Diamond-like amorphous carbon

J. Robertson*

Engineering Department, Cambridge University, Cambridge CB2 1PZ, UK

Abstract

Diamond-like carbon (DLC) is a metastable form of amorphous carbon with significant sp³ bonding. DLC is a semiconductor with a high mechanical hardness, chemical inertness, and optical transparency. This review will describe the deposition methods, deposition mechanisms, characterisation methods, electronic structure, gap states, defects, doping, luminescence, field emission, mechanical properties and some applications of DLCs. The films have widespread applications as protective coatings in areas, such as magnetic storage disks, optical windows and micro-electromechanical devices (MEMs). © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Diamond-like carbon; Properties; Applications

1. Introduction

Diamond-like carbon (DLC) is a metastable form of amorphous carbon containing a significant fraction of sp³ bonds. It can have a high mechanical hardness, chemical inertness, optical transparency, and it is a wide band gap semiconductor [1–10]. DLC films have widespread applications as protective coatings in areas such as optical windows, magnetic storage disks, car parts, biomedical coatings and as micro-electromechanical devices (MEMs). This review will describe the preparation, characterisation, bonding, electronic and mechanical properties, the deposition mechanism and some applications of DLCs.

There have recently been very important advances in the science of carbon such as the development of the chemical vapour deposition of diamond [11–13] and the discovery of C₆₀ and carbon nanotubes [14]. There have been parallel developments in the field of disordered carbons. New ways have been developed to synthesise DLC, while a range of disordered carbons with local fullerene-like order on a nanometre length scale have been discovered. The DLCs have now been characterised in great detail. The ways to grow the most diamond-like materials are now understood. Their growth mechanism is broadly understood in terms of the subplantation of incident ions.

Carbon forms a great variety of crystalline and disordered structures because it is able to exist in three hybridisations, sp³, sp² and sp¹ (Fig. 1) [3]. In the sp³ configuration, as in diamond, a carbon atom’s four valence electrons are each assigned to a tetrahedrally directed sp³ orbital, which makes a strong σ bond to an adjacent atom. In the three-fold coordinated sp² configuration as in graphite, three of the four valence electrons enter trigonally directed sp² orbitals, which form σ bonds in a plane. The fourth electron of the sp² atom lies in a π orbital, which lies normal to the σ bonding plane. This π orbital forms a weaker π bond with an π orbital on one or more neighbouring atoms. In
the sp\textsuperscript{1} configuration, two of the four valence electrons enter σ orbitals, each forming a σ bond directed along the ±x-axis, and the other two electrons enter π orbitals in the y and z directions.

The extreme physical properties of diamond derive from its strong, directional σ bonds [11]. Diamond has a wide 5.5 eV band gap, the largest bulk modulus of any solid, the highest atom density, the largest room temperature thermal conductivity, smallest thermal expansion coefficient, and largest limiting electron and hole velocities of any semiconductor [11]. Graphite has strong intra-layer σ bonding and weak van der Waals bonding between its layers. A single graphite plane is a zero band gap semiconductor, and in three dimensions it is an anisotropic metal [14,15].

DLC has some extreme properties similar to diamond, such as the hardness, elastic modulus and chemical inertness, but these are achieved in an isotropic disordered thin film with no grain boundaries. It is much cheaper to produce than diamond itself. This has great advantages for many applications. Typical properties of the various forms of DLC are compared to diamond and graphite in Table 1 [16–22].

The sp\textsuperscript{3} bonding of DLC confers on it many of the beneficial properties of diamond itself, such as its mechanical hardness, chemical and electrochemical inertness, and wide band gap. DLC consists not only of the amorphous carbons (a-C) but also of the hydrogenated alloys, a-C:H. It is convenient to display the compositions of the various forms of amorphous C–H alloys on a ternary phase diagram as in Fig. 2, as first used by Jacob and Moller [21]. There are many a-Cs with disordered graphitic ordering, such as soot, chars, glassy carbon, and evaporated a-C. These lie in the lower left hand corner. The two hydrocarbon polymers polyethylene (CH\textsubscript{2})\textsubscript{n} and polyacetylene (CH)\textsubscript{n} define the limits of a triangle in the right hand corner beyond which interconnecting C–C networks

![Diagram](https://example.com/diagram.png)

**Fig. 1.** The sp\textsuperscript{3}, sp\textsuperscript{2}, sp\textsuperscript{1} hybridised bonding.

<table>
<thead>
<tr>
<th></th>
<th>sp\textsuperscript{3} (%)</th>
<th>H (%)</th>
<th>Density (g cm\textsuperscript{-3})</th>
<th>Gap (eV)</th>
<th>Hardness (GPa)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diamond</td>
<td>100</td>
<td>0</td>
<td>3.515</td>
<td>55</td>
<td>100</td>
<td>[12]</td>
</tr>
<tr>
<td>Graphite</td>
<td>0</td>
<td>0</td>
<td>2.267</td>
<td>0</td>
<td></td>
<td>[15]</td>
</tr>
<tr>
<td>C\textsubscript{60}</td>
<td>0</td>
<td>0</td>
<td>1.6</td>
<td></td>
<td></td>
<td>[14]</td>
</tr>
<tr>
<td>Glassy C</td>
<td>0</td>
<td>0</td>
<td>1.3–1.55</td>
<td>0.01</td>
<td>3</td>
<td>[3]</td>
</tr>
<tr>
<td>Evaporated C</td>
<td>0</td>
<td>0</td>
<td>1.9</td>
<td>0.4–0.7</td>
<td>3</td>
<td>[3]</td>
</tr>
<tr>
<td>Sputtered C</td>
<td>5</td>
<td>0</td>
<td>2.2</td>
<td>0.5</td>
<td></td>
<td>[18]</td>
</tr>
<tr>
<td>ta-C</td>
<td>80–88</td>
<td>0</td>
<td>3.1</td>
<td>2.5</td>
<td>80</td>
<td>[6,17,18]</td>
</tr>
<tr>
<td>a-C:H hard</td>
<td>40</td>
<td>30–40</td>
<td>1.6–2.2</td>
<td>1.1–1.7</td>
<td>10–20</td>
<td>[4]</td>
</tr>
<tr>
<td>a-C:H soft</td>
<td>60</td>
<td>40–50</td>
<td>1.2–1.6</td>
<td>1.7–4</td>
<td>&lt;10</td>
<td>[4]</td>
</tr>
<tr>
<td>ta-C:H</td>
<td>70</td>
<td>30</td>
<td>2.4</td>
<td>2.0–2.5</td>
<td>50</td>
<td>[22]</td>
</tr>
<tr>
<td>Polyethylene</td>
<td>100</td>
<td>67</td>
<td>0.92</td>
<td>6</td>
<td>0.01</td>
<td>[20]</td>
</tr>
</tbody>
</table>
cannot form, and only molecules form. Deposition methods have been developed to produce a-Cs with increasing degrees of $sp^3$ bonding. Sputtering can extend from $sp^2$ bonding some way towards $sp^3$ bonding. If the fraction of $sp^3$ bonding reaches a high degree, McKenzie [6] suggested that the a-C is denoted as tetrahedral amorphous carbon (ta-C), to distinguish it from $sp^2$ a-C. A range of deposition methods, such as plasma enhanced chemical vapour deposition (PECVD) [4], is able to reach into the interior of the triangle. This produces a-C:H. Although this is diamond-like, it is seen from Fig. 2 that the content of $sp^3$ bonding is actually not so large, and its hydrogen content is rather large. Thus, a more $sp^3$ bonded material with less hydrogen which can be produced by high plasma density PECVD reactors is called hydrogenated tetrahedral amorphous carbon (ta-C:H) by Weiler et al. [22].

2. Deposition methods

2.1. Ion beam

The first DLCs were prepared as thin films by Aisenberg and Chabot [23] using ion beam deposition. It is possible to produce DLCs by a wide range of deposition methods [1,24]. The methods can be categorised as to whether they are most suitable for laboratory studies or industrial production. The common feature of these methods is that the DLC film is condensed from a beam containing medium energy (~100 eV) carbon or hydrocarbon ions. It is the impact of these ions on the growing film that induces the $sp^3$ bonding—a physical process. This contrasts with the chemical vapour deposition (CVD) of diamond, where a chemical process stabilises its $sp^3$ bonding [13]. The best deposition process for DLC will provide a carbon ion flux at about 100 eV per carbon atom, with a narrow energy distribution, a single energetic species and a minimum number of non-energetic (generally neutral) species [22].

Various deposition systems are shown schematically in Fig. 3. In a typical ion beam deposition system, carbon ions are produced by the plasma sputtering of a graphite cathode in an ion source [23–29]. Alternatively, as in the Kaufman source [30], a hydrocarbon gas such as methane is ionised in a plasma [30–32]. An ion beam is then extracted through a grid from the plasma source by a bias voltage. The carbon ions or hydrocarbon ions are then accelerated to form the ion beam in the high
vacuum deposition chamber. In both cases, the ion source runs at a finite pressure, so that the beam also contains a large flux of unionised neutral species. This can reduce the flux ratio of ions to neutrals to as low as 2–10%. Ion beam sources tend to run best at higher ion energies of 100–1000 eV.

A variant of ion beam deposition is the cascade arc source [33]. Here, a high-pressure source produces an intense plasma, which then expands supersonically into a high vacuum, giving rise to large fluxes of ions and radicals.
2.2. Mass selected ion beam

For laboratory work, it is desirable to have a controlled deposition from a single ion species at a well-defined ion energy. This is achieved by mass selected ion beam deposition (MSIB) [7,8,34–39]. Carbon ions are produced in an ion source from a graphite target, such that the spread of ion energies is small, 1–10 V. The ions are then accelerated to 5–40 kV and passed through a magnetic filter. This filters out any neutrals and selects ions with an \( e/m \) ratio of the \( \text{C}^+ \) ion. The ion beam will diverge because of its Coulombic repulsion. The ions are then decelerated to the desired ion energy by electrostatic lens, and the beam is focused onto the substrate in a vacuum of order \( 10^{-8} \) Torr to produce a ta-C film. The advantages of MSIB are that it gives a controllable deposition species and energy, a filtering out of non-energetic species, and the ability to dope by switching the ion species. The disadvantage is the low deposition rate of order 0.001 Ångstrom s\(^{-1}\) and the high cost and size of the apparatus. The MSIB method and its use have been reviewed by Lifshitz [7,8], Hofass and Ronning [37] and Hirvonen et al. [39].

2.3. Sputtering

The most common industrial process for the deposition of DLC is sputtering [40–49]. The most common form uses the dc or rf sputtering of a graphite electrode by an Ar plasma. Because of the low sputter yield of graphite, magnetron sputtering is often used to increase the deposition rate. Magnets are placed behind the target to cause the electrons to spiral and increase their path length, and thus to increase the degree of ionisation of the plasma. As ion bombardment helps the formation of \( \text{sp}^3 \) bonding, the magnetic field can be configured to pass across the substrate, so this causes the Ar ions to also bombard the substrate, to give an ‘unbalanced magnetron’. A dc bias can be applied to the substrate to vary the ion energy. The a-C:H can be produced by reactive sputtering, by using a plasma of Ar and hydrogen or methane, and a-CN\(_x\) can be produced using an argon–nitrogen plasma.

Alternatively, in ion beam sputtering, a beam of Ar ions can be used to sputter from the graphite target to create the carbon flux [47]. A second Ar ion beam can be used to bombard the growing film, to densify the film or encourage \( \text{sp}^3 \) bonding. This is called ion beam assisted deposition (IBAD) or ion plating.

Sputtering is preferred for industrial applications because of its versatility, its widespread use to sputter many materials, and its ease of scale up. Also, the deposition conditions can be controlled by the plasma power and gas pressure but they are reasonably independent of the substrate geometry or condition. A disadvantage of sputtering is, like ion beam deposition, that it can have a relatively low ratio of energetic ions to neutral species, so that it does not produce the hardest DLC films. However, sputtering methods with a very high fraction of ions have been developed by Schwan et al. [46] and Cuomo et al. [47] to produce a-C with a relatively large \( \text{sp}^3 \) fraction, but this is at the expense of a low growth rate.

2.4. Cathodic arc

An unusual method for laboratory and industrial use is the cathodic arc [6,17,50–73]. An arc is initiated in a high vacuum by touching the graphite cathode with a small carbon striker electrode and withdrawing the striker. This produces an energetic plasma with a high ion density of up to \( 10^{13} \) cm\(^{-3}\). A typical cathodic arc system is shown in Fig. 3 after Coll and Chhowalla [54]. The power supply is a low voltage, high current supply. The cathodic arc is also widely used to deposit hard coating materials, such as tin by the reactive deposition of Ti in a nitrogen atmosphere, as reviewed by Brown [55].
The cathode spot is small, 1–10 μm and it carries a very high current density of $10^6$–$10^8$ A cm$^{-2}$. The spot is formed by an explosive emission process. This creates particulates as well as the desired plasma. The particulates can be filtered by passing the plasma along a toroidal magnetic filter duct [50,55], as shown in Fig. 4. This is known as filtered cathodic vacuum arc (FCVA). The toroidal currents produce a magnetic field of about 0.1 T along the axis of the filter. The electrons of the plasma spiral around the magnetic field lines and so they follow them along the filter axis. This motion produces an electrostatic field, which causes the positive ions to follow the electrons around the filter. This produces an amipolar transport of the plasma around the filter. The particulates cannot follow the field and they hit the walls and baffle the walls. Alternatively, in the open filter system used by Brown [55] and Anders et al. [56], the particulates pass between the coils out of the filter zone into the chamber. The neutrals also hit the walls, so the filter raises the plasma ionisation from about 30% to nearly 100% at the filter exit. The plasma beam is condensed onto a substrate to produce the ta-C. At the cathode end, the plasma has a mean ion energy of 10–30 eV, depending on the arc current and it has a Gaussian distribution of ion energies [66] as shown in Fig. 5. A dc or rf self-bias voltage applied to the substrate is used to increase the incident ion energy [17]. The FCVA should be operated at a background pressure of $\sim 10^{-8}$ Torr, but this may rise to $10^{-3}$ when the plasma is running.

The advantages of the FCVA are that it produces a highly ionised plasma with an energetic species, a fairly narrow ion energy distribution, and high growth rates of 1 nm s$^{-1}$ for a low capital cost. It is used on an industrial scale. Unlike ion beam deposition, the depositing beam in FCVA is a neutral plasma beam so that it can deposit onto insulating substrates. The disadvantages are that the filtering is not sufficient for some applications, and that the cathode spot is unstable.

The particulates are generally submicron-size particles. These can still pass through the filter section by bouncing off the walls [69,73,74]. The filtering can be improved by a factor of 100, by adding baffles along the filter section, and by including a second bend to give a double bend or ‘S-bend’ filter [67–72], as shown in Fig. 4b.

The arc can run continuously or in a pulsed mode. The pulsed mode occurs by using a capacitor bank to strike the arc, or by laser initiation. The arc current can be passed through the filter coils. The pulsed mode allows better filtering, because the ions tend to be entrained in the plasma beam during the pulse, but fall out of the plasma when the beam stops [71].

On metals, the arc produces a molten spot of higher resistivity, so the arc tends to move sideways to another spot of lower resistivity. The arc spot therefore moves around the surface, giving a continuously running arc, which consumes the cathode evenly. On semiconductors or materials like
carbon, the arc spot has lower resistivity than its surroundings, so the arc spot tries to stay in the same place, and erodes deeper at the same place. The cathode stability can be improved by re-striking the arc or by using a magnetic field in the cathode region to steer the cathode spot around the cathode surface [6, 73]. Repetitive initiation of the arc by laser is also used [62].

Ions are lost to the wall during transport through the filter duct. This can be minimised by giving the filter section a positive bias of about +20 V [59]. It is often useful to allow the excess electron current from the cathode to flow to an anode at the entrance of the filter duct. It is possible to deposit uniform films over areas of 10 cm or more by using a magnetic bottle configuration to diverge the beam after the filter duct [69] or by scanning the beam by current coils. The deposition rate is proportional to the electron current from the cathode.

In the arc, the main current is an electron current to the anode, known as the arc current. The ion current away from the cathode is proportional to the arc current and about 10–100 times less. The mean ion energy rises slightly with arc current. Very high plasma currents can be generated by the high current arc as detailed by Scheibe and Schultrich [62].

2.5. Pulsed laser deposition

Pulsed excimer lasers such as ArF give very short, intense energy pulses, which can be used to vaporise materials as an intense plasma [9, 75–82]. The plasma then expands towards the substrate. The kinetic energy of this expansion gives an ion energy analogous to the ion energy of MSIB or the cathodic arc. The mean ion energy is proportional to the laser fluence concentrated at the target spot. It has been measured as a function of laser fluence [9] (Fig. 6). In this way pulsed laser deposition (PLD) produces ta-C films similar to those from the MSIB and FCVA methods [80, 82]. The dependence of properties on ion energy is similar [82]. The advantage of PLD is that it is versatile laboratory scale method, which can be used to deposit many different materials, from high temperature superconductors to hard coatings. The PLD method for carbon has been reviewed by Voevodin and Donley [9] and by Siegal et al. [81].

2.6. PECVD

The most popular laboratory deposition method is rf PECVD [4, 83–96, 98]. The reactor consists of two electrodes of different area. The rf power is usually capacitively coupled to the smaller
electrode on which the substrate is mounted, and the other electrode (often including the reactor walls) is earthed. The rf power produces a plasma between the electrodes. The higher mobility of electrons than ions in the plasma creates a sheath next to the electrodes with an excess of ions. This has a positive space charge, so the plasma develops a positive voltage with respect to the electrodes, which equalises the mean electron and ion current to the wall [91], as shown in Fig. 7.

The sheaths act as diodes, so that the electrodes acquire dc self-bias voltages equal to their peak rf voltage. The rf voltage is divided between the sheaths of the two electrodes as in a capacitive divider, according to their inverse capacitance. Thus, the dc self bias voltage varies inversely with the
electrode areas \[84,91\],

\[
\frac{V_1}{V_2} = \left(\frac{A_2}{A_1}\right)^2
\]

The smaller electrode with smaller capacitance acquires the larger bias voltage and becomes negative with respect to the larger electrode. This is made the substrate electrode. The negative sheath voltage accelerates the positive ions to give the bombardment needed to create the \(\text{sp}^3\) bonding. In low-pressure rf plasmas, the plasma is excited by an rf coupling to the sheaths. At higher pressures, the plasma is excited by a Joule heating of the bulk plasma.

For DLC deposition, the plasma should be operated at the lowest possible pressure, in order to maximise the ion to radical fraction of the plasma. However, even at 50 mTorr pressure, the ions are only about 10% of the film-forming flux. The ions can loose energy by collisions when being accelerated across the sheath. The ion energy is then no longer the sheath voltage. It is desirable to use a low pressure to minimise these collisions, to maintain a narrow ion energy distribution. The sheath thickness decreases with increasing pressure \(P\) as \[91\]

\[
d = kP^{-1/2}
\]

(from the Debye length) while the ion mean free path decreases as \(\lambda = k'/P\). Hence, the ratio \(\lambda/d\) scales as \(P^{-1/2}\) and the mean free path becomes less than the sheath thickness at low enough pressures. There are still significant ion collisions in the sheath at 50 mTorr, so that there is a wide ion energy distribution and the mean ion energy is only about 0.4 of the sheath voltage (roughly the bias voltage \(V_b\)) \[4,87,88\] as shown in Fig. 8a.

It is necessary to use lower pressures, but this is not possible for conventional PECVD as the plasma will no longer strike. A lower pressure plasma can be created by using a magnetic field to confine the plasma, to increase the electron path length and increase the ionisation efficiency. This

![Fig. 8. Ion energy distributions in conventional plasma deposition at 5 Pa pressure, and in the plasma beam source at 0.01 Pa pressure, from Wild and Koidl \[87,88\], and Weiler et al. \[22\].](image-url)
allows a capacitively coupled plasma to continue to operate at $5 \times 10^{-4}$ Torr. At this pressure, the ion mean free path exceeds the sheath thickness and the ion energy now becomes a narrow distribution, as shown in Fig. 8b.

This is a principle behind the plasma beam source (PBS) [95] shown in Fig. 9a. The PBS consists of a magnetically confined plasma in which the plasma exits through a grid at earth potential. The rf is applied to a moveable electrode whose area is larger than the grid, so that this electrode acquires the positive self-bias. This repels the positive ions through the grid, to form a plasma beam which then condenses on the substrate to form ta-C:H. The plasma beam is neutral so it can be used on insulating substrates.

A feature of the simple capacitively coupled PECVD system is that it is not possible to have independent control of the ion current and the ion energy, as they both vary with the rf drive power. The bias voltage for a resistive plasma will vary roughly with rf power ($W$) and pressure ($P$) as [86]

$$V_b = k' \frac{W}{P^{1/2}}$$

(3)

The mean ion energy $E_i$ is proportional to the bias voltage $V_b$, $E_i = kV_b$ ($k$ is ~0.4 at typical PECVD operating pressures of 5 Pa, as noted above [4]). The ion current is roughly

$$I \approx \frac{W}{V_b} = k'(WP)^{1/2}$$

(4)

so that both $V_b$ and $I$ depend on rf power, $W$. A separate control of ion energy and current is possible by having one power source for exciting the plasma (controlling the ion current) and a separate rf source for biasing the substrate electrode (controlling the ion energy).
The PBS achieves some separation by varying the effective electrode area to control $V_b$. The magnetic field confines the plasma to a certain area of the anode, so moving the anode will vary the bias voltage and thereby the ion energy.

In recent years, it has become clear that high-density plasma sources are possible [91]. The two fundamental properties of the plasma are the plasma density $n_0$ and the electron temperature $T_e$. One aim is to maximise $n_0$. The plasma electrons have a Maxwellian energy distribution, which defines the electron temperature, $T_e$

$$N(E) = n_0 \exp\left(\frac{-E}{kT_e}\right)$$  \hspace{1cm} (5)

Electrons with an energy above some threshold energy colliding with a neutral species will ionise or dissociate it according to the convolution

$$N_i = \int n_0(E, T_e)f_i(E)\,dE$$  \hspace{1cm} (6)

where $f_i$ is the ionisation probability. This is shown schematically in Fig. 10. Over a limited energy range, this gives

$$N_i \approx N_0 \exp\left(\frac{-E_i}{kT_e}\right)$$  \hspace{1cm} (7)

where $E_i$ is the ionisation potential. A high $T_e$ maximises the ionisation. Similar relations hold for dissociated atomic species and excited species.

$N_p$ and $T_e$ are set by the requirements of energy and particle balance [91]. The particle balance sets $T_e$ by equating the rate of generation of ions in the bulk plasma to the rate of loss of radicals and ions to the walls,

$$K_iN_g n_0 V = n_0\mu_B A$$  \hspace{1cm} (8)

![Fig. 10. Electron temperature $T_e$ and the dissociation/ionisation probability of a species.](image-url)
This gives

\[
\frac{K_i(T_e)}{u_B(T_e)} = \frac{1}{N_g l}
\]

where \( l \) is the effective plasma length, \( A \) and \( V \) are the surface area and volume of the plasma, \( n_0 \) is the plasma density \( N_g \) is the density of atoms in the gas, \( K_i \) is the ionisation rate constant and \( u_B \) is the Bohm velocity of the electrons. \( K_i \) and \( u_B \) are functions of \( T_e \) as shown. Recombination of radicals at the walls is an important factor for the radical density.

The energy balance sets the plasma density \( n_0 \) by equating the power absorbed by the plasma \( W \) to the energy loss per ion \( E_T \) as ion energy and as ion loss to the walls. This gives

\[
n_0 = \frac{W}{e u_B A E_T}
\]

Capacitive coupling at rf frequencies of 13.6 MHz provides a relatively inefficient excitation of the plasma and so the plasma density is quite low, at most \( 10^9 \) cm\(^{-3} \). More efficient excitation is obtained by raising the rf frequency to VHF or by using an inductive coupled plasma (ICP). The ICP is often employed for plasma etching. Another method is to use electron cyclotron resonance (ECR) systems using microwaves [92–94]. In an ECR, a large static axial magnetic field is used to cause electrons to oscillate at a cyclotron resonance frequency of 2.45 GHz. Microwave power is applied at this frequency and a carrier gas such as Ar is fed into the up-stream part of the chamber. The Ar atoms are excited and pass into the down stream part of the chamber, where they meet the reactant gas. The Ar atoms or ions transfer energy to the reactant gas by collisions, and these then deposit on the substrate. ICPs and ECRs usually have an additional rf voltage capacitively-coupled to the substrate electrode, to give a self-bias voltage. This is used to control the ion energy, separately from the ion current.

Perhaps the most compact, rf-powered, high plasma-density is the recently developed electron cyclotron wave resonance (ECWR) source [96] shown in Fig. 9b. The rf is inductively coupled to the plasma through a single turn coil. A transverse static magnetic field confines the plasma. This causes the rf electromagnetic wave in the plasma to form left and right hand circularly polarised waves. The refractive index of one of these waves increases dramatically. This decreases the wavelength of the rf, so the rf can form a half wavelength standing wave across the chamber, which allows a resonant coupling of power into the plasma bulk. An rf signal is also capacitively coupled to a rear electrode to provide a self-bias voltage to vary the ion energy. The plasma can exit the chamber as a neutral beam through a grounded grid electrode. The ECWR is equivalent to the helicon, except for a different orientation of the fields and antennas.

The ECWR source produces an extremely high-density plasma of \( 10^{12} \) cm\(^{-3} \) or over, with an independent control of the ion energy and ion current density [96]. It gives a narrow ion energy distribution with a width under 5%. The ECWR produces ta-C:H at a much higher growth rate (1.5 nm s\(^{-1} \)) than the PBS and gives uniform deposition over an diameter of 10 cm, which is scaleable to larger values. It is the first industrialised, high-density PECVD source for DLC.

The gas used in PECVD has a significant effect on the a-C:H properties. In the early days, precursors with low ionisation potentials such as benzene were chosen, as this gave a much higher growth rate. The deposition rate increases roughly exponentially with decreasing ionisation energy [4] as shown in Fig. 11. For mechanical applications, it is desirable to maximise the hardness, which we shall see means minimising the incorporation of hydrogen. This requires using a precursor with small H/C ratio, such as acetylene, as this strongly affects the H/C ratio of the resulting film.
It is now known that DLC properties depend on the ion energy per C atom. Thus, a benzene ion C<sub>6</sub>H<sub>6</sub><sup>+</sup> with six carbons requires high bias voltages to reach the desired 100 V per atom. Acetylene is more acceptable because 200 V is easier to handle. Acetylene is in fact a very useful source gas for low pressure deposition, because its strong C= C bond means it has a simple dissociation pattern, giving mainly C<sub>2</sub>H<sub>4</sub><sup>+</sup> ions [22,97], as seen in Fig. 12. There is also less plasma polymerisation than occurs with methane. Acetylene is the preferred source gas for mechanical applications. However, acetylene is unsatisfactory for electronic applications because it is not available in high purity form, and possesses a substantial nitrogen impurity [98], which can cause a doping effect particularly if it is used in high-density plasmas. Methane remains a popular choice for electronic applications because it is available in high purity, but the growth rate is lower and it gives too high a hydrogen content. Hydrogen dilution can be used to vary the hydrogen content.

3. Deposition mechanism

The key property of DLC is its sp<sup>3</sup> bonding. The deposition process which promotes sp<sup>3</sup> bonding is a physical process, ion bombardment [6–8,17,99,100]. The highest sp<sup>3</sup> fractions are formed by C<sup>+</sup> ions with ion energy around 100 eV. To see how this occurs, it is useful to first consider the hydrogen-free carbon, ta-C, and then a-C:H.

The atomistic description of the deposition mechanism has developed as follows. Fig. 13 compares the ion ranges and yields for various processes of C and H ions in carbon [34]. Spencer et al. [25] proposed various mechanisms such as that the sp<sup>3</sup> sites arise from the sp<sup>3</sup>/sp<sup>2</sup> mixture by a preferential sputtering of the sp<sup>2</sup> sites. However, Lifshitz et al. [34] noted that this process would not work because sputtering yield (efficiency) of carbon. In addition, the sputtering yield depends primarily on an atom’s cohesive energy, which is almost the same for the sp<sup>2</sup> and sp<sup>3</sup> sites, so there is little difference in sputtering yields. Weissmantel [101] suggested that the sp<sup>3</sup> bonding arose from a shock wave of the displacement spike of the ion cascade. This picture has some similarities to later models, but the shock wave is more appropriate to higher ion energies and higher ion masses.
Fig. 12. Mass spectra dissociation pattern of CH₄, C₂H₄ and C₂H₂ at different pressures. Data from Kleber et al. [97].

