evolution as a function of lattice parameter suggested a 32.5 meV *decrease* in the direct gap between 300 and 4.2 K (using the as-determined thermal expansion and correcting for the underestimation of the GGA functional). This anomalous band gap temperature dependence results as the DFT calculations ignore the electron-phonon interaction. A search for structural distortions found no total energy local minima for a candidate perovskite-like rigid-unit mode, which is not experimentally ruled-out by the XRD published here. The absence of stabilising rigid-unit modes is thought to be due to the lack of stereochemically active, lone electron pairs on the bridging copper atoms[226].

Fourier-transform infrared spectroscopy and spectroscopic ellipsometry suggested a strong absorption onset preceded by contributions from a weak indirect onset and exponential Urbach tail rising from 1 eV. A strong direct absorption edge at 1.68 eV strengthens continuously and reaches a magnitude of $3.5 \times 10^5\,cm^{-1}$ at 2.7 eV, suggesting potential for photovoltaic applications. Fitting the absorption edge as a direct onset suggested a redshift of 32 meV between 4.2 and 300 K, with a band gap temperature coefficient dE_g/dT of just $-0.11\,meV\,K^{-1}$, somewhat smaller than seen typically in semiconductors.

Fitting the temperature-dependence of the direct gap to the Varshni relation finds parameters well within the typical range. Fitting the band gap with a double Bose–Einstein oscillator, combined with the as-determined thermal expansion and the aforementioned DFT band gap calculations, suggests that the net effect of the electron-phonon interaction is to decrease the gap with increasing temperature. The average phonon energies fitted in this work, 98.5 and 8.73 meV, are rather close to a pair of three-fold degenerate optical photon modes (the Cu_3N stretching and bending modes predicted[230] respectively at 80.8 and 12.3 meV). The higher of these modes may be seen at 78 meV in the FTIR spectra of Figure 69.

Chapter 5

Temperature-dependent optical properties of copper antimony sulphide CuSbS$_2$

5.1 Introduction

In recent years the leading technologies for thin film photovoltaics (TF-PV) have achieved performance parity with polycrystalline silicon, with both types reaching record 21 % cell efficiencies[251]. The careful selection of excellent optical absorbers, some 100x better than silicon[252], combined with the application of deposition-to-glass processes matured in the display-industry, give thin film modules various cost benefits over crystalline silicon and three-times better energy payback times[253].

Yet the leading technologies for thin-film photovoltaics, cadmium telluride (CdTe) and copper indium gallium diselenide, Cu(In,Ga)Se$_2$ (or CIGS), each have commercial limitations which complicate increasing total installed capacity beyond 100 GWp, just a fraction[252] of the multi-terawatt PV market envisaged by 2050. Tellurium, gallium and indium are relatively expensive, indium and tellurium (particularly) are scarce with limited geographical availability*, and concerns remain over cadmium toxicity[255,256,257,258]. All this drives research into earth abundant and non-toxic absorber materials with desirable properties for thin-film PV: strong optical absorption with an optimal bandgap, limited non-radiative recombination and favourable carrier dynamics[259,260].

Rather than seeking entirely new systems, one approach is to adapt presently successful systems (such as CIGS) by replacing problem elements[252]. In particular, ternary I-V-VI chalcogenides such as Cu-(Sb,Bi)-(S,Se) (where chalcogen refers to a group VI anion) have received attention due to their chemical similarity to CIGS, i.e. the similar ionic radius of In and Sb, and for other beneficial factors, such as their strong absorption (somewhat better than CIGS[259,261]) and the low cost of antimony[262,263]. Reports suggest that the low-cost, earth-abundant and non-toxic semiconductor copper antimony disulphide (CuSbS$_2$ or chalcostibite) has a near-optimal 1.5 eV direct band gap and a strong optical absorption coefficient rising from 10^4 to 10^5 cm^{-1} between 1.6 and 1.8 eV[255,264,265]. The material shows intrinsic p-

---

* USGS estimated global Si, Ga, In and Te production in 2015 was as follows[254]: 8 100 000, 600, 755 and 120 t.

type conductivity, thought to be resultant of the dominant V_{Cu} vacancy which is a shallow acceptor[265], with carrier concentrations tunable between 10^{16} and 10^{18} cm^{-3} through growth parameters[255].

CuSbS$_2$ has a stable orthorhombic phase (space group $Pnma$, number 62) below $550\,°C$ with good thermal, environmental and electrical stability[264,265,266,267]. The CuSbS$_2$ unit cell has a 16-atom basis (see Figure 71) comprising four copper-centred CuS$_4$ tetrahedra and four rectangular pyramidal SbS$_5$ complexes, with two sulphur atoms inside the unit cell per polyhedron. Synthesis of phase-pure CuSbS$_2$ is complicated by the many phases of the Cu-Sb-S system[258,267]. Four ternary phases are seen in thermal equilibrium: CuSbS$_2$ (chalcostibite), Cu$_{12}$Sb$_4$S$_{13}$ (tetrahedrite), Cu$_3$SbS$_3$ (skinnerite) and Cu$_3$SbS$_4$ (famatinite). Other phases observed during growth[255] include: Sb$_2$S$_3$ (stibnite), Cu$_{1.8}$S (digenite), Cu$_2$S and CuS.

The long edges of the SbS$_5$ pyramidal bases are shared between adjacent pyramids in chalcostibite (see Figure 71), forming continuous columns through the material parallel to the primitive b-axis[268]. The pyramidal column packing is optimised through the bonding of an inverted pyramidal column in an offset position, leaving each column-pair bounded by the parallel pyramidal base planes with the lone electron-pairs of antimony projected into the void between these[256]. This accommodation of the lone pairs, combined with the close pyramidal column-packing, sets the antimony in an off-centre position in each pyramidal base plane. One short edge of each pyramidal base joins to a CuS$_4$ tetrahedron through two vertices; thus each pyramidal SbS$_5$ column is decorated by a parallel column of CuS$_4$ tetrahedra. The sulphur vertices in the common edge between the pyramids and tetrahedra occupy a five-coordinated chalcogen B-site, associated with three antimony and two copper atoms (sites 2, 4, 6 and 7 in Figure 71); the remaining sulphurs are four-coordinated with two antimony and two copper atoms (the chalcogen A-site)[256].
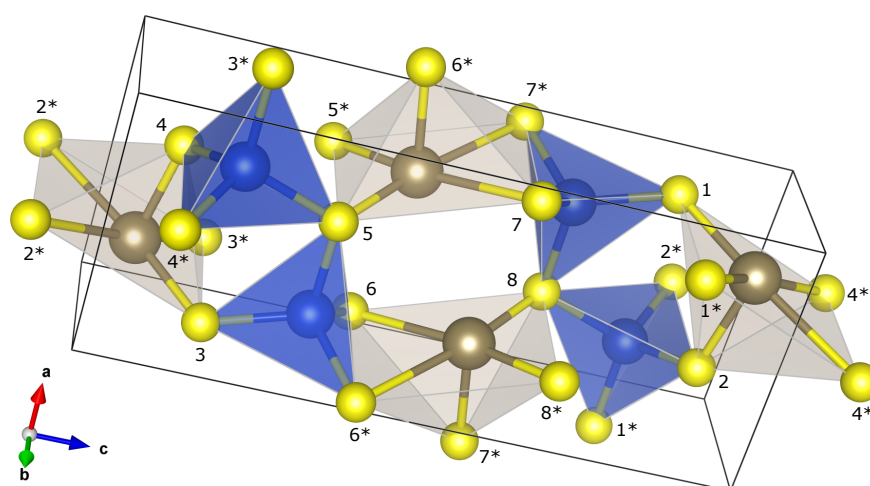
Figure 71: The CuSbS$_2$ unit cell comprises four copper (blue), four antimony (silver) and eight sulphur (yellow, numbered without stars) atoms, forming copper-centred CuS$_4$ tetrahedra and rectangular SbS$_5$ pyramids. To present the CuSbS$_2$ unit cell unambiguously, the figure includes 16 sulphurs in adjacent unit cells (yellow, numbered with stars) and omits bonds from interior sulphurs to adjacent unit cells. To visualise the full crystal geometry, superimpose the starred and unstarred atoms of identical number. For example, the sulphurs labelled 8 and 7 are respectively 4- and 5-coordinated: the A and B chalcogen sites.

Prior synthesis methods fall into two categories. In the first class, $CuSbS_2$ is formed directly and material quality may be further refined by subsequent annealing. Examples include spray pyrolysis[269], thermal evaporation of stoichiometric powder[262], chemical bath deposition[264,265], magnetron co-sputtering of Cu_2S and Sb_2S_3 targets[255] and low-temperature solvothermal synthesis[270]. In the second class, layers of Cu and Sb compounds are deposited and later transformed into homogeneous $CuSbS_2$ material in a secondary sulphurisation process. Examples include the chemical bath deposition (CBD) of CuS/Sb_2S_3 layers[271], or the combined CBD and thermal evaporation of Cu/Sb_2S_3 layers[263], where sulphur already resides within the layers, and the thermal evaporation (or electrodeposition) of Cu/Sb layers[272] or the deposition of a Cu/Sb stack by DC sputtering[258], where sulphurisation occurs as a post-process under vacuum or nitrogen atmosphere.

First-principles calculations in the density-functional theory indicate that Cu-d and Sb-p states dominate the respective density of states at the valence band maximum (VBM) and conduction band minimum (CBM)[273]: with the implication that Cu-Sb transitions are responsible for $CuSbS_2$'s impressive absorption strength; however, recent work finds that the absorption strength mainly results from the large matrix elements for S-S, S-Sb and Sb-Sb transitions[261]. Studies on intrinsic defects in $CuSbS_2$ find these in close proximity to the band edges, suggesting an attractively low concentration of recombination defect centres[265].

5.2 Research objectives and material information

This work reports the optical and electrical properties of polycrystalline $CuSbS_2$ films as a function of temperature via Fourier-transform infrared spectroscopy (FTIR) over 18 temperatures between 4.2 and 350 K. As far as can be determined, this is the first such report for this material.

The $CuSbS_2$ material under investigation was synthesised in equilibrium with Sb_2S_3 vapour on borosilicate glass substrates by radio-frequency (RF) magnetron co-sputtering of Cu_2S and Sb_2S_3 targets by Adam Welch and collaborators[255] at the National Renewable Energy Laboratory (Colorado, USA). This team reported a growth window (the 'self-regulated' regime) within which synthesis of stoichiometric, phase-pure $CuSbS_2$ material could be achieved[255]. Essentially, this window is bounded by an Sb_2S_3 secondary phase which develops at low temperature but sublimes away above $\sim 350\,°C$, and the upper limit for stoichiometric growth at $400\,°C$: where $CuSbS_2$ decomposes into $Cu_{12}Sb_4S_{13}$. More details of the magnetron sputtering configuration are discussed by Adam Welch[274].

After verifying the material quality the research aims are two-fold: first, to carefully fit the absorption onset and to determine any temperature-dependent evolution of the fundamental band gap; secondly, to investigate any other optical properties of the material as a function of temperature. Our FTIR data is complemented by room-temperature x-ray diffraction, spectroscopic ellipsometry, scanning electron microscopy (SEM) and energy dispersive spectrometry (EDS).
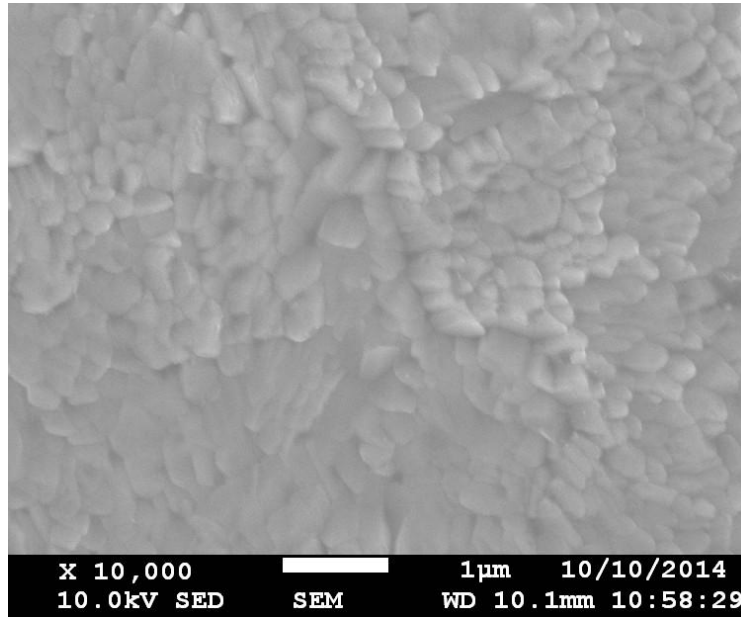
Figure 72: Scanning electron microscopy finds a surface covered uniformly with a non-specific granularity.

5.3 Experimental methods

Infrared transmission and specular reflection spectra for 18 temperatures between 4.2 and 350 K and photon energies of 0.3 to 2.4 eV were made at 11° angle of incidence using a Bruker Vertex 70v Fourier-transform infrared (FTIR) spectrometer with a combined reflection-transmission accessory and an Oxford Instruments CFV2 continuous-flow helium cryostat. The reduction to absorption spectra is discussed in Section 5.5.2.

To complement the FTIR absorption spectra, a Woollam M-2000-UI (rotating compensator) spectroscopic ellipsometer was used to independently evaluate the $CuSbS_2$ dielectric function from 0.7 to 5.1 eV (5650 to 41 100 cm^{-1} or 1770 to 243 nm) at 300 K using incident angles of 65, 70 and 75°. A film-substrate model comprising surface roughness was fitted to the reflection, transmission and depolarisation data in *CompleteEASE* v5.08.

Post-growth analysis by the NREL group[255] using x-ray diffraction (XRD) at low-resolution with a 2D detector suggested single phase material. Although 2D detectors produce impressive intensities, they can lead to broader peaks in comparison to those of point detectors, and the phase determination for this sample was somewhat unconvincing. For this reason, XRD was repeated at 300 K using a Panalytical X'pert Pro MRD equipped with an unmonochromated copper anode ($K_{\alpha1}$=1.540 592 9, $K_{\alpha2}$=1.544 427 4; running at 40 kV and 40 mA), a PW3050/65 goniometer and a point detector. An ω-2θ scan was performed between 10 and 55° 2θ with 0.05° step size, 20 s dwell time and respective divergence and receiving slit widths of 0.76 and 0.5 mm.

Hall measurements via the NREL group in van der Pauw geometry with indium contacts and an Accent HL5500 system found p-type material with a hole concentration of 10^{16} to 10^{17} cm^{-3} (accurate to an order of magnitude[274]), indicating negligible band-filling given the discussion in Section 3.7.

Scanning electron microscopy using a JEOL JSM 7001F with a 10 keV beam energy showed a film with uniform coverage and a non-specific granular morphology, see Figure 72. An average film thickness of 570 nm was determined in cross-sectional geometry. Energy dispersive spectrometry (EDS) using an Oxford Inca X-Act silicon drift detector showed prominent Cu-Lα, Cu-Lβ, Sb-Lα and S-Kα lines, with a weak C-Kα line comprising the only evident contaminant, supporting the XRD results in Section 5.4 for the phase-purity of the sample.

5.4 Structural investigations by x-ray diffraction

X-ray diffraction (XRD) was performed using a Panalytical X'pert Pro diffractometer to assess the phase purity of the material, and secondly, to establish any micro-structural parameters.

5.4.1 Analysis of phase-purity of films

Figure 73 shows an x-ray diffraction pattern of the as-deposited CuSbS$_2$ film, after subtraction of the amorphous background (via fitting and linear interpolation). A scale on the figure shows the positions of reflections associated with the CuSbS$_2$ phase, and reflection positions are also included for the potentially coexisting synthesis by-products: Cu$_{12}$Sb$_4$S$_{13}$ and Sb$_2$S$_3$ (as discussed in Section 5.2). Half-height markers in the figure indicate weak reflections, those whose intensities are below 50 % of that of the strongest reflection in their patterns.
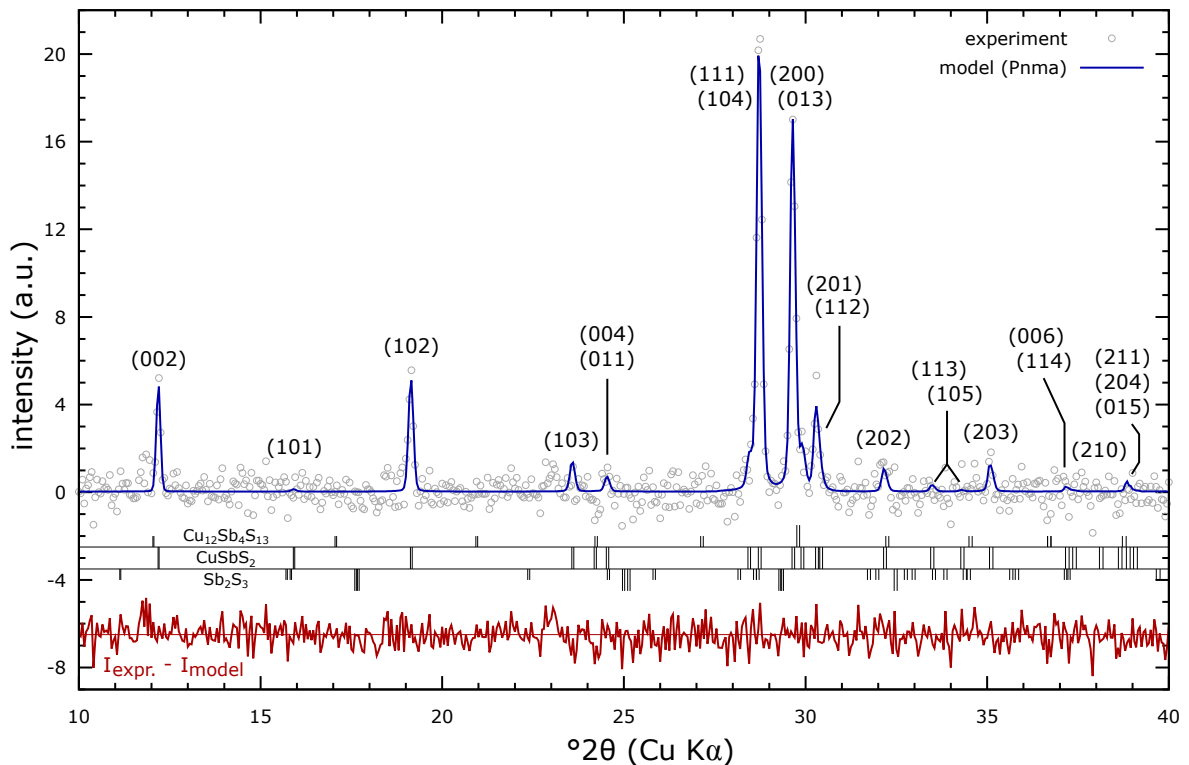
Figure 73: X-ray diffraction of CuSbS$_2$ and Rietveld refinement of a *Pnma* model. Experimental points are plotted after removal of the amorphous background. The reflection positions for CuSbS$_2$ and the nearby Cu$_{12}$Sb$_4$S$_{13}$ and Sb$_2$S$_3$ impurity phases are shown; only CuSbS$_2$ is convincingly supported by this data.

The patterns are well-fitted by Rietveld refinement[47] with the well-known $CuSbS_2$ structure (space group $Pnma$) in *FullProf 2k*[51] (v5.70). As seen in Figure 73, all the significant reflection features are strongly associated with the $CuSbS_2$ phase. The Sb_2S_3 phase is strongly ruled out: none of the strong Sb_2S_3 reflections are evident in the pattern of Figure 73. The $Cu_{12}Sb_4S_{13}$ phase is more difficult to exclude given the number of weak reflections, many of which coincide with those from $CuSbS_2$. While the strongest $Cu_{12}Sb_4S_{13}$ reflection (at $29.8°$ 2θ $K\alpha$) is just $0.1°$ from a strong $CuSbS_2$ reflection, the lack of any significant excess in the residual at $34.5°$ 2θ, where one would expect a $Cu_{12}Sb_4S_{13}$ (400) reflection if this phase was present, suggests that any contributions from this phase are minimal.

Further evidence against a $Cu_{12}Sb_4S_{13}$ phase was acquired by again refining the pattern (in *FullProf*) with a phase mixture of $CuSbS_2$ and $Cu_{12}Sb_4S_{13}$. The tetrahedrite structure as described by B.J. Wuensch[275] was used and introduced nine additional fitting parameters. Refinement of the phase mixture produced only minimal changes in the $CuSbS_2$ structural parameters: suggesting a $Cu_{12}Sb_4S_{13}$ relative phase content of $1.0 \pm 1.2\%$. Given the absence of significant $CuSbS_2$ structural changes (or improvements in the fit) despite many more parameters, and with a $Cu_{12}Sb_4S_{13}$ phase content uncertainty which exceeds the claimed value, such an approach was deemed to be less preferable than refinement with the $CuSbS_2$ phase alone. These results were taken as confirmation that the $CuSbS_2$ material is essentially phase pure for our purposes: with negligible contributions from impurity phases.

5.4.2   Refinement of CuSbS₂ phase in space group *Pnma*

Rietveld refinement[47] (in *FullProf*) of the $CuSbS_2$ structure used the space group $Pnma$ with each atom positioned in a Wyckoff $4c$ position (with two degrees of freedom). The pseudo-Voigt lineshape was then used with the usual Caglioti-like linewidth parameters V, W, η and X (for details see Albinati et al.[48]). The refinement results are presented in Table 1.

Quantity	Result
space group & formula units	$Pnma$ (number 62), 4 formula units per cell
cell volume & density	$331.1(4)$ Å^3 and 4.999 g/cm^3
lattice parameters (a, b, c)	$6.0210(23)$, $3.793(4)$, $14.496(6)$ Å
atomic positions (x, y, z)	Cu: $0.785(19)$, 0.75, $0.196(5)$
(relative cellular coords.)	Sb: $0.262(17)$, 0.25, $0.065(6)$
	S: $0.83(3)$, 0.25, $0.138(4)$
	S: $0.31(4)$, 0.75, $0.077(14)$
atomic displacement parameters	B_{Cu}: $22(1)$ Å^2, B_{Sb}: $23(1)$ Å^2 (isotropic)
preferential orientation	(200) with orientation degree $G=-0.55(8)$
fit parameters and data	19 parameters leaving 440 points in Bragg regions
background corrected metrics	$R_{wp} = 30.4\%$, $R_{Bragg} = 12.0\%$, reduced $\chi^2 = 1.19$

Table 1: Results of Rietveld refinement of CuSbS₂. The positions of the remaining 12 atoms in the unit cell are set by the symmetry elements of group $Pnma$. See the text for further details.

The fitted lattice parameters are consistent with prior literature, as discussed in the next section. Some excess residuals were observed, particularly near the (200) reflection. These were fitted with an elementary preferred-orientation model: a term $\exp(G\,\alpha^2)$ with G a scalar representing the degree of preferential orientation, and α the acute angle between the diffraction scattering vector and the normal to the (200) plane of the crystallites[51] (the Platy habit). Atomic displacement parameters (ADPs) were fitted where these improved the fit and lead to sensible results. In this case isotropic Cu and Sb parameters were beneficial, whilst anisotropic ADPs or isotropic ADPs for the sulphur atoms led to non-physical (negative) values or a divergent solution.

5.4.3 Comparison with prior work

The lattice parameters determined in this work are broadly consistent with those published previously: all of the published $CuSbS_2$ lattice parameters are listed in Table 2. Where inconsistencies arise, these are likely due to extreme values in the literature or specific differences, e.g. the nanolayer results of Ramasamy et al.[276] would be expected to differ somewhat.

Unfortunately, while experimental XRD work is often included in the literature, frequently the only experimental results reported are claimed matches to particular $CuSbS_2$ reference patterns (i.e. powder diffraction file serial numbers), rather than reporting actual computed lattice parameters or atomic coordinates with appropriate uncertainties. Without these, it is not possible to evaluate the results of different experiments for consistency.

The atomic coordinates in Table 1 are inconsistent with those found by Kyono et al.[268], who investigated single crystals to higher angles and accuracy. The results in Table 1 are possibly limited by preferential orientation and the signal-to-noise achieved in this study, which is perhaps not sufficiently accurate to fully assess such micro-structural details.

study	material	a (Å)	b (Å)	c (Å)	vol. (Å^3)	Q
Salem et al. (2013)[264]	thin film	6.0144(1)	3.768(1)	14.302(1)	324.12(9)	x
Hofmann (1933)[277]	natural cryst.	6.008(10)	3.784(10)	14.46(3)	328.6 ± 1.2	✓
Ramasamy et al. (2014)[276]	nanolayers	6.002(1)	3.793(1)	14.477(1)	329.58(11)	x
Zhou et al. (2009)[270]	single cryst.	6.014(2)	3.7882(12)	14.472(5)	329.70(19)	≈
(this work)	thin film	6.0210(23)	3.793(4)	14.496(6)	331.1(4)	·
Kyono et al. (2005)[268]	natural cryst.	6.018(1)	3.7958(6)	14.495(7)	331.11(17)	✓
Razmara et al. (1997)[278]	powder	6.016(1)	3.7968(6)	14.499(3)	331.18(9)	✓
Yang et al. (2014)[265]	thin film	6.024(1)	3.786(1)	14.701(1)	335.28(11)	x

Table 2: All published, experimentally determined $CuSbS_2$ lattice parameters and unit cell volumes. Studies listed in ascending volume order: all claimed orthorhombic structures in space group $Pnma$. *Natural* and *single cryst.* respectively label single crystal studies on naturally and artificially grown material. The Q column shows lattice parameters consistent (✓) and inconsistent (x) with this work: the parameters of Zhou et al.[270] are very close to those claimed here, only just inconsistent on the c lattice parameter.

5.5 Band gap temperature-dependence by FTIR spectroscopy

5.5.1 Reduction of FTIR reflectivity/transmissivity to absorption spectra

The measured FTIR reflectivity $R_t(\hbar\omega)$ and transmissivity $T_t(\hbar\omega)$ spectra are presented in Figure 74. The transmission spectra show a strong, temperature-dependent absorption edge centred at $1.6\,\text{eV}$. As discussed in Section 3.12.1, the reflectivity and transmissivity magnitudes in this region show opposite temperature evolution as expected for dielectric phenomena: in this particular case, the transmission magnitude at $1.6\,\text{eV}$ generally decreases with increasing temperature, while the reflection magnitude generally increases.

The inequality $R_t + T_t < 0.8$ holds across the entire photon energy range, suggesting a possibly sub-gap extinction contribution or systematic effect, such as (unmeasured) diffuse reflectivity. Below $1.6\,\text{eV}$ oscillations appear in the reflectivity spectra due to internal reflections (see the discussion in Section 3.12). For free-standing films with negligible sub-gap extinction, oscillations would normally also be expected in the transmissivity spectra; however, in the transmission of Figure 74, the oscillation amplitude is greatly reduced, by as much as a factor of 14. Possible explanations may be as follows:

1. oscillations in the transmission spectra would indeed be seen for free-standing $CuSbS_2$ films, but the coherence of the transmitted light is partially destroyed by inhomogeneities and thickness variations in the borosilicate substrate. This has the effect of averaging out the oscillations and may explain the observed spectra; however, contrary to this, similar oscillation amplitudes are seen (in this thesis) in the reflectivity and transmissivity spectra of copper nitride (with an identical substrate), see Figure 64.
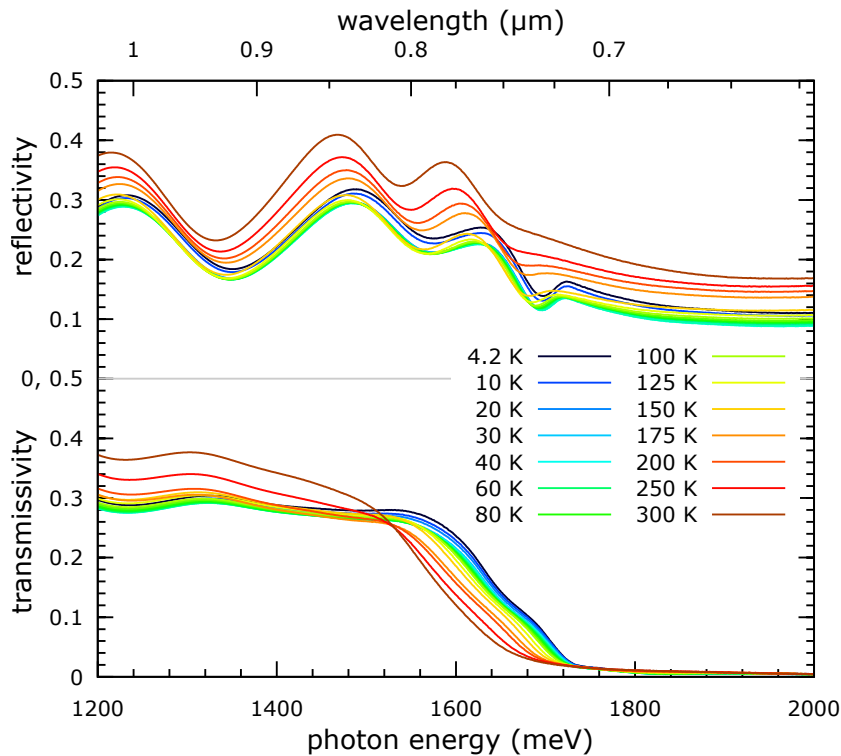
Figure 74: Temperature-dependent FTIR reflectivity and transmissivity spectra. The absorption edge shifts by some $70\,\text{meV}$ between 4.2 and $300\,\text{K}$. Negligible oscillations are seen in the transmissivity.
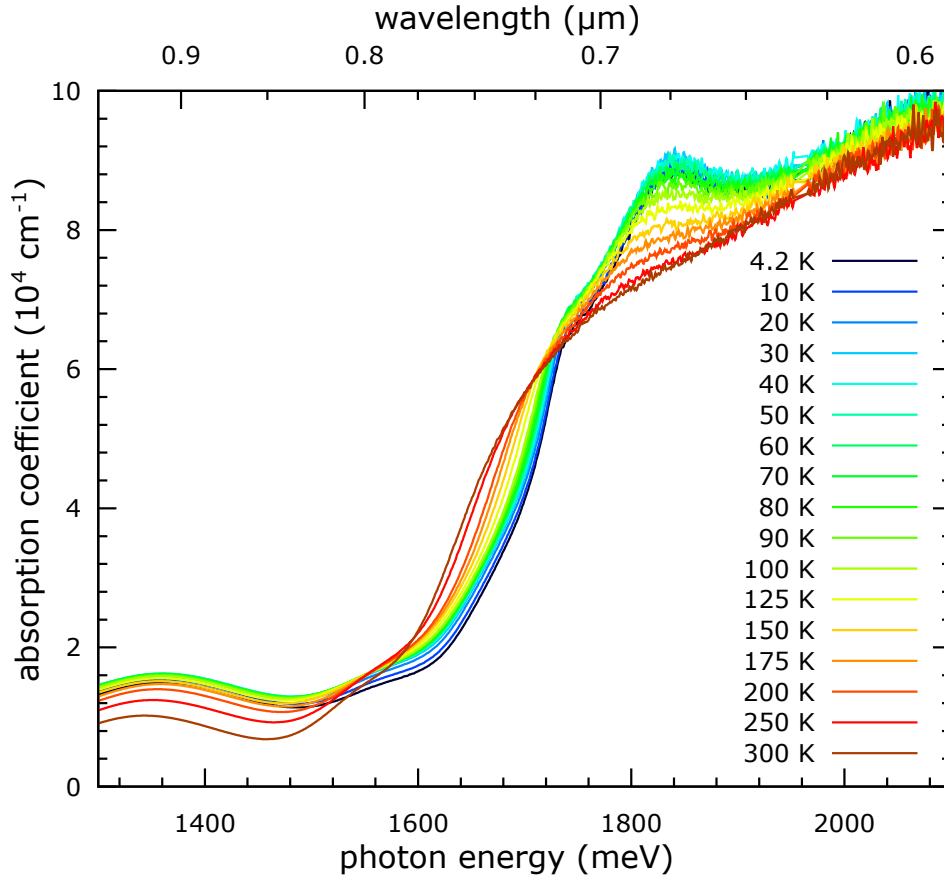
Figure 75: FTIR absorption spectra for CuSbS$_2$. Between 4.2 and $300\,\mathrm{K}$ the absorption edge redshifts by some $50\,\mathrm{meV}$. Significant sub-gap extinction is suggested and an interesting temperature-dependent feature is observed at low temperatures near $1.83\,\mathrm{eV}$, but vanishes with increasing temperature.

2. no transmission oscillations are observed due to sub-gap extinction in the film. At least two optical path lengths (twice the film thickness) are required to observe oscillations in the reflectivity, whilst at least three path lengths are required for oscillations to present in the transmissivity (see again Section 3.10). With significant sub-gap extinction, it is then feasible to observe oscillations only in the reflectivity.

These aspects are considered further in Section 3.12. The reflectivity and transmission spectra were reduced to absorption spectra $\alpha(\hbar\omega)$ using equations describing the reflection and transmission of light directed at normal incidence at a single incoherent optical layer, of thickness d with parallel faces

$$R_t = R\left[1 + \frac{(1-R)^2\exp(-2\alpha d)}{1 - R^2\exp(-2\alpha d)}\right] \qquad \text{and} \qquad T_t = \frac{(1-R)^2\exp(-\alpha d)}{1 - R^2\exp(-2\alpha d)} \qquad (5.1)$$

where $R(\hbar\omega)$ is the (unmeasured) intrinsic reflectivity at a semi-infinite boundary between the vacuum and the CuSbS$_2$ film (in contrast, the total reflectivity R_t includes the contributions from internal reflections at each face). The intrinsic reflectivity R must be determined before proceeding: this procedure has been discussed in detail in Section 3.13.4. This fitting used the sample thickness of $570\,\mathrm{nm}$, as indicated from SEM and ellipsometry.

Figure 75 shows the resulting absorption spectra. The absorption edge increases relatively
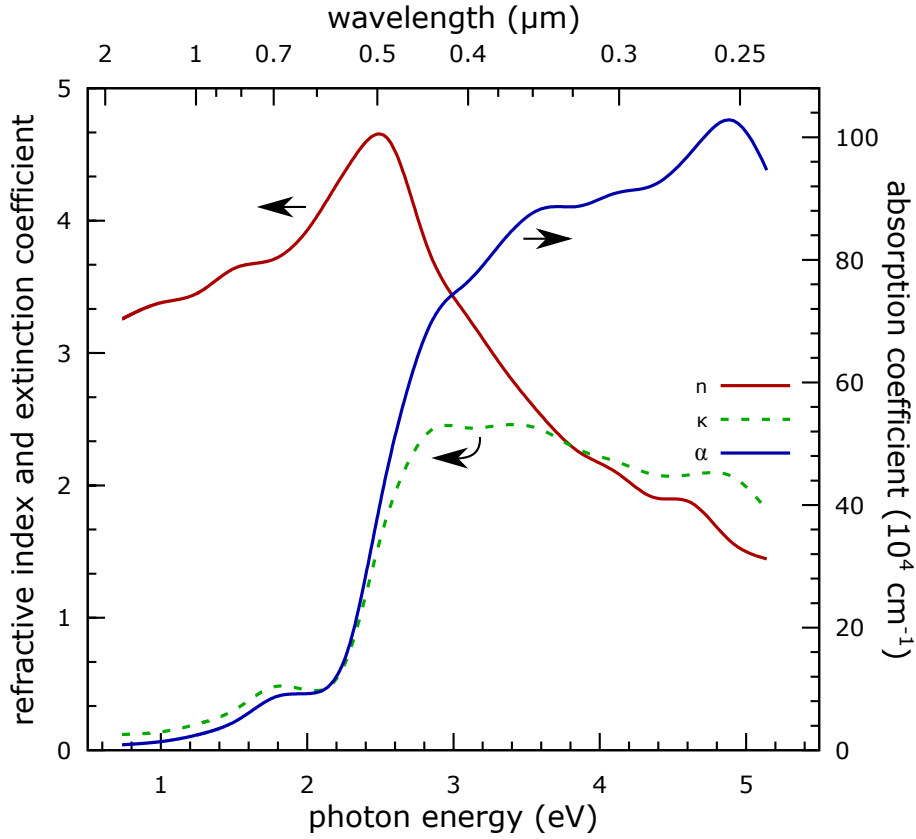
Figure 76: Absorption and refractive index spectra via spectroscopic ellipsometry. The CuSbS$_2$ complex refractive index $\tilde{n} = n - i\kappa$ and absorption coefficient are plotted at $300\,\mathrm{K}$.

steeply, as required for photovoltaic applications, and redshifts with increasing temperature. Linear extrapolation of the absorption edges below $1.7\,\mathrm{eV}$ to their intercepts with the abscissa suggests respective onsets at 1550 and $1600\,\mathrm{meV}$ for the spectra at 300 and $4.2\,\mathrm{K}$; the absorption onset is fitted more carefully in the next section. An interesting temperature-dependent absorption feature presents in Figure 75 at $1.82\,\mathrm{eV}$; this feature is analysed in Section 5.6.

The features seen in Figure 75 below $1.6\,\mathrm{eV}$ result from internal reflections and sub-gap extinction in the film. It is interesting to compare the FTIR absorption spectra with those from the $300\,\mathrm{K}$ spectroscopic ellipsometric (SE) investigation, presented in Figure 76. This fit achieved a mean-squared error (MSE) of 12.6, perhaps suggesting complexities beyond the model[†]. Below the absorption onset the difference is rather pronounced: with the FTIR and SE spectra claiming respective absorption coefficients of 1×10^4 and $4 \times 10^4\,\mathrm{cm}^{-1}$ at $1.4\,\mathrm{eV}$. Above the onset at $2\,\mathrm{eV}$ both methods agree on absorption coefficients of $9 \times 10^4\,\mathrm{cm}^{-1}$.

The difference at low photon energies may result due to issues with the SE model fit, or alternatively because the FTIR analysis neglects the substrate and treats the material as a free standing film. This omission leads to an overestimated refractive index contrast between the film and substrate and an overestimated reflection coefficient at the film-substrate boundary; hence less light is actually trapped by internal reflections in the CuSbS$_2$ film, and the claimed FTIR absorption coefficient should in fact be higher at low photon energies. Above the absorption onset, the discrepancy disappears as little light survives to reach the rear interface.

[†] When written as a fractional sum with typical accuracy[279], ideal and thicker films have respective MSEs of 1 and 10.

5.5.2 Fitting the absorption onset

The $CuSbS_2$ band structure resulting from DFT calculations (see Section 5.7.1) suggests that while the fundamental absorption is indirect, a direct onset (associated with an M_0 critical point) exists to only $50\,\mathrm{meV}$ higher energy. Furthermore, various additional transitions contribute within the next $80\,\mathrm{meV}$, so that the absorption edge comprises contributions from multiple transitions [266,273] (see Section 5.7.1). In light of this, and given that the absorption strength at indirect onsets is typically orders of magnitude weaker than that for direct transitions, the experimental $CuSbS_2$ absorption edge was fitted with functions describing absorption due to direct transitions with an Urbach tail at an M_0 critical point (see Section 3.6.3)

$$\alpha_{\mathrm{d}} = \frac{A}{\hbar\omega}\sqrt{R\left(\hbar\omega - E_{\mathrm{g}}^{\mathrm{d}}\right)} \quad \text{and} \quad \alpha_{\mathrm{U}}(\hbar\omega) = B\exp\left(\gamma\left\{\hbar\omega - E_{\mathrm{g}}^{\mathrm{d}} - E_{\mathrm{U}}\right\}\right) \quad (5.2)$$

where $E_{\mathrm{g}}^{\mathrm{d}}$ is the direct band gap, A and E_{U} are material-dependent parameters, B and γ are functions only of these parameters as discussed in Section 3.8, and $R(\cdot)$ denotes the ramp function: $R(x) = x$ for $x > 0$, else $R(x) = 0$. The total absorption associated with a direct onset is then

$$\alpha_{\mathrm{total}}(\hbar\omega) = \begin{cases} \alpha_{\mathrm{U}}(\hbar\omega), & \hbar\omega \leq \left(E_{\mathrm{g}}^{\mathrm{d}} + E_{\mathrm{U}}\right) \\ \alpha_{\mathrm{d}}(\hbar\omega), & \text{otherwise} \end{cases} \quad (5.3)$$

The results of fitting (5.3) to the experimental absorption edges via the Levenberg-Marquardt algorithm in MATLAB are presented in Figure 77. In this fitting, all spectra were fitted between 1.13 and $1.74\,\mathrm{eV}$ with a baseline evaluated for each spectrum using the average absorption coefficient below the fundamental absorption between 1.13 and $1.46\,\mathrm{eV}$. Constraints where set to ensure the solution converged to physically reasonable values. The direct gap and Urbach energy E_{U} where constrained to lie respectively between 1.5 and $2.0\,\mathrm{eV}$ and below $0.1\,\mathrm{eV}$, while the absorption strength was constrained to be positive.

Direct band gaps of $1.694\,\mathrm{eV}$ and $1.608\,\mathrm{eV}$ are seen respectively at 4.2 and $300\,\mathrm{K}$, with a redshift of $92\,\mathrm{meV}$ between 4.2 and $350\,\mathrm{K}$, leading to a band gap temperature coefficient $\mathrm{d}E_{\mathrm{g}}/\mathrm{d}T$ of $-0.26\,\mathrm{meV\,K^{-1}}$, typical of the magnitudes seen in semiconductors [244], and within the distribution seen for twelve common semiconductors [245] including Si, Ge, GaAs, GaP and CdTe, which exhibit mean (and standard deviation) band gap temperature coefficients of $-0.36(7)\,\mathrm{meV\,K^{-1}}$. The Urbach intercept E_{U} increases with temperature from 29.3 to $32.6\,\mathrm{meV}$ between 4.2 and $300\,\mathrm{K}$, see the inset in Fig. 77, with the growing phonon population prehaps accounting for the step change below (and near-linearity above) $15\,\mathrm{K}$ (see also Section 3.8); the absorption strength A shows a scatter of $15\,\%$ around $1.2 \times 10^9\,\mathrm{eV^{0.5}cm^{-1}}$. The vanishing temperature dependence of the band gap at low temperatures, as seen in Figure 77, is typical of the behaviour seen in semiconductors [244,246]. The $300\,\mathrm{K}$ and $350\,\mathrm{K}$ band gaps in the Figure 77 possibly suggest a relaxation of the temperature-dependence at higher temperatures, but there are too few datapoints available here to discriminate this from unidentified systematic experimental factors. Likewise, too few datapoints exist to confirm the apparent abrupt relaxation of the Urbach energy at $300\,\mathrm{K}$ as seen in the inset of Fig. 77.
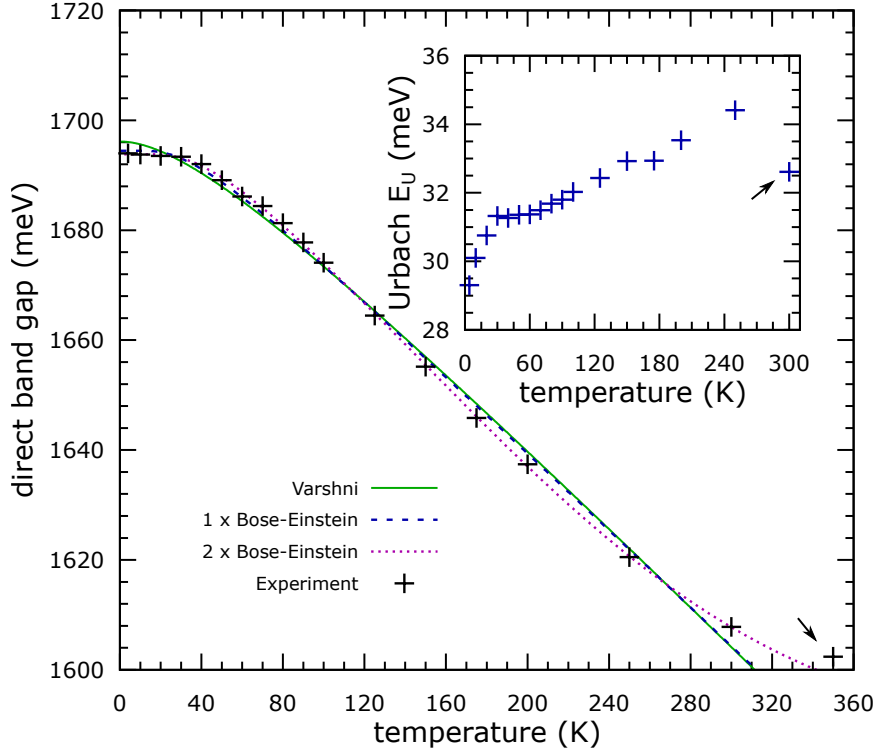
Figure 77: Temperature-dependence of the CuSbS$_2$ absorption edge, considered as a direct band gap. The temperature-dependence is fitted with the empirical Varshni relation, and one and two Bose–Einstein oscillators. The insert shows the familiar evolution of the Urbach tail, which grows with temperature. The $350\,\mathrm{K}$ datapoint is not fitted for temperature-dependence; the $300\,\mathrm{K}$ Urbach energy is discussed in the text.