Fig. 13. Ion ranges and the yields of ion processes in carbon. Data from Lifshitz et al. [34].
Lifshitz et al. [34] used Auger analysis of the depth profiles of medium energy C ions incident on Ni substrates to show that the growth was sub-surface. They denoted this process as ‘subplantation’ (low energy subsurface implantation). They then proposed that the sp³ sites accumulate by a preferential displacement of sp² sites. Moller [102] modelled this idea in more detail. The sp² and sp³ atoms would be displaced at certain rates into interstitial sites, and then fall back at similar rates into sp² and sp³ sites. The fraction of sp³ sites increases if there is a preferential displacement of sp² atoms. This idea arose from some early estimates of the displacement threshold in graphite and diamond of 25 and 80 eV, respectively [15,103]. However, more recent direct measurements of the displacement threshold find similar values for graphite (35 eV) and diamond (37–47 eV) [104–107], so preferential displacement is not correct. It should be noted that the displacement threshold of graphite is quite anisotropic because of its layer structure, and Banhart [108] still takes the displacement threshold for sp² sites to be lower.

McKenzie et al. [51], and McKenzie [109] noted that sp³ bonded graphite occupies 50% more volume than sp³ bonded diamond. This leads to the phase diagram of diamond and graphite shown in Fig. 14, with diamond stable at higher pressures above the Berman–Simon line. McKenzie [109] and Davis [110] proposed that the role of the ion beam is to create a compressive stress in the film, which will move the film above the Berman–Simon line and so stabilise the high pressure diamond (–like) phase. Of course, in all these models, the phase once created is quenched into the growing film.

Robertson [99,100,111–114] proposed that the subplantation created a metastable increase in density, which tends to cause the local bonding to change to sp³ Preferential displacement is not needed. Only subsurface growth in a restricted volume is needed to get sp³ bonding. The difference between these viewpoints is discussed shortly.

Various numerical and analytical simulations prove the basic idea of subplantation [115–120]. The unsolved problem is in the details of the relaxation process, which suppresses sp³ bonding at higher ion energies and higher deposition temperatures.

Let us consider the atomic scale processes in more detail [110–113]. In the energy range of interest, 10–1000 eV, the carbon ions have a range of a few nm and they loose their energy largely by elastic collisions with the target atoms (nuclear stopping). The elastic collisions of ions in solids can be simplified to the binary collision approximation, in which the collisions occur as a sequence of
independent pair collisions. The cross-section of the collisions decreases as the energy is raised, as this is the repulsive part of the inter-atomic potential. Thus, an ion of zero energy incident on a surface sees an impenetrable wall of touching spheres. At a higher ion energy, the atomic radii decrease, so the interstices look wider. At some energy, the ion can pass through an interstice and so penetrate the surface layer. This ion energy is called the penetration threshold, $E_p$.

Another important ion energy is the displacement threshold, $E_d$. This is the minimum energy of an incident ion needed to displace an atom from a bonded site and create a permanent vacancy-interstitial pair. The surface of a solid acts like an attractive potential barrier of height $E_B$, the surface binding energy. This raises the kinetic energy of an ion by $E_B$ when it enters the surface. Thus, the net penetration threshold for free ions is

$$E_p \sim E_d - E_B$$

The surface binding energy equals the sublimation energy or cohesive energy, or 7.4 eV for C, so $E_d$ of 25 eV gives $E_p \sim 32$ eV.

Now consider carbon ions incident on an amorphous carbon surface [112]. A low energy ion will not have enough energy to penetrate the surface, so it will just stick to the surface, and remain in its lowest energy state which is sp$^2$ If the ion energy is higher than $E_p$ it has a probability to penetrate the surface, and enter a subsurface interstitial site. This will increase the local density. The local bonding will then reform around that atom according to this new density. The solid is amorphous, so the incident atom and the target atoms are equivalent. We assume that in the highly energetic conditions of ion bombardment existing during film growth, atomic hybridisations will adjust easily to changes in the local density, and become more sp$^2$ if the density is low and more sp$^3$ if the density is high.

As the ion energy increases further, the ion range increases, and the ion penetrates deeper into the solid. A rather small fraction of this energy is used to penetrate the surface, and another fraction of about 30% is dissipated in atom displacements [121]. The ion must dissipate the rest of this energy ultimately as phonons (heat). This whole process consists of three stages: (a) a collisional stage of $10^{-13}$ s; (b) a thermalisation stage of $10^{-12}$ s; and (c) a relaxation stage after $10^{-10}$ s. Process (b) and/or (c) allow the excess density to relax to zero, and cause a loss of sp$^3$ bonding at higher ion energies.

Consider an incident beam of flux $F$ with a fraction $\phi$ of energetic ions of energy $E_i$ (Fig. 15). Let a fraction $f$ of the energetic ions penetrate the surface. The non-energetic fraction of atoms or ions $(1 - f\phi)$ will just stick on the outer surface. Some of the penetrating ions will relax back to the surface. This flux is proportional to a driving force, the fraction of interstitials below the surface, $n$.

![Fig. 15. Schematic diagram of densification by subplantation. A fraction of the incident ions penetrate the film and densify it, the remainder end up on the surface to give thickness growth.](image-url)
Thus, in the steady state, we write that the fraction of ions remaining at interstitial sites to give densification is \( n = f\phi - \beta n \), where \( \beta \) is a constant. This gives

\[
\frac{n}{1 + \beta} = f\phi
\]

Thus, a fraction \( n \) of the beam becomes subplanted inside the film and a fraction \( 1 - n \) is left on the surface, as \( \text{sp}^2 \) sites. The subplanted fraction creates a density increment of

\[
\frac{\Delta \rho}{\rho} = \frac{n}{1 - n}
\]

which gives

\[
\frac{\Delta \rho}{\rho} = \frac{f\phi}{1 - f\phi + \beta}
\]

where \( \rho \) is the density of \( \text{sp}^2 \) carbon, \( \Delta \rho \) the density increase.

Fig. 16 shows that penetration can occur in two ways, either directly or indirectly by knock-on. Only knock-on penetration occurs for the case of ion assisted deposition. The penetration probability can be estimated as a function of ion energy by TRIM [122], and follows roughly

\[
f = 1 - \exp\left(-\frac{E - E_p}{E_s}\right)
\]

where \( E_p \) is the penetration threshold, and \( E_s \) is a constant (spread) parameter.

The first numerical models assume that relaxation occurs in the thermal spike stage of \( \sim 10^{12} \) s [111,112]. This gives a relaxation rate of \( \beta \approx 0.016(E_i/E_0)^{5/3} \), with \( E_0 \) a diffusion activation energy, and

\[
\frac{\Delta \rho}{\rho} = \frac{f\phi}{1 - f\phi + 0.016(E_i/E_0)^{5/3}}
\]

Fig. 16. Schematic of the basic processes in subplantation; direct penetration, penetration by knock-on of a surface atom, and relaxation of a densified region.
The thermal spike model proposes that all the excess energy of the ion is converted to thermal energy at a point or along a line (ion trajectory) and then the energy diffuses outwards by thermal diffusion, as if in a bulk medium. This creates an expanding front in which all atoms inside the front have a similar thermal energy, by equipartition. The temperature profile of a spherical spike is [123]

\[ T(r, t) = \frac{Q}{c[4\pi Dt]^{3/2}} \exp \left( -\frac{r^2}{4Dt} \right) \] (17)

where \( r \) is the distance from the impact, \( t \) the time from the impact, \( D \) the thermal diffusivity and \( c \) the thermal capacity of the carbon. \( Q \) is the ion energy \( E_i \) minus that lost in displacements [124]. We take \( Q \approx E_i \) for now. We assume that the thermal spike allows atoms to diffuse back to the surface to relax the density to the lower, stable \( sp^2 \) density. We assume that relaxation occurs by thermally activated diffusion at a rate of

\[ v = v_0 \exp \left( -\frac{E_0}{kT} \right) \] (18)

with \( E_0 \) the activation energy of atomic diffusion. The total number of hops of all the atoms within one spike is the integral over the spike volume and time,

\[ \beta = \int_{r_1}^{r_2} 4\pi n_0 r^2 \int_{t_1}^{t_2} v \exp \left( -\frac{E_0}{kT(r,t)} \right) dr \] (19)

where \( k \) is Boltzmann’s constant, \( n_0 \) the atomic number density, \( v_0 \) the phonon frequency, \( r_1 \) the minimum radius of the spike and \( t_1 \) the minimum time of the spike. It is useful to put everything in dimensionless atomic units, with an atomic radius of \( a \). We define reduced distances \( r' = r/a \), times \( t' = v_0 t \), and a reduced temperature \( \tau = kT/E_0 \). The heat capacity per atom is \( 3k \) so \( c = 9k/4\pi a^3 \), \( D \approx v_0 a^2 \), and \( n_0 = 3/4\pi a^3 \). Integration over \( r' \) is carried out and the integral over \( t' \) is converted to one over \( \tau \) to give

\[ \beta = \int_{\tau_1}^{\tau_2} \tau^{-8/3} \exp \left( -\frac{1}{\tau} \right) d\tau \] (20)

In the limits \( \tau_1 \to 0 \) and \( \tau_2 \to \infty \) this gives

\[ \beta = 0.016 p \left( \frac{E_i}{E_0} \right)^{5/3} \] (21)

where \( p \) is a dimensionless constant of order 1. All the other quantities cancel out. This is the equation used in Eq. (16).

Fig. 17 shows that Eq. (16) gives a good representation of the variation of density or \( sp^3 \) fraction for ta-C deposited by FCVA [17,112] with \( E_0 = 3.1 \) eV. This value is consistent with the thermal stability of deposited ta-C up to 1000 °C from Eq. (65). In Fig. 17, the increasing \( sp^3 \) fraction at low ion energy is controlled by the penetration probability \( f \), and the decline in \( sp^3 \) fraction at high ion energy is controlled by the relaxation. However, Eq. (16) would require a lower value of \( E_0 \) to represent the much shallower decline of \( sp^3 \) fraction at large ion energy found by MSIB by Lifshitz [7,8].

The penetration/spike model can also account for the density dependence found in ta-C:H deposited from acetylene. Fig. 18 shows how an incident molecular ion will fragment into two C
atoms on impact with the surface, sharing its kinetic energy equally between them. The hydrogens take little energy because of conservation of momentum. The penetration and densification will occur for each C atom separately and independently according to Eq. (16). A small energy must be subtracted from $E_i$ to break the C≡C bond in acetylene. In contrast, the relaxation step will occur as a single event for the molecular ion, as the spikes of the two C atoms will overlap. This will be described by the total energy of the molecular ion. Thus, Fig. 19 plotted on a scale of energy per C atom shows a much sharper decline in densification at high energies at in ta-C, where the incident ion is monatomic.

The evidence that there is a common relaxation stage is that the optical gaps of a-C:H grown from methane, acetylene and benzene have a similar dependence on bias voltage, while the density does not (Section 6.5). The optical gap depends on the configuration of sp$^2$ sites, and this is determined by the relaxation stage.

![Subplantation of molecular ion](image1.png)

Fig. 18. Subplantation by a molecular ion. The ion fragments on impact, and has separate penetration and combined relaxation stages.
Despite these superficial agreements between theory and experiment, this model of deposition is still deficient. A satisfactory model of growth must account for

1. The different dependence of \( sp^3 \) fraction on ion energy in the cathodic arc and in the mass selected ion beam, which has been attributed to the slower growth rate in mass selected ion beam.
2. The transition temperature to \( sp^2 \) bonding \( T_1 \), being 400–500 K [7,8,66], despite the temperature in a thermal spike being \( 10^9 \) K.
3. The variation of the transition temperature \( T_1 \) for \( sp^3 \) formation with ion energy [66].
4. The variation of \( T_1 \) with instantaneous growth rate [124,125].

There are many faults with the thermal spike concept above when applied to carbon deposition. A thermal spike is only truly valid for much heavier ions and higher ion energies, where the energy loss rate per distance (stopping power) is much higher, so the energy density is greater [126–128]. In carbon, energy equipartition does not really hold, so that the spike volume consists of a few excited atoms among much less excited atoms.

Hofsass et al. [121] has completely rearranged the thermal spike concept of deposition. He finds numerous small errors in the above analysis, such as equating \( Q \) to \( E_i \), which are correct. More importantly, he suggests that the \( sp^3 \) fraction varies simply with the number of hops per atom within the spike. The model above proposes that there is a driving force of penetration, forcing densification and \( sp^3 \) bonding, which becomes ineffective at higher ion energies due to relaxation. Penetration itself has no major role in determining \( sp^3 \) content in the Hofsass model. There is no driving force in his model, just a ‘relaxation’ from some undefined state into \( sp^3 \) bonding allowed by the spike. This is unphysical, the undefined state matters.

Koponen et al. [129] noted that amorphous solids tend to possess a range of activation energies for transport, which can give rise to a stretched exponential relaxation behaviour, with relaxation following a law of

\[
R = \exp \left( -\left( \frac{t}{t_0} \right)^\beta \right)
\]  

(22)
Koskinen and co-workers [124] noted that the dependence of the transition temperature $T_1$ on the growth rate implies that there must be overlapping cascades or spikes. Any valid model must include this. The condition for this is that the product of flux density, area and time should exceed 1,

$$\phi A \tau > 1$$

(23)

Let us consider this requirement in the thermal spike model. The hot region of a spike expands so that the temperature $T$ and radius $r$ obeys

$$T cr^3 = E$$

(24)

and diffusion gives

$$r^2 = 4D t$$

(25)

We take parameter values $v_0 = 10^{13} \text{ s}^{-1}$, $a = 10^{-10}$ m, $E_0 = 3 \text{ eV}$ and $E_i = 100 \text{ eV}$. A typical growth rate of 1 Å s$^{-1}$ is $\phi = 10^{-16}$ in these reduced units. Cascade overlap will occur at $r' = 100$, and $t' = 10^4$ or $t = 10^{-9}$ s. This gives $\tau = 3 \times 10^{-4}$ corresponding to 1 K. This corresponds to after the relaxation stage. This shows that this model cannot explain cascade overlap with its present growth parameters.

Fig. 20 shows that $d\beta/dt$, the integrand of (20), is a strongly peaked function, and that most hops occur for reduced temperatures near $\tau = 3/8$, that is at 1 eV (11,000 K). That is, they scale with the hopping energy $E_0$ (3 eV) not the ion energy [113]. This is for a single activation energy of 3 eV. It would require a spectrum of very low activation energies to allow hops at much lower temperatures compatible with cascade overlap.

It was previously proposed that the temperature effects might by explained by radiation enhanced diffusion effects [7,8,124]. The idea is that incident ions would create long-lived vacancies and interstitials whose recombination would cause the cascade overlap effect. The problem with this description is the definition of vacancies or interstitials in an amorphous network, and there is less likelihood of long-lived defect species. A range of activation energies is a more credible model for the amorphous state.

![Graph](image)

Fig. 20. Hopping rate in a thermal spike vs. reduced temperature [113].
Finally, McKenzie [109], McKenzie and Bilek [130] and Davis [110] have proposed the pre-eminent role of compressive stress in the formation of sp$^3$ bonding. This is based on the idea that a-C is under quasi-thermodynamic equilibrium, so the stability of sp$^2$ and sp$^3$ bonding in a-C follows the phase diagram of crystalline carbon. A minimum pressure or compressive stress above the Berman–Simon line is needed to stabilise sp$^3$ bonding. Deposition is clearly a non-equilibrium process, but parts of it can be described by thermodynamics. However, once deposited, a film cannot globally change from sp$^3$ to sp$^2$, it is quenched. There is a clear link between sp$^2$ bonding fraction and stress, as shown later. Molecular dynamics simulations of networks find that high sp$^3$ networks tend to be under compression [118]. However, recent data shows much less correlation between compressive stress and sp$^3$ fraction for ta-C films deposited at room temperature [131]. Recently, McCarty [132] criticised the thermodynamic analysis of how stress may align anisotropic materials, and suggested that the grown-in anisotropies arise from preferential displacement (ion-channelling) along the beam direction. Ketires [133] carried out molecular dynamics simulations and found the distribution of local stress and each atom site. He found that sp$^3$ sites tend to be under local compression and sp$^2$ under tension.

3.1. a-C:H

There are many processes in the deposition of a-C:H, as shown in Fig. 21. The strong dependence of the properties of plasma deposited a-C:H on the bias voltage and hence on the ion energy indicates that ions play a critical role in the deposition of a-C:H. In contrast to ta-C, the ion flux fraction is much less than 100% and may be typically 10% [4,84]. The a-C:H can be deposited from different source gases such as CH$_4$, C$_2$H$_2$, C$_2$H$_4$ and C$_6$H$_6$. The variation of the film density with bias voltage for each source gas can be redrawn on a scale of bias voltage per C atom in the molecule. When this is done, the maxima in density lie at a similar energy (see Fig. 89). This indicates that the action of ions is still via subplantation. This can be understood as follows. An energetic molecular ion incident at the film surface will break up into atomic ions and the energy will be distributed evenly. Thus, each atomic ion will subplant independently with that energy.

![Component processes in the growth mechanism of a-C:H.](image-url)
A complete model of the growth of a-C:H requires us to also describe the chemical processes of neutral species and of dehydrogenation [134–146], as well as the physical process of subplantation. There are three general stages of the plasma deposition; the reactions in the plasma (dissociation, ionisation etc.), the plasma–surface interaction and the subsurface reactions in the film. The plasma reactions are driven by the energetic electrons, as defined by the electron energy distribution (EED), Eq. (6). There will be other species formed in secondary reactions such as polymerisation. These tend to be less important at the lower pressures used for DLC deposition. Except in the case of high-density plasmas, the mass spectrometer analyses (Fig. 12) show that undissociated source gas molecules are still the dominant species in the plasma.

The plasma species incident on the growing film will consist of ions and neutrals. The neutrals will be closed shell molecules such as undissociated precursor gas, mono-radicals such as CH₃, di-radicals and other unsaturated species such as C₂H₄ or C₂H₂. The plasma also contains significant amounts of atomic hydrogen, H'.

It is known that neutral species contribute to growth because the mass deposition rate exceeds the rate due to ions alone. The first effect of note is that the growth rate decreases with increasing temperature. This was first thought to be because the neutrals were weakly adsorbed on the surface, and which would desorb at higher temperatures [134,135]. It is now known that this temperature dependence is due to the etching of the film by atomic hydrogen [139–141]. Growth itself is independent of temperature. The etching rate increases with temperature, so the net growth rate decreases with temperature (Fig. 22).

The contribution of each neutral species to the growth rate depends on their sticking coefficient. The a-C:H surface is essentially fully covered with C–H bonds, so it is chemically passive. Di-radicals and other unsaturated species can insert directly into surface C–C or C–H bonds, so these species react strongly with the film and their sticking coefficients approach 1. On the other hand, closed shell neutrals like CH₄ have very low sticking coefficients of under 10⁻⁴ and their effect is negligible. The monoradicals have a moderate effect. They cannot insert directly into a bond, they can only react with the film if there is an existing dangling bond on the surface. They will add to this bond to form a C–C bond. The dangling bond must be created by removal of an H from a surface C–H bond. This can occur by an ion displacing an H from the bond, or by an H atom abstracting H from the bond,

\[ \equiv \text{C} - \text{H}^+ \rightarrow \equiv \text{C}^+ \text{+ H}_2 \]  

(26)

Fig. 22. Schematic temperature dependence of net growth rate, actual growth rate and etching rate.
or by another radical like CH₃ abstracting H from the C–H bond. Measurements have found that atomic H' is the most efficient species for abstraction (30 times faster than CH₃) [144]. The CH₃ then adds to this dangling bond,

\[ \equiv C-\text{CH₃} \rightarrow \equiv C-\text{CH₃} \]  

(27)

Thus, the effective sticking coefficient of CH₃ is small, but it is high in the presence of atomic hydrogen [143,144]. This leads to a synergistic effect of H' on the sticking probability of CH₃ [144].

Neutral hydrocarbon species can only react at the surface, they cannot penetrate the film. Hydrogen atom and ions are different. H atoms being so small, they can penetrate about 2 nm into the film [145]. There, they can again abstract H from C–H bonds and create subsurface dangling bonds and H₂ molecules. Some of these dangling bonds will be re-saturated by incoming atomic H.

Ions can also penetrate the film. Carbon and hydrocarbon ions can cause subplantation. A more typical role of ions in a-C:H is to displace H from C–H bonds. This H can then recombine with other H' to form H₂ molecules, and desorb from the film. This is the main process which causes the H content of PECVD a-C:H to decrease with increasing bias voltage as shown later. Some of the atomic H' does not recombine, but finds dangling bonds to re-saturate.

Because of their low mass, hydrogen ions interact weakly with C atoms. Thus, H⁺ ions have the longest range and penetrate deepest into the film (Fig. 13). They undergo the same reactions as atomic H', but to a greater depth. Thus, von Keudell [145] notes that a-C:H films have three characteristic depths; the surface itself is controlled by reactions of hydrocarbon and hydrogen species; the upper 2 nm in which chemistry is controlled by reactions of atomic H', and a larger depth depending on ion energy in which reactions are controlled by H⁺ ions.

4. Atomic structure and characterisation

4.1. Bonding

Carbon can exist in three hybridisations, sp³, sp² and sp¹, as shown in Fig. 1. In the sp³ configuration, a carbon forms four sp³ orbitals, which make a strong σ bond to an adjacent atom. In the sp² configuration, a carbon atom makes three sp² orbitals, to form σ bonds, and the fourth p π orbital forms a π bond with a neighbouring π orbital. In the sp¹ configuration, there are two σ bonds along the ±z-axis, and there are pπ bonds in the y and z planes.

The σ bonds of all carbon sites and C–H bonds form occupied σ states in the valence band and empty σ⁺ states in the conduction band, separated by a wide σ–σ⁺ gap (Fig. 23) [3]. The π bonds of sp² and sp¹ sites form filled π states and empty π⁺ states, with a much narrower π–π⁺ gap [147,148].

![Fig. 23. Schematic DOS of a carbon showing σ and π states [1].](image)
Table 2
Comparison of characterisation methods for bonding in amorphous carbons and their advantages and their disadvantages

<table>
<thead>
<tr>
<th>Method</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>NMR</td>
<td>Large sample needed, C(^{13}), dephasing</td>
</tr>
<tr>
<td>Diffraction</td>
<td>Time consuming</td>
</tr>
<tr>
<td>ESCA</td>
<td>Small peak shifts, due to homopolar bonding</td>
</tr>
<tr>
<td>C–H modes, IR</td>
<td>Only sites bonded to H</td>
</tr>
<tr>
<td>(\bar{V}<em>d/\bar{V}</em>{\text{eff}})</td>
<td>OK if wide spectral range</td>
</tr>
<tr>
<td>Spectroscopic ellipsometry</td>
<td>Useable in situ, but small spectral range</td>
</tr>
<tr>
<td>EELS</td>
<td>Present method of choice, but destructive, time-consuming</td>
</tr>
<tr>
<td>Visible Raman</td>
<td>Indirect, sp(^3) sites invisible</td>
</tr>
<tr>
<td>UV Raman</td>
<td>Future method of choice</td>
</tr>
</tbody>
</table>

A very simple model of the atomic structure was developed some years ago, based on the properties of \(\sigma\) and \(\pi\) bonds [147,148]. It was argued that maximising the \(\pi\) bonding energy tends to cause sp\(^2\) sites to form \(\pi\) bonded clusters within a sp\(^3\) bonded matrix. The cluster size determined the band gap. It turns out that the model over-estimates cluster size [149–151], as discussed in Section 5. Nevertheless, it is a useful model for understanding the bonding in amorphous carbons.

The ternary phase diagram of the C–H system (Fig. 2) emphasises that two key parameters determine the structure and properties of DLCs; the fraction of sp\(^3\) bonded carbon sites and the hydrogen content. Structural characterisation of DLCs focuses strongly on these two parameters. The ordering of sp\(^2\) sites is a third significant factor, particularly for the electronic properties.

Various characterisation methods have been used to determine those structural parameters. One should distinguish between methods for detailed studies such as diffraction, and more routine methods for repeated structural monitoring which concentrate on the sp\(^3\) content and hydrogen content. Table 2 compares the effectiveness and disadvantages of various routine methods to determine the sp\(^3\) content and hydrogen content.

4.2. Diffraction

Diffraction by electrons, X-rays and neutrons is used to determine the atomic order of carbons over longer distances [16,19,152–162]. The structure factor \(S(k)\) is derived from the diffracted intensity \(I(k)\) by

\[
I(k) = S(k)Nf^2(k)
\]

where \(N\) is the number of atoms and \(f(k)\) is the atomic form for electrons, X-rays or neutrons. The key quantity is the radial distribution function (RDF), which is the probability of finding an atom at a given distance from a given atom. The reduced RDF, \(G(r)\), is derived from \(S(k)\) by a fourier transform.

\[
G(r) = \int_0^{\infty} k[S(k) - 1] \sin(kr) dk
\]

At large \(k\), \(S(k)\) consists mainly of a single sine wave, due to the sharp first neighbour peak. In practice, \(S\) is measured to a finite wavenumber \(k_{\text{max}}\). If the sinusoidal \(S(k)\) is terminated abruptly at \(k_{\text{max}},\) this will lead to ‘termination ripples’ in the RDF to a sharp cut-off when \(S(k)\) is Fourier transformed back into \(r\)-space. It is better to terminate \(S(k)\) by a Gaussian function to prevent the introduction of the termination ripples [152]. Electron diffraction can only be measured to a low \(k_{\text{max}}\) so it has only a low resolution and it is used for guidance only. Neutron diffraction is able to produce
very high resolution studies out to \( k_{\text{max}} = 50 \, \text{Å}^{-1} \) with the availability of modern spallation neutron sources. The only disadvantage of neutron diffraction is that the inelastic scattering by hydrogen (even in ta-C!) gives problems in setting the baseline of the RDF.

The RDF consists of a series of peaks from each shell of neighbours. In a-Si, the first peak is due to the first, directly bonded neighbours. The position of this peak gives the bond length \( r_1 \), and the area under this peak gives the coordination number \( N_1 \). The second peak is due to the second neighbours at distance \( r_2 \) and its area is the number of second neighbours \( N_2 \). While the first and second peaks are usually clearly separated by a region with \( G = 0 \), the higher order peaks are never well separated from the second peak. It is necessary either to assign intensity to higher order peaks, or to define the second peak by a symmetric Gaussian fitted to its leading edge. With these definitions, we note that the bond angle is given by

\[
r_2 = 2r_1 \sin \left( \frac{\theta}{2} \right)
\]

and that \([155]\)

\[
N_2 = N_1 (N_1 - 1)
\]

unless there are closed rings of under five atoms.

The RDF and structure factor are only one-dimensional representations of the network structure, which miss many of the higher order correlations \([152]\). They are therefore only partial discriminators of possible structures. Diffraction was popular during the early identification of DLC phases formed by new deposition methods \([24]\). It was often claimed that a diffraction pattern of some sample was similar to that expected for diamond. This can be dangerous and was usually untrue. Table 3 lists the main peaks in the diffraction patterns of diamond and graphite. It is clear that there is not so much difference; there are common peaks around 3.0, 5.1, 5.8, and 7.7 Å\(^{-1}\). The first peak in \( S(k) \) of graphite labelled (002) at 1.9 Å\(^{-1}\) is due to inter-layer scattering. This is only seen if there is a clear layer ordering in a graphic structure. The lack of a real distinction is also clear from comparing the \( S(k) \) of ta-C \([155]\) and sputtered \((sp^2)\) a-C \([19]\) in Fig. 24. The two curves have a similar peak structure up to 13 Å\(^{-1}\) with common peaks at 3.0, 5.1, 8.7, 10.3 and 13 Å\(^{-1}\). The expected peaks at 5.9 and 7.8 Å\(^{-1}\) are washed out in both cases, while the (200) peak of diamond at 3.5 Å\(^{-1}\), which could distinguish between them, is not apparent in ta-C. McCulloch \([162]\) distinguishes a-C from ta-C as the 8.7 Å\(^{-1}\) peak is higher than the 10.3 Å\(^{-1}\) peak in ta-C and the reverse in a-C.