5.5.3 Fitting the band gap temperature-dependence with electron-phonon interaction

The temperature-dependence of the optically-determined band gaps $E_g(T)$ in Fig. 77 are fitted between 4.2 and $300\,\mathrm{K}$ with the empirical Varshni relation[247] (see Appendix C)

$$E_\mathrm{g}^{\mathrm{Varshni}} = E_0 - \frac{\alpha T^2}{T + \beta} \tag{5.4}$$

and to models comprising one and two Bose-Einstein oscillators[246,248,249]

$$E_\mathrm{g}^{BE} = E_0 + \sum_i \alpha_i \left[1 + \frac{2}{\exp\left(\frac{E_i}{k_\mathrm{B}T}\right) - 1} \right] \tag{5.5}$$

The results of the least-squares fits are shown as the solid and broken lines in Fig. 77. The Varshni fit suggests respective E_0, α and β parameters of $1.70\,\mathrm{eV}$, $0.37\,\mathrm{meV\,K^{-1}}$ and $63.2\,\mathrm{K}$; such values are well within the typical range seen for Varshni parameters: as seen by respective mean (and standard deviation) α and β values of $0.6(5)\,\mathrm{meV\,K^{-1}}$ and $400(500)\,\mathrm{K}$ for fits to 37 common semiconductors[247,250]. The single Bose–Einstein oscillator fit finds a zero Kelvin gap of $1.71\,\mathrm{eV}$ with reasonable respective average phonon and electron-phonon interaction energies of 8.4 and $-17.3\,\mathrm{meV}$. The double Bose–Einstein oscillator fit, which caters for the different energy scales expected for acoustic and optical phonons, suggests an E_0 of $1.15\,\mathrm{eV}$, and respective average phonon and interaction energies $\{E_i, \alpha_i\}$ of $\{11.3, -26.80\}$ and $\{117.9, 567.35\}\,\mathrm{meV}$. This fit achieves a particularly good reduced χ^2 statistic of 0.18, with a reasonable mix of small and large averaged phonon energies.

5.6 On potential excitonic feature in FTIR absorption spectra

5.6.1 Background subtraction, line-shape fitting and temperature-dependence

A temperature-dependent feature is evident at $1.82\,\mathrm{eV}$ in the absorption spectra of Figure 75. With decreasing temperature the feature narrows, increases in prominence and shifts to higher photon energies, becoming particularly obvious below $150\,\mathrm{K}$. For all temperatures, the feature sits at a photon energy just above the fundamental absorption edge (whose strength weakens somewhat above $1.72\,\mathrm{eV}$). Below $200\,\mathrm{K}$, a definite shoulder separates the feature from the absorption edge: suggesting a transition between states separated by more than the fundamental band gap. At $4.2\,\mathrm{K}$, the feature has raised by some $10\,\mathrm{meV}$ to $1.83\,\mathrm{eV}$; for comparison, in Section 5.5.2 a direct gap of $1.694\,\mathrm{eV}$ was determined at $4.2\,\mathrm{K}$.

To further examine the feature it is necessary to remove the absorption due to the earlier fundamental onset. To achieve this, some assumptions are necessary. Above the feature, similar absorption evolution is seen at all temperatures: with the absorption strength increasing linearly from $1.92\,\mathrm{eV}$ with a similar gradient at all temperatures (see Figure 75). In the $300\,\mathrm{K}$ trace, where no feature is discernible, the absorption gradient is constant between 1.74 and $2.1\,\mathrm{eV}$. Yet below $100\,\mathrm{K}$ the baseline absorption gradient (i.e. that ignoring the feature) changes slightly above and below $1.92\,\mathrm{eV}$. This necessitates some assumptions on the nature of the baseline absorption below $1.92\,\mathrm{eV}$. Assuming that, without the $1.82\,\mathrm{eV}$ feature, all the fundamental onsets would evolve similarly to the $300\,\mathrm{K}$ onset, background subtraction is achieved by projecting a line from the shoulder below the feature to intercept with the unambiguous linear absorption at $1.92\,\mathrm{eV}$. This approach relies on the assumptions that all contributions from the feature are confined between 1.74 and $1.92\,\mathrm{eV}$, and that the feature indeed makes negligible contributions to the $300\,\mathrm{K}$ absorption spectrum.

The background-subtracted spectra are shown in Figure 78. The feature has an asymmetrical line shape; whilst the low-energy edge is nearly linear, the high-energy edge is approximately Lorentzian. Asymmetric Lorentzian or Gaussian line shapes are typically seen in the discrete absorption line spectra of excitons[119,120]. To fit the temperature evolution, a quasi-Lorentzian was constructed with these line shape characteristics. Starting with the standard Lorentzian lineshape

$$L(x; x_0, \Gamma, A) = \frac{A}{\pi} \cdot \frac{\Gamma/2}{\left(x - x_0\right)^2 + \left(\Gamma/2\right)^2} \tag{5.6}$$

where x_0 is the line centre-position, Γ is the full-width at half-maximum and the amplitude A accounts for the line intensity. The quasi-Lorentzian is then

$$L_Q(x; x_0, \Gamma, \Lambda, A) = \begin{cases} L_\ell(x; x_0, \Gamma, \Lambda, A), & x < x_0 - \Lambda \\ L(x; x_0, \Gamma, A), & x \geq x_0 - \Lambda \end{cases} \tag{5.7}$$

where the parameter Λ sets the separation from the line centre-position x_0 to the transition point where the line shape changes smoothly and continuously from a Lorentzian into a
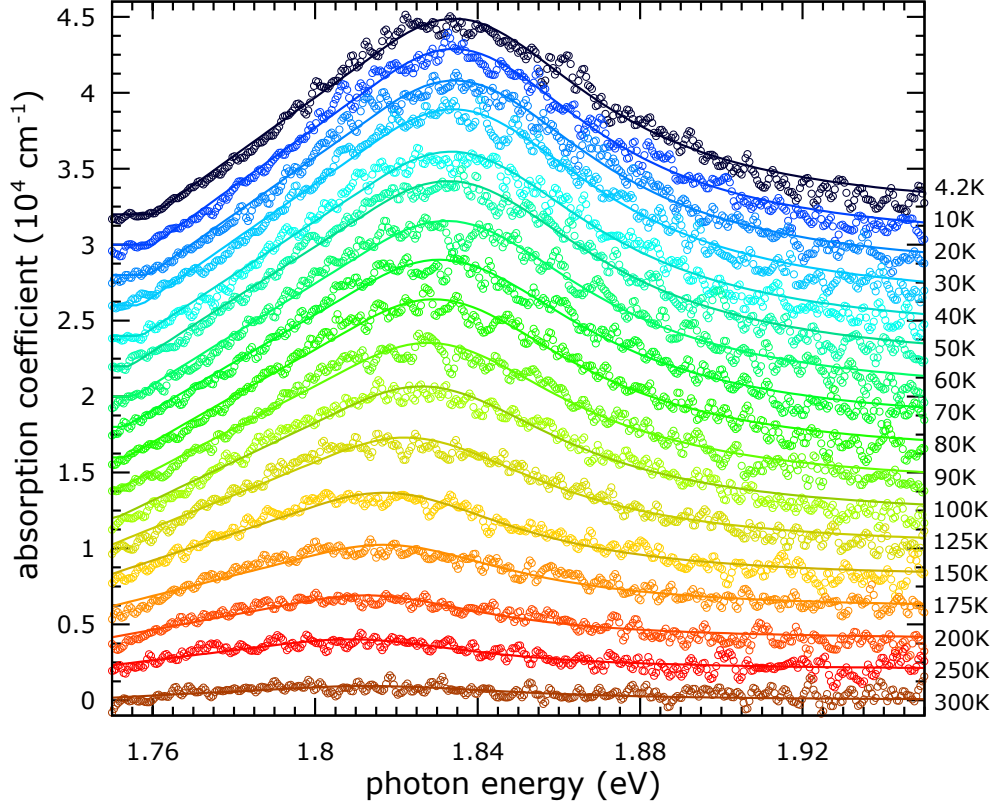
Figure 78: Temperature-dependent feature in CuSbS$_2$ absorption spectra. The figure expands on the feature seen in Figure 75. The fundamental absorption onset has been removed (as discussed in the text). For clarity, each trace is displaced vertical by 2×10^3 cm^{-1}. The line position and intensity both increase with decreasing temperature; no feature is seen at 300 K. Note the asymmetric line shape which comprises a Lorentzian high-energy edge (and peak) with a linear low-energy edge.

linear edge. Analysis suggested a parameter $\Lambda = \Gamma/4$ for this work. The linear edge is then

$$L_\ell(x; x_0, \Gamma, \Lambda, A) = R\left\{ L(x_0 - \Lambda; x_0, \Gamma, A) - (x_0 - \Lambda - x) \cdot \left.\frac{\mathrm{d}L}{\mathrm{d}x}\right|_{(x_0 - \Lambda; x_0, \Gamma, A)} \right\} \qquad (5.8)$$

where R again denotes the ramp function, $R(x)$=0 for $x < 0$ and $R(x)$=x for $x \geq 0$. The fits are shown by the solid lines in Figure 78, while the temperature-dependence of the line-position is seen more clearly in Figure 79. The feature and temperature evolution seen in Figures 78 and 79 may result from various phenomena coupled with a band structure temperature dependence due to thermal expansion and electron–phonon interactions (as discussed in Appendix C). Indeed, the behaviour seen in Figure 79 is highly typical of band structure temperature dependence. To investigate this, the temperature dependence is fitted with the Varshni relation (5.4) and a single Bose–Einstein oscillator (5.5), as depicted by the lines in Figure 79. This work suggests Varshi parameters $\{E_0, \alpha, \beta\}$ of $1.8349(4)$ eV, $0.18(3)$ meV K^{-1} and $120(50)$ K, and Bose–Einstein parameters $\{E_0, E_{\mathrm{ph}}, \alpha_{\mathrm{ep}}\}$ of $1.8406(16)$ eV, 7.6 ± 2.2 meV and $-12(4)$ meV. Of these, the Bose–Einstein oscillator best describes the data, with an impressively small reduced χ^2 statistic and a sensible averaged phonon energy; indeed, the fitted averaged phonon and electron–phonon energies are consistent with those determined for the double Bose–Einstein oscillator in Section 5.5.3 (assuming a 10% uncertainty for the latter). The fits in this section, as displayed in Figure 79, employ asymptotic standard errors (a measure of the residuals in Figure 78) rather than standard errors.
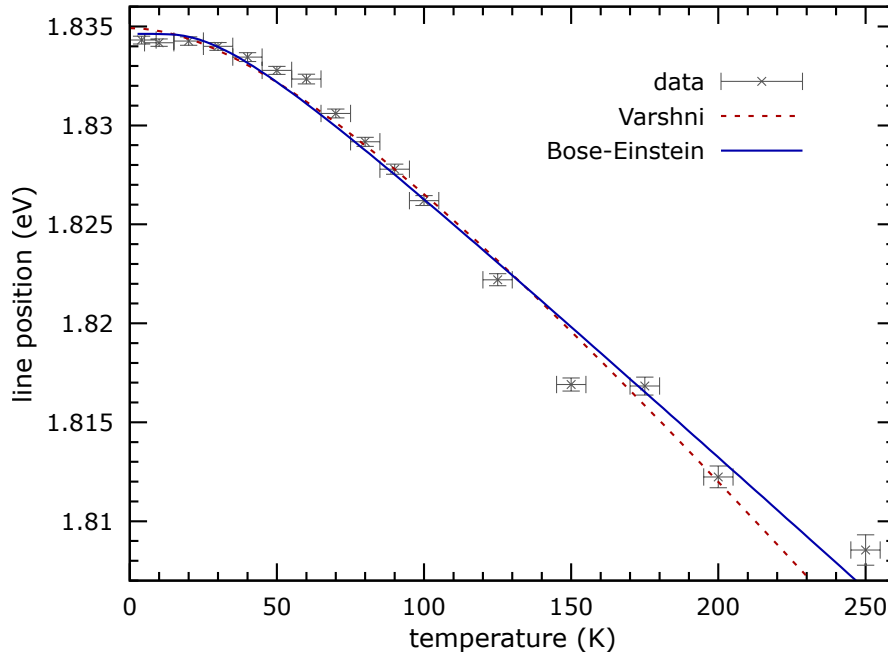
Figure 79: Temperature dependence of possible exciton line position, fitted as a quasi-Lorentzian and then with the Varshni relation and a single Bose–Einstein oscillator. The temperature dependence follows that typically seen for semiconductor bands. The error bars denote asymptotic standard errors.

5.6.2   Characterisation as excitonic phenomenon

To establish an excitonic hypothesis for the feature in Figure 78, the nature of the exciton must first be determined. Additionally, one must explain why no discrete excitonic lines are seen below $1.69\,\mathrm{eV}$ due to the interband transitions of the fundamental onset. Because only a single asymmetrical line is seen (a typical outcome in experimental work on excitons), the available evidence is somewhat scarce and one approach is to study candidate transitions in the band structure. Excitonic models differ according to the magnitude of the exciton's binding energy, the associated band gap and critical point type (direct or indirect; M_0 or M_1), and the symmetries of the excitonic envelope, perturbing operator and initial and final states (which give electric dipole selection rules). The familiar hydrogenic solution for the weak-bound Wannier exciton is developed for allowed and forbidden electric-dipole transitions across direct gaps at M_0 critical points. The Wannier exciton is often further simplified with isotropic or spherical bands at the Γ-point[96,115], where dipole-allowed or -forbidden transitions lead respectively to s- or p-type hydrogenic states exclusively.

Without experimental results on the dispersion of the electronic bands or symmetries of the wavefunctions, theoretical methods enable first steps and reasonable approximations to interpret observed behaviour. For this purpose, the HSE06 band structure and GGA symmetry analysis of Section 5.7 are used. As mentioned in that section, using the Kohn–Sham wavefunctions and eigenvalues in this manner is somewhat without rigorous justification, particularly for excited states or hybrid functionals; however, such usage seems standard practice in the community, where theoretical and experimental absorption spectra are regularly compared, and potential photovoltaic absorbers are assessed based on the symmetries of their conduction and valence bands[259,260] (respectively CB and VB).
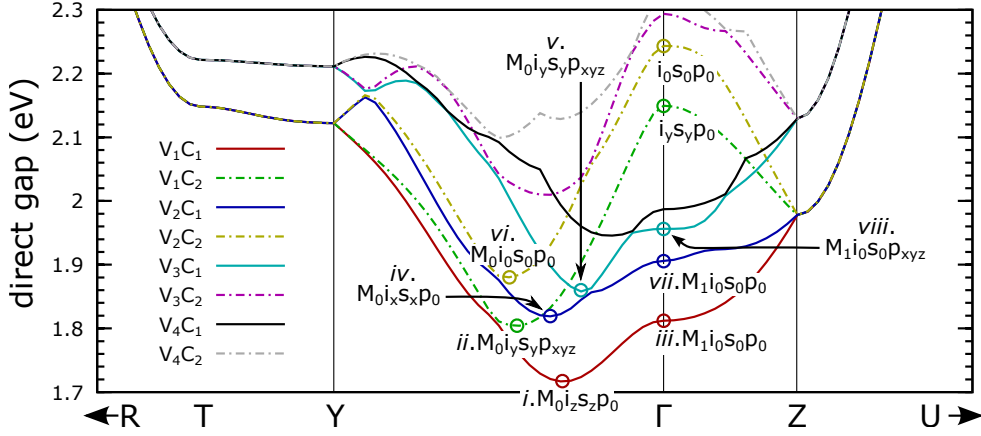
Figure 80: Lowest critical points in the HSE06 band structure and GGA selection rules. The topmost valence bands (V_1 descending through V_4) are subtracted from the lowest conduction bands (C_1 increasing through C_4) giving the direct gaps near Γ. The lowest critical points (M_0 and M_1 Van Hove singularities) and dipole selection rules are given: i, s and p respectively denote interband and excitonic s- and p-transitions, while the subscripts denote the required light-polarisation direction. See the discussion in Section 5.7.1.

For convenience, the critical points and dipole selection rules of Figure 84 from Section 5.7.3 are reproduced in Figure 80. As discussed in Section 5.7.1, the band structure of Figure 82 finds that the lowest transition in CuSbS$_2$ is indirect, but that a dipole-allowed direct gap (labelled i in Figure 80) exists between Γ-Y to only $50\,\mathrm{meV}$ higher energy. This band configuration suggests potentially interesting behaviour. Studying the figure for candidate direct transitions associated with the exciton, i.e. those lying above the dipole-allowed M_0 critical point (i) between C_1-V_1 along Γ-Y, suggests

- the M_0 critical points along Γ-Y associated with the secondary valence and conduction bands (respectively V_1-C_2, V_2-C_1, V_3-C_1 and V_2-C_2 in the nomenclature of Figure 84).

- the M_1 critical point (iii) between the CB and VB edges at the Γ-point.

Each of these transitions has a calculated energy of the required magnitude, $E_\mathrm{g} + 0.1\,\mathrm{eV}$. Excitonic transitions are then limited by the dipole selection rules as indicated in Figure 80. Assuming a hydrogenic Mott–Wannier model and using the critical points on the symmetry line between Γ and Y, **excitonic s-states are allowed for only one polarisation direction** of the incident light (due to the orthorhombic lattice) at each M_0 critical point (transitions i, ii, iv) while only above $2.1\,\mathrm{eV}$ at Γ. This polarisation restriction combined with the polycrystalline nature of the material have the result that s-type excitonic transitions are weakened to one third of their full intensity (assuming a random distribution of crystallites).

In contrast, **excitonic p-states are allowed for all polarisation directions** for transitions ii and v between Γ and Y, respectively just some $90\,\mathrm{meV}$ and $140\,\mathrm{meV}$ above the direct onset (i); these features may be inseparable in the absorption spectra of Figure 78. Further, due to the non-isotropic nature of the orthorhombic lattice, the (usually mutually exclusive) s- and p-type excitonic states are allowed simultaneously at the ii and v transitions: leading to an enhanced resonance which may explain the observation of excitonic phenomena in

the sputter-grown, polycrystalline $CuSbS_2$ material. As discussed in Section 3.9, the three-dimensional Wannier–Mott model suggests a series of discrete lines at energies E_n

$$E_n = E_0 - \frac{\mu\, e^4}{2\left(4\pi\epsilon_0\epsilon_r n\hbar\right)^2} = E_0 - \frac{\mu}{m_e\epsilon_r^2}\cdot\frac{R_H}{n^2} = E_0 - \frac{R_x}{n^2}\quad,\qquad n = 1, 2, \cdots \qquad (5.9)$$

where E_0 is the associated gap, μ is the reduced electron and hole mass ($1/\mu = 1/m_e^* + 1/m_h^*$), m_e^* and m_h^* are the effective masses of electrons and holes in their bands, R_H=13.6 eV and R_x are respectively the Rydberg energy for hydrogen and for the exciton, n is the principal quantum number and ϵ_r is the real part of the dielectric constant evaluated at E_n.

The reduced mass in each direction may be approximated by fitting the dispersion of the HSE06 bands given in Figure 82. For example, fitting these bands with fifth degree polynomials finds a reduced mass μ=0.47 m_e for the critical point (ii) of Figure 80. Using a dielectric constant $\epsilon_r(1.82\,\mathrm{eV})$=13.7 (from the ellipsometry of Figure 76) and asumming n=1 in Eqn. (5.9) gives a binding energy R_x of 34 meV for the exciton associated with (ii). These parameters suggest an excitonic Bohr radius of 1.5 nm, such that a spherical volume enclosing the n=1 state averages over 47 $CuSbS_2$ unit cells, supporting a Wannier–Mott treatment. In practice, such an approach is unlikely to be accurate for the bands between Γ-Y due to asymmetric dispersion around each critical point caused by anti-crossings at different wavevectors in both valence and conduction bands.

Another possible explanation suggested in Figure 80 is the possibility that **excitons may be somehow restricted to (or enhanced at) the** M_1 **critical points** at Γ. The literature is somewhat contradictory on observation criteria for excitons: some authors assume isotropy and essentially constrain any exciton to the Γ-point[96] (which would conveniently forbid excitons between Γ-Y); however, excitons are widely-reported[280,281] at critical points well away from Γ. Alternatively, excitons associated with M_1 transitions (saddle-point or hyperbolic excitons) are somewhat controversial due to the metastability of the excitonic state[282,283]; such excitons are expected to be viable in the case of large negative reduced masses, with solutions taking the form of two-dimensional hydrogenic states[44,284]. Hyperbolic excitons are now reported in various material systems[285], often with reduced lifetimes and broader lineshapes, e.g. a claimed asymmetric M_1 excitonic line in GaSe shows Varshni-like temperature-dependence, broadening and diminishing in intensity with increasing temperature[116]; such typical excitonic behaviour mirrors that seen in Figures 78 and 79.

One interesting aspect of the 2D exciton is the eight-fold intensity enhancement of the n=0 line[44] (compared with the 3D exciton), which may again account for the presence of excitonic features in polycrystalline $CuSbS_2$ material. In Figure 80, the the (viii) critical point has the largest negative reduced mass (μ_y=$-$1.4) of the M_1 transitions at Γ. Additionally, the selection rules here allow p-**like states for all light polarisations**. Thus hyperbolic excitonic states may be viable at Γ. The 2D excitonic solution[44] predicts discrete lines at energies E_n

$$E_n = E_0 - \frac{\mu}{m_e\epsilon_r^2}\cdot\frac{R_H}{(n + 1/2)^2} = E_0 - \frac{R_x}{(n + 1/2)^2}\quad,\qquad n = 0, 1, \cdots \qquad (5.10)$$

Assuming that the $n = 0$ line is seen in Figure 78, equation (5.10) suggests an exciton binding energy of 77 meV. The larger binding energy results from the form of the denominator; in comparison, a binding energy of 300 meV was claimed[116] for GaSe. The result requires some qualification: while two-dimensional hydrogenic solutions (and selection rules) hold for M_1 excitons under the constraint of large negative reduced masses[44], these are unlikely to apply generally for reduced masses of comparable magnitude. In conclusion, the band structure of Figure 80 **suggests distinctly enhanced absorption in the vicinity of the photon energy of** $E_{\mathrm{g}} + 0.1$ eV arising from specific excitons at M_0 or M_1 critical points, which dominate over potential excitonic lines associated with other transitions.

5.6.3 Discussion and context

Claims of excitons in $CuSbS_2$ are perhaps unsurprising given the rich display of exciton phenomena in the (tetragonal) Cu-III-VI$_2$ chalcopyrites[286], see Figure 81. An excitonic hypothesis for the absorption feature near 1.82 eV is motivated primarily by the lineshape and temperature evolution of the line intensity. The highest temperature for which the feature shows significant absorption is 200 K, the average thermal energy $3/2\, k_{\mathrm{B}}T$ here being 25.9 meV, 76 % of the claimed excitonic binding energy R_x of 34 meV; however, the viability of excitonic states is mainly determined by interactions with optical phonons[87]. With the phonon population given by Bose–Einstein statistics, and using a 11 meV phonon energy (from Section 5.5.3), the phonon occupation number exceeds unity before 200 K, which qualitatively supports the absence of excitonic features beyond this temperature. The significance of selection rules for excitonic transitions in orthorhombic systems is demonstrated in AgInS$_2$ (space group $Pna2_1$; number 33; point group $mm2$), whose three non-degenerate valence bands (Γ_4, Γ_2 and Γ_1) allow dipole transitions to the Γ_1 conduction band only from the Γ_4 and Γ_1 bands (respectively for z- and y-polarised light); similarly, transitions to excitonic s-states are allowed from all valence bands with z-polarised light only[287,288].

The appearance of a new absorption resonance above the band edges intuitively seems problematic, but no difficulty arises if the state lifetime is of similar order to the timescale of the fastest relaxation process. In a similar manner, electrons excited via interband absorption to higher states in the conduction band thermalise (via phonon interactions) back to the band edges in picosecond timescales. At room temperature these same thermalisation processes are sufficiently rapid to halt exciton formation; at lower temperatures, excitonic states are viable wherever they exist. As carrier density increases under strong illumination, exciton resonances are weakened due to Coloumb potential screening[289]; however, excitonic phenomena are still viable below the critical density: so this does not trivially restrict excitons in $CuSbS_2$. In fact, **it is not unusual to find exciton lines above the band gap**: particularly when associated with transitions from split-off valence bands, e.g. the absorption spectrum of the layered compound InSe has excitons associated with the 1.35 eV fundamental band gap, and also at 2.6 and 2.9 eV (lines E1′ and E1) associated with split off valence bands derived from Se $4p$ orbitals[290]. Above gap excitons are seen in the binary transition metal dichalcogenides
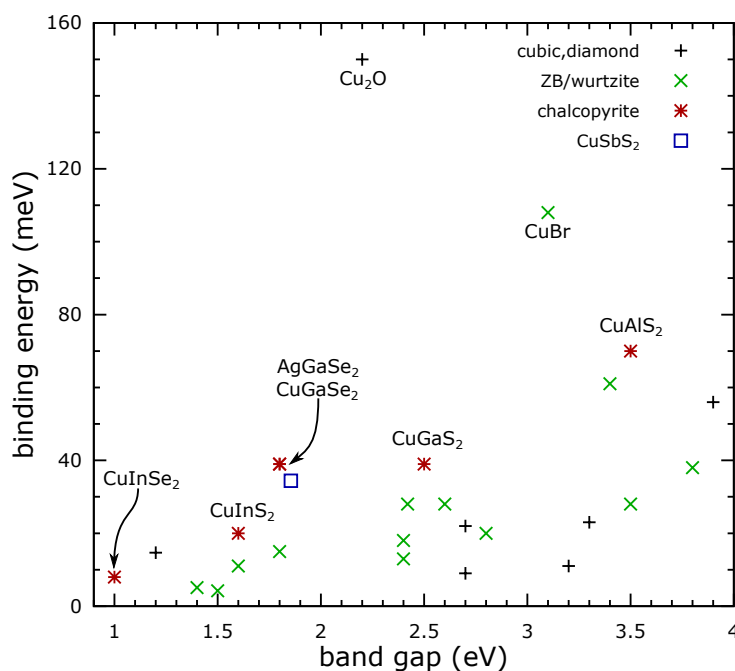
Figure 81: Exciton binding energies versus bandgap. The excitonic binding energy of $34\,\text{meV}$ for the $1.82\,\text{eV}$ feature in $CuSbS_2$ (considered as a 3D Wannier–Mott exciton with a $1.85\,\text{eV}$ gap at the M_0 critical points marked ii or v) is reasonably consistent with the trend found in chalcopyrite systems.

(TMDCs)[291], which share a layered structure with the ternary TMDC $CuSbS_2$. For example, two 2D exciton lines (A and B) show below and above the band gap in the absorption spectra of the bulk semiconductor $2H\text{-}MoS_2$; these are again associated with a $0.2\,\text{eV}$ spin-orbit split valence band[292]. Because optical transitions conserve the spin of the excited carrier, and the spacial parts of the wavefuctions may possess different symmetries, the excitonic selection rules involving spin-orbit split bands may be more restrictive[280].

The linewidth of the state (FWHM of $60\,\text{meV}$) suggests a $10\,\text{fs}$ lifetime via the uncertainty principle; however, such an elementary approach is likely invalid as the band structure suggests an envelope comprised of various components: including multiple interband transitions and different excitonic discrete features and continua. Further, due to the nature of the polycrystalline, sputter-grown material, broadening due to inhomogeneities (such as defects and impurities) may dominate over the homogeneous linewidth (mainly resultant of phonon interactions and the intrinsic width). Tightly bound Frenkel excitons were not considered as the evaluated binding energies are within the typical range for Wannier–Mott excitons (see Figure 81); metallic electron-hole drops (extreme excitons) were not conviningly supported. If an excitonic origin for the feature in Figure 78 is excluded, then other possibilities may include atypical interband absorption or optically-active impurity or defect states above the band edges. Impurity or defect states can share many of the same characteristics as seen in excitonic absorption. Alternatively, while M_1 critical points are expected near the absorption onset (see Figure 80) and interband (non-excitonic) absorption at such points creates absorption maxima directly at the associated band gap energy[44], the curvature of the contributing bands would need to evolve differently with temperature (effectively changing the per-band density of states) if such phenomena is to explain the temperature-dependence seen in Fig. 78.
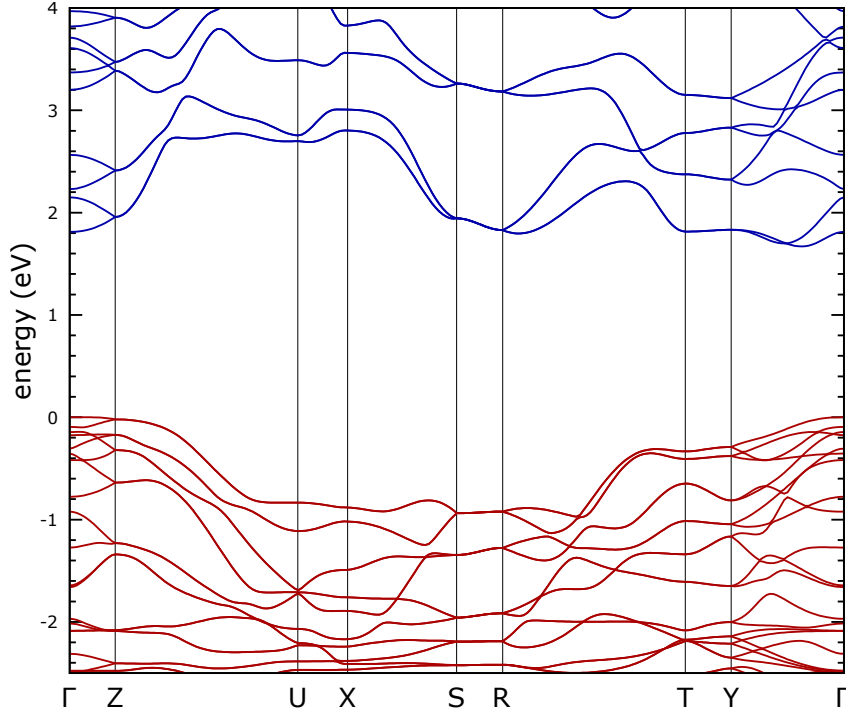
Figure 82: CuSbS$_2$ band structure by hybrid DFT (HSE06). Neglects spin-orbit effects. The smallest gap is indirect at $1.67\,\mathrm{eV}$ from Γ. The smallest direct gap is $1.72\,\mathrm{eV}$ at $(0, 0.156, 0)$.

5.7 Theoretical studies

Density functional theory (DFT) calculations were performed in the ABINIT[233,234] code, with the eigenfunctions expanded in a planewave basis with a $952\,\mathrm{eV}$ cut-off energy, the ionic potentials approximated by Troullier–Martins pseudopotentials[235] and the exchange–correlation potential treated with the generalised gradient approximation (GGA) functional of Perdew, Burke and Ernzerhof[78] (PBE96). Convergence studies on the planewave cut-off energy and appropriate k-point grid density using the experimental lattice parameters achieved total energy convergence to better than $0.01\,\%$; while the self-consistent field cycle converged upon energy differences below $0.03\,\mathrm{meV}$. The selected mesh was based on a shifted $13 \times 8 \times 4$ grid with 56 k-points in the irreducible Brillouin zone.

The GGA functional tends to underbind systems and always underestimates band gaps. For improved accuracy, DFT calculations were independently made by Chris Savory (UCL, UK) in the VASP[236,237] code using the screened hybrid density functional of Heyd, Scuseria and Ernzerhof[79,238]; a plane-wave basis was used for the valence states and the projector augmented wave (PAW) method[239,240] described interactions between core- and valence elections. Further details of similar calculations are found in Temple et al.[256]. DFT calculations often omit spin-orbit interactions. Given prior literature suggesting valence and conduction bands dominated respectively by (Cu-d, S-p) and (Sb-p, S-p) states, this interaction potentially has significant effects in the CuSbS$_2$ band structure: especially for antimony (atomic number, z=51) levels as the interaction goes as z^4. To test the impact of the spin-orbit interaction, the HSE06 band structure was additionally evaluated with the spin-orbit interaction with the same lattice parameters, cut-off energy and k-mesh.

Figure 83: Band structure and symmetries via GGA functional (neglects spin-orbit interaction).

5.7.1 Band structure analysis

The resulting band structure is shown in Figures 82 and 83 for the HSE06 and GGA functionals respectively. While the results from the HSE06 functional are preferred, finding a band gap in good agreement with experiment, the dispersion of the GGA bands in Figure 83 is strikingly similar. The GGA work is useful due to coverage of all symmetry lines in the irreducible Brillouin zone, confirming that no smaller gaps exist along symmetry lines such as T-Z or Γ-X, which are omitted from Figure 82. Furthermore, the similarity of the dispersions of the HSE06 and GGA bands means that a symmetry analysis of the GGA bands is potentially highly applicable to the HSE06 results, as discussed in Section 5.7.2.

The HSE06 valence band maximum lies at the Γ-point, while the conduction band minimum is found on the line between Γ and Y (see Section 5.7.2 for setting details). The fundamental absorption in $CuSbS_2$ is thus indirect across a gap of $1.67\,\text{eV}$. The smallest direct gap is some $50\,\text{meV}$ to higher energy, giving a direct gap of $1.72\,\text{eV}$ associated with a type M_0 critical point (Van Hove singularity) at the reciprocal space point $(0, 0.156, 0)$. This is more clearly seen in Figure 84, in which the topmost valence bands are subtracted from the lowest conduction bands to show the nature of the critical points. The direct gap at Γ is somewhat higher at $1.81\,\text{eV}$. Interestingly, the band structure at Γ has a saddle-point geometry, suggesting a M_1 critical point in the density of states and potentially distinct characteristics in optical spectra. Even before direct transitions become viable at the Γ-point, $1.80\,\text{eV}$ transitions to the second conduction band (another M_0 critical point) at $(0, 0.219, 0)$ may become important. When spin-orbit effects are included, all of the band gap magnitudes in this section are reduced by some $30\,\text{meV}$, but the form of the band structure remains essentially unchanged.

133

The atomic contributions to the HSE06 density of states support prior work suggesting valence and conduction bands comprised mainly of {Cu, S} and {Sb, S} states.

5.7.2 Symmetry analysis of bands near the Γ-point

Experimental spectroscopy may be greatly assisted with details of the symmetries of the initial and final states, as this information permits the writing of selection rules governing transitions between these states, e.g. electric-dipole selection rules in optical absorption. Additionally, an appreciation of these same band symmetries is important for photovoltaic applications: as dipole-forbidden bands lying below the first dipole-allowed transition potentially behave as non-radiative recombination centres[260].

Group theory is the study of symmetry and includes tools to determine the symmetries of the wavefunctions associated with each band and wavevector. The space group of the crystal designates the point-group R and translation τ operations which leave the crystal unchanged. In the case of the $Pnma$ space group of CuSbS$_2$, non-symmorphic symmetry elements are required to fully describe the crystal symmetry. These elements combine point-group and translation operations $\{R, \tau\}$, with the result that the origin choice for the unit cell is significant: affecting both the atomic coordinates and the matrices which describe the symmetry operations. The band structures of Figures 82 and 83 are plotted for the right-handed system with lattice parameter magnitudes $a < b < c$. This system differs from the conventional setting shown for $Pnma$ in the International Tables[293], which for CuSbS$_2$ gives the right-handed system $b < a < c$ and is often seen in published CuSbS$_2$ band structures (swapping the X and Y points in the Brillouin zone).

To determine the symmetries of the GGA bands around Γ, the characters of the equivalent representations of the wavefunctions are required. To evaluate these, the wavefunctions of the top three valence bands and the lowest two conduction bands were exported, each giving a set of probability amplitudes (complex numbers) over a realspace grid of $384\,000$ vectors covering the volume of the unit cell. Software written in C++ was then used to transform each wavefunction $|\psi\rangle$ using the eight symmetry operations $\{R, \tau\}$ of $Pnma$ (written in the unconventional setting) and to evaluate the modulus M and argument θ of the inner-product

$$\langle\psi|\{R, \tau\}\psi\rangle = Me^{i\theta} \qquad \text{with} \qquad |\psi|^2 = M \qquad (5.11)$$

where M is a positive real scalar. Vectors transformed outside the unit cell boundaries by symmetry operations were translated back into the unit cell via a lattice vector. Only for elements of the space group is the modulus M invariant for both equations in (5.11). The characters of the equivalent representation of the wavefunction are then given by $e^{i\theta}$, which in space group $Pnma$ is always ± 1 (achieved up to an numerical error below 10^{-9}).

Finally, the characters of each wavefunction and the character table of the point group are used with the orthogonality theorem of group theory[294] to decompose the equivalent representation of each wavefunction into irreducible representations (irreps) of the point
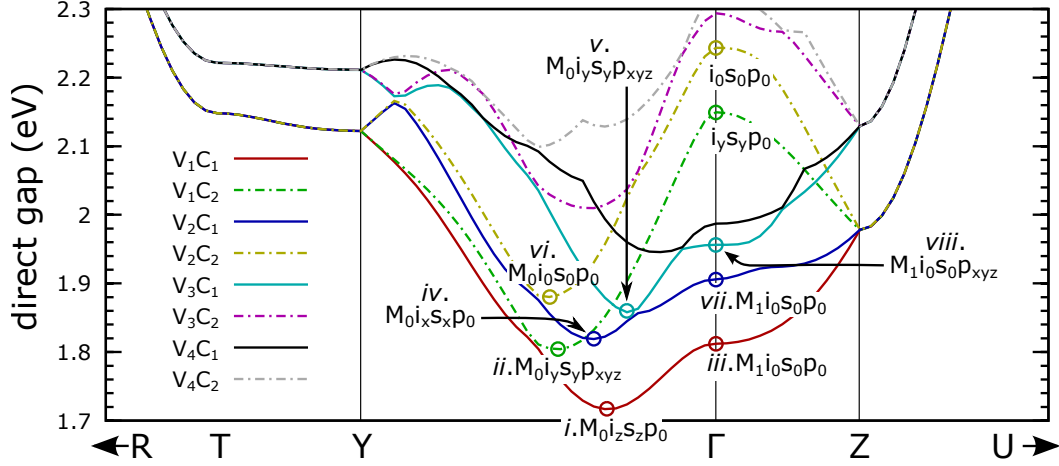
Figure 84: Lowest critical points in the HSE06 band structure. Neglects spin-orbit interaction. The topmost valence bands (V_1 descending through V_4) are subtracted from the lowest conduction bands (C_1 increasing through C_4), giving the direct energy gaps near the Γ-point. The lowest critical points (Van Hove singularities) are labelled: these should be associated with strong optical absorption.

group. All the wavefunctions evaluated were described by a single irrep of the point group. For wavevectors on the symmetry line Γ-Y, the $Pnma$ space group was reduced to the group of the wavevector.

Each band in Figure 83 is then labelled with its computed irreducible representation using the notation of Koster[44,288,295,296]. As noted by Cornwell[297], the labelling of the irreps depends on the choice of origin (but the physics is invariant); for the work performed on $CuSbS_2$, the irrep labels differ according to the $a < b < c <$ and $b < a < c$ origin choices. Relabelling the irreps as follows converts from the unconventional to the conventional ($b < a < c$) origin

$$\Gamma_2^- \to \Gamma_4^-, \qquad \Gamma_4^- \to \Gamma_2^-, \qquad \Gamma_4^+ \to \Gamma_2^+, \qquad \Gamma_4 \to \Gamma_2, \qquad \Gamma_2 \to \Gamma_3 \qquad (5.12)$$

Studying the Kohn–Sham wavefunctions and eigenvalues in this manner is somewhat not rigorously justifiable, particularly for excited states or hybrid functionals; while the Kohn–Sham (independent particle) wavefunctions and eigenvalues are not known to possess any direct physical reality (excepting the last eigenvalue, which gives the ionisation energy)[65,298,299] these often make good approximations to (interacting) many-particle wavefunctions and eigenenergies[83]. Further, precisely this usage seems standard practice in the community, where theoretical and experimental absorption spectra are regularly compared, and the symmetries of bands in potential photovoltaic absorbers are carefully scrutinised[259,260].

5.7.3 Selection rules for interband and excitonic transitions

Using the procedure described in Appendix D, the selection rules for the top three GGA valence bands (V_1, V_2 and V_3), and the bottom two conduction bands (C_1 and C_2) are labelled in Figure 84. In the figure, i, s and p denote respective interband and excitonic transitions to s- and p- hydrogenic states, whilst the subscript 0 or {x,y,z} respectively denotes forbidden transitions or those permitted for x-, y- or z-polarised light.

5.7.4 Comparison with prior results

The determination here of a dipole-allowed direct transition (critical point (i) in Figure 80) some $50\,\mathrm{meV}$ above the fundamental indirect onset, is supported by other theoretical studies evaluating the potential of $CuSbS_2$ for photovoltaic applications. Yu et al.[259] also claimed a dipole-allowed gap $50\,\mathrm{meV}$ above the fundamental indirect gap from HSE06 calculations, but did not elaborate on the approach used to evaluate the band symmetries.

Prior theoretical work on the $CuSbS_2$ band structure is summarised in Table 3. The table lists the magnitude of the direct and indirect band gaps, locations of band extrema and some details of calculations. Approaches such as DFU+U or HSE06 are required as the band gaps are significantly underestimated by the LDA and GGA exchange-correlation functionals. The GW work of Yu et al.[259] is perhaps the most computationally demanding published thus far. All of the work suggests that a direct gap lying above the indirect gap by just 50 to $100\,\mathrm{meV}$. Interestingly, the band structures of Table 3 all suggest that the smallest direct gap is offset from the Γ-point, with the direct gap at Γ some $100\,\mathrm{meV}$ to higher energy.

Notably, excepting the work of Gudelli[300], all the calculations published thus far neglect consideration of the spin-orbit interaction. As discussed, with valence and conduction bands dominated respectively by (Cu-d, S-p) and (Sb-p, S-p) states, this may be expected to be a significant omission; however, the results reported in this thesis suggest otherwise (indicating a shift of just $30\,\mathrm{meV}$).

$E_{g,i}$	$E_{g,d}$	$E_{g,\Gamma}$	VBM	CBM	E_{cut}	k-grid	Method
1.72	1.83	1.86	Γ-Z	U-R	400	$4\times6\times2$	HSE06[257,266]
1.69	1.79	–	–	–	300	$4\times6\times2$	HSE06[273]
1.68	1.73	1.83^b	Γ-Z	Γ-Y	400	$6\times6\times1$	HSE06[256]
1.56	1.61	–	–	–	300	$6\times8\times2$	HSE06+GW[259]
1.47	1.57	–	–	–	–	116	DFT+U[261]
1.19	1.24^b	1.30^b	Γ-Z	Γ-X	–	116	DFT+U[261]
1.05	1.1^b	1.2^b	Γ	Γ-X	–	$19\times31\times8$	TBmBJ,sp[300]
0.9	1.0	–	–	–	300	$4\times6\times2$	GGA[273]
0.79	0.9	1.01	–	–	400	$4\times6\times2$	GGA[257]
0.77	-	–	–	–	–	$19\times31\times8$	GGA,sp[300]

Table 3: Theoretical results for the $CuSbS_2$ band structure. The three leftmost columns are the lowest indirect-, direct- and Γ-point band gaps in eV; the b superscript indicates that the quoted value is measured directly from the band structure. The next columns give the respective reciprocal space positions of the valence and conduction band extrema.

The E_{cut} and k-grid columns give the energy cut-off in eV and the Monkhorst–Pack grid density where given (in the irreducible Brillouin zone). The last column gives the exchange-correlation functional used: HSE06= Heyd-Scuseria-Ernzerhof; GGA= Generalized gradient approximation of Perdew, Burke and Ernzerhof; TBmBJ= GGA/Tran-Blaha modified Becke-Johnson; sp indicates that the Hamiltonian included spin-orbit coupling.

5.8 Conclusions and future work

This chapter has detailed optical, structural and first-principles studies on the potential photovoltaic absorber $CuSbS_2$. The polycrystalline material was determined to be phase-pure in space group $Pnma$, with lattice parameters $\{a, b, c\}$ of 6.0210(23), 3.793(4) and 14.496(6) Å broadly consistent with prior studies. Direct band gaps of 1.69 and 1.61 eV were seen respectively at 4.2 and 300 K, with a redshift of 92 meV found between 4.2 and 350 K, leading to a typical band gap temperature coefficient dE_g/dT of $-0.26\,\text{meV}\,\text{K}^{-1}$. Strong absorption of $9 \times 10^4\,\text{cm}^{-1}$ was seen at 2 eV. The temperature-dependence was fitted with Varshni and Bose–Einstein relations, suggesting a average phonon energy of 11 meV (although note that parameter freedom in one fit reduced the Bose–Einstein E_0 parameter to 1.15 eV).