The RDF can be compared to specific structural models or analysed on a more ad-hoc basis. The first continuous random network models of amorphous carbons were the hand-built models of Beeman et al. \([163]\). These used the Keating \([164]\) valence force field. This expresses the distortion

<table>
<thead>
<tr>
<th>Peak position (Å(^{-1}))</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graphite ((h k l))</td>
<td>1.88 (002)</td>
</tr>
<tr>
<td>Diamond ((h k l))</td>
<td>3.06 (100)</td>
</tr>
<tr>
<td>Glassy C</td>
<td>1.8</td>
</tr>
<tr>
<td>a-C (evap)</td>
<td>1.0</td>
</tr>
<tr>
<td>a-C (sput)</td>
<td>1.6</td>
</tr>
<tr>
<td>ta-C</td>
<td>–</td>
</tr>
</tbody>
</table>

Table 3
Comparison of peak positions in the structure factor \( S(k) \) of various disordered carbons
energies $\Delta E$ of the network of bonds as the sum of nearest neighbour bond stretching and bond bending forces,

$$\Delta E = \sum_i k_i \Delta r_i^2 + \sum_{ij} k_{ij} \Delta r_i \Delta r_j$$  \hspace{1cm} (32)

The four parameters $N_1$, $r_1$, $\theta$ and the density $\rho_0$ are sufficient to establish the nature of any amorphous carbon. These are compared in Table 4 for the main types.

Table 4
Comparison of structure parameters (bond length ($r_1$), first coordination number ($N_1$), second neighbour distance ($r_2$), number of second neighbours ($N_2$) and density for various forms of carbon)

<table>
<thead>
<tr>
<th>Form</th>
<th>$r_1$</th>
<th>$N_1$</th>
<th>$r_2$</th>
<th>$N_2$</th>
<th>Density (g cm$^{-3}$)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graphite</td>
<td>1.42</td>
<td>3</td>
<td>2.45</td>
<td>6</td>
<td>2.267</td>
<td></td>
</tr>
<tr>
<td>Diamond</td>
<td>1.54</td>
<td>4</td>
<td>2.512</td>
<td>12</td>
<td>3.515</td>
<td></td>
</tr>
<tr>
<td>Glassy C</td>
<td>1.425</td>
<td>2.99</td>
<td>2.45</td>
<td>6.1</td>
<td>1.49</td>
<td>[154]</td>
</tr>
<tr>
<td>a-C (evap)</td>
<td>1.43</td>
<td>3.3</td>
<td>2.53</td>
<td>8.8</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>a-C (sput)</td>
<td>1.46</td>
<td>3.34</td>
<td>2.49</td>
<td>6.7</td>
<td>2.44</td>
<td>[19]</td>
</tr>
<tr>
<td>ta-C</td>
<td>1.52</td>
<td>3.9</td>
<td>2.48</td>
<td>7.66</td>
<td>3.2</td>
<td>[156,184]</td>
</tr>
</tbody>
</table>
Glassy carbon consists of warped planes of graphite [3,157,158]. The ordering within the planes is high, locally retaining its planar six-fold rings. The inter-layer spacing at the van der Waals distance is retained, so the (002) peak remains in the $S(k)$. However, the layer warping means that the orientational ordering between layers is lost, giving a turbostratic structure with all $(hk2)$ peaks washed out.

The RDF of sputtered a-C indicates that it is still mainly sp$^2$ bonded [19]. The $N_1$ of 3.3 shows that some sp$^3$ sites present, as is confirmed by the nuclear magnetic resonance (NMR) spectra [165]. However, its structure differs from simple graphitic ordering because of topological disorder within the sp$^2$ layer. This introduces odd-membered rings, and these wash out the cross-ring peak at 2.84 Å of the ideal six-fold ring [19]. The first neighbour peak is also relatively broad, due to the disorder.

The RDF of ta-C was measured to $k = 25$ Å$^{-1}$ by neutron diffraction (Fig. 25). It is analysed in great detail by Gilkes et al. [156]. The first neighbour distance of 1.52 Å of the RDF is consistent with a high sp$^3$ fraction, as is the $N_1$ value of 3.90 and bond angle $\theta$ of 109$^\circ$. The first peak is also wide compared to a-Si, indicating the presence of large disorder, and the contributions of both sp$^3$ and some sp$^2$ sites. The first peak is so wide that the oscillations of $S(k)$ decay rapidly compared to those in a-Si, so that $S(k)$ measurements out to 25 Å$^{-1}$ are sufficient. However, there are problems with the second neighbour peak. Generally, the area of the first peak $N_1$ is not known to high accuracy (for the small samples used) because this requires a good normalisation of the total scattered intensity. The number of second neighbours is $N_2 = N_1(N_1 - 1)$. The ratio $N_2/N_1 =$
$N_1 - 1$ is a more accurate way to evaluate $N_1$. However, if the second peak area is derived, as usual, by defining a Gaussian from the leading edge, an unusually small $N_2$ value is found compared to the expected 12. Gilkes et al. [156] argued that this was because part of a third peak centred at 2.48 Å should be counted as being second neighbours, because these might correspond to intermediate hybridisation states sp$^{2+\infty}$.

This problem has now been resolved, and is due to the presence of small three- and four-fold rings of sp$^3$ sites, which were first noticed in the molecular dynamics simulations of Marks et al. [160,161]. Low order rings reduce the number of second neighbours below $N_1(N_1 - 1)$. Such rings are forbidden in a-Si, because they possess too large bond angle distortion. However, the bonding in carbon is different. It allows the presence of low-order rings, as in the molecules cyclopropane and cyclobutane. These have a stress much below that expected from the bond angle stiffness, such as in the Keating formula [164].

The structure factor of a-C:H was recently measured by Walters et al. [157] up to 50 Å$^{-1}$. The resulting RDF is the first to show a split first neighbour peak and to distinguish between the sp$^2$ and sp$^3$ bond lengths. An important result is that the bond length is closer to the shorter 1.34 Å C=C bonds in ethylene, rather than the longer 1.42 Å bonds of benzene. This suggests that the sp$^2$ sites in this a-C:H favour an olefinic over aromatic ordering.

In a second experiment on deuterium substituted a-C:H by Burke et al. [159], the three partial structure factors C–C, C–H and H–H could be resolved, to directly measure the location of H in a-C:H. The C–H bond length was found to be 1.06 Å, and the H–H correlations showed a broad distribution with no specific clustering of hydrogen.

4.3. NMR

The most direct measurement of the sp$^3$ fraction is by C$^{13}$ NMR [165–177]. The advantage of NMR is that each hybridisation gives rise to a separate chemically-shifted peak, with the same weighting factor for each peak. The sp$^2$ and sp$^3$ peaks are identified from molecular standards. However, large sample sizes are needed, unless C$^{13}$ enriched samples are used. Magic angle spinning can be used to remove the broadening effects in the spectra.

Fig. 26 compares the NMR spectra of a-C and ta-C. Pan et al. [165] found that sputtered a-C contained $\sim$7% sp$^3$ sites. Golzan et al. [166] found that their ta-C contained 75% sp$^3$ sites. The assignments in a-Cs have recently been confirmed by simulations [177]. Friedman et al. [170] found a higher sp$^3$ content for their laser deposited ta-C than Golzan et al. [166]. This is presumably because the sample of Golzan containing some unfiltered graphite inclusions from the FCVA process which raised their sp$^2$ value.

NMR has also been used to determine the C sp$^3$ fraction in a-C:H films [167–176]. An important contribution of NMR is to use proton decoupling or cross-polarisation methods to distinguish between the C sites with different numbers of bonded hydrogen neighbours. This was first used by Tamor and coworkers [173,174] to determine the development of bonding in a-C:H as a function of the deposition bias voltage. It was recently extended by Donnet et al. [176], to determine the amount of each CH configuration.

4.4. Density and X-ray reflectivity

The mass density of diamond is 50% larger than graphite, because the inter-layer bonding of graphite takes considerable space. This difference carries over into the amorphous phase, and indeed a linear correlation of density and sp$^3$ fraction is expected for a mixed sp$^2$ sp$^3$ network. Thus, the density
Fig. 26. NMR spectra of sp² a-C and ta-C, after Pan et al. [165] and Golzan et al. [166].

is a useful indicator of sp³ fraction, particularly in hydrogen-free carbons. The presence of hydrogen complicates matters, as CHᵢ groups also occupy a lot of space, so CHᵢ groups reduce the density. The density can be measured by weight gain during the deposition, by floatation or by X-ray reflectivity.

X-ray reflectivity is a technique widely used to analyse thin films and multilayers in the optoelectronics industry. It can be used to give a parameter-free measurement of the density of DLC films [178–186]. The refractive index of a solid for X-rays is slightly less than 1, so that X-rays have a critical angle for total external reflection. The critical angle is a glancing angle, ~0.20° for an X-ray wavelength of 1.39 Å. The refractive index depends on the density of all electrons, valence and core. This allows the mass density of a carbon film to be derived from the critical angle. In practice, the C film is on a substrate such as silicon. For a high density film like ta-C, the X-rays are reflected at the outer surface and only a single critical angle is seen [184]. On the other hand, if the film has a rather low density like evaporated a-C or a-C:H, then the critical angles of both film and Si are observed, and the overall reflection edge must be simulated to derive the film’s density [187,188]. Above the film critical angle, the reflected waves from the top surface and the film-substrate interface produce interference fringes, whose periodicity gives the film thickness.

In detail, the critical angle is given by

\[ \theta_c^2 = \frac{\lambda^2 N_A r_0}{\pi} \rho \sum \frac{X_i Z_i}{M_i} \]  

where \( \lambda \) is the X-ray wavelength, \( N_A \) Avogadro’s number, \( r_0 = e^2/4\pi\varepsilon_0 mc^2 \) the classical electron radius, \( \rho \) the mass density, \( X_i \) the atomic fraction of species \( i \) (e.g. C or H), \( Z_i \) the atomic number and \( M_i \) the atomic mass of atom \( i \) (in kg).

Fig. 27 shows some X-ray reflection curves for various DLC films on Si, showing the single critical angle for ta-C and ta-C:H, and the double critical angle for the less dense a-C:H and sputtered a-C films [184]. The presence of layering can also be detected from a beats pattern above the critical angle in a ta-C film. These data allowed us to obtain densities of 3.26 g cm⁻³ for ta-C with 88% sp³, 2.36 g cm⁻³ for ta-C:H with 75% sp³ (sp³ fractions from electron energy loss spectroscopy (EELS)). These values have been used to check and normalise density values previously determined by EELS from the plasmon energy (Fig. 20). Notably, the density of ta-C:H is much lower than previously
believed [95,184]. The maximum density for any form of a-C:H is about 2.4 g cm$^{-3}$, much lower than that of ta-C ($\sim$3.3 g cm$^{-3}$).

Extremely thin films can be measured by XRR [186]. Fig. 28 shows reflected intensity for ta-C films of 5–8 nm thickness. The X-rays now penetrate the film and are reflected from the Si substrate. The system shows only the critical angle of the Si. The film properties can be extracted by a fitting to the intensity decay above this critical angle.

---

**Fig. 27.** X-ray reflectivity intensity vs. angle of incidence for ta-C, ta-C:H and an a-C:H film on a Si substrate, after Ferrari et al. [184].

**Fig. 28.** X-ray reflectivity intensity vs. angle of incidence for ultra-thin films of ta-C of different thickness on a Si substrate [186]. Intensity scaled with (angle)$^2$ to show the spectrum tail. The samples are labelled by the deposition time.
4.5. Electron energy loss spectroscopy

EELS is presently the preferred method to derive sp\(^3\) fraction in DLCs [17, 47, 66, 189–196]. The EELS technique is explained in detail by Egerton [189]. An electron beam passing through a DLC film can be inelastically scattered. There are two energy loss regions of interest, the low loss region from 0 to 40 eV and the high loss region at the C 1s (K edge) at 285 eV and beyond. The high loss region is also observed by X-ray absorption using synchrotron radiation, where it is called X-ray near edge absorption spectroscopy (XANES). The advantage of EELS is that it works on thin films (10–20 nm), the disadvantage is that the film must be removed from the substrate, so the method is destructive.

Fig. 29 compares the C 1s EELS spectra of various carbon phases. The spectrum is really the excitation spectrum of the C is states into the empty conduction band states. It is roughly the partial p conduction band density of states (DOS), due to the s → p selection rule. The near edge consists of two features, a peak at 285 eV due to excitation to \(\pi^*\) states of sp\(^2\) (and sp\(^1\)) states, and a step at 290 eV due to excitation to \(\pi^*\) states of sp\(^3\) and sp\(^2\) sites [190, 192]. Diamond has no peak at 285 eV as seen in Fig. 18. The sp\(^2\) fraction of an arbitrary film can be found by taking the ratio of the area of the 285 eV peak to the area of the 290 eV step within some energy window, and comparing it to the ratio for a reference 100% sp\(^2\) sample of disordered graphite. Berger et al. [192] chose a window to 305 eV. Some workers take the \(\sigma^*\) area by fitting the leading edge of the 290 eV step with a Gaussian instead. A disadvantage is that K edge EELS only sees directly the sp\(^2\) sites.

A number of factors should be noted. First, the films should be 10–20 nm thick so that there is less need for correction for multiple scattering. Second, the \(\pi^*\) peak in graphite is excitonically enhanced compared to the simple \(\pi^*\) conduction band DOS. This has the effect of bringing oscillator strength down from higher energies, which is good because the band structure of graphite shows that the \(\pi^*\) band extends well beyond the onset of the \(\sigma^*\) band. On the other hand, this is not a constant effect, as there may be less excitonic enhancement in a-C.
A third point is that scattering from the $\pi^*$ peak is strongly polarised [197], with a selection rule of E\perp c for graphite. Thus, the reference sample should be randomly oriented micro-crystalline graphite. Alternatively, if a finite collection angle is used, then there is a combination of convergence and collection angles for which the $\pi^*/\sigma^*$ ratio has no dependence on the local orientation of the graphite plane. This is an important condition called the magic angle [184,198]. This is the recommended set-up for collecting EELS data. Not using this condition can lead to overestimates of the sp$^2$ fraction [199].

Finally, carbonaceous films such as fullerenes and hydrocarbons can produce peaks between 285 and 290 eV. $C_{60}$ has a more complex conduction band structure than graphite, which in effect splits its $\pi^*$ bands into a distinctive triplet peak [200]. Its $\sigma^*$ states start at 291 eV. The XANES of hydrocarbon molecules show additional peaks at 287–288 eV [201–204]. These are attributed to outer-well resonances [201]. It is not clear if these features transfer to the solid state. In practice, pure carbon phases like sp$^2$ a-C and ta-C possess a well resolved separated 285 eV peak, whereas in a-C:H the 285 eV peak is less well separated and spreads asymmetrically on its high energy side up to the 290 eV step. This requires more careful fitting to produce sp$^3$ values for a-C:H [205].

The low energy EELS has a broad peak corresponding to the plasmon oscillation of the valence electrons, as seen in Fig. 30. This may be used to derive the mass density [17,184,206]. This varies closely with the fraction for ta-C. The plasmon energy of a free electron gas (metal) $E_{p0}$ is given by [206]

$$E_{p0} = \hbar \left( \frac{Ne^2}{\varepsilon_0 m} \right)^{1/2}$$

where $N$ is the valence electron density, $e$ the electronic charge, $\varepsilon_0$ the permittivity of free space and $m$ is the free electron mass. If we assume four valence electrons per C atom contribute to the plasmon, then this gives us the mass density from $N$. In practice, the observed plasmon energy of

![Graph](image-url)

Fig. 30. Low loss electron energy loss spectra of various carbon phases showing the valence plasmon energies, after Waidmann et al. [196].
diamond of 34 eV is higher than given by (34), so \( m \) is often replaced by an effective mass \( m^* \). Fitting gives \( m^* \approx 0.87 \) m, and this value is used as a scaling factor for all a-Cs [17]. A number of issues must be clarified when using this method [184].

Firstly, the plasmon loss peak is actually quite broad, and follows the Drude formula [207],

\[
\text{Im}\left( -\frac{1}{\varepsilon(E)} \right) = \frac{E_p^2 \Gamma}{(E^2 - E_p^2)^2 + (\Gamma)^2}
\]

where \( \Gamma \) is the peak width. If \( E_p \) denotes the true plasmon energy, then the maximum of the peak actually lies below the true \( E_p \) by \( E_{\text{max}} = (E_p^2 - (\Gamma/2)^2)^{1/2} \).

A second, more critical correction is that the plasmon energy in a semiconductor in the Drude–Lorentz model is moved from the free electron energy \( E_{p0} \) to a higher energy \( E_p \) by the presence of the bonding–antibonding gap \( E_g \) [207]

\[
E_p^2 = E_{p0}^2 + E_g^2
\]

\( E_g \) is the average or Penn gap (10–13 eV), not the minimum optical gap [208,209]. This correction is a 10–15% effect. Now, \( E_g \) is also given in terms of the zero frequency dielectric constant \( \varepsilon_1(0) \) and refractive index \( n \) by [208,209]

\[
\varepsilon_1(0) = n^2(0) = 1 + \frac{E_{p0}^2}{E_g^2}
\]

Using \( E_p/E_{p0} = (m/m^*)^{1/2} \) with Eqs. (36) and (37), we find that

\[
\frac{m^*}{m} = 1 - \frac{1}{n(0)^2}
\]

Thus, \( m^* \) is really an inter-band effective mass which corrects for the presence of a band gap and a finite dielectric constant on \( E_{p0} \). The observed \( E_p \) should be multiplied by \((m/m^*)^{1/2}\) to get the free-electron value \( E_{p0} \) from which \( N \) and the mass density can be derived [184].

The value of \( m^* \) for any a-C or a-C:H can be derived from its refractive index \( n \). A constant value of \( m^* = 0.87 \) has been found to be valid for all hydrogen-free a-Cs by comparing the density from EELS and from X-ray reflectivity [186] (Fig. 31). In contrast, in a-C:H the refractive index falls for high hydrogen content in polymeric a-C:H. The \( m^* \) is no longer constant, and it should be estimated using Eq. (37). This means that the observed plasmon energy of a-C:H seldom drops below 22 eV. This does not mean that the density is still high, it means that the correction factor \( m^* \) has become small.

Fig. 32 shows a correlation of the sp\(^3\) mass density and the fraction for a number of DLCs derived from EELS in this way [184]. Note the density rises continuously with sp\(^3\) fraction for a-C, but it passes through a maximum for a-C:H. The maximum is because the hydrogen content is also rising, and high hydrogen contents strongly reduce the density.

4.6. Electron spectroscopy for chemical analysis (ESCA)

ESCA consists of measuring the shifts of core levels by X-ray photoemission spectroscopy (XPS). ESCA is chemically specific as each element has its own core level. The shifts reflect the chemical bonding of each site. The largest shifts are caused by the Coulombic potential from ionic
Fig. 31. Correlation of the reduced electron mass from valence plasmon and from X-ray reflectivity, leading to $m^*/m_e = 0.87$ for various carbon phases, after Ferrari et al. [184].

charges on adjacent atoms. For example, in SiO$_2$ the Si–O bond is polar, the Si 2p level is shifted by 4 eV from its position in Si, and Si sites with 1, 2, 3 or 4 oxygen neighbours can be resolved. Carbon is homopolar, so the shifts are small. There is a small shift of the C is core level shift of order 0.9 eV between diamond and graphite. This is due to the slightly shorter bond length of sp$^2$ bonds than sp$^3$ bonds [210] so that sp$^2$ sites experience a slightly deeper potential than sp$^3$ sites. This difference can be used to determine the fraction of sp$^2$ and sp$^3$ sites in a-C. Like NMR, an advantage of ESCA is

Fig. 32. Variation of sp$^3$ fraction with density for a-C and a-C:H phases [184].
that both sp² and sp³ sites are seen, and that their weighting is the same. However, disorder broadens the core levels by more than the sp² to sp³ shift, so the overall core spectrum must be deconvoluted to separate the two levels and to derive an sp³ fraction. Haerle et al. [210] showed that the core spectra do consist of just two lines, and that the sp² to sp³ core shift is constant across the a-C to ta-C series, and does not depend on the mean sp³ content. They also showed that although XPS is surface sensitive, the spectra are not unduly affected by the sp² surface layer seen by Davis [195].

4.7. Optical spectra and ellipsometry

It is often attempted to derive the sp³ fraction from the optical spectra in the energy range 1–6 eV [41,42,211]. This is based on the idea that the excitations of π and σ electrons are separate, and that the π excitations occur at a lower energy than the σ excitations. These ideas are described in more detail in Section 5.7 on optical spectra. In principle, the method can work when wide energy range (0–40 eV) spectra are available, by energy loss or vacuum-UV measurements [191,196,212]. However, this is rarely true. The more usual case is that the optical data is available only over a more limited range of say 1–6 eV, in which case there is often an over-estimation of the sp³ fraction, as in the case of Savvides [41,42].

A variant of this method is the spectroscopic ellipsometry method, which could even give a real-time or in-situ measurement of the sp³ fraction and thickness of the growing film. It is possible to obtain reasonable estimates of the sp³ fraction when compared to EELS values [213,214].

4.8. Raman

Raman spectroscopy is the best way to obtain the detailed bonding structure of DLCs. Raman is widely used, being a routine, non-destructive way to characterise the structural quality of diamond [215–217], graphites, DLCs and carbon nanotubes [14]. The Raman spectra of diamond, graphite and some disordered carbons are compared in Fig. 33. Diamond has a single Raman active mode at 1332 cm⁻¹, which is a zone centre mode of T₂g symmetry. Single crystal graphite has a single Raman active mode, which is the zone–centre mode at 1580 cm⁻¹ of E₂g symmetry labelled ‘G’ for ‘graphite’. (There is a second Raman active E₂g mode at 42 cm⁻¹ due to interplane vibrations.) Disordered graphite has a second mode at around 1350 cm⁻¹ of A₁g symmetry labelled ‘D’ for ‘disorder’. It corresponds to breathing vibrations of rings at the K zone boundary [218]. An unusual and significant fact is that the Raman spectra of most disordered carbons remain dominated by these two G and D modes of graphite, even when the carbons do not have particular graphitic ordering [219]. It is therefore of interest to explain this fact, and then if possible to find how Raman can be used to derive the structural information of DLCs and if possible their sp³ fraction.

Raman is light scattering by the change in polarisability χ due to a lattice vibration [220],

\[
χ(k) = χ_0 + \frac{dχ}{dq} Q(k, q)
\]  

(39)

where χ is the polarisability at wavevector k, and Q the amplitude of a phonon of wavevector q. This change in polarisability causes an inelastic scattering of an incident photon (ω, k) into the scattered photon (ω’, k’). (ω is the phonon frequency. The Raman cross-section can be expressed as

\[
C = k \left( \frac{∂χ}{∂ω} \right)^2
\]  

(40)
The polarisation can occur by excitation of the electronic ground state into virtual states at energy $E$ or into real states at $E$. The latter case is called resonant Raman [220].

In an amorphous material, there is a complete loss of periodicity and a breakdown of the $k$ selection rule of optical and phonon transitions. In this case, the IR spectra and Raman spectra of an amorphous network correspond to the vibrational density of states (VDOS) $G(\omega)$ weighted by the appropriate matrix element $C(\omega)$. This is the Shuker–Gammon formula for the Raman spectrum [221],

$$I(\omega) = \frac{n(\omega) + 1}{\omega} C(\omega) G(\omega)$$

(41)

The Raman and IR spectra should be relatively smooth and will resemble each other. This occurs in a-Si [221–223], but it is not true for the Raman spectrum of a-C. One reason for the dominance of the G and D modes is that the Raman spectra of a-Cs are dominated by scattering by the sp² sites. The π states are lower energy than the σ states and so they are much more polarisable [147]. This gives the sp² sites a 50–230 times larger Raman cross-section than sp³ sites [224–226], so they dominate the Raman spectra of even ta-C, which only has a residual 10–15% sp² content. Nevertheless, the Raman does not simply follow the VDOS of the sp² sites. The deeper reason is that the matrix element has a much stronger effect than in σ bonded networks. The Raman spectrum becomes controlled by the order of the sp² sites, not by the sp² fraction [219].

We note that the G mode is actually the stretching vibration of any pair of sp² sites, whether in C=C chains or in aromatic rings [219], as shown in Fig. 34. This occurs in ethylene as well as graphite (but at a higher wavenumber). Thus, G does not only mean ‘graphite’! The D mode is the breathing mode of those sp² sites only in rings, not in chains.
There are three basic causes of the unusual Raman behaviour of a-C [219].

1. Conjugated π bonds create long range polarisability (see Section 5). While the phonon spectrum of Si can be fitted by a nearest neighbour force field of Eq. (32), graphite requires forces up to 12th neighbours in conventional fits [227]. Recently, Mapelli et al. [228] made a force field based on the π bond order and polarisability. This includes long range forces, but they are each directly derived from nearest-neighbour Hückel interactions. Thus, a short-range field gives rise to long-range forces. The polarisability of π states is also long-ranged, and this gives the modes quite large effective charges.

2. Both the G and D modes are bond stretching modes, which have the largest matrix element for π states. Long-range polarisability further enhances their intensity. The D mode is particularly intense because it is a breathing mode of six-fold rings. In a graphite layer, there is constructive interference of the eigenvectors of each ring, and destructive interference for rings of other orders.

3. The D mode is a double resonance [229], as described in detail shortly.

These factors intensify the G and D breathing modes and tend to suppress modes of other symmetries. The variation of the intensity of the D mode with atomic order is interesting. Some years ago, Tuinstra and Koenig [218] noted that the intensity ratio of the D and G modes, \( I(D)/I(G) \), varies inversely with the in-plane correlation length \( L_a \) or grain size of the graphite (Fig. 35),

\[
\frac{I(D)}{I(G)} = \frac{c}{L_a}
\]  (42)

This means that \( I(D)/I(G) \) is proportional to the number of rings at the edge of the grain. It is clear that this relationship cannot extend down to zero \( L_a \). Recent data on the high temperature deposition of ta-C suggests that for \( L_a \) below 2 nm, the ratio decreases according to (Fig. 35) [219, 230];

\[
\frac{I(D)}{I(G)} = cL_a^2
\]  (43)

The G peak is due to all \( sp^2 \) sites, but the D peak is only due to six-fold rings, so \( I(D)/I(G) \) falls as the number of rings per cluster falls and the fraction of chain groups rises [219]. The important factor for DLCs is that \( L_a \) is always less than 1 nm, so that the Tuinstra–Koenig relationship is never valid for them, and Eq. (43) should be used instead.

We now consider the overall Raman spectra of disordered carbons. The difficulty of this task is summarised in Fig. 36, which shows the various factors which can shift the G and D peaks in either
direction and alter their relative intensity. One nice feature of Raman for visible photons is that it does not see the C–H bonds.

Before proceeding with a classification of the Raman spectra, it is important to note how the spectra were fitted, as this affects the numerical values. We fit the spectra with a skew Lorentzian

Fig. 35. Variation of Raman $I(D)/I(G)$ peak intensity ratio with in-plane correlation length $L_a$. The Tuinstra–Koenig relationship only applies for $L_a$ over 2 nm.

Fig. 36. Schematic of the factors affecting the position and heights of the Raman G and D peaks of non-crystalline carbons.
(otherwise known as a Breit–Wigner–Fano, BWF) line shape for the G peak and a Lorentzian for the D peak [219]. The BWF is given by

\[ I(\omega) = \frac{I_0 [1 + 2(\omega - \omega_0)/Q\Gamma]^2}{1 + [2(\omega - \omega_0)/\Gamma]^2} \]  

(44)

where \( I_0 \) is the peak intensity, \( \Gamma \) is the full width half maximum (FWHM) and \( Q^{-1} \) is the coupling or skewness coefficient. A symmetric Lorentzian corresponds to \( Q = \infty \). Note that because of skewness the maximum of the BWF occurs at

\[ \omega_{\text{max}} = \omega_0 + \frac{\Gamma}{2Q} \]  

(45)

which is what we use in the following analysis, as the nominal position \( \omega_0 \) has no fundamental meaning in Raman.