A potential excitonic feature was observed at 1.82 eV at low temperatures with a 60 meV FWHM. This feature showed a characteristic excitonic line-shape and a typical semiconductor band temperature-dependence. Developing the exciton hypothesis, a binding energy of 34 meV and excitonic Bohr radius of 1.5 nm were determined. Analysis of the irreducible representations and selection rules for excitonic transitions via the GGA and HSE06 band structures suggest distinctly enhanced absorption in the vicinity of $E_g + 0.1$ eV. This work immediately suggests further interesting experiments on $CuSbS_2$. In particular, can the precise nature of the interband transitions and the excitonic feature at 1.82 eV be established?

The unexpected observation of excitonic features in the polycrystalline material investigated here, synthesised for envisaged photovoltaic (PV) applications, may actually be a desirable material characteristic for photovoltaic absorbers: as the electron-hole interaction significantly enhances the absorption onset even when no discrete excitonic features are observed, i.e. at room temperature[44]. As the electron-hole interaction is neglected in nearly all first-principles calculations, this work may contribute to explaining the impressive absorption strength seen in $CuSbS_2$, and perhaps cautions against relying on first-principles methods in isolation.

Because the orthorhombic crystal shows biaxial optical behaviour, i.e. electronic and optical properties which differ in all three spacial directions[301], further optical investigations with oriented single crystals and light polarised parallel and perpendicular to the crystallographic principal axes would be of great interest. Not only do the critical points of Figure 84 suggest four different interband transitions within 200 meV which may be discriminated as a function of light polarisation, but the excitonic feature at 1.82 eV should be easier to resolve under such experimental conditions. The feature has a good lineshape at 80 K, and so should be conveniently accessible at liquid nitrogen temperature.

If the optical absorption coefficient is found to change significantly with crystallite orientation, steps may be feasible at synthesis-time to promote growth of material with orientations favourable for photovoltaic applications. Much work remains.

Appendix A

Energy density of a Boson gas

The density of states $N(k)$ for waves of wavevector k in a region of space with volume V is

$$\frac{\mathrm{d}N}{\mathrm{d}k} = \frac{4\pi k^2 V}{(2\pi)^3} = \frac{k^2 V}{2\pi^2} \tag{A.1}$$

Using the dispersion relation for light $c = w/k$ and the Chain rule with $\mathrm{d}\omega/\mathrm{d}k = c$, equation (A.1) may be written

$$\frac{\mathrm{d}N}{\mathrm{d}\omega} = \frac{\mathrm{d}N}{\mathrm{d}k}\frac{\mathrm{d}k}{\mathrm{d}\omega} = \frac{w^2 V}{2\pi^2 c^3} \tag{A.2}$$

which can likewise be written in terms of photon energy E via $E = hw/2\pi$, giving

$$\frac{\mathrm{d}N}{\mathrm{d}E} = \frac{\mathrm{d}N}{\mathrm{d}\omega}\frac{\mathrm{d}\omega}{\mathrm{d}E} = \frac{4\pi E^2 V}{h^3 c^3} \tag{A.3}$$

The internal energy $\mathrm{d}U$ of a boson gas in the interval between the photon energy E and $E + \mathrm{d}E$ is the product of the density of states in this interval, the photon energy, and the average occupancy of each mode given by the Bose–Einstein distribution. Thus

$$dU = \frac{\mathrm{d}N}{\mathrm{d}E} \cdot \frac{2E}{e^{E/kT} - 1}\,\mathrm{d}E \tag{A.4}$$

where a factor of 2 is introduced as the density of states (A.1) omits the two tranverse polarisation modes of light. Expanding with (A.3) and differentiating by volume gives the gradient of the energy density u of a boson gas (the internal energy per unit volume) with respect to photon energy[14]

$$\frac{du}{dE} = \frac{8\pi}{h^3 c^3} \cdot \frac{E^3}{e^{E/kT} - 1} \qquad [\mathrm{m}^{-3}] \tag{A.5}$$

Alternatively this is the internal energy of the gas in the photon energy interval E to $E + \mathrm{d}E$, per unit volume, per photon energy step $\mathrm{d}E$. Equation (A.5) is called the Planck radiation (or blackbody) function.

Appendix B

Lattice heat capacity and internal energy

This section revisits the Debye and Einstein heat capacities (used in Chapter 4 for fitting thermal expansion). The approach follows Kittel[46] and Eisberg & Resnick[14]. The heat capacity C_V of a solid is the temperature gradient of internal energy U under constant volume V

$$C_V = \left(\frac{\partial U}{\partial T} \right)_V \tag{B.1}$$

The internal energy U includes contributions from electrons and phonons. Neglecting the electron contributions, the internal energy U_{lattice} due to phonon contributions may be written as a sum over the energies $U_{\mathbf{k},p}$ of all phonon modes[46]

$$U_{\text{lattice}} = \sum_{\mathbf{k},p} U_{\mathbf{k},p} = \sum_{\mathbf{k},p} \langle n(\mathbf{k},p) \rangle \, \hbar\omega_{\mathbf{k},p} \tag{B.2}$$

where $\mathbf{k}$ and p respectively label the wavevector and branch of each mode, while $\hbar\omega_{\mathbf{k},p}$ and $\langle n(\mathbf{k},p) \rangle$ give the energy and average quantum number for each state in thermal equilibrium. Phonons are bosons so follow the Planck distribution in thermal equilibrium.

B.1 Planck Distribution

Consider a hypothetical set of identical quantum harmonic oscillators, each with a quantum number n (an integer). Different microstates arise by distributing energy between the oscillators in units of $\hbar\omega$. Assuming that all microstates are equally likely[14], the Boltzmann factor $A\exp(-n\hbar\omega/k_{\text{B}}T)$ gives the average number of oscillators with quantum number n, and thus a number fraction $N(n)$

$$N(n) = \mathrm{e}^{-\frac{n\,\hbar\omega}{k_{\text{B}}T}} \Big/ \sum_{m=0}^{\infty} \mathrm{e}^{-\frac{m\,\hbar\omega}{k_{\text{B}}T}} \tag{B.3}$$

With this fraction (B.3) the average quantum number $\langle n \rangle$ is

$$\langle n \rangle = \sum_{n=0}^{\infty} n\,\mathrm{e}^{-\frac{n\,\hbar\omega}{k_{\text{B}}T}} \Big/ \sum_{m=0}^{\infty} \mathrm{e}^{-\frac{m\,\hbar\omega}{k_{\text{B}}T}} \tag{B.4}$$

The summations in (B.4) may be simplified using relations in geometric series (with $r < 1$)

$$\sum_{n}^{\infty} r^n = \frac{1}{1-r} \qquad \text{and} \qquad \sum_{n}^{\infty} n\,r^n = r\,\frac{\mathrm{d}}{\mathrm{d}r}\sum_{n=1}^{\infty} r^n = \frac{r}{(1-r)^2} \tag{B.5}$$

Writing the average quantum number (B.4) with (B.5) gives the Planck distribution

$$\langle n \rangle = \frac{e^{-\frac{\hbar\omega}{k_B T}}}{1 - e^{-\frac{\hbar\omega}{k_B T}}} = \frac{1}{e^{\frac{\hbar\omega}{k_B T}} - 1} \tag{B.6}$$

which is the expected quantum number for an oscillator in thermal equilibrium.

B.2 Debye Heat Capacity

The internal energy due to phonons (B.2) may then be written with (B.6) generally as[46]

$$U_{\text{lattice}} = \sum_{\mathbf{k},p} \frac{\hbar\omega_{\mathbf{k},p}}{e^{\frac{\hbar\omega_{\mathbf{k},p}}{k_B T}} - 1} \quad \longrightarrow \quad \sum_p \int_0^\infty \mathrm{d}\omega \, D_p(\omega) \frac{\hbar\omega}{e^{\frac{\hbar\omega}{k_B T}} - 1} \tag{B.7}$$

where in the right-hand side, the expression is rewritten so that the density of states $D_p(\omega)$ counts the numbers of states in phonon branch p with energy $\hbar\omega$. The heat capacity results by differentiating (B.7) with respect to temperature

$$C_V = \sum_p \int_0^\infty \mathrm{d}\omega \, D_p(\omega) \frac{\hbar\omega}{\left(e^{\frac{\hbar\omega}{k_B T}} - 1\right)^2} \cdot e^{\frac{\hbar\omega}{k_B T}} \cdot \frac{\hbar\omega}{k_B T^2} = k_B \sum_p \int_0^\infty \mathrm{d}\omega \, D_p(\omega) \frac{x^2 e^x}{(e^x - 1)^2} \tag{B.8}$$

where $x = \hbar\omega/k_B T$. The number of states N up to a wavevector of magnitude k in a crystal of volume V is

$$N(k) = \frac{4}{3}\pi k^3 \cdot \frac{V}{(2\pi)^3} = \frac{k^3 V}{6\pi^2} \tag{B.9}$$

The density of states is thus

$$D(k) = \frac{\mathrm{d}N}{\mathrm{d}k} = \frac{k^2 V}{2\pi^2} \tag{B.10}$$

Debye used a dispersion relation $v_s = \omega/k$, with angular frequency ω and wavevector k related by a constant sound velocity v_s. Using this dispersion relation to write equations (B.9) and (B.10) in frequency gives a density of states $D(\omega)$ and total number of states $N(\omega)$

$$D(\omega) = \frac{w^2 V}{2\pi^2 v_s^3} \qquad \text{and} \qquad N(\omega) = \frac{w^3 V}{6\pi^2 v_s^3} \tag{B.11}$$

Debye considered only acoustic phonon modes in a monatomic lattice; setting the total number of modes $N(\omega)$ equal to the number of cells N_{cells} gives a maximum frequency ω_{max}

$$\omega_{\text{max}}^3 = 6\pi^2 v_s^3 \frac{N_{\text{cells}}}{V} \tag{B.12}$$

Inserting the Debye density of states (B.11) and cut-off frequency (B.12) into the general expression for the internal energy due to phonons (B.7) gives

$$U_{\text{lattice}} = \sum_p \int_0^{\omega_{\text{max}}} \frac{w^2 V}{2\pi^2 v_s^3} \cdot \frac{\hbar\omega \, \mathrm{d}\omega}{e^{\frac{\hbar\omega}{k_B T}} - 1} \longrightarrow \frac{3\hbar V}{2\pi^2 v_s^3} \int_0^{\omega_{\text{max}}} \frac{w^3 \, \mathrm{d}\omega}{e^{\frac{\hbar\omega}{k_B T}} - 1} = \frac{9 N_{\text{cell}} k_B T^4}{\theta_D^3} \int_0^{\theta_D/T} \frac{x^3 \, \mathrm{d}x}{e^x - 1} \tag{B.13}$$

where the substitution $x = \hbar\omega/k_B T$ and the Debye temperature $\theta_D = \hbar\omega_{\text{max}}/k_B$ are used, and the factor 3 appears as the three acoustic phonon branches are assumed to share the same

density of states $D(\omega)$. Inserting (B.11) and (B.12) into the expression for the lattice heat capacity (B.8) gives

$$C_V = \frac{3\hbar^2 V}{2\pi^2 v_s^3 \, k_B T^2} \int_0^{\omega_{\text{max}}} \mathrm{d}\omega \; \frac{\omega^4 \mathrm{e}^{\frac{\hbar\omega}{k_B T}}}{\left(\mathrm{e}^{\frac{\hbar\omega}{k_B T}} - 1\right)^2} = 9\, k_B \, N_{\text{cells}} \left(\frac{T}{\theta_D}\right)^3 \int_0^{\theta_D/T} \frac{x^4 \mathrm{e}^x \, \mathrm{d}x}{(\mathrm{e}^x - 1)^2} \qquad \text{(B.14)}$$

B.3   Einstein Heat Capacity

Einstein considered a system of N quantum oscillators with a fixed natural frequency $\hbar\omega_0$, with a 3D density of states[46]

$$D(\omega) = 3N\, \delta(\omega - \omega_0) \qquad \text{(B.15)}$$

Using this density of states in the general expression for internal energy (B.7) gives

$$U_{\text{lattice}} = \frac{3N\, \hbar\omega_0}{\mathrm{e}^{\frac{\hbar\omega_0}{k_B T}} - 1} \qquad \text{(B.16)}$$

Differentiating (B.16) gives the Einstein heat capacity

$$C_V = \left(\frac{\partial U}{\partial T}\right)_V = 3N k_B \left(\frac{\hbar\omega_0}{k_B T}\right)^2 \frac{\mathrm{e}^{\frac{\hbar\omega_0}{k_B T}}}{\left(\mathrm{e}^{\frac{\hbar\omega_0}{k_B T}} - 1\right)^2} \qquad \text{(B.17)}$$

The Einstein heat capacity is often used to model optical phonons. Whilst a monatomic lattice has three phonon modes (one longitudinal and two transverse), a lattice with a Q-atom basis has an additional $3(Q - 1)$ optical phonon modes, which become important at higher temperatures[46]. Hence when using both Debye and Einstein models in systems with $Q > 1$, the Einstein model is weighted by a factor $(Q - 1)$.

Appendix C

Semiconductor band gap temperature dependence

The temperature-dependence of a semiconductor's band gap results from contributions from thermal expansion of the lattice, as well as effects from the electron–phonon interaction, written generally as[302]

$$\left(\frac{\mathrm{d}E_{\mathrm{g}}}{\mathrm{d}T}\right)_P = \left(\frac{\partial E_{\mathrm{g}}}{\partial T}\right)_{TE} + \left(\frac{\partial E_{\mathrm{g}}}{\partial T}\right)_{EP} \tag{C.1}$$

where on the left-hand side $(\mathrm{d}E_{\mathrm{g}}/\mathrm{d}T)_P$ denotes the isobaric band gap change as seen for example in temperature-dependent optical measurements, while the terms on the right-hand side require further explanation. The thermal expansion term may be written as[302]

$$\left(\frac{\partial E_{\mathrm{g}}}{\partial T}\right)_{TE} = \left(\frac{\partial E_{\mathrm{g}}}{\partial V}\right)_P \left(\frac{\partial V}{\partial T}\right)_P \tag{C.2}$$

where $(\mathrm{d}V/\mathrm{d}T)_P$ is the isobaric thermal (volume) expansion, as determined for example in x-ray diffraction. Alternatively, the chain rule allows the thermal expansion (C.2) to be formulated for pressure-dependent work

$$\left(\frac{\partial E_{\mathrm{g}}}{\partial T}\right)_{TE} = \left(\frac{\partial E_{\mathrm{g}}}{\partial p}\right)_T \left(\frac{\partial p}{\partial V}\right)_T \left(\frac{\partial V}{\partial T}\right)_P \tag{C.3}$$

which is often written in terms of the bulk modulus B and thermal (volume) expansion coefficient α_V

$$\left(\frac{\partial E_{\mathrm{g}}}{\partial T}\right)_{TE} = -\alpha_V B \left(\frac{\partial E_{\mathrm{g}}}{\partial p}\right)_T \tag{C.4}$$

The second term in (C.1) arises from the treatment of the electron–phonon interaction in perturbation theory, which creates further contributions (so-called renormalisations) to the band gap temperature dependence[302,303]

$$\left(\frac{\partial E_{\mathrm{g}}}{\partial T}\right)_{EP} = \frac{\partial}{\partial T} \sum_{b,\mathbf{q}}^{\mathrm{B.Z.}} \frac{\partial E_{\mathrm{g}}}{\partial n_{b,\mathbf{q}}} \left(n_{b,\mathbf{q}} + \frac{1}{2}\right) \tag{C.5}$$

where the sum is taken over the entire Brillouin zone and $n_{b,\mathbf{q}}$ is the Bose–Einstein factor for the phonon mode with energy $E_{b,\mathbf{q}}$ and wavevector $\mathbf{q}$ in branch b

$$n_{b,\mathbf{q}} = \left[\exp\left(\frac{E_{b,\mathbf{q}}}{k_{\mathrm{B}}T} \right) - 1 \right]^{-1} \tag{C.6}$$

The term $\partial E_{\mathrm{g}}/\partial n_{b,\mathbf{q}}$ in equation (C.5) includes contributions from two dominant electron–phonon processes. Phonon emission followed by successive absorption (and vice versa) are considered in the self-energy (SE or Fan[304]) process; while the Debye–Waller (DW) diagram treats the simultaneous emission and absorption of two phonons[305]. These interactions are considered in perturbation theory[98] to respectively second and first orders, leading to a perturbed energy for an electron state $|n\mathbf{k}\rangle$

$$E_{n\mathbf{k}} = E_{n\mathbf{k}}^{(0)} + \langle n\mathbf{k}|H_{\mathrm{DW}}|n\mathbf{k}\rangle + \sum_{m,\,\mathbf{k}'}\sum_{b,\,\mathbf{q}} \frac{|\langle m\mathbf{k}'|H_{\mathrm{SE}}|n\mathbf{k}\rangle|^2}{E_{n\mathbf{k}}^{(0)} - \left(E_{m\mathbf{k}'}^{(0)} \pm \hbar\omega_{\mathbf{q},b} \right) - i\delta} \tag{C.7}$$

where the first sum runs over all intermediate electron states $|m\mathbf{k}'\rangle$, with the unperturbed energies for states $|n\mathbf{k}\rangle$ and $|m\mathbf{k}'\rangle$ denoted respectively by $E_{n\mathbf{k}}^{(0)}$ and $E_{m\mathbf{k}'}^{(0)}$, while the second sum runs over all phonon branches b and wavevectors $\mathbf{q}$ under the constraint of conservation of momentum, $\mathbf{k} = \mathbf{q}+\mathbf{k}'$. If the real part of the denominator in (C.7) should vanish, e.g. for transitions to nearby intraband states, the infinitesimal δ leads to an attenuated wavefunction[305]. Otherwise, the imaginary part is essentially negligible and the real part leads to a band gap change. For example, the interaction of states at the conduction band minimum with higher levels or those in the valence band leads respectively to negative or positive contributions to the band gap (due to the energy denominator). A similar analysis holds for electrons in the valence band, with the band gap typically taking a negative temperature dependence due to the enhanced availability of intermediate states in the conduction band.

Each of the DW and SE terms is second degree in the root-mean square atomic displacements $\sqrt{\langle x^2\rangle}$, which leads to a proportionality to the Bose–Einstein factor[306,307]

$$\langle x^2\rangle = \frac{\hbar}{2\,M\,\omega_{b,\mathbf{q}}} \left(2n_{b,\mathbf{q}} + 1 \right) \tag{C.8}$$

where M and $\omega_{b,\mathbf{q}}$ are respectively the oscillator's atomic mass and angular frequency. In equation (C.5), the derivative $\partial E_{\mathrm{g}}/\partial n_{b,\mathbf{q}}$ serves to divide out the Bose–Einstein factors from (C.8), so that the Bose–Einstein dependence is instead explicit in (C.5).

In principle, if the matrix elements of the electron–phonon interaction are known, e.g. from pseudopotential[306] or tight-binding considerations[302], then any band gap renormalisation due to the electron–phonon interaction may be calculated from (C.5). However, such calculations are particularly involved; instead, workers typically examine the temperature-dependence of band gaps by fitting experimentally determined band gaps $E_{\mathrm{g}}(T)$ to the empirical Varshni relation[247]

$$E_{\mathrm{g}}^{\mathrm{Varshni}} = E_0 - \frac{\alpha T^2}{T + \beta} \tag{C.9}$$

and to models comprising one or more Bose-Einstein oscillators[246,248,249]

$$E_{\mathrm{g}}^{BE} = E_0 + \sum_i \alpha_i \left[1 + \frac{2}{\exp\left(\frac{E_i}{k_{\mathrm{B}}T}\right) - 1} \right] \tag{C.10}$$

where the sum in $(C.10)$ counts over the number of oscillators, E_i and α_i are respectively an averaged phonon energy and a value related to the strength of the electron–phonon interaction for oscillator i, and E_0 is the zero Kelvin band gap. The term in brackets is occasionally seen without the unity value which accounts for the zero-point energy. Physically, the sign of the electron-phonon interaction energy α_i corresponds to the net attraction or repulsion of the band edges due to an average over all displacements of ions involved in all phonon modes modelled by the Bose–Einstein oscillator i. See the discussion below Equation (C.7).

Appendix D

Selection rules for electric dipole transitions

With the irreducible representations (irreps) of an operator and those of the initial and final states, group theory allows statements to be made about the values of certain integrals. In particular, we are interested in the matrix element for interband (electric dipole) transitions

$$\mathbf{p}_{mn\mathbf{k}} = \int_{\text{space}} d\mathbf{r}\, \psi_{m\mathbf{k}}^{\dagger}(\mathbf{r})\, P\, \psi_{n\mathbf{k}}(\mathbf{r}) \tag{D.1}$$

Expanding the integrand in irreducible representations of the space group, the integral may be shown to vanish[294] (due to the orthogonality of the basis functions) unless the decomposition of the direct product of the integrand contains the completely symmetric irrep Γ_1^+

$$\Gamma_{\text{equiv}} = \Gamma_{\text{final state}} \otimes \Gamma_{\text{operator}} \otimes \Gamma_{\text{initial state}} = \Gamma_1^+ \oplus \Gamma_a \oplus \Gamma_b \oplus \cdots \tag{D.2}$$

where Γ_a and Γ_b denote some other irreps associated with the integral. For example, considering the irreps of the first valence (V_1) and first and second conduction bands (C_1 and C_2) at the Γ-point in Figure 83, respectively Γ_3^-, Γ_2^- and Γ_4^+. The electric dipole operator transforms with the irrep of a vector, which in the mmm point group is $(\Gamma_4^-, \Gamma_2^-, \Gamma_3^-)$ (where the notation used is due to Koster[44,295]). This gives the direct products (and selection rules)

$$\Gamma_{C_1} \otimes \Gamma_{\text{dipole}} \otimes \Gamma_{V_1} = \Gamma_2^- \otimes (\Gamma_4^-, \Gamma_2^-, \Gamma_3^-) \otimes \Gamma_3^- = (\Gamma_1^-, \Gamma_3^-, \Gamma_2^-) \rightarrow \text{forbidden} \tag{D.3}$$

$$\Gamma_{C_2} \otimes \Gamma_{\text{dipole}} \otimes \Gamma_{V_1} = \Gamma_4^+ \otimes (\Gamma_4^-, \Gamma_2^-, \Gamma_3^-) \otimes \Gamma_3^- = (\Gamma_3^+, \Gamma_1^+, \Gamma_4^+) \rightarrow \text{allowed-y} \tag{D.4}$$

where we have simplified using the multiplication table for the mmm point group and the presence of the Γ_1^+ irrep in the y-component of (D.4) permits transitions only for y-polarised light. For excitons, the equivalent representation of the matrix element is[114,308]

$$\Gamma_{\text{equiv}} = \Gamma_{\text{exciton, final state}} \otimes \Gamma_{\text{electric dipole}} \otimes \Gamma_{\text{initial}}$$

$$= \Gamma_{\text{conduction band}} \otimes \Gamma_{\text{orbital}} \otimes \Gamma_{\text{valence band}} \otimes \Gamma_{\text{dipole}} \otimes \Gamma_1^+ \tag{D.5}$$

where the initial state of the crystal Γ_{initial} is taken as the symmetric irrep Γ_1^+, and Γ_{orbital} describes the functional form of the excitonic orbital: either a vector-like irrep for p-type wavefunctions (as for the dipole operator) or the completely symmetric irrep (Γ_1^+) for s-type wavefunctions.

Bibliography

[1] Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.). Climate change 2014: Synthesis report, summary for policymakers. contribution of working groups I, II and III to the Fifth Assessment Report of the IPCC. Tech. Rep., The Intergovernmental Panel on Climate Change (IPCC) (2014). URL `http://ar5-syr.ipcc.ch/`.

[2] Davenport, C. Nations approve landmark climate accord in Paris. *The New York Times* (13th Dec 2015).

[3] Hemingway, J. & Waters, L. National statistics, energy trends: renewables. Tech. Rep., UK Department of Energy & Climate Change (2016). URL `https://www.gov.uk/government/statistics/energy-trends-section-6-renewables`.

[4] Clark, P. Britain leads charge in renewables. *The Financial Times* (3rd May 2016).

[5] Press Association. UK energy from coal hits zero for first time in over 100 years. *The Guardian* (13th May 2016).

[6] Neslen, A. Portugal runs for four days straight on renewable energy alone. *The Guardian* (18th May 2016).

[7] Watts, J. Uruguay makes dramatic shift to nearly 95% electricity from clean energy. *The Guardian* (3rd Dec 2015).

[8] Neslen, A. Renewable energy smashes global records in 2015, report shows. *The Guardian* (1st Jun 2016).

[9] Renewable Energy Division. Technology roadmap: Solar photovoltaic energy. Tech. Rep., International Energy Agency (2014). URL `http://bit.ly/28Ou6mz`.

[10] UK Dept. Energy & Climate Change. Digest of United Kingdom energy statistics (DUKES) 2015, Chapter 6: Renewable sources of energy. Tech. Rep., UK Department of Energy & Climate Change (2015). URL `http://bit.ly/1NQzR2s`.

[11] Nelson, J. *The Physics of Solar Cells* (Imperial College Press, 2004), 1st edn. ISBN 9781860943492.

[12] Würfel, P. *Physics of Solar Cells: From Principles to New Concepts*, chap. 4, 4.1–4.29 (Wiley, 2005), 1st edn. ISBN 9783527618545.

[13] Soga, T. *Nanostructured Materials for Solar Energy Conversion*, chap. 1, 3–43 (Elsevier, 2006), 1st edn. ISBN 9780444528445.

[14] Eisberg, R. & Resnick, R. *Quantum Physics of Atoms, Molecules, Solids, Nuclei, and Particles* (Wiley, 1985), 2nd edn. ISBN 9780471873730.

[15] Araújo, G. L. & Martí, A. Absolute limiting efficiencies for photovoltaic energy conversion. *Solar Energy Materials and Solar Cells* **33**, 213–240 (1994). doi: 10.1016/0927-0248(94)90209-7.

[16] Luque, A. & Martí, A. *Handbook of Photovoltaic Science and Engineering*, chap. 5, 130–168 (Wiley, 2011), 2nd edn. ISBN 9780470974704.

[17] Shockley, W. & Queisser, H. J. Detailed balance limit of efficiency of p-n junction solar cells. *Journal of Applied Physics* **32**, 510–519 (1961). doi: 10.1063/1.1736034.

[18] Saga, T. Advances in crystalline silicon solar cell technology for industrial mass production. *NPG Asia Materials* **2**, 96–102 (2010). doi: 10.1038/asiamat.2010.82.

[19] Chirilă, A. *et al.* Highly efficient Cu(In,Ga)Se$_2$ solar cells grown on flexible polymer films. *Nature Materials* **10**, 857–861 (2011). doi: 10.1038/nmat3122.

[20] Hsu, C.-P. S. *Handbook of Instrumental Techniques for Analytical Chemistry*, chap. 15, 247 (Prentice Hall, 1997), 1st edn. ISBN 0131773380.

[21] Herres, W. & Gronholz, J. Understanding FT-IR data processing: Parts 1-3. *Computer Applications in the Laboratory* **4**, 216 (1984).

[22] Jaggi, N. & Vij, D. R. *Handbook of Applied Solid State Spectroscopy*, chap. 9, 411–447 (Springer, 2006), 1st edn. ISBN 0387324976.

[23] Griffiths, P. R. & de Haseth, J. A. *Fourier Transform Infrared Spectrometry* (Wiley, 2007), 2nd edn. ISBN 9780471194040.

[24] Connes, J., Connes, P. & Maillard, J.-P. Atlas des spectres dans le proche infrarouge de Venus, Mars, Jupiter et Saturne. *Editions du Centre Nationale de la Recherche Scientifique* 471 (1969).

[25] Sablinskas, V. *Handbook of Spectroscopy*, chap. 4, 50–54 (Wiley, 2003), 1st edn. ISBN 3527605029.

[26] Jacquinot, P. In *Proceedings of the 17th Congress du GAMS, Paris* (1954).

[27] Fellgett, P. I. General principles of the new methods in interference spectroscopy - on the theory of interference multiplex spectrometer. *J. Phys. Radium* **19**, 187–191 (1958). doi: 10.1051/jphysrad:01958001903018700.

[28] Kauppinen, J. & Partanen, J. *Fourier Transforms in Spectroscopy* (Wiley, 2001), 1st edn. ISBN 3527600299.

[29] Wartewig, S. *IR and Raman Spectroscopy: Fundamental Processing*, chap. 5, 47 (Wiley, 2006), 1st edn. ISBN 9783527606436.

[30] Harris, F. J. On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE* **66**, 51–83 (1978). doi: 10.1109/PROC.1978.10837.

[31] Shimadzu Corp. Fourier transform and apodization. Tech. Rep., Shimadzu Corp. (2012). URL http://www.shimadzu.com/an/ftir/support/tips/letter15/apodization.html.

[32] Mertz, L. Auxiliary computation for Fourier spectrometry. *Infrared Physics* **7**, 17–23 (1967). doi: 10.1016/0020-0891(67)90026-7.

[33] Shimadzu Corp. Tips for FTIR analysis - resolution and aperture. Tech. Rep., Shimadzu Corp. (2015). URL http://www.shimadzu.com/an/ftir/support/tips/letter8/tech.html.

[34] Rakić, A. D., Djurišić, A. B., Elazar, J. M. & Majewski, M. L. Optical properties of metallic films for vertical-cavity optoelectronic devices. *Applied Optics* **37**, 5271–5283 (1998). doi: 10.1364/AO.37.005271.

[35] Bruker Corp. VERTEX 70 & 70v FTIR spectrometer. Tech. Rep., Bruker Corp. (2016). URL http://bit.ly/28PBDDS.

[36] Azzam, R. M. A. *Handbook of Optics: Volume I - Geometrical and Physical Optics, Polarized Light, Components and Instruments*, chap. 16, 16.1–16.13 (McGraw-Hill, 2010), 3rd edn. ISBN 9780071498890.

[37] Woollam, J. A. *et al.* Overview of variable angle spectroscopic ellipsometry (VASE), part I: basic theory and typical applications. In *SPIE Proceedings Vol. CR72* (1999).

[38] Tompkins, H. G. & Irene, E. A. *Handbook of Ellipsometry* (Springer, 2005), 1st edn. ISBN 3540222936.

[39] Fujiwara, H. *Spectroscopic Ellipsometry: Principles and Applications* (Wiley, 2007), 1st edn. ISBN 9780470016084.

[40] Woollam, J. A. Application note: Compensators. Tech. Rep., J.A. Woollam Co., Inc. (2009). URL http://www.jawjapan.com/pdf/compensators.pdf.

[41] Born, M. & Wolf, E. *Principles of Optics* (Pergamon Press, 1970), 4th edn. ISBN 9780521784498.

[42] Fujiwara, H., Koh, J., Rovira, P. I. & Collins, R. W. Assessment of effective-medium theories in the analysis of nucleation and microscopic surface roughness evolution for semiconductor thin films. *Physical Review B* **61**, 10832 (2000). doi: 10.1103/PhysRevB.61.10832.

[43] Basu, P. K. *Theory of Optical Processes in Semiconductors: Bulk and Microstructures* (Oxford, 2002), 1st edn. ISBN 9780198526209.

[44] Yu, P. Y. & Cardona, M. *Fundamentals of Semiconductors* (Springer, 2010), 4th edn. ISBN 9783642007101.

[45] Cullity, B. D. & Stock, S. R. *Elements of X-Ray Diffraction* (Prentice Hall, 2001), 3rd edn. ISBN 0201610914.

[46] Kittel, C. *Introduction to Solid State Physics* (Wiley, 2005), 8th edn. ISBN 9780471415268.

[47] Rietveld, H. M. A profile refinement method for nuclear and magnetic structures. *Journal of Applied Crystallography* **2**, 65–71 (1969). doi: 10.1107/S0021889869006558.

[48] Albinati, A. & Willis, B. T. M. *International Tables for Crystallography Volume C: Mathematical, physical and chemical tables*, chap. 8.6, 710–712 (International Union of Crystallography, 2006), 2006 edn. ISBN 9781402019005.

[49] Toby, B. H. & Dreele, R. B. V. GSAS-II: the genesis of a modern open-source all purpose crystallography software package. *Journal of Applied Crystallography* **46**, 544–549 (2013). doi: 10.1107/S0021889813003531.

[50] Boultif, A. & Louër, D. Powder pattern indexing with the dichotomy method. *Journal of Applied Crystallography* **37**, 724–731 (2004). doi: 10.1107/S0021889804014876.

[51] Rodríguez-Carvajal, J. Recent advances in magnetic structure determination by neutron powder diffraction. *Physica B: Condensed Matter* **192**, 55–69 (1993). doi: 10.1016/0921-4526(93)90108-I.

[52] Pawley, G. S. *EDINP*, the Edinburgh powder profile refinement program. *Journal of Applied Crystallography* **13**, 630–633 (1980). doi: 10.1107/S0021889880012964.

[53] Pawley, G. S. Unit-cell refinement from powder diffraction scans. *Journal of Applied Crystallography* **14**, 357–361 (1981). doi: 10.1107/S0021889881009618.

[54] Bail, A. L., Duroy, H. & Fourquet, J. L. Ab-initio structure determination of LiSbWO$_6$ by x-ray powder diffraction. *Materials Research Bulletin* **23**, 447–452 (1988). doi: 10.1016/0025-5408(88)90019-0.

[55] Bail, A. L. Whole powder pattern decomposition methods and applications: A retrospection. *Powder Diffraction* **20**, 316–326 (2005). doi: 10.1154/1.2135315.

[56] XMaS. X-ray source: The collimating mirror. Tech. Rep., XMaS: The UK Materials Science Facility at the ESRF (2016). URL http://bit.ly/28O2jQ4.

[57] Parrish, W., Wilson, A. J. C. & Langford, J. I. *International Tables for Crystallography Volume C: Mathematical, physical and chemical tables*, chap. 5.2, 491–544 (International Union of Crystallography, 2006), 2006 edn. ISBN 9781402019005.

[58] Meyer, M. Crysalispro version 38 powder diffraction features. Tech. Rep., Rigaku Oxford Diffraction (2015). URL `(privatecommunicationfromauthor)`.

[59] van der Heide, P. *X-ray photoelectron spectroscopy: an intoduction to principles and practices* (Wiley), 1st edn. ISBN 9781118062531.

[60] Ley, L., Pollak, R. A., McFeely, F. R., Kowalczyk, S. P. & Shirley, D. A. Total valence-band densities of states of III-V and II-VI compounds from x-ray photoemission spectroscopy. *Physical Review B* **9**, 600–621 (1974). doi: 10.1103/PhysRevB.9.600.

[61] Khursheed, A. *Scanning Electron Microscope Optics and Spectrometers* (World Scientific, 2011), 1st edn. ISBN 9789812836670.

[62] Macková, A., Morton, S. A., Walker, C. G. H. & Volka, K. *Handbook of Spectroscopy*, chap. 13, 567–570 (Wiley, 2003). ISBN 3527297820.

[63] Hafner, B. Scanning electron microscopy primer. Tech. Rep., Characterization Facility, University of Minnesota (2007). URL `http://www.charfac.umn.edu/sem_primer.pdf`.

[64] Hafner, B. Energy dispersive spectroscopy on the SEM: A primer. Tech. Rep., Characterization Facility, University of Minnesota (2007). URL `http://www.charfac.umn.edu/instruments/eds_on_sem_primer.pdf`.

[65] Martin, R. M. *Electronic Structure: Basic Theory and Practical Methods* (Cambridge, 2004), 1st edn. ISBN 9780521782852.

[66] Jones, R. & Gunnarsson, O. The density functional formalism, its applications and prospects. *Reviews of Modern Physics* **61**, 689–746 (1989). doi: 10.1103/RevModPhys.61.689.

[67] Engel, E. & Dreizler, R. M. *Density Functional Theory: An Advanced Course* (Springer, 2011). ISBN 9783642140891.

[68] Hartree, D. R. The wave mechanics of an atom with a non-Coulomb central field. parts I, II and III. *Mathematical Proceedings of the Cambridge Philosophical Society* **24**, 89, 111, 426 (1928). doi: 10.1017/S0305004100011919.

[69] Fock, V. Näherungsmethode zur lösung des quantenmechanischen mehrkörperproblems. *Zeitschrift für Physik* **61**, 126–148 (1930). doi: 10.1007/BF01340294.

[70] Ashcroft, N. W. & Mermin, N. D. *Solid State Physics* (Thomson, 1976), 1st edn. ISBN 0030839939.

[71] Perdew, J. P. & Ruzsinszky, A. Fourteen easy lessons in density functional theory. *International Journal of Quantum Chemistry* **110**, 2801 (2010). doi: 10.1002/qua.22829.

[72] Kohn, W. Nobel lecture: Electronic structure of matter, wave functions and density functionals. *Reviews of Modern Physics* **71**, 1253–1266 (1999). doi: 10.1103/RevModPhys.71.1253.

[73] Thomas, L. The calculation of atomic fields. *Mathematical Proceedings of the Cambridge Philosophical Society* **23**, 542–548 (1927). doi: 10.1017/S0305004100011683.

[74] Fermi, E. Un metodo statistico per la determinazione di alcune priorieta dell'atome. *Rend. Accad. Naz. Lincei* **6**, 602–607 (1927).

[75] Dirac, P. Note on exchange phenomena in the Thomas atom. *Mathematical Proceedings of the Cambridge Philosophical Society* **26**, 376–385 (1930). doi: 10.1017/S0305004100016108.

[76] Hohenberg, P. & Kohn, W. Inhomogeneous electron gas. *Physical Review* **136**, B864–B871 (1964). doi: 10.1103/PhysRev.136.B864.

[77] Kohn, W. & Sham, L. Self-consistent equations including exchange and correlation effects. *Physical Review* **140**, A1133–A1138 (1965). doi: 10.1103/PhysRev.140.A1133.

[78] Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Physical Review Letters* **77**, 3865–3868 (1996). doi: 10.1103/PhysRevLett.77.3865.

[79] Heyd, J., Scuseria, G. E. & Ernzerhof, M. Hybrid functionals based on a screened Coulomb potential. *The Journal of Chemical Physics* **118**, 8207–8215 (2003). doi: 10.1063/1.1564060.

[80] Imada, M., Fujimori, A. & Tokura, Y. Metal-insulator transitions. *Reviews of Modern Physics* 1039–1263 (1998). doi: 10.1103/RevModPhys.70.1039.

[81] Vettier, C., Burlet, P. & Rossat-Mignod, J. Neutron scattering studies of anomalous rare earth compounds. *Journal of Magnetism and Magnetic Materials* **63-64**, 18 (1987). doi: 10.1016/0304-8853(87)90509-9.

[82] Phusittrakool, A. DFT: Basic idea and practical calculations. Tech. Rep., National Electronics and Computer Technology Center, Pathum Thani, Thailand (2007). URL http://www.lsr.nectec.or.th/images/e/e9/DFT_lecture_abinitworkshop.pdf.

[83] Stowasser, R. & Hoffmann, R. What do the Kohn-Sham orbitals and eigenvalues mean? *Journal of the American Chemical Society* **121**, 3414–3420 (1999). doi: 10.1021/ja9826892.

[84] Monkhorst, H. J. & Pack, J. D. Special points for Brillouin-zone integrations. *Physical Review B* **13**, 5188–5192 (1976). doi: 10.1103/PhysRevB.13.5188.

[85] Moreno, J. & Soler, J. M. Optimal meshes for integrals in real- and reciprocal-space unit cells. *Physical Review B* **45**, 13891–13898 (1992). doi: 10.1103/PhysRevB.45.13891.

[86] Setyawan, W. & Curtarolo, S. High-throughput electronic band structure calculations: Challenges and tools. *Computational Materials Science* **49**, 299–312 (2010). doi: 10.1016/j.commatsci.2010.05.010.

[87] Fox, M. *Optical Properties of Solids* (Oxford, 2010), 2nd edn. ISBN 9780199573363.

[88] Pankove, J. I. Absorption edge of impure gallium arsenide. *Physical Review* **140**, A2059–A2065 (1965). doi: 10.1103/PhysRev.140.A2059.

[89] Ibáñez, J. & Cuscó, R. *Semiconductor Research: Experimental Techniques*, chap. 9, 260 (Springer, 2012), 1st edn. ISBN 9783642233517.

[90] McCluskey, M. D. & Haller, E. E. *Dopants and Defects in Semiconductors* (CRC Press, Taylor & Francis, 2012), 1st edn. ISBN 9781439831526.

[91] Jackson, J. D. *Classical Electrodynamics* (Wiley, 1998), 3rd edn. ISBN 9780471309321.

[92] Grant, I. S. & Phillips, W. R. *Electromagnetism* (Wiley, 1990), 2nd edn. ISBN 9781118723357.

[93] Macleod, H. A. *Thin-Film Optical Filters* (Institute of Physics Publishing, 2001), 3rd edn. ISBN 0750306882.

[94] Macleod, A. The mixed poynting vector. Tech. Rep., The Society of Vacuum Coaters (2014). URL http://www.svc.org/DigitalLibrary/documents/2014_Summer_AMacleod.pdf.

[95] Byrnes, S. J. Multilayer optical calculations. *arXiv Computational Physics* arXiv:1603.02720v1 (2016).

[96] Hamaguchi, C. *Basic Semiconductor Physics* (Springer, 2010), 2nd edn. ISBN 9783642033032.

[97] Bassani, F. & Parravicini, G. P. *Electronic States and Optical Transitions in Solids* (Pergamon Press, 1975), 1st edn. ISBN 9780080168463.

[98] Zettili, N. *Quantum Mechanics* (Wiley, 2009), 2nd edn. ISBN 9780470026793.

[99] Sakurai, J. J. & Napolitano, J. *Modern Quantum Mechanics* (Addison-Wesley, 2011), 2nd edn. ISBN 9780805382914.

[100] Helrich, C. S. *The Classical Theory of Fields: Electromagnetism* (Springer, 2012), 1st edn. ISBN 9783642232046.

[101] Chuang, S. L. *Physics of Optoelectronic Devices* (Wiley, 1995), 1st edn. ISBN 0471109398.

[102] Pankove, J. I. *Optical Processes in Semiconductors* (Dover, 1976), 2nd edn. ISBN 9780486602752.

[103] Burstein, E. Anomalous optical absorption limit in insb. *Physical Review* **93**, 632–633 (1954). doi: 10.1103/PhysRev.93.632.

[104] Moss, T. S. The interpretation of the properties of indium antimonide. *Proceedings of the Physical Society. Section B* **67**, 775 (1954). doi: 10.1088/0370-1301/67/10/306.

[105] Smith, R. A. *Semiconductors* (Cambridge, 1978), 2nd edn. ISBN 0521218241.

[106] Pankove, J. I. & Aigrain, P. Optical absorption of arsenic-doped degenerate germanium. *Physical Review* **126**, 956–962 (1962). doi: 10.1103/PhysRev.126.956.

[107] Singh, J. *Optical Properties of Condensed Matter and Applications* (Wiley, 2006), 1st edn. ISBN 9780470021927.

[108] John, S., Soukoulis, C., Cohen, M. H. & Economou, E. N. Theory of electron band tails and the Urbach optical-absorption edge. *Physical Review Letters* **57**, 1777–1780 (1986). doi: 10.1103/PhysRevLett.57.1777.

[109] Urbach, F. The long-wavelength edge of photographic sensitivity and of the electronic absorption of solids. *Physical Review* **92**, 1324–1324 (1953). doi: 10.1103/PhysRev.92.1324.

[110] Tropf, W. J., Thomas, M. E. & Harris, T. J. *Handbook of Optics: Volume IV - Optical Properties of Materials*, chap. 33, 33.17 (McGraw-Hill, 2010), 2nd edn. ISBN 9780071498920.

[111] Balkanski, M. & Wallis, R. F. *Semiconductor Physics and Applications* (Oxford, 2007), 1st edn. ISBN 9780198517405.

[112] Studenyak, I., Kranjec, M. & Kurik, M. Urbach rule in solid state physics. *International Journal of Optics and Applications* **4**, 76–83 (2014). doi: 10.5923/j.optics.20140403.02.

[113] Johnson, S. R. & Tiedje, T. Temperature dependence of the Urbach edge in GaAs. *Journal of Applied Physics* **78**, 5609 (1995). doi: 10.1063/1.359683.

[114] Pelant, I. & Valenta, J. *Luminescence Spectroscopy of Semiconductors* (Oxford, 2012), 1st edn. ISBN 9780199588336.

[115] Elliott, R. J. Intensity of optical absorption by excitons. *Physical Review* **108**, 1384–1389 (1957). doi: 10.1103/PhysRev.108.1384.

[116] Subashiev, V. K., Le-khac-Binh & Chertkova, L. S. Hyperbolic excitons in GaSe. *Solid State Communications* **9**, 369–372 (1971). doi: 10.1016/0038-1098(71)90257-2.

[117] Welkowsky, M. & Braunstein, R. Interband transitions and exciton effects in semiconductors. *Physical Review B* **5**, 497–509 (1972). doi: 10.1103/PhysRevB.5.497.

[118] Simmons, J. H. & Potter, K. S. *Optical Materials* (Academic Press, 2000), 1st edn. ISBN 9780126441406.