Ferrari [219] found that it is possible to classify the Raman spectra of all disordered carbons within a three-stage model of increasing disorder. The three stages are as follows (Fig. 37):

1. perfect graphite to nano-crystalline graphite;
2. nano-crystalline graphite to sp\(^2\) a-C;
3. sp\(^2\) a-C to sp\(^3\) a-C.

Stage 1 corresponds to the progressive reduction in grain size of ordered graphite layers, while keeping aromatic rings. The VDOS is of ideal graphite. As the grain size decreases, phonon confinement causes phonons away from \( \Gamma \) to participate with \( \Delta q = 1/L_q \). The phonon bands of graphite (Fig. 38) disperse upwards from 1580 cm\(^{-1}\) at \( \Gamma \) [227,231], so this causes an up-shift of the G peak to 1600 cm\(^{-1}\). The D mode is forbidden in an ideal graphite layer, but the disorder causes it

![Graph](image)

Fig. 37. Schematic variation of the Raman G peak wavenumber, and \( I(D)/I(G) \) ratio with degree of disorder, showing how this defines three stages of amorphisation of carbon, after Ferrari et al. [184].
to appear, and its intensity rises with decreasing $L_a$ according to the Tuinstra–Koenig relation (42). The crossover from stage 1 to 2 is seen in ion-irradiated graphite [232] as seen in Fig. 39.

Stage 2 corresponds to the topological disordering of a graphite layer (odd membered rings) and loss of aromatic bonding, but with a purely sp$^2$ network. The disorder and loss of aromaticity weakens the bonds and lowers the VDOS compared to that of perfect graphite [233]. This causes the G peak to shift downwards (Fig. 37). The $L_a$ is below 2 nm, so the $I(D)/I(G)$ ratio falls continuously to zero. The VDOS at the end of stage 2 corresponds to sputtered a-C [234].

In stage 3, the sp$^3$ content increases from 0 to 100%. This changes the sp$^2$ configuration from mainly rings to short chains [219,233]. The bond length of chains (olefins) is shorter than that of rings, so their vibrational frequency is higher, 1640 cm$^{-1}$ compared to 1580–1600 cm$^{-1}$. These changes are seen in the VDOS [235]. Thus, in stage 3, the G mode rises, while the D peak remains at zero intensity [219] (Fig. 40). The line-shape of the G peak becomes more symmetric as the sp$^3$ reaches high values [236]. Note that the maximum of the G peak shifts up with increasing sp$^3$ content, but it would apparently shift downwards if a symmetric fit or the uncorrected $\omega_0$ in (44) is used. The maximum sp$^3$ content corresponds to the most symmetric G peak [236].

This analysis allows us to say that if $I(D)/I(G)$ is near zero, we are in stage 3. The G position then varies with sp$^3$ fraction according to Fig. 40. This is a unique relationship, which can be used to derive the sp$^3$ fraction from the Raman spectrum.

The width of the G and D peaks scales with disorder. One way to know the correct regime is that if the FWHM of the G peak exceeds 50 cm$^{-1}$, then $L_a$ is below 1 nm. Fig. 41 shows that the G width varies in a power law fashion with $L_a$, using data for micro-crystalline graphite, disordered C and a-C, using data from Knight [215], Lespade [237] and Schwan [238].

Raman is widely used to characterise the bonding in a-C:H [4,239–244]. Typical spectra are shown in Fig. 42 for a-C:H deposited at increasing bias voltage. The bonding in a-C:H is described in
detail in Section 5. In conventional PECVD a-C:H, an increasing sp³ content correlates with an increasing H content. H saturates the C=C bonds, mainly converting sp² C sites into sp³ =CH₂ and =CH sites, rather than increasing the number of C–C bonds, as occurs when sp² a-C becomes ta-C. Thus, the change in the sp² configuration is not the same, and it is the sp² configuration, which sets the Raman spectrum [219]. The a-C:H with a narrow band gap has a large sp² content [1,4]. The sp² sites in this material are still predominantly organised in aromatic rings, as is evidenced by the sizeable D peak in their spectra. The a-C:H with maximum H content has a high sp³ content, and a wide gap of ~3 eV. The D peak is now absent, so the sp² sites are organised in very small clusters. Thus, a-C:H spreads from stage 2 to 3, and its Raman spectrum can be understood from this. Firstly, as a-C:H retains some aromatic ordering of sp³ sites, the increase in sp³ fraction has an alloying effect, which causes the G wavenumber to move down. Secondly, there is also a down-shift in the G peak at olefinic sites because H bonded to chains begins to resemble polyacetylene. This lowers the G frequency. Thus, the G peak moves down with increasing sp³ content in a-C:H (Fig. 43), the opposite direction to that in ta-C.

Tamor and Vassell [240] have observed the monotonic fall in the G peak with increasing optical gap. This is shown in Fig. 44. The optical gap of as-deposited a-C:H varies in a uniform way with its
Fig. 40. Variation of the Raman G position and \( I(D)/I(G) \) ratio for ta-C as a function of fraction, using the refitted data of Prawer et al. [236] and Ferrari et al. [184].

Fig. 41. Variation of the Raman G peak width with in-plane correlation length \( L_a \), using data from Lespade et al. [237] and Schwan et al. [238].
sp³ content [151]. It is therefore possible to combine these variations to give a variation of G position and I(D)/I(G) ratio with sp³ fraction as shown in Fig. 43. The four crosses represent reference a-C:H films whose sp³ fraction has been measured independently by EELS. It is seen that Raman has provided a good estimate of the sp³ fraction of these films [219].

The spectra of ta-C:H differ from a-C:H. The G peak of ta-C:H lies above that of a-C:H of a similar sp³ content or optical gap. This is because the sp² sites in ta-C:H are arranged more in short olefinic configurations than in a-C:H. This raises the G peak position, as seen in Fig. 45. The D peak is also very small in ta-C:H, showing the absence of many aromatic groups. Thus, the high degree of ionisation of the deposition process of ta-C:H has not only produced a highly sp³ bonded network like ta-C, but also a network with a similar sp² configuration to ta-C. Although it has not been proven directly by NMR, it is likely that unhydrogenated C sp³ sites exist in ta-C:H, whereas only hydrogenated sp³ sites exist in a-C:H.

4.9. UV Raman

Raman excited by visible photons at 514 or 488 nm is only sensitive to the sp² sites, because of their much greater cross-section than sp³ sites [224]. This is because visible light does not excite the
higher lying $\sigma$ states. This needs a higher energy. UV Raman at 244 nm (5.1 eV) provides an excitation of $\sigma$ states both $sp^2$ and $sp^3$ sites [245,246]. It at last allows the $sp^3$ sites to be directly observed [244–250].

Fig. 45 compares the UV Raman spectra of ta-C to its visible Raman spectrum. A new peak labelled T has appeared at 1050–1100 cm$^{-1}$. This wavenumber is close to the expected maximum of the VDOS of a fully $sp^3$ bonded random network. It is also close to the VDOS maximum found by inelastic neutron scattering of a PLD sample of ta-C [235]. This suggests that the T peak corresponds to the average vibrational DOS of the network, of the $\sigma$ bonds [245]. The simulations of Profeta and Mauri [250] on ta-C confirm that the T peak comes from $sp^3$ sites, and show the switch in excitation from $sp^2$ to $sp^3$ sites as the photon energy exceeds 5 eV.

The G peak is still present and strong. This intensity shows that the $\pi$ states of $sp^2$ sites still have a larger cross-section than $\sigma$ states, even for UV excitation. This is consistent with their high polarisability. The G peak is still present, but it has moved to higher wavenumber. The higher photon energy means not only that both $\sigma$ and $\pi$ states can be excited, but also that all types of $\pi$ states can be excited, those belonging to both wider gap clusters and narrower gap clusters. The G peak at 1620 cm$^{-1}$ lies above the band maximum of the graphite VDOS, and so it cannot be due to aromatic groups. It must originate from olefinic groups with their higher vibrational frequencies. This is the
first unequivocal evidence of the presence of olefinic groups in ta-C, first predicted some years earlier by simulations [233].

It is useful to correlate the UV Raman spectrum with the sp³ fraction. Three parameters have been studied; the position of the T peak, the position of the G peak and the intensity ratio of T and G peaks [247]. The T position could follow the mean VDOS with increasing sp² fraction and move up in energy. The G peak could move down in wavenumber as increasing sp² fraction changes the dominant sp² configuration from chains to rings. Both these ideas are borne out in practice. The simplest variation is that the I(T)/I(G) ratio increases in systematic fashion with increasing sp³ ratio. This is expected as the G peak is due to all sp² sites, rings or chains, while the T peak comes from all states. The variation is shown in Fig. 46 and it is non-linear [244,247]. The T peak apparently moves up with sp² bonding. In fact, it is more accurate to say that the T peak stays fixed and decreases in intensity, while a new peak at ~1380 cm⁻¹ appears and increases as the sp² fraction rises [244]. It turns out that the new peak is effectively a D peak, but due to all sp² sites (see Section 5.7).

The analysis has been extended to ta-C:H and a-C:H (Fig. 45). The UV Raman spectra of these phases show a broader T peak that has shifted down to 980 cm⁻¹. The residual of a D peak is still apparent at 1450 cm⁻¹. The I(T)/I(G) ratio of this falls reasonably well on the existing trend [244]. However, more work is needed on this aspect.
Fig. 45. Typical Raman spectra of ta-C, ta-C:H and an a-C:H film, taken at 514 and 244 nm excitation.

Fig. 46. Variation of the Raman peak ratio $I(T)/I(G)$ with sp$^3$ fraction for ta-C, after Gilkes et al. [247].
4.10. Resonant Raman and mode dispersion

In most semiconductors, the photon excitation is above the minimum gap, so most electronic states are excited. Carbon is really a wide-gap material; visible photons do not excite the \( \sigma \) states and they only excite the \( \pi \) states over a limited part of the Brillouin zone (see Section 5).

An interesting property of Raman modes in \( sp^2 \) carbons is that they are dispersive, that is their wavenumbers vary with the photon excitation energy. This is evidence that a mode is not simply a density-of-states feature, but that the matrix element changes with excitation energy because the excitation energy resonates with a real electronic transition. The Raman matrix element (42) is actually [220]

\[
C = \frac{\langle 0 | H_e | \omega_i \rangle | 0 | a \rangle \langle a | H_{e-ion} | a \rangle < a | H_{eR} | a \rangle^2}{(E_a - \hbar \omega_i)(E_a - \hbar \omega_k)}
\]

(46)

where \( H_{eR} \) is the interaction of electron–photon interaction, and \( H_{e-ion} \) is the electron–phonon interaction. The photon creates an electron–hole pair in state \( \omega \) at energy \( E_a \) above the ground state. There is a double resonance in (46) when \( E_a \) coincides with either the incoming or outgoing photon energy \( \hbar \omega_i \) or \( \hbar \omega_k \). Often \( E_a \) is a virtual state in the gap, or in a broad continuum of states in the conduction band. In graphite, the lower dimensionality means that the resonance condition will select specific states.

It has been realised that resonance causes the dispersion of the D peak [219,251–253]. It is now realised that the D mode of graphite is a double resonance condition [229]. A layer of graphite is a semi-metal, with the valence and conduction \( \pi \) bands touching at the K zone boundary. The \( \pi \) bands disperse linearly away from K, as shown schematically in Fig. 47. A photon of energy \( E \) incident on graphite will create an electron–hole pair at a wavevector \( k \) where the \( \pi \) gap equals \( E_i \). The electron can then emit a phonon of wavevector \( q \), scattering the electron from \( k \) to \( k - q \). If \( k - q = -k \), then the electron can be scattered back to \( k \) by the finite grain size or disorder and decay by emitting the outgoing photon. The resonant condition is therefore \( 2k = q \). Both the incident and outgoing photon correspond to the \( \pi \) gap. Raising the photon energy selects \( k \) vectors further away from K. At \( E_i = 0 \), the D mode is the singly degenerate \( A_{1g} \) mode. As the photon excitation energy increases from 0, the D mode moves away from K and its wavevector rises as shown in Fig. 42.

Resonance carries over into \( \pi \) bonded clusters. In general terms, incident photons will excite those clusters whose gap equals the photon energy. In fact the D mode of graphitic clusters behaves exactly like in graphite. A graphitic cluster is a finite piece of graphite. Its electron states are those of

![Fig. 47. Processes in the second-order resonant Raman effect creating the D peak in graphite; the linear electron band structure of graphite around K point, electron–hole excitations, and the phonon scattering, from Thomsen and Reich [229].](image-url)
graphite folded about the K point [147]. The phonon modes are found similarly. A cluster of grain size $L_a$ maps onto a wavevector $k$ along the $\Gamma$KM direction of an ideal graphite layer by

$$1 + 2 \cos(ka) = \frac{a}{L_a}$$

which is equivalent to $\sin(\Delta k a) = a/\sqrt{3}L_a$, where $\Delta k$ is the distance away from K. This means that the dispersion behaviour of an ensemble of graphitic clusters is exactly like that of a graphite layer [219].

We now turn to examples. Fig. 48 shows the example of the dispersion of the Raman spectra of ta-C with photon energy. Similar spectra are found for other a-C(H) phases. A strong dispersion of the G and D modes in a-C:H was previously found by Wagner [241] and Yoshikawa [243].

Fig. 49 summarises the dispersion of the G and D modes in various carbons [241–244], including data for graphitic carbons from Pocsik [252] and Vidano [254]. The G mode of graphite or even glassy carbon does not disperse. The G mode only disperses in stages 2 and 3 carbons. Indeed, the dispersion increases with disorder and is highest in ta-C. For sp$^2$ bonded a-C, the G mode rises with increasing photon energy and saturates at 1600 cm$^{-1}$, the band limit of the graphite VDOS. The dispersion of the G mode in ta-C is much more extreme [244]. It starts at low wavenumber 1500 cm$^{-1}$ for infra-red (IR) excitation, and rises all the way to 1690 cm$^{-1}$ for deep UV excitation at 215 nm. This means that 633 nm photons selectively excite clusters with weak $\pi$ bonding and relatively low vibrational modes. These may be aromatic in character, even in ta-C. Raising the
photon energy selects clusters of increasing olefinic (chain-like) groups with higher C–C bond stretching modes. These will be C=C pairs at the highest excitation energy. The G dispersion is roughly linear with excitation wavelength.

Various forms of a-C:H fall between these two cases [244]. A polymeric a-C:H film with a wide gap follows ta-C, with its G peak moving up to 1630 cm$^{-1}$ for 215 nm excitation. This shows that the sp$^3$ sites in polymeric a-C:H are mainly olefinic (chain-like). On the other hand, in diamond-like a-C:H with a 1.5 eV gap, the G peak rises only just above 1600 cm$^{-1}$, showing it has some olefinic content. In more graphitic a-C:H, the G peak will saturate at 1600 cm$^{-1}$, showing it has only aromatic sp$^2$ sites.

The D mode is very different to the G mode. The data is summarised in Fig. 49. The strongest dispersion is in the most well-ordered, graphitic carbons. The dispersion of the D mode increases with order, the opposite of the G mode. Of course the D peak is not found in the most random a-C. In diamond-like a-C:H, the D mode dispersion is quite small. The D peak has a second harmonic at 2700 cm$^{-1}$, which exists even in single crystalline graphite, and this also disperses.

The dispersion of the $I(D)/I(G)$ intensity ratio is also revealing [244]. The $I(D)/I(G)$ ratio decreases fastest for ordered carbons. Thus, the D peak is absent for UV excitation. However, the
ratio hardly disperses for very disordered a-C. What is happening is that, at high photon energies, the D peak now arises from all sp² sites in rings, not just aromatic clusters. Thus, for UV excitation, we can say that the T peak shifts up as the bonding changes from sp³ to sp²; but a better description is that the T peak decays and a ‘D-like’ peak appears at higher wavenumber, as discussed in Section 4.9.

In summary, dispersion of Raman features occurs because the excitation resonates with the band gap of a particular cluster, and this selects the corresponding vibrational mode. A shorter wavelength selects a smaller cluster with a wider band gap. Smaller clusters have a higher D and G mode, and this causes the modes to disperse. Shorter wavelength also reduces the intensity of the D peak, causing dispersion of the I(D)/I(G) ratio.

4.11. IR spectra

IR spectroscopy is another widely used technique to characterise the bonding in a-C:H. The IR absorption consists of C–H stretching modes at 2800–3300 cm⁻¹ and C–C modes and C–H bending modes below 2000 cm⁻¹, as shown in Fig. 50. The second derivative of the spectrum can be used to separate the peaks. The C–H bond bending and wagging modes overlay the C–C modes. Deuterium substitution can be used to confirm if modes are due to C–H vibrations—they should shift by a factor of square root of the mass ratio, or 1.4. This is useful for modes in the 1500 cm⁻¹ range.

The features in the C–H IR modes follow closely those of hydrocarbon molecules. The modes, particularly above 1340 cm⁻¹ are quite localised, so that their wavenumbers can be assigned reliably by comparison to their values in molecules. The standard assignment is that of Dischler [255,256].
Inelastic neutron scattering can be used to locate the inactive modes and check the assignments [1,257]. IR ellipsometry is a sensitive way to separate the components of the various bands [258]. Ristein [259] has traced the variation of mode wavenumbers with H content, and found little variation. However, a few corrections to the original assignments were noted. Table 5 gives the accepted assignments of modes in a-C:H, including these corrections.

We see that the C–H stretching modes break down into three regions, the \( \text{sp}^1 \equiv \text{C–H} \) modes centre on 3300 cm\(^{-1}\), the \( \text{sp}^2 = \text{CH}_n \) modes lie from 2975 to 3085 cm\(^{-1}\), and \( \text{sp}^3 \equiv \text{C–H}_n \) modes lie from 2850 to 2955 cm\(^{-1}\). It has been popular to fit the C–H bands with Gaussians to derive \( \text{sp}^3 \) fractions. The bands can be broad and the centres can drift, so the decomposition of the C–H stretching modes into the individual bands is not unique [259].

Note that the oscillator strength of the various C–H stretching modes is not constant. Thus, the C–H modes are not a reliable way to obtain \( \text{sp}^3 \) content [21,265], despite being popular [255].

Deuterium substitution is used to move the C–H modes out of the 1500 cm\(^{-1}\) range to reveal the underlying C–C stretching modes [211]. These are visible. The IR spectra showed clear evidence of both C=\( \equiv \) olefinic bonds in more polymeric a-C:H and more aromatic bonds in a-C:H deposited at higher bias voltages.

### 4.12. Hydrogen content

The hydrogen content of a-C:H can be determined by nuclear reaction analysis (NRA), elastic recoil detection (ERD), NMR, combustion, hydrogen evolution or IR spectroscopy. Nuclear methods
require special facilities but are widely used. They have the advantage that if used to determine the areal density of both H and C, then the desired H:C ratio can be derived without a measurement of the film thickness. The determination of H content by NMR requires the use of proton decoupling to separate the carbons bonded to C and H [173,176]. This method is rarely used because of the effort involved, but it has the unique ability to determine the fraction of each type of C elastic recoil detection H site, its hybridisation and number of bonded hydrogens.

The two methods mostly commonly used to determine hydrogen content in a-Si:H, thermal evolution [260,261] and IR, both have difficulty in a-C:H. In thermal evolution, a known weight of a-C:H is heated and the amount of H2 evolved is measured. This is complicated by the fact that H evolves from a-C:H not just as H2 but also as hydrocarbons particularly CH4 and C3H8, so that a complete mass spectrometer analysis is required [259,262]. This is then an accurate method.

In IR, the IR spectrum of the C–H bond stretching modes around 2900 cm−1 is measured, the background subtracted, the integrated weight under the curve is derived, and the number of H atoms per unit area N is determined by multiplying by an absorption strength, A.

\[ N = A \int \frac{\alpha(\omega)}{\omega} \, d\omega \]  

(48)

This is complicated because the absorption strength of the various types of C–H bonds can vary. There could also be a local field enhancement for bonds at voids as in a-Si:H. The smaller dielectric constant means that local field effects are not important in a-C:H. In a-Si:H, A was determined by calibrating the absorption against a hydrogen content known by NRA or ERDA. In a-C:H, Jacob and Unger [263] calibrated A against hydrogen contents derived from ERDA and found A to increase strongly from \( 2 \times 10^{20} \) cm−2 for polymeric a-C:H to \( 8.5 \times 10^{20} \) cm−2 for diamond-like a-C:H. This data is recommended for simple use. However, strictly, A varies because different C–H stretching modes have different absorption strengths and the fraction of each varies with total H content. The absorption strengths in hydrocarbon molecules are known to vary by an order of magnitude [259] with C–H bonds at aromatic sites having four times the absorption of sp3 C–H bonds. Thus, the C–H bond stretching band should be decomposed into its component absorption bands, and these each multiplied by the appropriate absorption strength \( A_i \). Ristein et al. [259] did this, using bond absorption strengths scaled from molecules. They found that H contents from IR compared well with those found from H evolution.

There is finally the question of whether all hydrogen is bound at C–H bonds. Some interstitial H2 molecules have indeed been observed by inelastic neutron scattering [257]. Grill and Patel [265] argued that in extreme cases, 50% of the hydrogen could be unbound. However more recent NMR data by Donnet et al. [176] found a much lower unbound H content and neutron scattering only found a small molecular hydrogen component. Thus, in general, this is a less critical issue.

The most optimistic use of the IR spectra is to use the C–H bond stretching modes to derive the fraction of sp3 to sp2 C bonding [255,266]. This is not recommended [263–265] as it will not detect sites not bonded to H. It is now known that essentially all sp3 sites in a-C are hydrogenated, but many sp2 sites are not [173,176]. The original mode assignments may have also caused an overestimate of sp3 content, because the modes at 2970 and 2945 cm−1 have reversed in assignment [259]. Now all modes below 2960 cm−1 are sp3.

4.13. Summary

The advantages and disadvantages of the various methods of structural characterisation of DLCS are summarised in Table 3. EELS is the preferred method for sp3 fraction, Raman can be used
with care, and UV Raman will be the preferred non-destructive method in the future. The hydrogen content should be measured by nuclear methods such as ERDA, or by gas evolution but only if all the evolving species are properly accounted for.

5. Electronic structure

5.1. Diamond and graphite

Diamond consists of a four-fold coordinated lattice of sp³ sites. Its band structure is shown in Fig. 51b. It has a large band gap of 5.5 eV, which is indirect. The direct gap at Γ is 6 eV.

Graphite consists of three-fold coordinated planar layers of sp³ sites, in an AB stacking sequence. The band structure of a single layer of graphite is shown in Fig. 51a. The σ and π states separate by symmetry, in a single layer. The σ states form valence and conduction bands which are separated by a 6 eV band gap at Γ. The π states form a zero gap semiconductor, in which the valence and conduction bands touch at the zone boundary, K. There is also a low lying conduction state at 4 eV at Γ, which is due to 3s-like higher lying inter-layer states [267].

Only carbon forms a stable graphite structure. It is often thought that this arises because the energy gained by stronger bonding out-weights the s–p promotion energy cost of forming the sp³ state. In fact, sp² and sp³ states both involve the promotion of one s electron to a p state, from the s²p² ground state to s¹p²π¹ in sp² and to s¹p³ in sp³. Thus, the s–p energy difference is not the cause. The cause is the absence of a 1p core state in the first row elements, so the 2p orbital lies closer to the nucleus, which makes 2pπ bonds proportionately stronger. The difference is also seen in the shape of the bond charge—it has a single peak in Si but is double-peaked in diamond [268].

5.2. The σ and π states

Before discussing the electronic structure in detail, it is important to notice a general difference in behaviour of σ and π states in carbon. The σ states form a network of two-centre bonds. It is always possible to express the occupied wavefunctions of such a set as the sum of

![Tight-binding band structures of graphite and diamond](image-url)
bond orbitals. A bond orbital is the symmetric combination of the two hybrids, which make up each bond [269],

$$\sigma = 2^{-1/2}(\varphi_1 + \varphi_2)$$  \hspace{1cm} (49)

The total energy and other properties of the occupied states of σ bonds such as the charge density and polarisability can then be written as the sum of independent, short-range terms for each bond [269].

The π states form π bonds. The bonds can be two-centre bonds like in ethylene. However, often a π orbital will form π bonds with more than one neighbour, like in benzene, or it can form a continuous linkage of π bonds as in polyacetylene or graphite. These bonds are called conjugated bonds. It is no longer possible to separate conjugated bonds into two-centre bonds [270]. The occupied states \( \phi_j \) can only be written in the general form

$$\phi = \sum_i c_i \varphi_i$$  \hspace{1cm} (50)

This situation gives rise to longer range forces and longer range polarisabilities. The long range polarisability causes the strong enhancement of the IR and Raman matrix elements for π states.

5.3. Cluster model

There have been numerous electronic structure simulations of models of amorphous carbon networks. Nevertheless, it is first useful to give a very simple description of the electronic structure and bonding. This is done by taking the Hückel approximation of treating σ and π states separately [3]. This is based on the observation that the σ and π states lie at different energies and that the π states lie in an orthogonal plane to the σ bonds at sp\(^2\) sites, so that their interaction is small [147,148].

The main effect is that π bonding favours a clustering of sp\(^2\) sites because of the energy gain compared to a random distribution of sp\(^2\) and sp\(^3\) sites. An arbitrary network of π states forms a half-filled band. Clearly, the binding energy of the occupied states is lowered if one creates a gap in this band at the Fermi level, \( E_F \). Further effects can be deduced by analogy to the behaviour of organic molecules [3,147].

- Any cluster will tend to have an even number of π orbitals, or otherwise there is a half-filled state at \( E_F \), which lowers the binding energy. The unpaired state at \( E_F \) is a coordination defect.
- A pair of sp\(^2\) sites tend to align their π states parallel, to maximise the π interaction, as in ethylene.
- The sp\(^2\) sites tend to form planar clusters, to maximise the π interaction.
- Three pairs of ethylene-like C=C units tend to group into a six-fold ring, as in benzene. This gains an aromatic stabilisation energy, as the π electrons are now delocalised over the ring.
- A further small energy gain occurs if the six-fold benzene rings condense into graphitic sheets.
- Odd-member or eight-fold rings on the other hand are not favoured, as these usually give rise to states near \( E_F \) which reduce the binding energy.

These rules led to the cluster model [3]. The cluster model gives a very simple view of the electronic structure. The model proposes that the sp\(^2\) sites tend to be arranged in planar, π bonded clusters of certain size, embedded in a sp\(^3\) bonded matrix. The arrangement of the sp\(^2\) sites controls the electronic properties and the optical gap, because their π states lie closest to the Fermi level, while the sp\(^3\) matrix controls the mechanical properties. The cluster model arises from the weak, long-range attraction of π states.
The band gap of the aromatic clusters has a very simple form. The most stable planar, compact clusters have a band gap given by \[3,147,\]

\[E_g \approx \frac{2\gamma}{M^{1/2}} \quad (51)\]

where \(\gamma\) is the nearest neighbour V(ππ*) interaction, and \(M\) the number of six-fold rings in the cluster. \(M\) can be related to the cluster diameter or in-plane correlation length \(L_a\) as \(M \approx L_a^2\), so that

\[E_g \approx \frac{2\gamma}{L_a} \quad (52)\]

Eq. (51) says that the band gap depends only on the \(\pi\) states of sp\(^2\) states and it varies inversely with the cluster diameter. This applies only to undistorted clusters. Thus, the gap depends only on the sp\(^2\) configuration, not on that of other atoms such as hydrogen [147].

The general view-point of the cluster model is borne out. However, the problem with the simple cluster model is that the clusters must be quite large to give band gaps of the size found by experiment. Such clusters are not found experimentally in as-deposited a-C and a-C:H. It is found that the sp\(^2\) sites were arranged mainly in small olefinic (chain-like) rather than graphitic clusters [271]. A more accurate schematic of a network is Fig. 52. It turns out that the cluster model provides a reasonable description of micro-crystalline graphite and thermally annealed DLCs. However, it greatly over-estimated the degree of order in as-deposited DLCs [150]. This is because the ion-dominated deposition processes needed to stabilise sp\(^3\) bonding also introduce a large disorder energy. This disorder energy is comparable to the stabilisation energy of forming larger clusters. This contrasts to the small disorder in bonded networks such as a-Si or a-Se.