[119] Toyozawa, Y. Theory of line-shapes of the exciton absorption bands. *Progress of Theoretical Physics* **20**, 53–81 (1958). doi: 10.1143/PTP.20.53.

[120] Schreiber, M. & Toyozawa, Y. Numerical experiments on the absorption lineshape of the exciton under lattice vibrations. i. the overall lineshape. *Journal of the Physical Society of Japan* **51**, 1528–1536 (1982). doi: 10.1143/JPSJ.51.1528.

[121] Abelès, F. Sur la propagation des ondes electromagnetiques dans les milieux stratifies. *Annales de Physique* **3**, 504–520 (1948).

[122] Ohta, K. & Ishida, H. Matrix formalism for calculation of electric field intensity of light in stratified multilayered films. *Applied Optics* **29**, 1952–1959 (1990). doi: 10.1364/AO.29.001952.

[123] Heavens, O. S. Optical properties of thin films. *Reports on Progress in Physics* **23**, 1 (1960). doi: 10.1088/0034-4885/23/1/301.

[124] Centurioni, E. Generalized matrix method for calculation of internal light energy flux in mixed coherent and incoherent multilayers. *Applied Optics* **44**, 7532–7539 (2005). doi: 10.1364/AO.44.007532.

[125] Crook, A. The reflection and transmission of light by any system of parallel isotropic films. *Journal of the Optical Society of America* **38**, 954–964 (1948). doi: 10.1364/JOSA.38.000954.

[126] Heavens, O. S. *Optical Properties of Thin Solid Films* (Dover Publications, 1955), 1st edn. ISBN 9780486669243.

[127] Katsidis, C. C. & Siapkas, D. I. General transfer-matrix method for optical multilayer systems with coherent, partially coherent, and incoherent interference. *Applied Optics* **41**, 3978–3987 (2002). doi: 10.1364/AO.41.003978.

[128] Swanepoel, R. Determination of the thickness and optical constants of amorphous silicon. *Journal of Physics E: Scientific Instruments* **16**, 1214 (1983). doi: 10.1088/0022-3735/16/12/023.

[129] Minkov, D. A. Calculation of the optical constants of a thin layer upon a transparent substrate from the reflection spectrum. *Journal of Physics D: Applied Physics* **22**, 1157 (1989). doi: 10.1088/0022-3727/22/8/021.

[130] Ruíz-Pérez, J. J., González-Leal, J. M., Minkov, D. A. & Márquez, E. Method for determining the optical constants of thin dielectric films with variable thickness using only their shrunk reflection spectra. *Journal of Physics D: Applied Physics* **34**, 2489 (2001). doi: 10.1088/0022-3727/34/16/314.

[131] Birgin, E. G., Chambouleyron, I. & Martínez, J. M. Estimation of the optical constants and the thickness of thin films using unconstrained optimization. *Journal of Computational Physics* **151**, 862–880 (1999). doi: 10.1006/jcph.1999.6224.

[132] Chambouleyron, I., Martínez, J. M., Moretti, A. C. & Mulato, M. Retrieval of optical constants and thickness of thin films from transmission spectra. *Applied Optics* **36**, 8238–8247 (1997). doi: 10.1364/AO.36.008238.

[133] Mulato, M., Chambouleyron, I., Birgin, E. G. & Martínez, J. M. Determination of thickness and optical constants of amorphous silicon films from transmittance data. *Applied Physics Letters* **77**, 2133–2135 (2000). doi: 10.1063/1.1314299.

[134] Swanepoel, R. Determination of surface roughness and optical constants of inhomogeneous amorphous silicon films. *Journal of Physics E: Scientific Instruments* **17**, 896 (1984). doi: 10.1088/0022-3735/17/10/023.

[135] Poelman, D. & Smet, P. F. Methods for the determination of the optical constants of thin films from single transmission measurements: a critical review. *Journal of Physics D: Applied Physics* **36**, 1850 (2003). doi: 10.1088/0022-3727/36/15/316.

[136] Paulick, T. C. Inversion of normal-incidence (R,T) measurements to obtain $n + ik$ for thin films. *Applied Optics* **25**, 562–564 (1986). doi: 10.1364/AO.25.000562.

[137] Nilsson, P.-O. Determination of optical constants from intensity measurements at normal incidence. *Applied Optics* **7**, 435–442 (1968). doi: 10.1364/AO.7.000435.

[138] Tomlin, S. G. Optical reflection and transmission formulae for thin films. *Journal of Physics D: Applied Physics* **1**, 1667 (1968). doi: 10.1088/0022-3727/1/12/312.

[139] Vriens, L. & Rippens, W. Optical constants of absorbing thin solid films on a substrate. *Applied Optics* **22**, 4105–4110 (1983). doi: 10.1364/AO.22.004105.

[140] Hjortsberg, A. Determination of optical constants of absorbing materials using transmission and reflection of thin films on partially metallized substrates: analysis of the new (T,Rm) technique. *Applied Optics* **20**, 1254–1263 (1981). doi: 10.1364/AO.20.001254.

[141] Aqili, A. K. S. & Maqsood, A. Determination of thickness, refractive index, and thickness irregularity for semiconductor thin films from transmission spectra. *Applied Optics* **41**, 218–224 (2002). doi: 10.1364/AO.41.000218.

[142] Phillips, R. T. A numerical method for determining the complex refractive index from reflectance and transmittance of supported thin films. *Journal of Physics D: Applied Physics* **16**, 489 (1983). doi: 10.1088/0022-3727/16/4/010.

[143] Manifacier, J. C., Gasiot, J. & Fillard, J. P. A simple method for the determination of the optical constants n, k and the thickness of a weakly absorbing thin film. *Journal of Physics E: Scientific Instruments* **9**, 1002 (1976). doi: 10.1088/0022-3735/9/11/032.

[144] Chiao, S.-C., Bovard, B. G. & Macleod, H. A. Optical-constant calculation over an extended spectral region: application to titanium dioxide film. *Applied Optics* **34**, 7355–7360 (1995). doi: 10.1364/AO.34.007355.

[145] Dorranian, D., Dejam, L. & Mosayebian, G. Optical characterization of Cu_3N thin film with Swanepoel method. *Journal of Theoretical and Applied Physics* **6**, 1–9 (2012). doi: 10.1186/2251-7235-6-13.

[146] Treharne, R. E. *et al.* Optical design and fabrication of fully sputtered CdTe/CdS solar cells. *Journal of Physics: Conference Series* **286**, 012038 (2011). doi: 10.1088/1742-6596/286/1/012038.

[147] Hosseini, S. M. Optical properties of cadmium telluride in zinc-blende and wurzite structure. *Physica B: Condensed Matter* **403**, 1907–1915 (2008). doi: 10.1016/j.physb.2007.10.370.

[148] El-Kadry, N., Ashour, A., Ahmed, M. F. & Abdel-Hady, K. Variation of optical constants of cadmium telluride thin films with deposition conditions. *Thin Solid Films* **259**, 194–202 (1995). doi: 10.1016/0040-6090(94)06416-4.

[149] Jun-feng, H., Liu, X., Li-mei, C., Hamon, J. & Besland, M. P. Investigation of oxide layer on CdTe film surface and its effect on the device performance. *Materials Science in Semiconductor Processing* **40**, 402–406 (2015). doi: 10.1016/j.mssp.2015.06.086.

[150] Bai, Z. & Wang, D. Oxidation of CdTe thin film in air coated with and without a $CdCl_2$ layer. *Physica Status Solidi (a)* **209**, 1982–1987 (2012). doi: 10.1002/pssa.201228107.

[151] Ghorannevis, Z., Akbarnejad, E. & Ghoranneviss, M. Effects of various deposition times and RF powers on CdTe thin film growth using magnetron sputtering. *Journal of Theoretical and Applied Physics* 1–7 (2016). doi: 10.1007/s40094-016-0219-7.

[152] Kwon, D. *Studies of sputtered CdTe and CdSe solar cells*. Ph.D. thesis, The University of Toledo (2012).

[153] Ferrini, R., Patrini, M. & Franchi, S. Optical functions from 0.02 to $6\,eV$ of Al(x)Ga(1−x)Sb/GaSb epitaxial layers. *Journal of Applied Physics* **84**, 4517–4524 (1998). doi: 10.1063/1.368677.

[154] Skauli, T. *et al.* Improved dispersion relations for GaAs and applications to nonlinear optics. *Journal of Applied Physics* **94**, 6447–6455 (2003). doi: 10.1063/1.1621740.

[155] Casey, H. C., Sell, D. D. & Wecht, K. W. Concentration dependence of the absorption coefficient for n- and p-type GaAs between 1.3 and $1.6\,eV$. *Journal of Applied Physics* **46**, 250–257 (1975). doi: 10.1063/1.321330.

[156] Aspnes, D. E., Kelso, S. M., Logan, R. A. & Bhat, R. Optical properties of Al(x)Ga(1−x)As. *Journal of Applied Physics* **60**, 754–767 (1986). doi: 10.1063/1.337426.

[157] Blakemore, J. S. Semiconducting and other major properties of gallium arsenide. *Journal of Applied Physics* **53**, R123–R181 (1982). doi: 10.1063/1.331665.

[158] Asano, M., Umeda, K. & Tasaki, A. Cu_3N thin film for a new light recording media. *Japanese Journal of Applied Physics* **29**, 1985 (1990). doi: 10.1143/JJAP.29.1985.

[159] Maruyama, T. & Morishita, T. Copper nitride and tin nitride thin films for write-once optical recording media. *Applied Physics Letters* **69**, 890–891 (1996). doi: 10.1063/1.117978.

[160] Cremer, R. *et al.* Deposition and characterization of metastable Cu_3N layers for applications in optical data storage. *Microchimica Acta* **133**, 299–302 (2000). doi: 10.1007/s006040070109.

[161] Caskey, C. M., Richards, R. M., Ginley, D. S. & Zakutayev, A. Thin film synthesis and properties of copper nitride, a metastable semiconductor. *Materials Horizons* **1**, 424–430 (2014). doi: 10.1039/C4MH00049H.

[162] Navío, C., Capitán, M. J., Álvarez, J., Yndurain, F. & Miranda, R. Intrinsic surface band bending in $Cu_3N(100)$ ultrathin films. *Physical Review B* **76**, 085105 (2007). doi: 10.1103/PhysRevB.76.085105.

[163] Borsa, D. M., Grachev, S., Presura, C. & Boerma, D. O. Growth and properties of Cu_3N films and Cu_3N/γ'-Fe_4N bilayers. *Applied Physics Letters* **80**, 1823–1825 (2002). doi: 10.1063/1.1459116.

[164] Maya, L. Deposition of crystalline binary nitride films of tin, copper, and nickel by reactive sputtering. *Journal of Vacuum Science & Technology A* **11**, 604–608 (1993). doi: 10.1116/1.578778.

[165] Maya, L. Covalent nitrides for maskless laser writing of microscopic metal lines. *Materials Research Society Symposium Proceedings* **282**, 203 (1992). doi: 10.1557/PROC-282-203.

[166] Nosaka, T., Yoshitake, M., Okamoto, A., Ogawa, S. & Nakayama, Y. Thermal decomposition of copper nitride thin films and dots formation by electron beam writing. *Applied Surface Science* **169-170**, 358–361 (2001). doi: 10.1016/S0169-4332(00)00681-4.

[167] Zakutayev, A. *et al.* Defect tolerant semiconductors for solar energy conversion. *J. Phys. Chem. Lett.* **5**, 1117–1125 (2014). doi: 10.1021/jz5001787.

[168] Green, M. A., Emery, K., Hishikawa, Y., Warta, W. & Dunlop, E. D. Solar cell efficiency tables (version 47). *Progress in Photovoltaics: Research and Applications* **24**, 3–11 (2016). doi: 10.1002/pip.2728.

[169] Chopra, K. L., Paulson, P. D. & Dutta, V. Thin-film solar cells: An overview. *Progress in Photovoltaics: Research and Applications* **12**, 69–92 (2004). doi: 10.1002/pip.541.

[170] Borsa, D. M. & Boerma, D. O. Growth, structural and optical properties of Cu_3N films. *Surface Science* **548**, 95–105 (2004). doi: 10.1016/j.susc.2003.10.053.

[171] Gonzalez-Arrabal, R., Gordillo, N., Martin-Gonzalez, M. S., Ruiz-Bustos, R. & Agulló-López, F. Thermal stability of copper nitride thin films: The role of nitrogen migration. *Journal of Applied Physics* **107**, 103513 (2010). doi: 10.1063/1.3369450.

[172] Juza, R. & Hahn, H. über die kristallstrukturen von Cu_3N, GaN und InN. *Zeitschrift für anorganische und allgemeine Chemie* **239**, 282–287 (1938). doi: 10.1002/zaac.19382390307.

[173] Du, Y., Ji, A., Ma, L., Wang, Y. & Cao, Z. Electrical conductivity and photoreflectance of nanocrystalline copper nitride thin films deposited at low temperature. *Journal of Crystal Growth* **280**, 490–494 (2005). doi: 10.1016/j.jcrysgro.2005.03.077.

[174] Dorranian, D., Dejam, L., Sari, A. H. & Hojabri, A. Structural and optical properties of copper nitride thin films in a reactive Ar/N_2 magnetron sputtering system. *The European Physical Journal Applied Phyics* **50**, 20503 (2010). doi: 10.1051/epjap/2010040.

[175] Odeh, I. M. Fabrication and optical constants of amorphous copper nitride thin films prepared by ion beam assisted dc magnetron reactive sputtering. *Journal of Alloys and Compounds* **454**, 102–105 (2008). doi: 10.1016/j.jallcom.2006.12.020.

[176] Gordillo, N., Gonzalez-Arrabal, R., Álvarez-Herrero, A. & Agulló-López, F. Free-carrier contribution to the optical response of N-rich Cu_3N thin films. *Journal of Physics D: Applied Physics* **42**, 165101 (2009). doi: 10.1088/0022-3727/42/16/165101.

[177] Maruyama, T. & Morishita, T. Copper nitride thin films prepared by radio-frequency reactive sputtering. *Journal of Applied Physics* **78**, 4104–4107 (1995). doi: 10.1063/1.359868.

[178] Ji, A. L. *et al.* Growth of stoichiometric Cu_3N thin films by reactive magnetron sputtering. *Journal of Crystal Growth* **295**, 79–83 (2006). doi: 10.1016/j.jcrysgro.2006.07.007.

[179] Hadian, F., Rahmati, A., Movla, H. & Khaksar, M. Reactive DC magnetron sputter deposited copper nitride nano-crystalline thin films: Growth and characterization. *Vacuum* **86**, 1067–1072 (2012). doi: 10.1016/j.vacuum.2011.09.001.

[180] Reddy, K. V. S., Reddy, A. S., Reddy, P. S. & Uthanna, S. Copper nitride films deposited by dc reactive magnetron sputtering. *Journal of Materials Science: Materials in Electronics* **18**, 1003–1008 (2007). doi: 10.1007/s10854-007-9120-0.

[181] Yuan, X. M., Yan, P. X. & Liu, J. Z. Preparation and characterization of copper nitride films at various nitrogen contents by reactive radio-frequency magnetron sputtering. *Materials Letters* **60**, 1809–1812 (2006). doi: 10.1016/j.matlet.2005.12.028.

[182] Zuli, L. *et al.* Effect of N_2-gas partial pressure on the structure and properties of copper nitride films by dc reactive magnetron sputtering. *Plasma Science and Technology* **9**, 147 (2007). doi: 10.1088/1009-0630/9/2/06.

[183] Du, Y. *et al.* Effect of oxygen inclusion on microstructure and thermal stability of copper nitride thin films. *Journal of Materials Research* **22**, 3052–3057 (2007). doi: 10.1557/JMR.2007.0414.

[184] Oliver, D. J. Electrical properties of n-type gallium arsenide. *Physical Review* **127**, 1045–1052 (1962). doi: 10.1103/PhysRev.127.1045.

[185] Sze, S. M. & Irvin, J. C. Resistivity, mobility and impurity levels in GaAs, Ge, and Si at 300 K. *Solid-State Electronics* **11**, 599–602 (1968). doi: 10.1016/0038-1101(68)90012-9.

[186] Haisty, R. W., Mehal, E. W. & Stratton, R. Preparation and characterization of high resistivity GaAs. *Journal of Physics and Chemistry of Solids* **23**, 829–836 (1962). doi: 10.1016/0022-3697(62)90140-3.

[187] Pierson, J. Structure and properties of copper nitride films formed by reactive magnetron sputtering. *Vacuum* **66**, 59–64 (2002). doi: 10.1016/S0042-207X(01)00425-0.

[188] Blucher, J., Bang, K. & Giessen, B. C. Preparation of the metastable interstitial copper nitride, Cu_4N, by d.c. plasma ion nitriding. *Materials Science and Engineering: A* **117**, L1–L3 (1989). doi: 10.1016/0921-5093(89)90110-X.

[189] Fallberg, A., Ottosson, M. & Carlsson, J.-O. CVD of copper(i) nitride. *Chemical Vapor Deposition* **15**, 300–305 (2009). doi: 10.1002/cvde.200906794.

[190] Kim, K. J., Kim, J. H. & Kang, J. H. Structural and optical characterization of Cu_3N films prepared by reactive RF magnetron sputtering. *Journal of Crystal Growth* **222**, 767–772 (2001). doi: 10.1016/S0022-0248(00)00968-4.

[191] Yue, G. H., Yan, P. X. & Wang, J. Study on the preparation and properties of copper nitride thin films. *Journal of Crystal Growth* **274**, 464–468 (2005). doi: 10.1016/j.jcrysgro.2004.10.032.

[192] Ji, Z., Zhang, Y., Yuan, Y. & Wang, C. Reactive DC magnetron deposition of copper nitride films for write-once optical recording. *Materials Letters* **60**, 3758–3760 (2006). doi: 10.1016/j.matlet.2006.03.107.

[193] Wosylus, A. *et al.* High pressure phase transition and properties of Cu_3N: An experimental and theoretical study. *Z. Anorg. Allg. Chem.* **635**, 1959–1968 (2009). doi: 10.1002/zaac.200900369.

[194] Ghosh, S. *et al.* Effect of substrate temperature on the physical properties of copper nitride films by r.f. reactive sputtering. *Surface and Coatings Technology* **142-144**, 1034–1039 (2001). doi: 10.1016/S0257-8972(01)01091-X.

[195] Chwa, S. O. & Kim, K. H. Adhesion property of copper nitride film to silicon oxide substrate. *Journal of Materials Science Letters* **17**, 1835–1838 (1998). doi: 10.1023/A:1006626124472.

[196] Paniconi, G. *et al.* Structural chemistry of Cu_3N powders obtained by ammonolysis reactions. *Solid State Sciences* **9**, 907–913 (2007). doi: 10.1016/j.solidstatesciences.2007.03.017.

[197] Zachwieja, U. & Jacobs, H. Ammonothermalsynthese von kupfernitrid, Cu_3N. *Journal of the Less Common Metals* **161**, 175–184 (1990). doi: 10.1016/0022-5088(90)90327-G.

[198] Soto, G., Díaz, J. A. & de la Cruz, W. Copper nitride films produced by reactive pulsed laser deposition. *Materials Letters* **57**, 4130–4133 (2003). doi: 10.1016/S0167-577X(03)00277-5.

[199] Törndahl, T. *Atomic Layer Deposition of Copper, Copper(I) Oxide and Copper(I) Nitride on Oxide Substrates.* Ph.D. thesis, Department of Materials Chemistry, Uppsala University (2004).

[200] Soukup, L. *et al.* Copper nitride thin films prepared by the RF plasma chemical reactor with low pressure supersonic single and multi-plasma jet system. *Surface and Coatings Technology* **116-119**, 321–326 (1999). doi: 10.1016/S0257-8972(99)00129-2.

[201] Wang, J. *et al.* Copper nitride (Cu_3N) thin films deposited by RF magnetron sputtering. *Journal of Crystal Growth* **286**, 407–412 (2006). doi: 10.1016/j.jcrysgro.2005.10.107.

[202] Corning Inc. Corning Eagle XG product information sheet. Tech. Rep., Corning Display Technologies (2013). URL http://bit.ly/292BUPK.

[203] Kelly, P. & Arnell, R. Magnetron sputtering: a review of recent developments and applications. *Vacuum* **56**, 159–172 (2000). doi: 10.1016/S0042-207X(99)00189-X.

[204] Bräuer, G., Szyszka, B., Vergöhl, M. & Bandorf, R. Magnetron sputtering - milestones of 30 years. *Vacuum* **84**, 1354–1359 (2010). doi: 10.1016/j.vacuum.2009.12.014.

[205] Kuzmin, A., Kalinko, A., Anspoks, A., Timoshenko, J. & Kalendarev, R. EXAFS study of local structure and lattice dynamics in copper nitride Cu_3N. Tech. Rep., Institute of Solid State Physics, University of Latvia (2012). URL http://photon-science.desy.de/annual_report/files/2012/20122294.pdf.

[206] Deslattes, R. D., Kessler Jr, E. G., Indelicato, P. & Lindroth, E. *International Tables for Crystallography Volume C: Mathematical, physical and chemical tables*, chap. 4.2.2, 200–212 (International Union of Crystallography, 2006), 2006 edn. ISBN 9781402019005.

[207] Lindley, P. F. *International Tables for Crystallography Volume C: Mathematical, physical and chemical tables*, chap. 3.4.1.3.1, 164 (International Union of Crystallography, 2006). ISBN 9781402019005.

[208] de Wolff, P. M. A simplified criterion for the reliability of a powder pattern indexing. *Journal of Applied Crystallography* **1**, 108–113 (1968). doi: 10.1107/S002188986800508X.

[209] Meredith, C. C., Moberly, J. W. & Barlow, M. Integrated x-ray diffraction measurements of beryllium. *Journal of the Less Common Metals* **18**, 423–425 (1969). doi: 10.1016/0022-5088(69)90012-5.

[210] Gordon, P. A high temperature precision x-ray camera: Some measurements of the thermal coefficients of expansion of beryllium. *Journal of Applied Physics* **20**, 908 (1949). doi: 10.1063/1.1698252.

[211] Swenson, C. A. HIP beryllium: Thermal expansivity from 4 to 300 k and heat capacity from 1 to 108 k. *Journal of Applied Physics* **70**, 3046 (1991). doi: 10.1063/1.349336.

[212] Parrish, W. & Langford, J. I. *International Tables for Crystallography Volume C: Mathematical, physical and chemical tables*, chap. 2.3.1.1.6, 42–79 (International Union of Crystallography, 2006), 2006 edn. ISBN 9781402019005.

[213] Sayetat, F., Fertey, P. & Kessler, M. An easy method for the determination of Debye temperature from thermal expansion analyses. *Journal of Applied Crystallography* **31**, 121–127 (1998). doi: 10.1107/S0021889897006936.

[214] Chen, H. *et al.* First principles study of anti-ReO$_3$ type Cu$_3$N and Sc-doped Cu$_3$N on structural, elastic and electronic properties. *Computational and Theoretical Chemistry* **1018**, 71–76 (2013). doi: 10.1016/j.comptc.2013.06.009.

[215] Kong, F., Hu, Y., Wang, Y., Wang, B. & Tang, L. Structural, elastic and thermodynamic properties of anti-ReO$_3$ type Cu$_3$N under pressure from first principles. *Computational Materials Science* **65**, 247–253 (2012). doi: 10.1016/j.commatsci.2012.07.025.

[216] Moreno-Armenta, M. G., Martínez-Ruiz, A. & Takeuchi, N. Ab initio total energy calculations of copper nitride: the effect of lattice parameters and Cu content in the electronic properties. *Solid State Sciences* **6**, 9–14 (2004). doi: 10.1016/j.solidstatesciences.2003.10.014.

[217] Wu, Z. *et al.* Ab initio calculations of the structural, elastic, electronic and optical properties of Cu$_3$N as well as Cu$_3$NLa and Cu$_3$NCe compounds. *Computational Materials Science* **95**, 221–227 (2014). doi: 10.1016/j.commatsci.2014.07.035.

[218] Rahmati, A., Ghoohestani, M., Badehian, H. & Baizaee, M. Ab. initio study of the structural, elastic, electronic and optical properties of Cu$_3$N. *Materials Research* **17**, 303–310 (2014). doi: 10.1590/S1516-14392014005000039.

[219] Zhou, X., Gall, D. & Khare, S. V. Mechanical properties and electronic structure of anti-ReO$_3$ structured cubic nitrides, M$_3$N, of d block transition metals M: An ab initio study. *Journal of Alloys and Compounds* **595**, 80–86 (2014). doi: 10.1016/j.jallcom.2014.01.116.

[220] Yim, W. M. & Paff, R. J. Thermal expansion of AlN, sapphire, and silicon. *Journal of Applied Physics* **45**, 1456–1457 (1974). doi: 10.1063/1.1663432.

[221] Champion, Y., Bernard, F., Millot, N. & Perriat, P. Surface adsorption effects on the lattice expansion of copper nanocrystals. *Applied Physics Letters* **86**, 231914 (2005). doi: 10.1063/1.1947887.

[222] Corp., D. C. Dow corning high vacuum grease, msds 756289-00004. Tech. Rep., Dow Corning Corporation (2015). URL http://www.dowcorning.com/DataFiles/0902770182495fca.pdf.

[223] Alexander, M. R., Mair, F. S., Pritchard, R. G. & Warren, J. E. Mild depolymerization of silicone grease using aluminum(III)chloride: high-yield synthesis and crystal structure of [ClSiMe2OAlCl22], and its controlled hydrolysis on aluminum surfaces. *Applied Organometallic Chemistry* **17**, 730–734 (2003). doi: 10.1002/aoc.500.

[224] Moskalenko, V. A. & Tsvankin, D. Y. X-ray method of determining the crystallinity of polysiloxanes. *Polymer Science USSR* **11**, 429–435 (1969). doi: 10.1016/0032-3950(69)90185-3.

[225] Mehta, A., Tembe, G., Parikh, P. & Mehta, G. Titanasilsesquioxane-alkylaluminum catalyst system for ethylene polymerization. *Modern Research in Catalysis* **1**, 29–42 (2012). doi: 10.4236/mrc.2012.13005.

[226] Vajenine, G. V., Hoch, C., Dinnebier, R. E., Senyshyn, A. & Niewa, R. A temperature-dependent structural study of anti-ReO_3-type Na_3N: to distort or not to distort? *Z. Anorg. Allg. Chem.* **636**, 94–99 (2010). doi: 10.1002/zaac.200900488.

[227] Moreno-Armenta, M. G., Pérez, W. L. & Takeuchi, N. First-principles calculations of the structural and electronic properties of Cu_3MN compounds with M = Ni, Cu, Zn, Pd, Ag, and Cd. *Solid State Sciences* **9**, 166–172 (2007). doi: 10.1016/j.solidstatesciences.2006.12.002.

[228] Moreno-Armenta, M. G., Soto, G. & Takeuchi, N. Ab initio calculations of non-stoichiometric copper nitride, pure and with palladium. *Journal of Alloys and Compounds* **509**, 1471–1476 (2011). doi: 10.1016/j.jallcom.2010.09.192.

[229] Suleiman, M. S. H., Molepo, M. P. & Joubert, D. P. A theoretical investigation of structural, electronic and optical properties of bulk copper nitrides. *arXiv: Materials Science* arXiv:1211.0179 (2012).

[230] Yu, W., Zhao, J. & Jin, C. Simultaneous softening of Cu_3N phonon modes along the T_2 line under pressure: A first-principles calculation. *Physical Review B* **72**, 214116 (2005). doi: 10.1103/PhysRevB.72.214116.

[231] Fioretti, A. N. *et al.* Understanding and control of bipolar doping in copper nitride. *arXiv: Materials Science* arXiv:1601.03362 (2016).

[232] Peng, H. *et al.* Convergence of density and hybrid functional defect calculations for compound semiconductors. *Physical Review B* **88**, 115201 (2013). doi: 10.1103/PhysRevB.88.115201.

[233] Gonze, X. *et al.* ABINIT : First-principles approach of materials and nanosystem properties. *Computer Phys. Comm.* **180**, 2582–2615 (2009). doi: 10.1016/j.cpc.2009.07.007.

[234] Fuchs, M. & Scheffler, M. Ab initio pseudopotentials for electronic structure calculations of poly-atomic systems, using density-functional theory. *Comput. Phys. Commun.* **119**, 67 (1999). doi: 10.1016/S0010-4655(98)00201-X.

[235] Troullier, N. & Martins, J. L. Efficient pseudopotentials for plane-wave calculations. *Physical Review B* **43**, 1993–2006 (1991). doi: 10.1103/PhysRevB.43.1993.

[236] Kresse, G. & Furthmüller, J. Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set. *Physical Review B* **54**, 11169–11186 (1996). doi: 10.1103/PhysRevB.54.11169.

[237] Kresse, G. & Hafner, J. *Ab initio* molecular-dynamics simulation of the liquid-metal amorphous-semiconductor transition in germanium. *Physical Review B* **49**, 14251–14269 (1994). doi: 10.1103/PhysRevB.49.14251.

[238] Peralta, J. E., Heyd, J., Scuseria, G. E. & Martin, R. L. Spin-orbit splittings and energy band gaps calculated with the Heyd-Scuseria-Ernzerhof screened hybrid functional. *Physical Review B* **74**, 073101 (2006). doi: 10.1103/PhysRevB.74.073101.

[239] Blöchl, P. Projector augmented-wave method. *Physical Review B* **50**, 17953–17979 (1994). doi: 10.1103/PhysRevB.50.17953.

[240] Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical Review B* **59**, 1758–1775 (1999). doi: 10.1103/PhysRevB.59.1758.

[241] King, P. D. C. *et al.* Valence-band electronic structure of CdO, ZnO, and MgO from x-ray photoemission spectroscopy and quasi-particle-corrected density-functional theory calculations. *Physical Review B* **79**, 205205 (2009). doi: 10.1103/PhysRevB.79.205205.

[242] Yeh, J. J. & Lindau, I. Atomic subshell photoionization cross sections and asymmetry parameters: $1 \leq Z \leq 103$. *Atomic Data and Nuclear Data Tables* **32**, 1–155 (1985). doi: 10.1016/0092-640X(85)90016-6.

[243] Veal, T. D. *et al.* Optical and photoemission studies of earth abundant solar absorbers CuSbS$_2$ and Cu$_3$N. In *European Materials Research Society Spring Meeting 2015* (2015).

[244] O'Donnell, K. P. & Chen, X. Temperature dependence of semiconductor band gaps. *Applied Physics Letters* **58**, 2924 (1991). doi: 10.1063/1.104723.

[245] Amirtharaj, P. M. & Seiler, D. G. *Handbook of Optics, Volume 2*, chap. 36, Table 13 (McGraw-Hill, 1995), 2nd edn. ISBN 0070479747.

[246] Bhosale, J. *et al.* Temperature dependence of band gaps in semiconductors: Electron-phonon interaction. *Physical Review B* **86**, 195208 (2012). doi: 10.1103/PhysRevB.86.195208.

[247] Varshni, Y. P. Temperature dependence of the energy gap in semiconductors. *Physica* **34**, 149–154 (1967). doi: 10.1016/0031-8914(67)90062-6.

[248] Yu, C. *et al.* Temperature dependence of the band gap of perovskite semiconductor compound CsSnI$_3$. *Journal of Applied Physics* **110**, 063526 (2011). doi: 10.1063/1.3638699.

[249] Göbel, A. *et al.* Effects of the isotopic composition on the fundamental gap of CuCl. *Physical Review B* **57**, 15183–15190 (1998). doi: 10.1103/PhysRevB.57.15183.

[250] Adachi, S. *Properties of Group-IV, III-V and II-VI Semiconductors* (Wiley, 2005), 1st edn. ISBN 9780470090336.

[251] Green, M. A., Emery, K., Hishikawa, Y., Warta, W. & Dunlop, E. D. Solar cell efficiency tables (version 45). *Progress in Photovoltaics: Research and Applications* **23**, 1–9 (2015). doi: 10.1002/pip.2573.

[252] Lane, D. W., Hutchings, K. J., McCracken, R. & Forbes, I. *Materials Challenges: Inorganic Photovoltaic Solar Energy*, chap. 6, 160–208 (Royal Society of Chemistry, 2015), 1st edn. ISBN 9781849733465.

[253] Mattson, B. Seven reasons thin film is alive and set to win in solar. Tech. Rep., Siva Power (2014). URL http://bit.ly/28NZnTw.

[254] USGS. Mineral commodity summaries 2016: Silicon, gallium, tellurium, indium. Tech. Rep., U.S. Geological Survey (2016). URL http://on.doi.gov/299KPm3.

[255] Welch, A. W., Zawadzki, P. P., Lany, S., Wolden, C. A. & Zakutayev, A. Self-regulated growth and tunable properties of CuSbS$_2$ solar absorbers. *Solar Energy Materials and Solar Cells* **132**, 499–506 (2015). doi: 10.1016/j.solmat.2014.09.041.

[256] Temple, D. J., Kehoe, A. B., Allen, J. P., Watson, G. W. & Scanlon, D. O. Geometry, electronic structure, and bonding in CuMCh$_2$ (M = Sb, Bi; Ch = S, Se): Alternative solar cell absorber materials? *J. Phys. Chem. C* **116**, 7334–7340 (2012). doi: 10.1021/jp300862v.

[257] Kumar, M. & Persson, C. CuSbS$_2$ and CuBiS$_2$ as potential absorber materials for thin-film solar cells. *Journal of Renewable and Sustainable Energy* **5**, 031616 (2013). doi: 10.1063/1.4812448.

[258] Peccerillo, E. *et al.* Characterization of sulfurized CuSbS$_2$ thin films for PV applications. In *Photovoltaic Specialist Conference (PVSC), 2014* (2014).

[259] Yu, L., Kokenyesi, R. S., Keszler, D. A. & Zunger, A. Inverse design of high absorption thin-film photovoltaic materials. *Advanced Energy Materials* **3**, 43–48 (2013). doi: 10.1002/aenm.201200538.

[260] Yu, L. & Zunger, A. Identification of potential photovoltaic absorbers based on first-principles spectroscopic screening of materials. *Physical Review Letters* **108**, 068701 (2012). doi: 10.1103/PhysRevLett.108.068701.

[261] Tablero, C. Microscopic analysis and applications of the Cu(Sb$_B$i)S$_2$ high optical absorption. *J. Phys. Chem. C* **119**, 8857–8863 (2015). doi: 10.1021/acs.jpcc.5b00850.

[262] Rabhi, A., Kanzari, M. & Rezig, B. Optical and structural properties of $CuSbS_2$ thin films grown by thermal evaporation method. *Thin Solid Films* **517**, 2477–2480 (2009). doi: 10.1016/j.tsf.2008.11.021.

[263] Ornelas-Acosta, R. *et al.* Thin films of copper antimony sulfide: A photovoltaic absorber material. *Materials Research Bulletin* **61**, 215–225 (2015). doi: 10.1016/j.materresbull.2014.10.027.

[264] Salem, S. M. *et al.* Influence of deposition time on structural and optical properties of chemically deposited $CuSbS_2$ thin films. *Journal of Applied Sciences Research* **9(2)**, 1083–1091 (2013).

[265] Yang, B. *et al.* $CuSbS_2$ as a promising earth-abundant photovoltaic absorber material: A combined theoretical and experimental study. *Chemistry of Materials* **26**, 3135–3143 (2014). doi: 10.1021/cm500516v.

[266] Kumar, M. & Persson, C. $Cu(Sb,Bi)(S,Se)_2$ as indium-free absorber material with high optical efficiency. *Energy Procedia* **44**, 176–183 (2014). doi: 10.1016/j.egypro.2013.12.025.

[267] Sugaki, A., Shima, H. & Kitakaze, A. Phase relations of the Cu_2S-Sb_2S_3 system. Tech. Rep., Yamaguchi University (1973).

[268] Kyono, A. & Kimata, M. Crystal structures of chalcostibite ($CuSbS_2$) and emplectite ($CuBiS_2$): Structural relationship of stereochemical activity between chalcostibite and emplectite. *American Mineralogist* **90**, 162–165 (2005). doi: 10.2138/am.2005.1585.

[269] Manolache, S. *et al.* The influence of the precursor concentration on $CuSbS_2$ thin films deposited from aqueous solutions. *Thin Solid Films* **515**, 5957–5960 (2007). doi: 10.1016/j.tsf.2006.12.046.

[270] Zhou, J. *et al.* Solvothermal crystal growth of $CuSbQ_2$ (Q=S, Se) and the correlation between macroscopic morphology and microscopic structure. *Journal of Solid State Chemistry* **182**, 259–264 (2009). doi: 10.1016/j.jssc.2008.10.025.

[271] Rodriguez-Lazcano, Y., Nair, M. & Nair, P. $CuSbS_2$ thin film formed through annealing chemically deposited Sb_2S_3-CuS thin films. *Journal of Crystal Growth* **223**, 399–406 (2001). doi: 10.1016/S0022-0248(01)00672-8.

[272] Colombara, D. *Investigation of Chalcogenide Absorber Materials for Photovoltaic Applications*. Ph.D. thesis, Department of Chemistry, University of Bath (2012).

[273] Dufton, J. T. R. *et al.* Structural and electronic properties of $CuSbS_2$ and $CuBiS_2$: potential absorber materials for thin-film solar cells. *Phys. Chem. Chem. Phys.* **14**, 7229–7233 (2012). doi: 10.1039/c2cp40916j.

[274] Welch, A. W. *CuSb(S_Se)_2 Thin Film Heterojunction Photovoltaic Devices*. Ph.D. thesis, Colorado School of Mines, USA (2015).

[275] Wuensch, B. J. The crystal structure of tetrahedrite, $Cu_{12}Sb_4S_{13}$. *Zeitschrift für Kristallographie - Crystalline Materials* **119**, 437–453 (1964). doi: 10.1524/zkri.1964.119.16.437.

[276] Ramasamy, K., Sims, H., Butler, W. H. & Gupta, A. Mono-, few-, and multiple layers of copper antimony sulfide ($CuSbS_2$): A ternary layered sulfide. *Journal of the American Chemical Society* **136**, 1587–1598 (2014). doi: 10.1021/ja411748g.

[277] Hofmann, W. Strukturelle und morphologische zusammenhänge bei erzen vom formeltyp ABC_2 I. die struktur von wolfsbergit $CuSbS_2$ und emplektit $CuBiS_2$ und deren beziehungen zu der struktur von antimonit Sb_2S_3. *Zeitschrift für Kristallographie - Crystalline Materials* **84**, 177–203 (1933). doi: 10.1524/zkri.1933.84.1.177.

[278] Razmara, M. F., Henderson, C. M. B., Pattrick, R. A. D., Bell, A. M. T. & Charnock, J. M. The crystal chemistry of the solid solution series between chalcostibite ($CuSbS_2$) and emplectite ($CuBiS_2$). *Mineralogical Magazine* **61**, 79–88 (1997).

[279] J. A. Woollam Co. CompleteEASE data analysis manual v4.63. Tech. Rep., J. A. Woollam Co., Inc. (2011).

[280] Yu, H., Cui, X., Xu, X. & Yao, W. Valley excitons in two-dimensional semiconductors. *National Science Review* **2**, 57–70 (2015). doi: 10.1093/nsr/nwu078.

[281] Moody, G. *et al.* Intrinsic homogeneous linewidth and broadening mechanisms of excitons in monolayer transition metal dichalcogenides. *Nature Communications* **6**, 8315 (2015). doi: 10.1038/ncomms9315.

[282] Duke, C. B. & Segall, B. Nonexistence of hyperbolic excitons. *Physical Review Letters* **17**, 19–22 (1966). doi: 10.1103/PhysRevLett.17.19.

[283] Hermanson, J. Existence of hyperbolic excitons. *Physical Review Letters* **18**, 170–174 (1967). doi: 10.1103/PhysRevLett.18.170.

[284] Kane, E. O. Coulomb effects at saddle-type critical points. *Physical Review* **180**, 852–858 (1969). doi: 10.1103/PhysRev.180.852.

[285] Nelkomsky, M. & Braunstein, R. Interband transitions and exciton effects in semiconductors. *Physical Review B* **5** (1972). doi: 10.1103/PhysRevB.5.497.

[286] Gil, B., Felbacq, D. & Chichibu, S. F. Exciton binding energies in chalcopyrite semiconductors. *Physical Review B* **85**, 075205 (2012). doi: 10.1103/PhysRevB.85.075205.

[287] Shay, J. L., Tell, B., Schiavone, L. M., Kasper, H. M. & Thiel, F. Energy bands of $AlInS_2$ in the chalcopyrite and orthorhombic structures. *Physical Review B* **9**, 1719–1723 (1974). doi: 10.1103/PhysRevB.9.1719.

[288] Aroyo, M. I., Kirov, A., Capillas, C., Perez-Mato, J. M. & Wondratschek, H. Bilbao crystallographic server II: Representations of crystallographic point groups and space groups. *Acta Cryst.* **A62**, 115–128 (2006). doi: 10.1107/S0108767305040286.

[289] Haug, H. & Koch, S. W. *Quantum Theory of the Optical and Electronic Properties of Semiconductor* (World Scientific, 2009), 1st edn. ISBN 9789812838834.

[290] Kuroda, N., Munakata, I. & Nishina, Y. Exciton transitions from spin-orbit split off valence bands in layer compound InSe. *Solid State Communications* **33**, 687–691 (1980).

[291] Beal, A. R., Knights, J. C. & Liang, W. Y. Transmission spectra of some transition metal dichalcogenides. II. group VIA: trigonal prismatic coordination. *Journal of Physics C: Solid State Physics* **5**, 3540 (1972). doi: 10.1088/0022-3719/5/24/016.

[292] Saigal, N., Sugunakar, V. & Ghosh, S. Exciton binding energy in bulk MoS_2: A reassessment. *Applied Physics Letters* **108**, 132105 (2016). doi: 10.1063/1.4945047.

[293] Hahn, T. & Aroyo, M. I. *International Tables for Crystallography, Vol. A*, chap. 7.1, 112–717 (International Union of Crystallography, 2006), 1st edn.

[294] Dresselhaus, M. S., Dresselhaus, G. & Jorio, A. *Group Theory: Application to the Physics of Condensed Matter* (Springer, 2008), 1st edn. ISBN 9783540328998.

[295] Koster, G. F., Dimmock, J. O., Wheeler, R. G. & Statz, H. *Properties of the thirty-two point groups* (MIT press, 1963), 1st edn.

[296] Aroyo, M. I. *et al.* Crystallography online: Bilbao crystallographic server. *Bulgarian Chemical Communications* **43(2)**, 183–197 (2011).

[297] Cornwell, J. F. *Group Theory in Physics: An Introduction* (Academic Press, 1997). ISBN 0121898008.

[298] Janak, J. Proof that $\partial e/\partial n_i = \epsilon$ in density-functional theory. *Physical Review B* **18**, 7165–7168 (1978). doi: 10.1103/PhysRevB.18.7165.

[299] Perdew, J. P. & Levy, M. Comment on "significance of the highest occupied Kohn-Sham eigenvalue". *Physical Review B* **56**, 16021–16028 (1997). doi: 10.1103/PhysRevB.56.16021.

[300] Gudelli, V. K., Kanchana, V., Vaitheeswaran, G., Svane, A. & Christensen, N. E. Thermoelectric properties of chalcopyrite type CuGaTe$_2$ and chalcostibite CuSbS$_2$. *Journal of Applied Physics* **114**, 223707 (2013). doi: 10.1063/1.4842095.

[301] Born, M. & Wolf, E. *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light* (Cambridge University Press, 2000), 7th edn. ISBN 9780521784498.

[302] Olguín, D., Cardona, M. & Cantarero, A. Electron-phonon effects on the direct band gap in semiconductors: LCAO calculations. *Solid State Communications* **122**, 575–589 (2002). doi: 10.1016/S0038-1098(02)00225-9.

[303] Allen, P. B. & Cardona, M. Temperature dependence of the direct gap of Si and Ge. *Physical Review B* **27**, 4760–4769 (1983). doi: 10.1103/PhysRevB.27.4760.

[304] Fan, H. Y. Temperature dependence of the energy gap in semiconductors. *Physical Review* **82**, 900–905 (1951). doi: 10.1103/PhysRev.82.900.

[305] Cardona, M. Electron-phonon interaction in tetrahedral semiconductors. *Solid State Communications* **133**, 3–18 (2005). doi: 10.1016/j.ssc.2004.10.028.

[306] Allen, P. B. & Cardona, M. Theory of the temperature dependence of the direct gap of germanium. *Physical Review B* **23**, 1495–1505 (1981). doi: 10.1103/PhysRevB.23.1495.

[307] Cardona, M. & Thewalt, M. L. W. Isotope effects on the optical spectra of semiconductors. *Reviews of Modern Physics* **77**, 1173–1224 (2005). doi: 10.1103/RevModPhys.77.1173.

[308] Kazimierczuk, T., Fröhlich, D., Scheel, S., Stolz, H. & Bayer, M. Giant Rydberg excitons in the copper oxide Cu$_2$O. *Nature* **514**, 343–347 (2014). doi: 10.1038/nature13832.