5.4. Network calculations

Direct simulations on networks have produced valuable ideas of the electronic and atomic structure of amorphous carbons. The simulations can be divided into three types.

- Those using only inter-atomic potentials that do not explicitly contain the electrons.
- Those using empirical electronic potentials.
- Those with first principles electronic potentials.

The molecular dynamics calculations typically quench a periodic super-cell of atoms of given size from the liquid state. The cell size and atom density is fixed. The atoms vary their configuration—the bond lengths, number of neighbours, etc.—to determine the network structure.

![Fig. 52. Schematic of sp\(^2\) clusters in a-C:H.](image-url)
For carbon, simulations are carried out as a function of density, to determine how the sp³ fraction varies with density.

The first continuous random network models of amorphous carbons were those of Beeman et al. [163] using the Keating [164] valence force field of nearest neighbour bond stretching and bond bending forces,

\[ \Delta E = \sum_i k_i \Delta r_i^2 + \sum_{ij} k_{ij} \Delta r_i \Delta r_j \]  

(53)

This force-field describes well σ-bonded random networks such as a-Si, in which the network tries to minimise the bond length and bond angle distortions. Indeed, networks like a-Si are surprisingly close to the minimum energy configuration. For a-C, Beeman et al. [163] constructed models with three different sp³ and sp² fractions, in order to compare its effect on the RDF and structure factor. Robertson and O’Reilly [147] subsequently calculated the electronic DOS for these networks and found that the networks with sp² sites were metallic, in contradiction to experiment. This is because the Keating potential does not represent the π interactions between sp² sites correctly.

Following Stillinger and Weber [272], a number of atomic potentials were introduced for Si and C, which explicitly allow the coordination to vary, the so-called environmentally-dependent potentials. The first for carbon was due to Tersoff [273,274], and he used this to produce the first large-scale model of an a-C sp² structure. This model showed good agreement with the experimental RDF. The Tersoff potential is widely used in simulations of ion beam growth. However, it should be recognised that it still does not describe π–π interactions correctly. Recall that orientation of the π orbitals is needed to get a band gap. Brenner [275,276] was the first to correctly include the π–π forces, which tend to align the π orbitals. However, no networks have been built with this potential to the author’s knowledge. The Tersoff potential also does not include a repulsive term between different sp² layers, needed to give a graphitic inter-layer spacing. This was corrected by Heggie [277]. The Tersoff potential was widely used by Kelires [278,279] to produce a number of models with different sp³ fractions, and to calculate various elastic properties of the networks. The networks of Kelires [278,279] have produced valuable information on the elastic properties and stress in a-C, but because of this error, they still will not give correct electronic spectra.

There have been numerous simulations of the atomic structure using empirical electronic potentials. The simplest type is those of Wang and coworkers [280,281] based on tight-binding. These use a simple four orbital orthogonal basis set for the valence electrons, and retaining only two-centre interactions between first neighbour atoms. The notable feature is that the values of the two-centre interactions vary with the coordination number and distance following the recipe of Pettifor [282]. They calibrated their potentials on crystalline carbon phases calculated where necessary by first-principles pseudopotentials. Wang et al. [280] carried out their calculations on an sp² network and on a higher density, more sp³ bonded network [281]. Their sp³ network is generally regarded to have too many sp² sites for its density, and have too much sp² order.

The next degree of sophistication comes in the extensive simulations of Frauenheim and coworkers [271,283–285]. These use empirical potentials with a basis of four atomic orbitals per atom for the valence electrons and they retain all two-centre interactions, which are calculated within the local density approximation (LDA). The LDA is a way of converting the true many-electron exchange-correlation potential acting on the electrons into an effective one-electron potential. Frauenheim et al. used an empirical core-core repulsive term to give the correct bond length. This method is a good level of approximation to retain sufficient accuracy, while being able to treat large enough networks. These calculations were the first to show that the sp² sites in ta-C and a-C:H were
mainly organised in chains (olefinic) and not in rings (aromatic), in contrast to the cluster model [271]. They nevertheless found band gaps comparable with experiment. These results emphasised that it was possible to have small gaps even for small clusters. The smaller gaps arise because they are controlled by the distortions of the π bonding from the aligned configuration. Simulations were carried out for networks with a wide range of densities and hydrogen contents. A recent simulation has now found aromatic rings in low density a-C:H.

Chen and Robertson [286], and Robertson [287] carried out a similar style of simulation. This used a four orbital, numerical basis set of contracted atomic orbitals for the valence electrons, and it computed the exchange-correlation terms using the Harris functional in the Fast-Structure code. It covered a range of densities. The sp³ values found are similar to those found in the Frauenheim study.

Drabold et al. [288,289] carried out local orbital LDA calculations with a four orbital basis set. They focused on networks with high sp³ contents. They compared the behaviour of the four-fold coordinated sites in a-C to those in a-Si, and they showed how the pairing of sp² sites and the π bonding caused states to be cleared from the gap. In contrast, three-fold sites in a-Si tend to remain isolated and so always give rise to states at the Fermi level.

The highest level of calculation uses a large basis set, of either plane waves or localised orbitals. The first plane-wave pseudopotential LDA calculations were carried out by Galli et al. [290,291] on a-C at relatively low density. They used the Car and Parrinello [292] molecular dynamics method in which electrons and nuclei are treated on an equal footing, to increase the calculation speed. They found that a mixture of some sp¹ and sp³ sites in addition to the sp² sites, and that the sp² sites retained some distorted aromatic behaviour. They also studied liquid C [290,291], and later a-C:H [293].

Marks et al. [160,161] was the first to carry out high quality plane wave Car–Parrinello calculations on ta-C. Although these calculations were restricted to a small cell size, they found the unexpected result that a few sp³ sites formed low order (three-, four-fold) rings. Such rings are generally assumed to be forbidden in random networks of a-Si, as being too energetic. This is clearly true for the Keating potential. However, the calculations clearly show they exist in ta-C. Small closed rings do exist in the molecules cyclopropane and cyclobutane, so clearly some specific small bond angles are allowed in C systems. The main effect of low order rings is to increase the density for a given sp³ content. The small rings have little or no effect on the electronic structure, as they still give rise to states well away from E₀. These calculations were extended to larger networks, greater accuracy and other densities by McCulloch [162]. These networks should be considered to be the most reliable. Recently, Haerle et al. [210] generated networks by tight-binding and then optimised them by plane wave methods. They found similar results.

In a molecular dynamics simulation, a network must be able to sample many different configurations to reach the configuration of lowest energy, without being trapped in a metastable configuration by high energy barriers. This is very difficult for carbon, as the under- and over-coordinated transition states between metastable states have very high energies (because of strong σ bonding and the lack of d states). In theory or experiment, diffusion is difficult in carbon. Schultz and Stechel [294], and Schultz et al. [295] carried out LDA calculations of high accuracy on ta-C by using a large localised basis set. They showed that a large basis set was useful and was necessary to lower the energy of transition states for a network to pass between different configurations. The basis set has a strong effect on whether strained sp³ bonds remain or break to form sp² sites, and especially on whether order sp³ rings form or break to form sp² sites. Thus, the basis set quality has a significant effect on the key quantity derived from the simulations—the sp³ fraction. They found that the networks of Drabold et al. [288] could be relaxed further, with the conversion of some sp³ sites to sp² sites. It is likely that some other networks would suffer from these problems.
With this in mind, Fig. 53 compares the calculated sp$^3$ fraction versus density for the various simulations. This plot is to be compared with the experimental result shown in Fig. 32. The agreement is not too bad. The points of McCulloch [162] should be the most accurate. Frauenheim and coworkers [271,283] points are reasonable. Chen and Robertson [286], and Robertson [287] values have a slightly higher sp$^2$ fraction. The networks of Wang et al. [280] have too low a sp$^3$ content compared to other calculations. Those of Drabold [288] have a little too high sp$^3$ content (as expected from earlier comments). McKenzie [6] has previously suggested that EELS was overestimating sp$^3$ fractions compared to the NMR values. This may be so, but the new XRR data add to our confidence that the sp$^3$ versus density plot of Fig. 32 is now reliable.

There have been few simulations of the structure of a-C:H [293]. Recently Bilek et al. [296] found a network with surprisingly graphitic ordering of sp$^2$ sites.

5.5. Photoemission

The electronic DOS for three random networks with different density and sp$^3$ content are shown in Fig. 54, as calculated by Chen and Robertson [286], and Robertson [287] using the local orbital method. It is seen that the σ and σ* states are separated by a wide band gap while the π and π* states lie around the Fermi level, as required in the simple description of the DOS in Fig. 23. These densities of states spectra can be compared to experiment. The valence band DOS was measured by photoemission. Photoemission spectroscopy gives the valence DOS weighted by the appropriate orbital cross-section for s or p states. In a-C:H, the spectra are dominated by C states because the cross-section of H states is small. Varying the photon energy in photoemission varies the relative cross-section of carbon 2s and 2p states. Ultra-violet photoemission (UPS) with 20 or 40 eV photons is dominated by the C 2p states. X-ray photoemission with 1486 eV photons is dominated by the C 2s states. Synchrotron photoemission at 100 eV gives equal weight to 2s and 2p states.

Fig. 55 shows the spectra with 20 eV photons for diamond, graphite, a-C and a-C:H from Wesner et al. [297]. The a-C:H sample was plasma deposited. The a-C sample is an annealed a-C:H sample. The spectra show a large upper peak centred on −7 eV and a wide sometimes split peak of s
Fig. 54. Calculated DOS spectra for networks of 20, 69 and 92% sp³ bonding. Data from Chen and Robertson [286].

states at $-11$ to $-20$ eV. The spectra show the splitting of the s band in graphite and diamond at $-15$ to $-20$ eV, also found in the calculated DOS. This splitting is retained in a-C, which shows that this a-C network contains mainly six-fold rings. Odd membered rings are known to wash out the split s-band spectra for all networks.

The upper p-like peak centred on -7 eV is more notable. It mainly contains the mass of σ-like p states. It also contains the π states of sp² states, which can be seen as a knee in the leading edge of the spectra of graphite and a-C. The states are less obvious in the a-C:H spectrum.

Ugolini et al. [298, 299] found the UPS spectra of a-C:H deposited by ion beam from various source gases as a function of ion energy (Fig. 56). Panels 1 and 2 show the spectra against ion energy for films deposited from ethane and benzene. They show the increasing size of the knee feature at $-2$ eV, which is due to π states. The spectra of films deposited from benzene and styrene show strong peaks from the precursor molecules. This is a strong indication that the molecules do not get dissociated at the film surface at low ion energy.

Schafer et al. [300] measured the XPS spectra of ta-C, ta-C:H and a-C:H and compared these to calculated spectra. The ta-C and ta-C:H showed some splitting of their s band peak at $-16$ eV. Schafer et al. [301], like Ugolini et al. [298, 299], also studied the development of the UPS spectra of a-C:H as a function of the bias voltage during deposition. They were able to extract the π state DOS from the 2p peak, and they could correlate this with the bias voltage. This showed the expected increase in π states and sp² fraction with bias voltage.
The conduction band DOS is seen by EELS or near edge X-ray absorption spectroscopy (NEXAFS). These spectra have already been discussed in Section 4.5.

5.6. Optical gap

The band gap is determined by the configuration of $\pi$ states on the $sp^2$ sites. In the cluster model with planar clusters, the band gap of a given cluster is given by \[ E_g = \frac{2\gamma}{M^{1/2}} \] (54)

where $\gamma$ is the $V(\pi\pi)$ interaction and $M$ the number of rings in the cluster.

It is now known that the cluster model grossly over-estimates the size of the clusters. The band gap is now controlled by the distortions of the cluster. It is no longer possible to give a simple formula for the gap, corresponding to (54). In calculations on networks, Frauenheim and coworkers [283], Chen and Robertson [286], and McCulloch et al. [162] all found that the band gap increased with decreasing $sp^2$ content as shown in Fig. 57. The presence of hydrogen has little direct effect as the C–H states lie well away from the band gap region.

The band gap in a crystal is defined as the minimum energy gap between the occupied and empty states, which can be either a direct or indirect gap. In an amorphous semiconductor where there is no true gap, an arbitrary definition must be used for the optical gap. Experimentally, the two most common are the $E_{0\alpha}$ gap defined as the energy at which the optical absorption coefficient $\alpha$ is $10^{-4}$ cm$^{-1}$, and the Tauc gap defined as the intercept $E_g$ found from plotting \[ \alpha E = B(E - E_g)^2 \] (55)
Fig. 56. Photoemission spectra of a-C:H films deposited at different ion energies after Ugolini et al. [298,299].

Fig. 57. Calculated variation of band gap with sp² fraction, after Chen and Robertson [286].
Fig. 58. Experimental variation of optical gap vs. sp² fraction for a-C:H [98,168,173,240], ta-C:H [95] and ta-C [66].

Fig. 58 plots the variation of experimental values of Tauc gap against sp² fraction for a-C:H, ta-C:H and ta-C films [151]. This is seen to have a similar dependence as the theory in Fig. 57. Note that ta-C, ta-C:H and a-C:H show a similar dependence. The gap tends to high values at low sp² content.

Tagliaferro and coworkers [302–305] developed a simple model of the optical transitions. They attributed the increase in gap with decreasing sp² content in Fig. 58 to a decrease in the width of the π and π’ bands, with a constant π–π’ separation. In detail, of course, the gap does depend on the configuration of sp² sites, which need not vary just with the sp³ content. Thus, the gap does not vary with sp² content in every situation, such as in annealing, or for high temperature deposition. Raman best measures the sp² configuration, and this allows this dependence to be determined in detail.

Although many of the details of the cluster model are wrong, it is still correct in viewing the network of an amorphous carbon as consisting of an inhomogeneous mixture of sp² and sp³ sites. Consider the local band gap in this network as a function of distance [306], as shown in Fig. 59. The sp³ sites have a wide gap, between the σ and σ’ states. The sp² sites have a variable gap, which depends on the configuration of each sp² cluster. The sp³ matrix acts as a tunnel barrier between each sp² cluster. The distribution of sp² gaps creates inhomogeneous disorder. The band gap of a sample is given by the average band gap of the clusters in it. The range of cluster sizes will give rise to a broad absorption tail, not a sharp band gap. This is indeed seen experimentally.

Fig. 59. Schematic variation of the local band gap with distance in cluster model.
Fig. 60. Optical absorption edges of a sputtered a-C, ta-C and an a-C:H film [211], compared to the sharper edge of a-Si:H.

The optical absorption spectrum of all forms of amorphous carbon shows a long tail, as seen in Fig. 60. The edge can be analysed in different ways. It has recently been found that it fits a Gaussian model [211,302–305,311]. In this, the valence and conduction $\pi$ band DOS are represented by Gaussians, the optical absorption is calculated and fitted to the experimental curves. The separation of two bands is denoted by $2E_{\pi}$. Values of this average $\pi$ gap, the $E_{04}$ gap and the Tauc gap are compared in Fig. 61 for some a-C:H films [211]. It is seen that $2E_{\pi}$, roughly equals twice the Tauc gap. This result is because the fit was weighted towards the absorption edge. If the fit is to the band states, then a more constant value of $2E_{\pi}$ is found [302–304]. A constant gap is more consistent with the calculated DOS of $\pi$ states [283,287], as in Fig. 54.

An alternative way to analyse the absorption edge is to fit it with an exponential slope, to extract the so-called Urbach energy $E_0$. The Urbach energy is one of the standard measures of electronic disorder in amorphous semiconductors. Although the optical edge does not always fit an exponential slope accurately, $E_0$ is a useful value for comparisons. Fig. 62 shows values of $E_0$ plotted against the Tauc gap for a-C:H films deposited by various groups [307–312]. It is clear that a-C:H has a large

Fig. 61. Comparison of optical gaps in a-C:H, the $E_{04}$ gap, Tauc gap and $\pi$ gaps, from Theye et al [211].
disorder, as the Urbach energy of good a-Si:H is about 55 meV, whereas that of a-C:H never falls below 150 meV.

Another measure of disorder in amorphous semiconductors is the bond angle distortion. In a-Si, the width of the main Raman TO band was found to be proportional to the bond angle distortion [313]. The equivalent in a-C is the Raman G mode. The Raman G width has been found to vary linearly with the compressive stress in as-deposited films [240], and so it is proportional to the bond angle distortion. Fig. 63 compares the variation of Urbach energy and the width of the Raman G peak [240] with the optical gap. It is seen that both factors increase together for small optical gap.

The Urbach energy then continues to increase, while the Raman G width decreases (together with the stress) in the wide gap, polymeric a-C:H films.

It is clear from Fig. 63 that the Urbach energy and the Raman G width are different measures of disorder. The G width is a measure of the homogeneous disorder and bond angle distortions within

Fig. 62. Variation of Urbach energy $E_0$ with $E_{04}$ optical gap for a-C:H and ta-C: a-C:H data from Dischler et al. [307], Schutte et al. [308], Xu et al. [309], Rusli et al. [310], and Paret et al. [311]; ta-C data from Kleinsorge, private communication.

Fig. 63. Variation of Urbach energy $E_0$ with Raman G linewidth for a-C:H against optical (Tauc) gap. Raman data from Pan et al. [165].
the network. The Urbach energy is a measure of the inhomogeneous disorder in the two-phase network, as it measures the range of \( sp^2 \) cluster sizes present. Thus, the homogeneous disorder in a-C:H reaches a maximum in the most ‘diamond-like’ a-C:II at a band gap of about 1.4 eV, where the density, C–C bonding, mechanical hardness and stress all have a maximum. The homogeneous disorder then decreases to quite small values in polymeric a-C:H. In contrast, the inhomogenous disorder in a-C:H increases continuously with increasing optical gap. This means that the cluster size gets small as the gap gets wider, but the range of cluster sizes remains large in the wide gap a-C:H.

5.7. Optical and energy loss spectra

The optical spectra of amorphous carbons provide valuable information on their local electronic structure. The complex dielectric function is defined in terms of its real and imaginary parts as follows:

\[ \varepsilon = \varepsilon_1 + i\varepsilon_2 \]

\( \varepsilon_1 \) and \( \varepsilon_2 \) are then given by

\[ \varepsilon_1 = n^2 - k^2 \]

and

\[ \varepsilon_2 = 2nk \]

where \( n \) is the refractive index, \( k \) the extinction coefficient, \( k = \alpha/4\pi\lambda \), where \( \alpha \) is the optical absorption coefficient and \( \lambda \) the wavelength. The variables \( \varepsilon_1 \) and \( \varepsilon_2 \) are related by the Kramers–Kronig relationship

\[ \varepsilon_1(E) = 1 + \frac{1}{\pi} \int_0^\infty \frac{\varepsilon_2(E')}{E - E'} \, dE' \]  

and \( \varepsilon_2 \) is given by

\[ \varepsilon_2(E) = \frac{2\pi e^2}{N} \int_0^\infty R^2(E)N_v(E')N_c(E + E') \, dE' \]

where \( N \) is the atomic density, \( R(E) \) is the distance dipole matrix element, and \( N_v \) and \( N_c \) are the valence and conduction band DOS, respectively. \( R(E) \) is generally reasonably independent of energy \( E \).

The optical spectra for most materials are acquired by direct transmission–reflectivity experiments. In Si, the main bonding–antibonding transitions give rise to the \( E_2 \) peak at around 4.2 eV. The peak falls to 3.3 eV in a-Si due to the loss of the k-selection rule. The spectral features fall in the accessible 1–5 eV energy range of transmission–reflectivity measurements. However, carbon is unusual in that its strong bonds cause its spectra to span a much higher energy range than usual, 2–25 eV, and the important features lie above 5 eV. This means that the spectra must be measured on vacuum UV [211,314] or energy loss systems [196,206]. Energy loss measures

\[ \text{Im} \left( \frac{-1}{\varepsilon} \right) = \frac{\varepsilon_2}{\varepsilon_1 + \varepsilon_2} \]
Just like $\varepsilon$, the real and imaginary parts of $\varepsilon^{-1}$ are related by Kramers–Kronig transformations. This allows the calculation of $\text{Re}(\varepsilon^{-1})$, and from these $\varepsilon_1$ and $\varepsilon_2$ can be found, following Raether [207]. This requires that the integration limits are known, so the refractive index is also needed.

Fig. 30 compared the energy loss spectra of diamond, graphite, ta-C and a more sp$^3$ a-C deposited at 10 eV [196]. The spectrum of diamond shows the plasmon of all the valence electrons at 34 eV. The spectrum of graphite shows the plasmon of all its valence electrons at 27 eV, plus a plasmon of its $\pi$ electrons at 7 eV. The plasmon of all the valence electrons has been discussed previously. It can be used to obtain the valence electron density, and thereby the mass density [184]. The $\pi$ plasmon at $\sim$7 eV only appears if the sp$^3$ sites are well ordered in for example layers, otherwise it is washed out. The energy of the $\pi$ plasmon is not a good way to obtain the density of $\pi$ electrons [206,314], and thus the sp$^3$ fraction, because it is displaced upwards by a zero-crossing of $\varepsilon_1$.

Figs. 64 and 65 compare the $\varepsilon_1$ and $\varepsilon_2$ spectra of diamond, graphite, ta-C and an a-C:H film [191,196]. The $\varepsilon_2$ spectra of graphite and a-C:H show two peaks, one around 4 eV due to $\pi$ excitations and one about 13 eV due to excitations. Diamond shows only a single peak at 13 eV due to $\sigma$ excitations. The ta-C shows evidence of its high sp$^3$ fraction, as it has only one main peak at 9 eV due to $\sigma$ excitations [77,196], although there is now a weak shoulder at 6 eV due to $\pi$
excitations. The energy of the main $\varepsilon_2$ peak corresponds to the ‘average’ or Penn gap [208,209]. The peak at 9 eV is lower than in diamond (3.6 eV) because of the loss of the k-selection rule which allow transitions to occur at lower energy. This also occurs in a-Si where the main peak lies below the $E_2$ peak of c-Si. The position of the $\sigma$ peak for ta-C depends sensitively on the value of the refractive index used in the Kramers–Kronig analysis, if the spectrum is derived from energy loss.

The $\varepsilon_2$ spectrum of graphite shows the expected divergence at zero energy of a metal. The $\varepsilon_2$ spectra of the other carbons, diamond, ta-C and a-C:H is zero below their band gap.

The excitations of the $\sigma$ and $\pi$ states show up as two relatively separate contributions in the $\varepsilon_2$ spectra of carbons with sp$^3$ sites [191,211]. This is because $\sigma \rightarrow \pi^*$ and $\pi \rightarrow \sigma^*$ transitions are weak for planar sp$^2$ sites, so the $\sigma$ and $\pi$ excitations are effectively decoupled in the optical spectra. This separation is particularly marked for graphite [314]. The separation is further shown by calculating the effective number of electrons participating in the spectrum up to a given energy $E$,

$$N_{\text{eff}}(E) = \frac{m}{2\pi^2N\varepsilon_2^2\hbar^2} \int_0^E E' \varepsilon_2(E') \, dE'$$  \hspace{1cm} (62)

The $N_{\text{eff}}$ spectra of graphite, diamond, ta-C and an a-C:H film are compared [191,196] in Fig. 66. The $N_{\text{eff}}$ of graphite shows a plateau at around 8 eV, which marks the separation of $\pi$ and $\sigma$ excitations.

The idea is to find the ratio of sp$^2$ to sp$^3$ sites by finding the ratio of $\pi$ to $\sigma + \pi$ excitations evaluating $N_{\text{eff}}$ at 8 and at say 40 eV, well beyond the completion of $\sigma$ excitations [191]. The problem is that the $N_{\text{eff}}$ plateau at 8 eV in graphite corresponds to only 0.6 electrons per atom, not one electron per atom as it should [315,316]. Nevertheless, the idea of deriving an sp$^2$ fraction from the wide band optical spectra is appealing. Instead of taking the absolute value of $N_{\text{eff}}$ (6 eV), Fig. 67 shows the ratio of $N_{\text{eff}}$ (6 eV) for a ta-C film with its ratio in graphite, and compares it with the sp$^2$ fraction determined directly from the K edge EELS [196]. It is seen that this definition provides a reasonable estimate of the sp$^2$ fraction.

This method appears to work satisfactorily when the wide band spectra are available. The problem is that the more usual situation is that only the narrow band optical spectra up to say 5.5 eV are available, and groups attempt to derive an sp$^2$ fraction from these [41,42,212]. Such spectra only cover the $\pi$ excitations. The size of $\sigma$ excitations is not known.

![Graph](image)

Fig. 66. Comparison of effective number of electrons, $N_{\text{eff}}$, spectra for graphite, a-C:H, ta-C and diamond, after Waidmann et al. [196].
6. Bulk properties of DLCs

6.1. The ta-C

The properties of ta-C have been studied intensively in the literature [6, 17, 18, 51, 52, 63–82, 317–341]. The ta-C is usually deposited at room temperature. The properties are then primarily a function of ion energy. Fig. 68 shows how the sp$^3$ fraction, density derived from the valence plasmon, intrinsic stress, and optical band gap are found to vary with ion energy for ta-C deposited by Fallon et al. [17] on a 90° filter FCVA. It is clear that the sp$^3$ fraction is the key parameter that sets the other properties. The maximum in sp$^3$ fraction occurs at an ion energy of about 100 eV. The films become more sp$^2$-like at low and high ion energies. The density is found to vary linearly with the sp$^3$ fraction, as was seen earlier in Fig. 31.

The optical gap depends on the ion energy [66, 67, 213, 214, 322], as shown in Fig. 68. It was recently found that optical absorption measurements were affected by residual particulates in ta-C films [328]. The band gap is larger once this scattering is removed. The ta-C films deposited on S-bond arcs with very low particulate content have wide gaps of around 3.5 eV [329, 330].

It is seen that the optical gap, using the Tauc or $E_{04}$ definitions, varies in proportion to the sp$^3$ fraction. This is expected from the dependence on the sp$^3$ fraction for all forms of a-C:H, ta-C:H and ta-C shown in Fig. 58. The band gap is controlled by the ordering and distortions of $\pi$ states on sp$^2$ sites [151], and this result confirms the findings of Raman that the sp$^2$ ordering depends in a consistent way on the sp$^3$ fraction [219].

The bonding in ta-C is described as follows. Its network is defined by the connectivity of its sp$^3$ atoms. They form a network similar to that of a-Si, except that it contains a small fraction of low order (three and four membered) rings [160, 161], which would be unstable in a-Si. The sp$^2$ atoms act as defects within this network. They form small clusters with an even number of atoms. The sp$^2$ clusters tend to be chains (olefinic) rather than aromatic, in that they do not give rise to a D peak in Raman and the C peak is above the VDOS band limit of graphite in UV Raman [219, 244]. The sp$^3$ network strains and distorts the sp$^2$ clusters. The band gap is determined by the size of the clusters and the degree of their distortion [151, 283]. The Raman and UV Raman spectra of ta-C as a function of sp$^3$ content were discussed earlier in Section 4.
An important property for applications of ta-C in coatings is the surface roughness. The small value of the roughness and good coverage quality of ta-C and other DLC films is a critical quality for their applications. There are various definitions of roughness; the modulus is

$$R = \sum_{n} \frac{|z - \overline{z}|}{n}$$  \hspace{1cm} (63)

where $z$ is the height at a certain position and $\overline{z}$ is the mean height. Their RMS roughness is much lower than the mean roughness because of the peaky quality of roughness. Precise atomic force microscopy (AFM) measurements have found that the roughness varies in a consistent way with the ion energy (Fig. 69), so that it reaches a minimum roughness at the ion energy of maximum sp$^3$ content [67,322,323]. The simplicity of deposition only from ions and the theory of its deposition (Section 3) suggested that ion energy is the single important parameter, which determines the properties of ta-C. However, it now appears that ta-C prepared by FCVA depends on other factors as well. The most important is probably growth rate. Fig. 70 compares the variation of sp$^3$ fraction of
ta-C prepared by different FCVA systems such as the original data of McKenzie et al. [51], and Fallon et al. [17], the S-bend FCVA in Cambridge [72] and the S-bend FCVA in Singapore [67]. This choice is only to illustrate the range of dependencies found. The newer FCVA systems give ta-C with higher sp³ fraction that extends over a wider range of deposition ion energies. Indeed, a high sp³ fraction is possible even at zero bias voltage, using only the plasma self-energy to give the ion energy. A high sp³ fraction without biasing is useful for industrial applications. Xu et al. [309] also found that the band gap increased with slower growth rate. Overall, it is clear that the sp³ fraction is not a single valued function of the ion energy. Generally, the maximum sp³ fraction occurs at an ion energy of about 100 eV, except for the data of McKenzie et al. [51].

Fig. 70. Comparison of the variation of the sp³ fraction with ion energy for ta-C. Data from Fallon et al. [17], McKenzie et al. [51], Polo et al. [72], and Shi et al. [67], showing the range of variations possible with the different FCVA filter systems and plasma duct sizes.
Fig. 71 shows the variation of sp³ fraction with mean ion energy for ta-C made by pulsed laser deposition, after Merkulov et al. [82]. It is clear that the variation is similar to that found for FCVA, with a maximum occurring at about 100 eV. This supports the idea that the ta-C formed by both processes is similar.

Fig. 72 shows the variation of sp³ fraction, density from Rutherford backscattering (RBS), and surface roughness for ta-C prepared by MSIB by Lifshitz and coworkers [7,8,321]. The sp³ data for other systems is also available [325]. The overall behaviour at lower ion energies is similar to that found for FCVA. However, there is a strong difference in that the sp³ fraction remains at a high value until the ion energy reaches over 1000 V in MSIB, whereas it starts to decline at 200 V or less in some FCVA systems. It is assumed that the ta-C prepared by both deposition systems is essentially the same. The different dependence on ion energy is tentatively attributed to the smaller growth rate on the MSIB [7,8].

Fig. 72 shows that the roughness of ta-C films grown by MSIB [7,8] also shows a broader minimum against ion energy than that grown by FCVA [67].

Fig. 73 emphasises this difference. It compares the properties of ta-C prepared by FCVA by Fallon et al. [17] and ta-C from MSIB. It is seen that ta-C prepared by MSIB retains a high sp³ fraction to a much higher ion energy. Lifshitz [7,8] has suggested that MSIB gives the wider sp³ window, because of its slower growth rate.

The ta-C films possess an intrinsic compressive stress, which arises from the ion bombardment during deposition, used to form the metastable sp³ bonding. The stress reaches high values of the order of 10 GPa. Even higher values have been reported in the sputtered ta-C with additional Ar bombardment, studied by Schwan et al. [46]. The compressive stress limits the maximum thickness of adhesive films, and this is a significant problem when trying to use a-C and indeed any DLC films as a protective coating material. It is therefore interesting to know if stress really is a necessary by product of sp³ bonding or not. The original data of McKenzie et al. [51] and Fallon et al. [17] showed a significant correlation between the stress and the sp³ fraction. However, a compilation of recent data [131] shown in Fig. 74 including data from Polo et al. [72] suggests that very high sp³ fractions can be achieved with stresses below 5 GPa. Therefore, there is hope that stress values can be lowered. The role of stress will also be discussed under deposition mechanism.
Fig. 72. Variation of $sp^3$ fraction, plasmon energy, density from Rutherford backscattering, and surface roughness with ion energy for ta-C deposited at room temperature by mass selected ion beam method, after Lifshitz [7,8].

Fig. 73. Comparison of the variation of $sp^3$ fraction with ion energy for ta-C deposited by FCVA [17] and mass selected ion beam [7,8].
Fig. 74. Correlation of compressive stress and sp$^3$ fraction for ta-C. Data from Fallon et al. [17], McKenzie et al. [51], and Polo et al. [72].

Fig. 75 shows the variation of refractive index, $n$, of ta-C with ion energy found for MSIB [322]. There is only a small change in $n$. This is principally because $n$ is given by Eq. (64),

$$n^2 = 1 + \frac{E_p^2}{E_g}$$

where $E_p$ is the plasmon energy and $E_g$ the Penn gap or average bonding-antibonding splitting. The plasmon energy $E_p$ decreases as the sp$^3$ fraction increases, but the Penn gap also decreases, and the two effects partly cancel. Thus, the refractive index does not change too much with ion energy.

It is sometimes implied that DLC has a high thermal conductivity like diamond. This is not true; amorphous materials generally do not have a very high thermal conductivity because of phonon scattering by disorder. For example, silica glass is not a particularly good thermal conductor.

Fig. 75. Variation of refractive index for ta-C with ion energy, from the data of Lifshitz et al. [323].
Table 6
Comparison of thermal conductivity at 25 °C of ta-C and a-C:H with diamond and other insulators

<table>
<thead>
<tr>
<th>Material</th>
<th>Thermal conductivity (W m⁻¹ K⁻¹) at 25 °C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diamond [12]</td>
<td>600–1000</td>
</tr>
<tr>
<td>CVD diamond</td>
<td>400</td>
</tr>
<tr>
<td>ta-C [326]</td>
<td>6–10</td>
</tr>
<tr>
<td>a-C:H [326]</td>
<td>1</td>
</tr>
<tr>
<td>Quartz</td>
<td>6.5–10</td>
</tr>
<tr>
<td>Silica</td>
<td>1.6</td>
</tr>
</tbody>
</table>

However, of the amorphous solids, ta-C has one of the highest thermal conductivities of 10 W/m K, with a-C:H at ~1 W/m K [326,327], as summarised in Table 6.

6.2. The ta-C microstructure

MacKenzie and coworkers [6,51] first noted that ta-C appears to possess a surface plasmon of lower energy than the bulk plasmon, suggesting that ta-C possesses a thin surface layer of lower density than the bulk. Subsequently, Davis et al. [195] used cross-sectional EELS in a scanning transmission electron microscope (STEM) to show that a low density, sp² rich surface layer existed (Fig. 76). Its thickness is about 1–2 nm for 100 eV deposition, and this roughly corresponds to the ion range at that ion energy. This is an important experiment because it confirms the deposition mechanism; the sp² rich surface layer shows that density and sp³ bonding increases in that part of the film in which subsurface deposition—subplantation—has occurred.

Recently, Siegal et al. [80,331] and Xu et al. [330] observed a more complex layered structure in PLD ta-C films by cross-sectional TEM. They find the low density layer at the surface, and a lower and a high density layer next to the film–substrate interface (Fig. 77). This layering was found in ta-C grown by PLD. Siegal et al. [80] suggested that the layering is an intrinsic property of the deposition mechanism. He suggests that the extra density of ta-C next to the substrate arises from C ions reflected back from the substrate atoms. The layering has also been found by TEM in some ta-C

![Fig. 76. Cross-sectional profile of carbon content and sp³ fraction across a ta-C film, showing the sp² rich surface layer, after Davis et al. [195].](image-url)
films grown on our single bend FCVA. XRR can also detect density layering. It has confirmed the density layering in some single-bend FCVA films. However, it has been possible to grow ta-C films on for example the S-bend FCVA without any detectable layering [184]. Thus, it is not clear that the effect is intrinsic to ta-C.

6.3. Effect of deposition temperature on ta-C

The ta-C is only formed if the deposition is carried out a lower temperatures, below a critical temperature of around 150–200 °C, which we all $T_i$ [35,62,66]. Fig. 78 shows the variation of $sp^3$...
fraction, density from the valence plasmon, stress and roughness with the deposition temperature, for ta-C deposited by FCVA at an ion energy of 80 and 130 eV, as found by Chhowalla et al. [66]. Fig. 79 shows a similar variation of density and roughness with deposition temperature for films deposited by MSIB after Lifshitz [7,8]. There is a sharp drop in the sp$^3$ fraction and density at $T_1$. The stress also drops sharply. The value $T_1$ is seen to decrease with increasing ion energy. The value of $T_1$ also depends slightly on the deposition rate or ion current.

Koskinen et al. [125] found a remarkable increase in $T_1$ if the ta-C is deposited by pulsed cathodic arc. $T_1$ increases with instantaneous deposition rate, and $T_1$ values up to 450 °C were achieved (Fig. 80).

Fig. 79 shows the corresponding variations of density and roughness for ta-C deposited by MSIB [7,8,35]. The same basic behaviour is found, with a sharp drop in the density and a sharp increase in the surface roughness at $T_1$. It is notable that the value of $T_1$ is lower than found for FCVA. It may be that this is due to the lower growth rate in MSIB than in FCVA.

Fig. 81 shows the variation of the electronic parameters optical gap, refractive index, resistivity activation energy and resistivity with the deposition temperature for ta-C deposited by FCVA at an ion energy of 100 eV [66]. It is notable that the optical gap and resistivity both decrease gradually with deposition temperature, compared to the sharp fall of sp$^3$ fraction and density at $T_1$ (Fig. 50). The fall in optical gap already begins below $T_1$ and continues above it.
The ta-C deposited at room temperature has no D peak in its Raman spectrum. As the deposition temperature is raised, the D peak appears [230]. This indicates that aromatic clusters are beginning to appear in ta-C. Fig. 82 plots the variation of the $I(D)/I(G)$ ratio with deposition temperature. The D peak is seen to appear below $T_1$, just like the fall in the optical gap. Thus, the aromatic clusters are beginning to appear below $T_1$, when the total sp$^2$ content has not changed. This indicates that the sp$^2$
sites are mobile during deposition below $T_1$, before sp$^2$ sites begin to convert to sp$^3$. The motion of sp$^2$ sites to form clusters even below $T_1$ indicates that the sp$^2$ sites can be regarded as defects in the sp$^3$ network. It is the clustering of the sp$^2$ sites, which causes the optical gap and the resistivity to fall. The optical gap really does depend only on the configuration of sp$^2$ sites. An empirical relation was found between optical gap and total sp$^2$ content for room temperature deposited ta-C [151] (Fig. 58), but Fig. 82 shows that no general relationship exists for ta-C deposited at any temperature.

A plot of $I(D)/I(G)$ against (optical gap)$^{-2}$ shows a linear dependence [230], as seen in Fig. 83. We assume that the sp$^2$ clusters created by annealing are aromatic clusters whose optical gap varies

---

Fig. 81. Variation of optical gap, refractive index, conductivity activation energy and room temperature resistivity with deposition temperature, for deposition by FCVA, after Chhowalla et al. [66].

Fig. 82. Variation of the Raman spectra ratio $I(D)/I(G)$ of ta-C with deposition temperature, for deposition by FCVA, after Chhowalla et al. [230].
as \( E \approx 2\gamma(a/L_a) \) as in Eq. (52). Fig. 83 then indicates that \( I(D)/I(G) \) varies as \( L_a^2 \), as noted earlier, Eq. (43).

Kulik and coworkers [323] has recently observed by transmission electron microscopy (TEM) and EELS that the graphitic clusters in films deposited above \( T_1 \) are oriented with their planes perpendicular to the film, that is with the graphite \( c \)-axis parallel to the film plane, Fig. 84a. This is consistent with a wide range of observations of planes perpendicular to the film, in c-BN [342], ion irradiated glassy carbon [343], bias-enhanced nucleation films of diamond [344], and in ta-C:H [345].

6.4. Thermal annealing of ta-C

There have been numerous studies of the thermal stability of ta-C [334–341]. Although ta-C is only metastable and it is necessary to keep the deposition temperature below 200 °C to form ta-C, it is stable after deposition up to 1000–1100 °C if annealed in a vacuum. If heated in air, it will oxidise at 450–500 °C. EELS and the size of the \( T \) peak in UV Raman shows that there is only a small
Fig. 85. Variation of UV Raman $I(T)/I(G)$ ratio, and visible Raman $I(D)/I(G)$ ratio and G peak shift, with annealing temperature, for ta-C deposited at room temperature by FCVA [340].

decrease in the total sp$^3$ content under annealing to 1100 °C [340] (Fig. 85). The ta-C films with a lower sp$^3$ content or with added nitrogen however are less stable [334,341].

Annealing ta-C up to 500–600 °C allows a relaxation of the compressive stress to essentially zero [335–340] (Fig. 86). This is an important observation, as stress limits the production of thick, adherent films. Cyclic deposition and annealing to 600 °C allows the production of ta-C films 1 μm thick [337]. The annealed films retain their large hardness and elastic modulus, as the sp$^3$ bonding is retained. Films produced in this way will be valuable for mechanical applications as discussed in Section 9. In fact, the stress can even go slightly tensile with annealing, although it is not understood why.

Sullivan and coworkers [335,336] proposed that the reduction in compressive stress occurs because a small fraction of sp$^3$ sites convert to sp$^2$ sites, and they have their π orbitals oriented perpendicular to the film surface, Fig. 84b. As the sp$^2$ bond is shorter than the sp$^3$ bond, this process would reduce the strain in the film plane. It is estimated that only about 2% of sp$^3$ sites need to convert to sp$^2$ for this to occur. Detailed density measurements by XRR suggest that this occurs [186–188].

Annealing also causes the optical gap to fall to zero and the resistivity to decrease strongly [340]. The visible Raman spectrum changes with the appearance of a D peak. As the total number of
sp³ sites has not changed much, according to EELS or UV Raman, the appearance of the D peak suggests that the existing sp³ sites have become mobile within the sp³ matrix, and that they are segregating into larger aromatic clusters [340]. This causes the optical gap to decrease. Thus, as in deposition at higher temperatures, existing sp² sites are mobile at lower temperatures than needed for sp³ to sp² conversion. The activation energy of sp³ to sp² conversion can be estimated from

$$E = kT \ln(\nu t)$$

as \(\sim 3.5\) eV, taking \(T = 1300\) K, the attempt frequency as \(10^{14}\) s\(^{-1}\) and \(t = 1\) s. The D peak appears at 650 °C, which gives an activation energy for the diffusion of sp² sites of 2.5 eV.

Siegal et al. [331] have recently found that the annealing of ta-C to 600 °C creates nano-phase inclusions which are visible by cross-sectional TEM. These regions cause an increase in hardness of about 15% as measured by a nano-indentor. It appears that these inclusions grow on existing nuclei, already present in the as-deposited films [333]. These may arise because of a stress-reduction mechanism in the as-deposited film.

6.5. The a-C:H

The properties of a-C:H have been studied by many workers [4,89,90,167–176,184,219,239–244,255–266,302–311,346–353]. Fig. 87 shows the variation of the main properties, the sp³ fraction, hydrogen content, mass density, and optical gap with the bias voltage \(V_b\) for a-C:H films deposited from different source gases: methane, acetylene and benzene [4,89,173,259]. Like ta-C, most properties of a-C:H depend on the incident ion energy per C atom [112]. For conventional PECVD, the incident ion energy is about 0.4 times the bias voltage \(V_b\) [4,87,88].

The bonding configurations in a-C:H deposited from methane have been studied in detail by Tamor et al. [173] using NMR to derive the fractions of the various C:H configurations as a function of \(V_b\). These results are summarised in Fig. 88. This shows how the total sp³ fraction and hydrogen content both decrease continuously with increasing \(V_b\), and also how the hydrogen is preferentially bonded to the sp³ sites.
The bonding and properties of a-C:H films fall into three regimes defined by the ion energy or bias voltage $V_b$ used in deposition [1,4]. The actual value of $V_b$ for these regimes depends on the precursor gas and the deposition pressure. At low bias voltages, the films have a large hydrogen content, a large sp$^3$ content and a low density. The films are called polymeric a-C:H or soft a-C:H. The optical gap is over 1.8 eV and it can even extend up to 3.5 or 4 eV [310]. This is a very wide gap. At intermediate bias voltages, the H content has fallen, the sp$^3$ content is less and the films have a maximum in density. In this regime, the amount of C–C sp$^3$ bonding reaches its maximum and the films have their highest diamond-like character. The optical gap lies in the range 1.2–1.7 eV. At high bias voltages, the H content has fallen further and the bonding has become increasing sp$^2$ like. The Raman spectra shown earlier now possess a D peak, and this shows that some sp$^2$ sites are aromatic [240]. The films can be called graphitic.

The bonding in a-C:H can be described as follows. The C sp$^3$ sites form a continuous network of C–C bonds. Most sp$^3$ sites are bonded to one or more hydrogens. A large part of the sp$^3$ bonding in

![Graph](image-url)
a-C:H is due to the saturation of bonding by hydrogens. The sp\(^2\) sites in a-C:H form small clusters in this matrix. The cluster sizes or sp\(^2\) site distortions increase with increasing \(V_b\), which causes the band gap to decrease. In soft a-C:H the clusters tend to be olefinic and do not give a Raman D peak [219]. In a-C:H deposited at high bias, the sp\(^2\) clusters become larger and increasingly aromatic, and so give rise to the observed D peak. The dispersive Raman spectra confirm this. The G peak position saturates at 1600 cm\(^{-1}\) at high photon energies for high bias samples, while it continues above 1600 cm\(^{-1}\) for polymeric a-C:H.

The properties of a-C:H depend on the ion energy per C atom, \(E\) [112]. Often the depositing species is a molecular ion \(\text{C}_m\text{H}_n^+\). When the molecular ion hits the film surface, it breaks up and shares their kinetic energy between the individual daughter carbon atoms. The effective energy per C atom is then [22]

\[
E = \frac{E_i}{m}
\]

This means that the properties of a-C:H will depend on the precursor molecule, when scaled according to this reduced ion energy per C atom as in Fig. 89. Thus the maximum in density for a-C:H prepared from acetylene will occur at about twice the bias voltage of that from methane, and that from benzene at about six times it. This accounts for the different variations of sp\(^3\) fraction and density with precursor gas seen in Fig. 89. The hydrogen content of the a-C:H film is always slightly lower than that of the precursor molecules. Hydrogen is lost during growth by chemical sputtering—the displacement of H from C–H bonds by incoming ions. This process is roughly proportional to the molecular ion energy. The H content of a-C:H prepared from methane is larger than that prepared from other precursors, largely because the H/C ratio of methane is so large.
On the other hand, the optical gap is seen to vary in the same manner against bias voltage for all the precursor gases [112] (Fig. 87)—that is not as in Eq. (66). This suggests that the gap depends just on the molecular ion energy \( E_i \), not on the number of carbons in the ion. It is proposed that this is because the configuration of the sp\(^2\) sites is set in the thermal spike phase of the ion deposition [112] (see Section 3).

Fig. 90 shows the variation of refractive index against bias voltage. This tends to reach a maximum at the bias voltage where the density reaches a maximum. The maximum lies above 1000 V for benzene derived a-C:H.

It was found by von Keudell and coworkers [264] that the refractive index varied in a uniform fashion against density, with variation of source gas or bias voltage, as shown in Fig. 91. This shows that density and refractive index do correlate well with each other.
6.6. The ta-C:H

The ta-C:H is deposited from high plasma density reactors such as the PBS [95] or the ECWR reactor [96]. The source gas should be acetylene. Deposition from methane gives films of less tetrahedral character which are not really 'ta-C:H' [98].

Fig. 62 shows the variation of sp$^3$ fraction, the plasmon energy, compressive stress and hydrogen content of ta-C:H as a function of the ion energy per C atom, of ta-C:H deposited by the PBS [95]. The sp$^3$ fraction is seen to reach a maximum at an ion energy of ~100 eV per C atom. This energy is remarkably close to the optimum ion energy found for ta-C from the FCVA, which suggests that a common deposition mechanism is responsible. The plasmon energy and thus the density are seen to follow the sp$^3$ fraction. The hydrogen content is seen to be below 30% for all ion energies and to fall slowly with increasing ion energy. The H content is much less than in a-C:H. The combination of a maximum sp$^3$ fraction of ~75% and a lower hydrogen content than in conventional a-C:H means that there is a much greater fraction of C=C sp$^3$ bonds in ta-C:H films than in a-C:H.

Figs. 92 and 93 shows the variation of the optical gap with the ion energy per C atom [95]. The stress is seen to follow the variation of the sp$^3$ fraction. This is to be expected as the stress often correlates with the fraction. The maximum optical gap is slightly higher than in ta-C. The optical gap also varies in the same way with the sp$^3$ fraction as other forms of DLC in Fig. 58. This variation with ion energy is unusual in that the optical gap in a-C:H decreases continuously with increasing ion energy (bias voltage), whereas a maximum occurs in ta-C:H. This emphasises the main difference between ta-C:H and a-C:H—there is no polymeric regime in ta-C:H as in a-C:H, because its H content is too low. At low ion energy, the ta-C:H network is more sp$^2$-like than polymer-like. At 100 eV, the network is highly tetrahedral with considerable C=C bonding. At large ion energy the network again becomes more sp$^2$-like, which is like conventional a-C:H. The Urbach energy of ta-C:H is quite large, ~0.25 eV, and there is only a small variation with ion energy [95].

The properties of ta-C:H have also been studied as a function of the deposition temperature [348]. Fig. 94 shows the variation of the plasmon energy, compressive stress, hydrogen content and the surface roughness as a function of the deposition temperature, for three ion energies. The sp$^3$ fraction was not measured in this case, but it is assumed to follow the variation of the plasmon energy. It is seen that the plasmon energy and the stress fall sharply above a critical deposition temperature of about 250 °C, which we label $T_i$ by analogy to that in ta-C. The plasmon energy falls
Fig. 92. Variation of sp³ fraction, plasmon energy, compressive stress and hydrogen content vs. ion energy per C atom for ta-C:H, after Weiler et al. [95].

Fig. 93. Optical gap vs. ion energy per C atom for ta-C:H, after Weiler et al. [95].
Fig. 94. Variation of density, compressive stress, surface roughness and hydrogen content of ta-C:H with deposition temperature, after Sattel et al. [348].

to values characteristic of sp$^2$ bonding. The surface roughness rises sharply at $T_1$. This is similar to the behaviour found by Lifshitz [7,8] for ta-C. Note that $T_1$ is slightly higher in ta-C:H than in ta-C, and that it falls with increasing ion energy, as was also found in ta-C.

We also see a second transition at around 450 °C, where the surface roughness falls back to low values and the H content falls to low values [348]. This transition is labelled $T_2$. It does not depend on ion energy. We summarise the transitions as $T_1$ is where the sp$^3$ bonding of ta-C:H is lost and $T_2$ is where the H content falls to near 0.

Fig. 95 shows the corresponding changes in the optical gap, and refractive index [348]. It is seen that the optical gap and resistivity decrease gradually with increasing deposition temperature. The decrease begins at temperatures below the transition $T_1$ and continues above it. This is the same behaviour as was observed in ta-C [66]. It is attributed, as in ta-C, to the motion of existing sp$^2$ sites to form larger aromatic clusters [230]. Then when sp$^2$ sites are formed above $T_1$, further clustering occurs. The refractive index on the other hand shows a more sharp transition at $T_1$ because it tends to follow the density rather than the minimum gap.

6.7. Thermal annealing of a-C:H and ta-C:H

The properties of ta-C:H and a-C:H have been studied as a function of thermal annealing [98,255,259,262,349–353]. Thermal annealing causes the evolution of hydrogen. Unlike in a-Si:H,
Fig. 95. Variation of optical gap (Tauc), extinction coefficient $k$, and refractive index of ta-C:H with deposition temperature, after Sattel et al. [348].

hydrogen can evolve as hydrocarbon molecules as well as molecular $H_2$. Fig. 96 shows the hydrogen evolution spectra of a-C:H deposited at different bias voltages after Wild and Koidl [352], for a-C:H deposited from methane and benzene at various bias voltages. The main evolution temperature for a-C:H is quite low at 350 °C for more polymeric a-C:H and it increases to a reasonable 550 °C for hard a-C:H deposited at a bias voltage of 500 V. As the evolution temperature increases, the fraction of hydrogen evolved as hydrocarbons goes down [259]. This shows that the network becomes denser.

Fig. 97 shows the hydrogen evolution spectra for ta-C:H deposited from methane and acetylene [98]. The ta-C:H deposited from methane is not so different to a-C:H. There is still some evolution of hydrocarbon molecules (note the log scale). The main evolution of $H_2$ is centred on 550 °C. The ta-C:H from acetylene is more dense. There is little evolution as hydrocarbons. The main evolution of $H_2$ is now centred on 700 °C [98,259] (Fig. 97). This material is much more stable than simple a-
C:H. The addition of Si to a-C:H is also able to improve the thermal stability of a-C:H, raising the H evolution temperature to ~700 °C [353].

The development of the carbon network during thermal annealing of ta-C:H can best be studied by Raman [98] or IR [255]. The Raman spectra are analysed into D and G peaks. The G position and the $I(D)/I(G)$ ratio are plotted as a function of annealing temperature in Fig. 98. It is seen that there is little D peak in the as-deposited films, as previously mentioned. The $I(D)/I(G)$ ratio starts to increase slightly before the main part of the H evolution. Thus, once the hydrogen is mobile, but still within the film, the
C–C network has started to rearrange. This rearrangement causes the optical gap to start to close up. The hydrogen motion also allows the density of defects to be reduced as seen by Conway et al. [349].

The development of bonding with annealing in a-C:H has also been studied using the IR spectra and the optical spectra. Although deducing bonding from C–H stretching modes is not recommended, Dischler et al. [255] has shown how the H evolves and the network transforms from more sp$^3$ to sp$^2$ bonded. The main change of bonding occurs at 500 °C for a-C:H deposited from benzene. The optical properties show an increase in number of π states and a gradual fall of the optical gap [191,255,550].

Thermal annealing of a-C:H also reduces the stress [98], as in ta-C [334]. However, as the bonding in a-C:H is less stable during annealing, annealing is less useful in this case.

The changes of properties of ta-C:H with deposition temperature and thermal annealing are slightly more clear cut than those in a-C:H. The changes in a-C:H are basically similar, but they occur over a lower and broader temperature range, rather than for example showing the rather sharp transitions $T_1$ and $T_2$. This is presumably because a-C:H has a wider range of bonding configurations.

6.8. Surface atomic structure

The surface atoms of a covalent solid will have broken or dangling bonds. These cost energy. In crystalline silicon, the surface reconstructs to remove some of these dangling bonds. On the ideal Si(111) surface, the dangling bonds are on second neighbour atoms. The surface reconstructs so that the dangling bonds move to nearest neighbour atoms, which then form bonds to each other. This also happens on the diamond (111) surface. On the Si(100) surface, each atom has two dangling bonds. This surface reconstructs to rebond pairs of dangling bonds.

The surfaces of ta-C also reduce the number of broken bonds. Simulations show that they usually do this by forming six-fold rings of three-fold coordinated sites [286,354–357]. This surface
Fig. 97. Hydrogen evolution spectra for ta-C:H deposited from (a) methane and (b) acetylene, after Conway et al. [98].

Graphitic bonding reduces the surface energy by surface $\pi$ bonding. This reconstruction affects the top atom layer. It should not be confused with the sp$^2$-rich surface layer of ta-C. This is a product of the deposition mechanism and can be wider, as seen by EELS [195].

7. Localised states

7.1. Localised and extended states

Fig. 99a shows a schematic of the DOS of the gap region of a typical amorphous semiconductor like a-Si. The electron states near the gap in a disordered system are either extended or localised
Here, states in the middle of bands are localised, while states next to the band edge called ‘tail states’ are localised. The localised and extended states are separated by an energy called the mobility edge. The band gap is no longer a perfectly forbidden gap, but contains some localised states. The imperfect gap is now called a ‘pseudogap’ or mobility gap, because the localised states have zero mobility at 0 K.

We now consider the origin of localisation in a-C. Fig. 100 shows the calculated electronic DOS, with the wider σ–σ* gap and narrow π–π* gap. A significant result is that in a-C the π states within the σ band gap are highly localised. Localisation is quantified by expanding the eigenstates in terms of the s, p basis orbitals as $\psi = \sum c_i \phi_i$ and then plotting the inverse participation ratio $Q$, which is defined as follows:

$$Q = \frac{\sum c_i^4}{\sum c_i^2}$$

(67)

$Q$ is seen to be above 0.1 for π states, signifying that they are localised, but approaches 1/$N$ for σ states where $N$ is the total number of states, signifying that they are extended states.

The strong localisation of the π states in a-Cs arises from the dihedral angle disorder. In a disordered system consisting of a band of s states with interactions in the range $V \pm \Delta V$, Mott and Davis [358] and Elliot [359] showed that the criterion for the localisation of the entire band is that
Fig. 99. Schematic of extended and localised state distribution in a typical amorphous semiconductor like (a) a-Si and (b) a-C.

The ratio of the disorder potential $\Delta V$ to the band width $B$ is less the same fraction of order 1,

$$\frac{\Delta V}{B} < 1$$

(68)

$B$ is given by the coordination number $Z$ times the interaction, $B = 2ZV$ for s states, so that

$$\frac{\Delta V}{V} \approx \frac{4}{Z}$$

(69)

In a typical $\sigma$ bonded network like a-Si, a degree of disorder below the critical value localises only states at the edges of each band as in Fig. 101.

The cause of the extremely effective localisation in a-C [286,287], shown schematically in Fig. 101, is the orientation dependence of the $\pi$ interaction. The Mott criterion applies to the isotropic interactions of s or $\sigma$ states. In contrast, $\pi$ interactions depend on their projected dihedral angle $\phi$, and in some cases the $\pi$ orbitals lie in nearly orthogonal planes, as shown in Fig. 101. This reduces the interaction $V$ to close to 0, so the disorder potential $\Delta V = V_0 - V$ is now a considerable fraction of $V_0$, the interaction of aligned $\pi$ orbitals which determines $B$. $\Delta V/B$ then becomes large, and this causes the localisation of the entire $\pi$ bands within the $\sigma-\sigma^*$ gap. This occurs even in the $sp^3$ rich networks, as seen in Fig. 100, so it is not just a DOS effect.

The mobility gap is defined as the energy difference between the valence and conduction band mobility edges. The optical gap is an energy difference between DOSs. In a-Si:H, the definitions of $E_{\text{opt}}$ gap, Tauc gap and mobility gap all lead to closely the same value, Fig. 99a. This is very
Fig. 100. Inverse participation ratio spectra for three a-C networks, showing strong localisation of states [286].
convenient. This is not so in a-C:H or ta-C. Firstly, the \( E_{04} \) gap is always 20–25% wider than the Tauc gap. More importantly, the mobility gap is much wider than the optical gap. At a fundamental level, the optical gaps are controlled only by the DOS. The optical matrix element does not depend on localisation, so the optical absorption reaches 10^5 cm\(^{-1}\) when the DOS reach some fraction of their maximum free electron value. However, in a-C the states at this energy are still localised. Thus, in a-C the mobility gap is much larger than the optical gap, as in Fig. 99b.

The strong localisation of states in a-C has a significant effect on the electronic properties and devices, as it lowers the carrier mobility. This is seen in the field effect mobility measured by field effect transistors (FETs) \([360]\) or in the photoluminescence behaviour, discussed shortly.

The cluster model gives rise to a very simple band diagram, as discussed previously in Fig. 59. This diagram shows that the disorder fluctuations are antisymmetric \([306]\). In general, fluctuations can be symmetric or antisymmetric about the centre of the gap. Coulombic potentials from charged impurities cause band edges to go up and down in parallel—or symmetrically. Deformation potentials (strain) cause antisymmetric fluctuations. The cluster size variation is a variety of deformation potential.

Optical transitions are vertical in real space, because there must be some spatial overlap between the initial and final state \([306]\). In crystals, the Bloch sums of both initial and final states convert this into a transition vertical in \( k \)-space, but it originates from a vertical real space transition. In conventional amorphous semiconductors like a-Si:H, the need for overlap between initial and final states means that the only allowed optical transitions are extended \( \rightarrow \) extended, localised \( \rightarrow \) extended, and extended \( \rightarrow \) localised. Localised \( \rightarrow \) localised transitions are forbidden, as the initial and final states are usually localised in different parts of space. On the other hand, when there are strong antisymmetric fluctuations as in a-C, then localised \( \rightarrow \) localised transitions are also allowed, because the initial and final states can be localised on the same \( \pi \) cluster.

It is important to understand the general layout of gap states in ta-C, ta-C:H and a-C:H. Fig. 102 shows an exponential distribution of tail states and some defect states around the centre of the gap. In a-C:H, conductivity data suggest that \( E_F \) lies near midgap. In ta-C and ta-C:H, field effect transistors suggest that hole transport dominates and \( E_F \) lies nearer the valence band \([360]\). This is confirmed by photoemission \([361]\). Scanning tunnelling spectroscopy measurements find that the slope of the valence band tail is sharper than the conduction band tail \([362]\). We have also analysed the slope of
the valence band tail as seen in photoluminescence and compared it to Urbach energy $E_u$ of the same samples [363]. These results are shown in Fig. 103. The Urbach energy roughly equals the larger of the tail widths. The valence band slope energy $E_u$ is less than the Urbach energy $E_0$. This means that the conduction band tail is the wider tail. This is the opposite of the situation in a-Si:H [364], but similar to that in many organic conductors. Thus, ta-C and ta-C:H are predominantly hole conductors.

7.2. Defects

Fig. 99b shows a schematic of the DOS for the gap region of a-C. In a real amorphous semiconductor, the localised states are of two types, tail states and defect states. The tail states lie close to the band edges, are moderately localised, and are created by weak forms of disorder such as bond angle distortions. Defect states are created by bonding configurations different to the bulk and they are ‘deep’ states strongly localised on a few sites. The states often but not always lie near the
centre of the gap. A typical defect is a broken (dangling) bond. In a-Si:H, this results in defect states and tail states lying at different energies in the gap, which is a convenient simplification. In a-C:H, the valence and conduction band tails are much broader, and they overlap in the centre of the gap, and also overlap the defect states. Thus, there is no clear separation on the basis of energy. It turns out there is also no distinction in character, as we show in Fig. 102.

A defect in a random network is defined as any configuration that gives rise to an electronic state near the Fermi level [1,3]. It is a deep level in the language of crystalline semiconductors. In a fully sp\(^3\) bonded network, the defects are isolated sp\(^3\) sites or ‘dangling bonds’. In a mixed sp\(^3\)–sp\(^2\) network, there are two main types of defects, Fig. 104. First there are isolated sp\(^3\) dangling bonds. Secondly, any cluster with an odd number of sp\(^2\) sites will give a state at \(E_F\). This includes a single sp\(^2\) site. In practice, unlike in a-Si, a sp\(^3\) dangling bond in carbon will tend to relax to the planar sp\(^2\) configuration, as in the methyl radical CH\(_3\), so the isolated sp\(^3\) and isolated sp\(^2\) site are in fact both sp\(^2\). Finally, the π bonds in small sp\(^2\) clusters can be so distorted that they give states near \(E_F\) [284].

Valence band tail states mostly lie below \(E_F\) and are doubly occupied and diamagnetic. Conduction band tail states mostly lie above \(E_F\) and are empty and diamagnetic. Defect states lie near \(E_F\) can be empty, filled with one electron or with two. They are neutral and paramagnetic when filled with one electron.

Generally, electron spin resonance (ESR) of the unpaired electron at a defect allows us to identify the defect configuration chemically and geometrically. In a-C, defects give rise to a resonance at \(g = 2.0028\). However, ESR is relatively uninformative in a-Cs. Firstly, the low atomic number means the spin orbit coupling is small, so the anisotropy of the \(g\) factor is small. Secondly, \(^{12}\)C is an even/even nucleus, so there is little hyperfine interaction between the electrons and unpaired nuclear spins. Thirdly, the large spin concentration in a-C means that a large exchange interaction between the electron spins dominates the line shape and creates a broad Lorentzian line. Even \(^{13}\)C enrichment has so far been unable to provide structural information [365].

The line shape has provided some information on the defect. Hoinkis et al. [366] used multifrequency ESR at 3–30 GHz to show that the ESR line consisted of a single line. Although the disorder in DLC is large, disorder does not control the line shape, because it is Lorentzian not Gaussian. Sadki et al. [367] proposed that the line shape was due to exchange, rather than dipole broadening due to hydrogen, which both give a Lorentzian shape. This was confirmed by Chhowalla et al. [66] who found a Lorentzian line shape in ta-C, which has no hydrogen. Various authors determined the line width as a function of deposition conditions [175,300,308–310,368,369]. The various data is plotted against optical gap in Fig. 105. Bardeleben and coworkers [368,369] recently noted that exchange controlled the line shape in most cases, except in the very polymeric (high hydrogen content) a-C:H films, where the dipolar broadening dominated. These effects are also seen in the spin-lattice relaxation times.

The nature of the defect in a-C:H has only recently been derived. ESR was carried out by von Bardeleben et al. [370] at 90 GHz, where the anisotropy of the \(g\)-tensor exceeds the exchange line.
width. Two components of the line were found. Most importantly, they found that the $g$-tensor has two low values and one larger value (Fig. 106). This splitting of $g$ components arises from a planar rather than an axial defect. Thus, ESR suggests that the defect is indeed an $sp^2$ bonded cluster, Fig. 103, not a dangling $sp^3$ bond or a distorted π bond. This is an important result that will affect our understanding of all amorphous carbons. It is expected that high frequency ESR will allow the determination of defects in other types of DLC.

The spin density in a-Cs is very large. Fig. 107 shows the ESR defect density of various a-C:H and ta-C:H films plotted against their $E_{04}$ optical gaps [95,175,205,300,308–310]. The spin density is of order $10^{20}$ cm$^{-3}$ in ta-C [66] and ta-C:H [95], and between $10^{17}$ and $10^{20}$ cm$^{-3}$ in a-C:H [175,308–310]. The plot emphasises how large the defect density is in a-C:H compared to, e.g. a-Si:H [364] where it is generally below $10^{16}$ cm$^{-3}$ in good material. It is seen that the defect density in DLCs tends to decrease as the band gap increases. Nevertheless, the defect density is too large for most electronic device applications.

It was noted that the defect level could be occupied by 0, 1 or 2 electrons. The localisation of the state means that the first electron will tend to repel the second electron. This repulsion is quantified by the correlation energy, $U$. The more localised the state, the larger the $U$. Dangling bonds in a-Si:H have

![Graph](image1.png)

**Fig. 105.** Variation of ESR linewidth of defects in a-C:H vs. the optical gap, data from Barklie [369] and Sattel et al. [348].

![Graph](image2.png)

**Fig. 106.** Hyperfine spectra of defects in a-C:H [370].
a correlation energy of +0.3 eV. In some special cases, the electron–lattice interaction can introduce a negative term into the correlation energy \([371,372]\). It can even make the ‘effective correlation energy’ negative, as in defects in a-Se. In this case, the electrons prefer to pair up at defects. This situation usually occurs by reforming a chemical bond. The defect states in a-C tend to be very localised and they have a quite large positive \(U\), of at least 0.5 eV. Ristein et al. \([373]\) have evaluated \(U\) experimentally for a-C:H as a function of band gap by comparing ESR densities with state densities derived from photoyield. Later, Fancini et al. \([374]\) compared ESR densities with state densities derived from optical measurements, and found a more expected increased of \(U\) with gap as in Fig. 108.

The large defect density is due to the wide band tails and the large Urbach energies. The DOS in an exponential tail is

\[
N(E) = N_0 \exp\left(-\frac{E}{E_0}\right)
\]

(70)

![Diagram](https://via.placeholder.com/150)

**Fig. 107.** ESR defect density vs. optical gap for a-C:H and ta-C:H \([98,308,309,374]\).

![Diagram](https://via.placeholder.com/150)

**Fig. 108.** Variation of effective correlation energy of defects vs. optical gap. Data from Tagliaferro, private communication.
where $E_0$ is the Urbach energy and $E$ the energy of the state from the top of the tail. The number of unpaired spins $N_s$ is given by the number of states, which lie within an energy range of $U$ of mid-gap. This is the product of $U$ and the DOS around mid-gap,

$$N_s \approx UN_0 \exp \left(-\frac{E_g}{2E_0}\right)$$  \hspace{1cm} (71)

Fig. 108 shows that this expression is a fair approximation. It is likely that the wide tails are the underlying cause of the large defect density, so any attempt to decrease defect density must also attempt to reduce $E_0$.

7.3. Photoluminescence

When a material is excited optically to create electron–hole pairs, these can either recombine radiatively with the emission of light as photoluminescence (PL) or non-radiatively to give phonons and heat. If the electron and hole are separated by an applied electric field and not recombine immediately, they can give rise to photoconductivity.

The wider band gap forms of a-C:H exhibit a strong photoluminescence which survives to room temperature. The PL forms a rather wide, Gaussian-shaped band near or just below the optical gap energy, and is excited by photon energies at or above the gap. Fig. 109 shows schematically the optical absorption spectrum, the PL spectrum, and the PL excitation spectrum (the PL efficiency as a function of the excitation energy) for a typical case [375].

Fig. 110 shows how the average PL energy of different a-C:H films varies with their optical gap [308–310,375–377]. The average PL energy increases as the gap widens, but with a slope of about 50%. It is interesting that the PL energy is slightly above the gap for gaps below 1.7 eV [377]. In the wide gap a-C:H, if the excitation energy is too low, this cuts off the upper part of the PL spectrum.

The PL is not affected by electric fields [375], it shows polarisation memory [378,379], and it is fast with a lifetime of $10^{-8}$ s [375,380,381]. These features indicate that the PL centre is quite localised. The PL in a-C:H coexists with a surprising large defect density compared to a-Si:H. In some cases, the PL efficiency does not correlate with the paramagnetic defect density, but generally there is a correlation. The PL efficiency appears to vary exponentially with optical gap.

![Fig. 109. Schematic diagram of the optical absorption, PL emission spectra and PL excitation (PLE) spectra in a-C:H.](image-url)
The PL mechanism in solids generally falls into four types. In direct gap semiconductors, electron–hole pairs in band edge states will form a bound electron–hole pair or (Wannier) exciton. This recombines to give sharp PL lines with an energy just below the band gap. This PL is fast.

In organic solids, usually with low dielectric constants and rather high effective masses, the electron–hole pairs form Frenkel excitons. The exciton can be long lived, and it can tunnel from one molecule to another by dipolar coupling known as the Forster mechanism. The exciton can recombine to give fast PL.

In deep defects in semiconductors, insulators or in some amorphous solids like a-Se, PL occurs at defects with a strong electron–lattice coupling. This is recognised by the large Stokes shift between the excitation band and the PL band. The Stokes shift occurs because the equilibrium position for the ground and excited states are different, so that PL occurs with a large emission of phonons, see Fig. 111c.

Finally, in a semiconductor containing both donors and acceptors, PL can occur by radiative tunnelling. The photo-excited electrons become trapped in donor levels and the holes trap into acceptor levels. These carriers can be quite widely separated, and at low temperatures they recombine by radiative tunnelling. This PL is slow because of its low probability. A generalisation of this mechanism occurs in a-Si:H (Fig. 111a).

There is debate about the exact PL mechanism in a-C:H. In one model, the PL mechanism can be described by a variant of that used for a-Si:H [151]. The non-radiative centres are mid-gap states including the defects. The alternative model is that PL occurs by a variant of that in organic molecules [381–383].

We first describe the PL mechanism in a-Si:H [364,384], Fig. 111a. In this case, the effect of disorder dominates the exciton binding. Photo-excitation creates electrons and holes in the extended states. The carriers loose energy (thermalise) and fall into the localised tail states. The electron now looses energy more slowly, by hopping within the tail states. Assume that one carrier, say the hole, has a lower mobility and is effectively stationary at $R = 0$. At low temperatures, where there are no phonons, the electron can only hop to a state lower in energy. This is called energy loss hopping [385,386]. The hopping rate $v$ is then given by

$$v = v_0 \exp \left( -\frac{2r}{a} \right)$$  \hspace{1cm} (72)
where \( v_0 \) is the attempt frequency (phonon frequency, \( 10^{13} \) Hz), \( a \) the Bohr radius of the tail state and \( r \) the hopping distance. As the distribution of states decreases exponentially into the gap (Eq. (70)), the density of available states decreases rapidly, and the average tunnelling distance increases,

\[
r \sim \left( N_0 \exp\left(-\frac{E}{E_0}\right) \right)^{1/3}
\]

(73)

Thus, the tunnelling probability decreases. The electron can recombine radiatively by tunnelling back to the trapped hole. The rate is

\[
v_r = \tau_0^{-1} \exp\left(-\frac{2R}{a}\right)
\]

(74)

This is the same equation as (72) except the prefactor is the dipole-allowed rate \( \tau_0 \sim 10^{-9} \) s

Hopping creates a distribution of electron–hole distances, \( R \). This maps onto a distribution of PL lifetimes, from (74). There is a broad distribution of PL lifetimes in a-Si:H, from \( 10^{-9} \) to over 1 s, with an average lifetime of [386]

\[
\tau_{av} \approx v_0 \tau_0^2
\]

(75)
Table 7
Conduction tail slope localisation radius and critical temperature from the low temperature photoconductivity

<table>
<thead>
<tr>
<th>Material</th>
<th>$E_{oc}$ (eV)</th>
<th>$a$ (nm)</th>
<th>$T$ (K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ta-C</td>
<td>0.28</td>
<td>0.9</td>
<td>70</td>
</tr>
<tr>
<td>ta-C:H</td>
<td>0.25</td>
<td>0.019</td>
<td>160</td>
</tr>
<tr>
<td>ta-C:H:N</td>
<td>0.25</td>
<td>0.034</td>
<td>200</td>
</tr>
<tr>
<td>a-Si:H</td>
<td>0.025</td>
<td>1.2</td>
<td>30</td>
</tr>
</tbody>
</table>

or $10^{-3}$ s. More distant electrons recombine more slowly. Electrons that have thermalised for a longer time have fallen deeper into the tail and are more widely separated. Thus, the PL energy gradually falls with time [364]. This effect is seen experimentally and confirms the mechanism.

Now consider a-C:H [151]. PL in a-C:H occurs from a rather localised centre, because it is unaffected by strong electric fields, it shows polarisation memory and it has a short lifetime of about $10^{-8}$ s. Generally, the relative size of disorder and exciton binding is found from the exciton energy

$$E_{ex} = -R \left( \frac{m^*}{\varepsilon^2} \right)$$

where $R$ is the Rydberg, 13.6 eV. Using say $m^* = 0.5$, $\varepsilon \sim 3.5$ for polymeric a-C:H gives $E_{ex} \approx 0.5$ eV. This is larger than the disorder (Urbach) energy of $\approx 0.3$ eV. However, the exciton radius

$$r = a_0 \left( \frac{\varepsilon}{m^*} \right) \approx 4 \text{ Å}$$

where $a_0 = 0.53$ Å more than the localisation radius of 1.9 Å for ta-C:H (see Table 7), which suggests that localisation by disorder is at least as important.

Thus, we assume that recombination occurs by a mechanism similar to a-Si H. The electron–hole pair is excited within a single $n$ bonded cluster as in Fig. 111a. Assume for simplicity one of the carriers (the hole) remains trapped in the cluster. The other (the electron) thermalises by energy loss hopping.

The hopping causes the electron to fall progressively deeper into the tail. At a finite temperature, the electron can hop out of the tail states by thermal excitation. We define a demarkation energy $E_d$ below the top of the tail, as in Fig. 112b. Electrons below $E_d$ are too deep to hop out, and must eventually recombine radiatively to give PL. Electrons above $E_d$ will eventually hop away and recombine somewhere non-radiatively. $E_d$ is given by

$$E_d = kT \ln(v_0 t)$$

So, $E_d$ increases with time. The time $t$ is set equal to the radiative lifetime $\tau_r$. If the band tail is exponential, the PL efficiency equals the fraction of electrons bound below $E_d$,

$$\eta \sim \exp \left( \frac{-E_d}{E_0} \right) = \exp \left( \frac{-kT \ln(v_0 \tau_r)}{E_0} \right)$$

(79)

The PL efficiency experimentally follows the relationship

$$\eta = \eta_0 \exp \left( -\frac{T}{T_L} \right)$$

(80)
This equals (79) if
\[ kT_L = \frac{E_0}{\ln(v_0\tau_r)} \]  
(81)

Fig. 112 plots the PL thermal quenching temperature \( T_L \) for the a-Si\(_{1-x}\)C\(_x\):H alloys [387–390]. \( T_L \) is seen to increase sharply from about 50 K for a-Si:H to 700 K for a-C:H. Fig. 112 also shows that \( E_0 \) increases from 0.05 to 0.4 eV. Fig. 111 shows the calculated value of \( T_L \) from (81) follows the experimental variation well. Thus, the higher quenching temperature in a-C:H than in a-Si:H occurs because of its broader tails and its faster luminescence.

The PL is also quenched if the electron–hole pair is created within a hopping distance of a non-radiative centre. We define a capture radius \( R_c \) around the centre [384]. Electrons created within \( R_c \) of the centre will be captured by it and recombine non-radiatively. Those created outside \( R_c \) are free to recombine radiatively and give PL. \( R_c \) is defined as where the rate of radiative PL tunnelling equals the rate of non-radiative tunnelling to the centre. From (72, 74), this gives
\[ R_c = \frac{a}{2} \ln\left(\frac{v_0\tau_r}{\nu}\right) \]  
(82)

The PL efficiency is then given by the probability that excitation occurs outside of spheres of radius \( R_c \) for centres of density \( N \), that is
\[ \eta = \exp\left(-\frac{4}{3}\pi R_c^3 N\right) \]  
(83)

In a-Si:H the main non-radiative recombination centres are the dangling bonds, which are half-filled and thus paramagnetic. Fig. 113a plots the PL efficiency of a-C:H against the density of paramagnetic defects for various sets of data [205, 308–310], and compares them to expression (83).
Eq. (83) will reproduce the trend of the results, if the Bohr radius takes a value

\[ a = 2.5 \text{ Å} \]  \hspace{1cm} (84)

The PL in a-C:H co-exists with a high defect density because of the small Bohr radius. This is consistent with the strong disorder.

However, there are certain key exceptions to the agreement shown in Fig. 113a. Principally, the data of Schutte et al. [308] show that the PL efficiency changes little for a large change of paramagnetic defect density. This led to the suggestion that the main recombination centres were not defects [388,389].

The answer to this problem is that in a-Si:H the tail and defect states are well separated in energy in the gap [384]. Thus, only defects act as recombination centres. In contrast, in a-C:H there is no clear distinction. There is a continuous distribution of localised states across the gap of a-C:H (Fig. 102). Singly occupied states near the gap centre are paramagnetic. However, a different set of states could be the recombination centres [391,392]. These are defined in the photoconductivity study, described later. States near in the upper gap act as electron traps or as recombination centres depending on whether the excitation probability to the conduction band edge exceeds the probability.
of recombination with a hole. This allows us to draw two demarkation energies as in Fig. 102. States between the two energies act as recombination centres. These centres are not necessarily the paramagnetic states, unlike in a-Si:H. This is the formal basis of the observation of Giorgis et al. [393] for which states quench PL. The density of recombination centres will vary in a monotonic way with optical gap, because the Urbach energy does (Fig. 62).

A second aspect is that the localisation radius of localised states, a, is not constant, as it is in a-Si:H. It could vary with the band gap, and also with hydrogen content. In the cluster model, a is the localisation radius of the $\pi$ states into the $sp^3$ bonded matrix. The PL efficiencies show a near exponential dependence when plotted against the optical gap, as in Fig. 113b. The absolute efficiencies are not known, so the various data sets are aligned vertically to emphasise the common dependence. We expect that the Bohr radius $a$ may vary with the optical gap as $a = cE_g - d$. Thus, the PL efficiency varies monotonically with optical gap because $a$ and the density of recombination centres varies in this way.

The major difference between PL in a-C:H and a-Si:H is the short lifetime in a-C:H, which is about $10^{-9}$ s [375]. This led to the suggestion that the PL centre in a-C:H is an exciton, and that the exciton hops as a unit by Forster tunnelling [382]. In this model, the electron–hole pair forms a well-defined exciton, which is not broken by the disorder. The exciton can diffuse by dipolar (Forster) coupling from cluster to cluster. Eventually, it will decay by PL, or it can hit a recombination centre and decay non-radiatively.

The main arguments in favour of the exciton model are the short lifetime and that the lifetime does not vary across the PL emission band [380]. This shift is a key signature of thermalisation effects. However, in fact the lifetime is unlikely to vary across the PL band, as the lifetime is already at the minimum for dipolar coupling, $10^{-9}$ s. A different way to study thermalisation effects is to study the polarisation memory or ‘PL anisotropy’ across the PL band. The PL emitted immediately after excitation or from a localised centre conserves the polarisation of the excitation. However, if the electron or hole thermalise into the tails by emitting phonons, they will lose some polarisation with each emission. An advantage of using PL memory is that it covers an exponential range of lifetimes across the PL band from (78), if thermalisation is occurring.

Polarisation memory is defined in terms of the degree of polarisation (DP),

$$DP = \frac{I_s - I_p}{I_s + I_p}$$

where $I_s$ and $I_p$ are the PL intensity parallel and perpendicular to that of the excitation.

The DP is shown in Fig. 114a for a-C:H films of different optical gaps as a function of the difference between the PL energy $E_{PL}$ and the excitation energy $E_{ex}$ [379]. We see that the degree of polarisation decreases continuously as $E_{PL}$ moves below $E_{ex}$. Berberan-Santos et al. [381] measured polarisation memory in as a function of delay time instead (Fig. 114b), and found that DP had a plateau at above 0.1. This was given as evidence of an exciton [381,382]. There are sometimes plateaus in the data in Fig. 114. However, in general, and particularly for the films with smaller gaps, there is a continuous decrease IP to 0. We take this as evidence of thermalisation in the general case. It is likely that the exciton model dominates for some polymeric a-C:H, particularly for a-C:H deposited from molecular precursors, and for high excitation energies. It is found that PL in some polymeric a-C:H shows three PL peaks, especially when excited at higher energies [383,394]. This is evidence of internal transitions.

We believe the band tail model is still basically correct, because an exciton band is not observed in the optical absorption spectra of a-C:H, unlike in organic solids, and because the disorder still
Fig. 114. Photoluminescence memory in a-C:H: (a) degree of polarisation vs. PL energy–excitation energy [379] and (b) degree of polarisation vs. delay time [381].

dominates electron–hole attraction. On the other hand, the electron–hole attraction is more important than in a-Si:H, and in a-C:H this sets the maximum separation of geminate electron–hole pairs, which can recombine radiatively. We use the disorder potential $E_0$ to define a maximum separation in the Coulombic potential following a type of Onsager model [364],

$$R_c = a_0 \left( \frac{R}{\varepsilon E_0} \right)$$  \hspace{1cm} (86)
This sets the hopping lifetime via Eq. (72), to be \( \tau = 10^{-9} \) s, which is the experimentally observed value. Thus, the electron–hole binding shortens the lifetime in a-C:H unlike in a-Si:H. Thermalisation effects in times less than \( 10^{-9} \) s cannot be seen in the spectrum, because this is shorter than the transition rate \( v_0 \). This means that the band tail model, with some excitonic binding, can explain PL in a-C:H.

Heitz et al. [383] also suggested that states in the band acted as non-radiative centres, again using the concept of capture radius (82). This is not correct. A non-radiative centre dissipates the energy as phonons, by a change in configuration coordinate, as in Fig. 111d. The condition for non-radiative transition is that point E lies below point B, so the electron moves from the excited state to the ground state without passing over a barrier [358,359]. This is likely for a defect state, as it is weakly bonded and lies in the gap. It is not likely for a band state. Nevertheless, this is a subsidiary argument on the exciton model.

7.4. Photoconductivity

Amorphous carbons have a low photoconductivity [391,392]. The reason to study it is the information it gives on transport and recombination processes. The photoconductivity of ta-C and ta-C:H shows two regimes (Fig. 115). Above \( \sim 200 \) K, the photoconductivity is thermally activated with an activation energy of about 0.2 eV, while below 200 K it becomes weakly dependent or independent of temperature.

Conductivity is given by the product of the carrier concentration \( n \), electron charge \( e \) and mobility \( \mu \) and \( n \) is given by generation rate \( G \) times the carrier lifetime \( \tau \), so the photoconductivity is

\[
\sigma = G e \mu \tau
\]

At higher temperatures, the photoconductivity is limited by recombination at recombination centres, so the \( \mu \tau \) product should vary inversely with their density \( N_d \) [364]

\[
\mu \tau N_d = \frac{e}{4\pi k T^{1/2}}
\]

Fig. 115. Photo and dark conductivity of ta-C, and ta-C:H deposited from acetylene and methane, after Ilie et al. [391].
where $1/\tau = 4\pi dDN_d$ for diffusive transport, $D = \text{carrier diffusion coefficient}$, $\mu = eD/kT$, $d = \text{capture radius}$, and the capture cross-section $s \sim d^2$. This holds in a-Si:H, where the main recombination centre is the dangling bond [364]. However, the photoconductivity of ta-C and ta-C:H does not vary inversely with defect density, as required by Eq. (88). This shows that other centres are also involved. It is because the band tails are so wide that there is no distinction between defect and tail states in the gap.

Instead, we must define demarcation levels to separate the trap states from the recombination centres [392]. The electron demarcation level $D_e$ lies where the probability of an electron hopping up to conduction states at $E_c$ equals its probability of recombining with a hole,

$$v_0 \exp \left( \frac{-E_c - D_e}{kT} \right) = (p + p_0)v_{th}s$$

where $p$ and $p_0$ are the photo-hole and dark hole concentration, respectively, $v_{th}$ is the electron’s thermal velocity, and $s$ is capture cross-section. A similar equation defines the hole level $D_h$. All states between $D_e$ and $D_h$ are recombination centres, those above $D_e$ or below $D_h$ are traps. Fig. 116 compares the calculated energies of demarcation levels in different types of ta-C and ta-C:H [392]. The electron and hole demarcation levels lie somewhat into the band tails in ta-C:H, so that some tail states as well as mid-gap defects will act as recombination centres. This explains why the photoconductivity and indeed the PL efficiency do not scale with the paramagnetic defect density, as expected in the simple analysis. It is also the formal explanation of why Giorgis et al. [393] found that states within some energy range act as recombination centres for PL. In contrast, in ta-C both demarcation energies are above midgap. This is because undoped ta-C has a high gap state density and is strongly p-type.

The photoconductivity varies with generation rate as $\sigma = kG^\gamma$. At high temperatures, the power $\gamma \sim 0.5$. In the low temperature regime, $\gamma$ approaches 1. A combination of a temperature-independent photoconductivity (Fig. 115) and $\gamma$ near 1 is the signature of energy loss hopping [384]. In this regime, electrons can only tunnel to lower states in the tail, or recombine radiatively by tunnelling to a hole. No carriers are lost to defects.

Photo-excitation creates a certain carrier density $n = 1/r^3$, due to the balance between generation and recombination. Defining a normalised inter-carrier separation as $L = r/a$, it is found that $L$ obeys [385]

$$L = -\ln(3G\tau_0 a^3 L^2 \ln(v_0\tau_0))$$

![Fig. 116. Calculated demarcation energies $D_e$ and $D_h$ in (a) ta-C; (b) ta-C:H (methane); and (c) ta-C:H(N).](image-url)
and also

\[ \gamma = 1 - L^{-1} \]  

(91)

The photoconductivity in the energy loss hopping regime is [385,386]

\[ \frac{\sigma}{eG} = \left( \frac{a^2 e L}{12 E_0} \right) \ln \left( \frac{\nu_0}{\gamma} \right) \]  

(92)

where \( E_0 \) is the energy width of the majority carrier exponential band tail. This formula allows us to calculate the localisation radius, \( a \), directly from the normalised photoconductivity, the Urbach energy \( E_0 \), the lifetime \( \tau_0 \), the phonon frequency \( \nu_0 \) and \( L \) or \( \gamma \). Values summarised in Table 7 compare well with those used in analysing space PL. Consistent values of \( L \) and \( a \) are found from Eqs. (90) and (92), showing that energy loss hopping gives a consistent description of low temperature photoconductivity in ta-C:H.

Fig. 117 compares the normalised photoconductivity \( \sigma/eG \) of ta-C and ta-C:H(meth) with that of other amorphous semiconductors like a-Si:H [395]. It shows that ta-C:H(meth) has the lowest photoconductivity of these materials. The small size of \( a \) is the main reason for this low photoconductivity, not the relatively large Urbach energy. A consequence of this low photoconductivity is that the transition temperature to the defect-limited regime is the highest of these semiconductors, as can be seen in Fig. 117. This transition is called the Adler–Monroe temperature and is given by [396]

\[ kT_c = \frac{3E_0}{L} \]  

(93)

Thus, the high disorder of ta-C:H causes energy loss hopping and luminescence to be observable at relatively high temperatures compared to a-Si:H [392].

![Fig. 117. Comparison of normalised photoconductivity of ta-C:H and a-Si:H, after Illie et al. [392].](image-url)
7.5. Doping

Any semiconductor is only really useful when it can be doped both n- and p-type. This is a particular problem in all wide-gap semiconductors. It is often found that they can only be doped to one polarity, either because the dopant levels are too deep, or the low dopant solubility, or because of auto-compensation. Diamond for example is easily doped p-type by boron. But there has been a long search for shallow n-type dopants. Dopants must satisfy two conditions in diamond, moderate solubility and having a shallow level. Nitrogen has a similar size to carbon, so nitrogen is soluble. However, substitutional nitrogen site undergoes a distortion, which sends its level deep to 1.7 eV below the conduction band edge. Substitutional phosphorus is difficult to incorporate, but has recently been found [397] to have a slightly deep level at about 0.55 eV below \( E_c \).

The ability to substitutionally dope a-Si:H with P and B has made it the most important large-area semiconductor in use today [364,398]. Solubility is not an issue, as the flexibility of a random network allows atoms of any size to be incorporated. The main problem in amorphous semiconductors is that network flexibility allows atoms to also exert their chemically preferred valence [399], and also form a trivalent non-doping site, which has the lower energy [3,400]. This is a form of auto-compensation. It turns out that there is an equilibrium, which allows some fraction of dopant atoms to be in doping sites [3,400].

A second question in amorphous semiconductors is if the density of gap states is low enough [364]. Doping must move \( E_F \) through the gap states by filling them up. Evaporated or sputtered a-Si contains a large density of gap states from dangling bonds. Hydrogenated a-Si is necessary, to passivate the dangling bonds and remove the gap states, to allow a viable doping effect. The final question is if the dopant state is shallow. Nitrogen is an obvious candidate as a donor in ta-C. It was noted that ta-C has a narrower gap than diamond, so that levels like N which a deep in diamond can be shallow in ta-C [401] (Fig. 118).

First consider nitrogen in ta-C, experimentally. The ta-C is made by FCVA. The nitrogen can be introduced as a gas into the cathode or duct region where it becomes ionised by the carbon plasma [312,402,403], or it can be introduced by a separate ion gun [404]. Undoped ta-C is slightly p-type with \( E_F \) lying just below mid gap [312,362,404]. This is consistent with the valence band tail being the sharper tail [362]. There are three regimes of N behaviour [196,312], as seen in Fig. 119. At low

![Fig. 118. Schematic of the levels of substitutional nitrogen in diamond and ta-C.](image-url)
N content up to 0.1%, the content and the optical gap remains sp\(^3\) constant. The conductivity rises perhaps with a first initial minimum, as \(E_F\) rises through the gap. At intermediate N contents up to about 10%, the sp\(^3\) content still remains high, but the optical gap falls. The conductivity continues to rise, but probably mainly due to the fall in band gap, not due to a Fermi level shift. Above 10% N, the sp\(^3\) content falls to low values, and the network becomes sp\(^2\) bonded. Conductivity continues to rise. Thus, the doping effect of N is not too good. There are two main reasons. At intermediate N contents, N encourages the existing sp\(^2\) sites to cluster and so it reduces the band gap. At high N contents over 10%, N induces the formation of new sp\(^2\) sites. This is a general property of N in carbon systems. A third reason is that nitrogen forms mainly non-doping (compensated) sites.

The poor doping efficiency of nitrogen is partly because N can adopt so many bonding configurations in a carbon network [401,405] (Fig. 120). Apart from the sp\(^3\) substituional sites (which is doping), these are the sp\(^2\) substituional site (doping) and many non-doping sites such as pyridine, pyrrole and nitrile. It is because it encourages carbon to form sp\(^2\) bonding [406]. Note that the extra electron of N goes in a non-bonding lone pair in the pyridine group, while it makes the six \(\pi\) electrons needed for aromatic stability in the pyrrole group.

A further reason for the poor doping of ta-C is the large density of gap states. The gap state density of ta-C is about \(10^{20}\) cm\(^{-3}\) (compared to only \(10^{18}\) cm\(^{-3}\) in a-Si:H). This requires a 10% doping efficiency for 1% N addition, to give observable doping. This probably does not occur. This difficulty of doping is consistent with the difficulty of shifting \(E_F\) in an FET [360].

Nitrogen displays a similar or worse behaviour in other cases. There have been numerous studies of nitrogen addition to a-C:H prepared by PECVD [407–411]. These often find an increase in conductivity. Usually, this is because nitrogen has decreased the band gap [409], rather than because of a shift in Fermi level. Only a few cases show a minimum in conductivity of as a function of doping, showing that a true shift of \(E_F\) across mid gap [392,410].

It is interesting that the addition of nitrogen to a-C:H reduces the spin density and the stress [410]. The reason for the reduced defect density is not clearly known, it could be due to a Fermi level...
shift. Boron is found to be a weak p-type dopant in some cases of ta-C [412], but boron has been much less studied than nitrogen.

There have been extensive studies of nitrogen incorporation in all forms of DLC. This work relates not only to doping, but also to a wider interest in the synthesis of carbon nitride, ever since it was proposed that C\textsubscript{3}N\textsubscript{4} could have a bulk modulus comparable to diamond [413,414]. C\textsubscript{3}N\textsubscript{4} has not been conclusively observed, whereas a-CN\textsubscript{x} is usually made [415,416]. This large body of work can only be briefly summarised here [415–432]. In C\textsubscript{3}N\textsubscript{4}, carbon is sp\textsuperscript{3} bonded and N is sp\textsuperscript{2} bonded [413]. There are two requirements to get C\textsubscript{3}N\textsubscript{4}, to incorporate enough nitrogen to reach the 4/3 stoichiometry and to keep the carbon sp\textsuperscript{3} bonded. Attaining a 4/3 stoichiometry appears to be impossible. C\textsubscript{3}N\textsubscript{4} is less stable than graphite plus N\textsubscript{2}, because of the stability of the N≡N bond. Atomic N tends to recombine as N\textsubscript{2} even within the solid. Atomic nitrogen and nitrogen ions are also a good etchant of carbon [426]. This causes the release of carbon as CN groups.

On the carbon bonding, the effect of nitrogen is to encourage C sp\textsuperscript{2} bonding [406]. This tendency must be overcome to form C\textsubscript{3}N\textsubscript{4}. The synthesis routes can be divided into CVD, PECVD, ion beam and PLD methods. CVD methods follow the recipes of diamond growth, using atomic hydrogen to promote sp\textsuperscript{3} carbon. This does not seem to be successful. PECVD allows a range of diverse conditions to be tried. Noting that C\textsubscript{3}N\textsubscript{4} is a dense phase compared to graphite plus N\textsubscript{2}, its formation is favoured by ion subplantation. It is preferable to have separate ion beams for the C and N ions, to have independent control. Rodil and coworkers [131] tried this route, and were able to maintain sp\textsuperscript{3} carbon to higher N contents than in previous works.

New types of local bonding are found in sputtered a-CN\textsubscript{x}. At higher temperatures this gives rise to a partly ordered structure with fullerene-like inclusions [425]. A phase diagram for this a-CN\textsubscript{x} has been introduced [426]. Nitrogen has one more electron than carbon, so that nitrogen in a five-fold pyrrole ring now gives the six π electrons needed for aromatic stability. These rings can introduce warping in a graphitic layer (Fig. 121). Another effect of nitrogen is to favour interlayer bonding. Substituting a nitrogen for C breaks a π bond and leaves an unpaired electron of the remaining C which is available to make a σ bond to a similar atom on an adjacent layer (Fig. 121). The same bonding occurs in the dimerisation of C\textsubscript{50}N where it can be proved [433].
Fig. 121. Diagram showing how nitrogen modifies bonding in graphitic layers, by introducing five-membered rings and warping, or inter-layer bonding.

The study of these various sites has been partly limited by how to identify the various sites spectroscopically [429–432]. The nitrile site is easily distinguished as it gives rise to vibrations at about 2100 cm$^{-1}$. However, this is the site we don’t really want! The core level shifts can be studied by XPS [418,429], but this is not favoured by the low ionicity of the C–N bond. The assignments of core levels in XPS have been debated [429]. EELS (or equivalently XANES) gives more direct information on hybridisation of the C and N sites [196,430]. A combination of EELS and X-ray emission spectroscopy on the C and N is edges appears to be a better means to distinguish the different sites [430].

8. Electronic applications

8.1. Field effect transistors

The most important economic application of a-Si:H is as thin film transistors (TFTs) in active matrix liquid crystal displays and in large area imaging systems [364]. The TFTs are unipolar devices in which the source-drain current is controlled by a gate electrode. The gate field shifts $E_f$ in the semiconductor layer immediately under the gate. In a-Si:H TFTs, the a-Si:H is undoped and the TFT is operated in enhancement mode. With no gate field, ideally there is a flat band condition with $E_f$ near mid gap and a very low off-current. Applying a gate field moves $E_f$ towards the conduction band, inducing electrons in the channel, which increases the conductivity between source and drain. Various device structures can be used (Fig. 122). The inverted staggered configuration is most popular. In this, the gate is deposited first, then the gate insulator a-Si$_3$N$_4$:H, then a-Si:H and then the source and drain electrodes. A thin layer of doped a-Si:H is used below the source and drain contacts to make them Ohmic rather than Schottky barriers.

It was desired to make similar TFTs from DLC. The ta-C was chosen because of its better stability. A major problem in ta-C is the sp$^2$ rich surface layer formed on the outer surface due to the subplantation mechanism [195]. This layer would be too conductive to operate in enhancement mode. There are two ways around this: use a top gate structure and etch off the sp$^2$ layer before depositing the gate electrode, or use a bottom gate electrode, and hope that the deposition of the ta-C does not cause ion damage of the gate insulator. Both have been done.
A coplanar top gate structure was used by Clough et al. [434] (Fig. 122a). The surface layer of ta-C was etched off by an oxygen plasma (using N2O). STS subsequently showed that the new surface is stable and does not convert back to sp² [362]. Silicon nitride gate insulator is then deposited immediately afterwards in the same chamber. Finally a gate electrode was evaporated on top. The channel length is 10 µm and channel width is 0.5 mm. The gate transfer characteristic is shown in Fig. 123a. It shows that the device is p-type. This is consistent with $E_F$ lying below mid gap in undoped ta-C. The field effect mobility is calculated to be $10^{-6} \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$. For reference, the mobility of a-Si:H is 0.5–1 in these units. Thus, it has been demonstrated that it is possible to make a TFT using ta-C, but the mobility is very low. In addition, the source-drain characteristics in Fig. 123b show ‘current-crowding’, and do not reach the saturated regime where the drain current becomes independent of drain voltage.

The bottom gate, inverted staggered TFT was made by Maeng et al. (Fig. 122b) [360]. A doped Si wafer was used as a gate body. Thermal silicon dioxide was grown on it. The ta-C was deposited onto this using the S-bend FCVA at floating potential. The absence of extra bias voltage minimised any ion damage (or chemical etching) of the oxide. Metallisation for source and drain top contacts was then applied. A gate transfer characteristic in Fig. 124a again shows p-type enhancement operation. The drain characteristics in Fig. 124b show less bad current crowding than the top gate device. Maeng found that the mobility can be increased by using ta-C of narrower band gap, but clearly this is at the expense of a higher off-current.

In summary, the TFTs have poor gate performance because ta-C possesses a large density of gap states, so it is difficult to move $E_F$ through the gap. Secondly, the mobility is very low because of the stronger localisation of states in amorphous carbons. In a-Si:H, $E_F$ is moved close to $E_c$ so that

![Fig. 122. Diagram of the (a) coplanar top gate and (b) inverted staggered bottom gate TFF structures.](image)

![Fig. 123. Gate and drain transfer characteristics for coplanar top gate TFT [434].](image)
conduction occurs by the fraction of electrons in extended states. This stage is never attained in ta-C, nor would it be in a-C:H.

8.2. Antifuses

Fuses are metallic links that go open circuit when they pass a large current. In contrast, antifuses change from a high to low resistance when they pass a large current. A number of oxide and carbon systems can be used as antifuses. Amorphous carbons make useful antifuses [435,436]. Generally antifuses act by a thermal runaway mechanism, as the conductivity of semiconductors increases as temperature rises. The current is concentrated in a narrow channel, to maximise the temperature increase [437,438]. The process in carbon likely involves a change to more sp² bonding [436]. McKenzie et al. [436] finds that nitrogen doping of ta-C makes the antifuse transition easier.

8.3. Low dielectric constant films for ULSI

In silicon complementary metal-oxide-semiconductor (CMOS) device technology, the device dimensions are continually shrinking, in a process known as Moore’s Law. At present, the insulator used between data lines is silicon dioxide. However, RD associated with the capacitance of these lines is becoming too significant. It is necessary to replace SiO₂ with an insulator with a lower dielectric constant (κ) to reduce this capacitance. The (low frequency) κ of SiO₂ is 3.9. Various carbon-based films are candidate dielectrics [439–444]. These materials must satisfy other criteria as well as low κ. They must be thermally stable to 400 °C, have adequate rigidity, low mechanical stress, low dissipation, low leakage, good adhesion and be processable using acceptable means. The industry first used fluorinated SiO₂ but there is a lower limit to its κ. Various carbon-based materials have been tried, such as a-C:H, fluorinated amorphous carbon (a-C:F), carbon nitrides, and methyl-substituted siloxanes, in addition to a wide range of organosilane derivatives. The aim is to find materials to reach κ values of 1.5–2.1.
There are two components to κ, the electronic and the lattice part. The lattice part is minimised by avoiding ionic bonds. The electronic part is minimised by using low atom densities and strong, less polarisable bonds. Fig. 91 show that a-C:H can reach κ values down to 2.6 by deposition at low bias voltages and higher pressures. The main problem with these polymeric a-C:H is that they have low thermal stability and evolve hydrogen and hydrocarbons at below 400 °C. The κ of a-C:H follows the same trend as the loss of thermal stability, and only films with κ above 3.3 are sufficiently stable [441].

PECVD a-C:F has been tested in a similar way [440–443]. A-C:F has a lower κ than a-C:H because of the stronger C–F bonds. However, a-C:F is also limited by its thermal stability. It has the added disadvantage that any evolving F is very aggressive on other parts of the device. It is possible this could be solved with diffusion barrier layers, but for added cost.

PECVD carbon nitrides have κ values below 3.9. In this case, the aim is not to maximise the density as in the case of trying to synthesise C3N4 [444]. At present, the most likely candidates are organo-silicates prepared by condensation of hydrogen silsequioxan (HSQ). After curing these are stable to 400 °C [439].

8.4. Field emission

Field emission is the emission of electrons from a solid under an intense electric field, usually at ambient temperature—a ‘cold cathode’. Emission occurs by tunnelling of electrons through the surface potential barrier. The typical local field required for emission from solids is 1000 V μm⁻¹. The simplest way to achieve such fields is to use sharp shapes such as Spindt tips [445]. Various types of carbon such as carbon nanotubes, diamond and DLC show field emission at much lower applied fields of order 10 V μm⁻¹. This could open up opportunities for electronic devices.

The most widely considered device is the field emission display (FED). This is a flat panel display in which the image is formed from a large array of pixels, each addressed by a controllable cathode [446] (Fig. 125). The use of thin film emitters rather than Spindt tips simplifies the manufacturing process. FEDs require emission current densities of 1 mA cm⁻² and emission site densities of the order of 10⁶ cm⁻². Other possible applications are in vacuum MEMSs such as microwave amplifiers, or in cathodes for parallel e-beam lithography.

The advantages of a carbon cathode are its chemical and physical stability compared to the Si and Mo used in Spindt tips [446]. The carbon is more resistant to poisoning by elements such as sulphur, which can be released by the phosphors. Carbon has a low sputter yield when bombarded by residual ions in the device. In addition, DLC devices could be very low cost as DLC can be deposited at room temperature.

![Fig. 125. Field emission display, with (a) Spindt tips and (b) carbon film cathodes.](image-url)
The need is for any low field emitting material, so it does not really matter if it is diamond, DLC or carbon nanotubes. All three must be compared. The advantage of DLC over diamond or CNTs is its low temperature deposition. This led to the intensive study of field emission in DLC. It is now realised that field emission from DLC has serious problems. Nanotubes are now viewed as the best carbon system for field emission [447]. It is interesting to see how this happened.

The field emission current density $J$ through a potential barrier obeys the Fowler–Nordheim equation [445],

$$J = a \frac{(\beta F)^2}{\phi} \exp \left( - \frac{b\phi^{3/2}}{\beta F} \right)$$

(94)

where $\phi$ (eV) is the barrier height, $F$ the applied field, $\beta$ a dimensionless geometric field enhancement factor, $\beta F$ the local field at the emission site and $a$ and $b$ are constants. Easy emission requires a high local field or a low barrier.

Field emission is usually tested in a parallel plate system, often with a transparent phosphor anode like indium–tin oxide, or else by scanning a small ball anode over the film. It is found that the emission occurs in spots or emission sites, and the site density is a critical parameter for applications as a display.

Emission in spots is the sign of a high local field. However, it also means that there are high local current densities, even when the average current density is small. Emission from diamond or DLC is often improved by a conditioning process in which the applied field is cycled a few times. This conditioning and indeed ‘normal’ emission produces sample modification in diamond or DLC. This is because diamond and DLC have low conductivity, but the spotty emission requires a high local current density, and this is initiated by damaging. The damage can be macroscopic (craters) formed by a vacuum discharge, and easily visible by scanning electron microscope [448,449]. Or it can be nanoscale and ‘soft’ and consist of nanometre sized conductive paths across the film [450]. STM experiments find conditioning causes a local conversion of $sp^3$ to $sp^2$ bonding [450]. Whatever the size, damage is nearly always there, even if it is difficult to find by microscopy because the emission site is so small and their number is low. This extrinsic nature of emission is a serious disadvantage of these systems. In contrast, carbon nanotubes are metallic, can carry very high current densities, and they are now the favoured system for field emission.

The original motivation for studying field emission in diamond is its negative electron affinity (NEA) [451]. An NEA is when the conduction band edge lies above the vacuum level, so that any electrons in its conduction band can pass into the vacuum with no energy barrier [452] (Fig. 126). A diamond surface must be terminated by hydrogen to have an NEA [453]. The surface C–H bonds create a dipole layer which raises the energy of the conduction band compared to the vacuum level [454].

However, field emission requires electrons to travel round a complete circuit. Diamond has a wide band gap (5.5 eV) so its resistivity is very high. Therefore, the electrons have to be injected over a large barrier at the back contact of order 4 eV [455,456]. To date, the best emitter from diamond is nano-crystalline diamond [457–460]. In this case, the emission varies inversely with the grain size [457,459]. This is because grain boundaries provide a conduction path to the front surface. The origin of field emission is also found by measuring the electron energy distribution (EED) of the emitted electrons (Fig. 127). Groning et al. [460] found that the electrons originate from the Fermi level, which lies about 1 eV above the valence band edge. The grain boundaries are $sp^2$ bonded, more conducting and have states about 1.4 eV above $E_F$. The EED width is also quite large, showing that a large local field is present.
Fig. 126. Band energies for (a) positive electron affinity; (b) effective negative electron affinity due to Cs layer and its dipole layer; and (c) true negative electron affinity systems.

Emission from grain boundaries can only be explained by a large enhancement of the local field at the emission site (Fig. 128). The grain boundary is sp² bonded, and has π states around $E_F$. These states act as localised states in the diamond band gap. The states are neutral when half filled. When filled, they are negatively charged and when empty they are positively charged.

If an external field is applied to the diamond surface, and the diamond has a uniform distribution of defect states, the field will charge the defects and form a uniform, negatively charged depletion layer over the surface. However, this field penetration only lowers the surface potential by a modest 0.1 V, compared to the total barrier of 4–5 V. If instead the defects are concentrated in grain boundaries, the field lines focus onto these regions. This produces strong local fields, which extend along the grain boundary. The grain boundary acts like an internal tip. This mechanism can provide an adequate narrowing and lowering of the 5 V barrier, and this allows the observed easy emission.

Fig. 127. Schematic of how the energy distribution of emitted electrons shows the emission mechanism: (a) emission from the Fermi level and (b) hot electron effects.
The simplest analysis is to consider what field is needed for a certain $J$ in (94). This is equivalent to

$$\frac{\phi^{3/2}}{\beta F} = \text{constant}$$

(95)

$$F_{\text{loc}} = \beta F \approx \text{constant}$$

(96)

if the barrier $\phi \sim 5$ eV, as it usually is. The local field is a constant of order 1000 V m$^{-1}$. This is indeed the field observed from the width of the EED [460], from which the local field can be derived.

In summary, various other diamond-based devices have been tried but, to date, no diamond field emission has truly utilised the negative electron affinity of diamond.

DLC has been extensively studied for field emission. However, the low emission site density (ESD), relatively low maximum current density and extrinsic nature of emission of DLC limits its use in field emission applications [461–474]. Some types such as undoped a-C:H have poor emission because their resistivity is too high (10$^{14} \Omega$ cm). This is improved by adding nitrogen [461]. The hydrogen can also lead to rather unstable emission, so many groups have focused on hydrogen-free a-C [462]. Emission from ta-C was found to depend strongly on its deposition conditions, the incorporation of nitrogen, and on plasma surface treatment [465–467]. However, it is becoming clear that much of the emission is extrinsic, linked to either particulate inclusions, or to the conditioning process.

The early field emission results tended to characterise emission in terms of a threshold field for an emission current density of say 10$^{-6}$ A cm$^{-2}$. Various factors such as changes in the sp$^3$ fraction or nitrogen content appeared to provide quite low threshold fields. A typical plot of emission current density for different nitrogen contents [464] is shown in Fig. 129. Fig. 130 shows the variation of threshold field of ta-C films deposited on a single-bend FCVA with the ion deposition energy (sp$^3$ fraction) and with the nitrogen content. It then became clear that the ESD was a more important parameter for many applications. Sometimes, ta-C from S-bend FCVA gives very weak emission [473]. The ESD of DLCs is too low. Treating the ta-C surface with an Ar, hydrogen or oxygen
plasma lowers the threshold field, as shown in Fig. 130a, because it increases the ESD [467]. However, the ESD is still too low. This then led to an interest in field emission from nanostructured forms of carbon, which are mainly sp² bonded [475–477]. These explicitly have tip-like surface features and can be produced at room temperature on the same DLC deposition systems. It was possible to increase the ESD and emission uniformity by using a resistive ballast layer to share the current between emitting sites [478]. Fig. 131 shows the emission site distribution for a film with a ballast resistor, as a function of increasing applied field [479]. The ESD has a roughly exponential dependence on applied field [479].

The emission mechanism in DLCs has been difficult to understand. The slope of the Fowler–Nordheim plot of 0.01 to 0.4 eV derived [464] from the Fowler–Nordheim plots by taking β = 1 is clearly not the real barrier, as such small barriers would allow temperature-dependent thermionic emission or Schottky emission, which is not observed.

Fig. 129. Emission current density of ta-C:N as a function of (a) N content vs. electric field and (b) in Fowler Nordheim plot [464].
We have constructed a band diagram for DLCs to understand the emission mechanism [480]. The a-C:H films have a band gap which varies with their sp$^3$ content. The energy of their valence band below the vacuum level is known from photoyield experiments [373]. The resulting band diagram (Fig. 132) shows that a-C:H always has positive electron affinity.

This is then combined into a band diagram for the substrate, DLC, vacuum system, Fig. 133. This requires knowledge of the band offset between the DLC and substrate, which is usually doped Si. Schafer et al. [481] found that there was a large valence band offset for a-C:H on Si, whereas Rupersinghe et al. [482] found a sizeable conduction band offset for ta-C on Si.

Photoemission [482] and Kelvin probe [467] have recently been used to find the valence band energy and work function of ta-C. The work function is measured to be 4.9 eV and so the electron affinity is 3.4–3.6 eV [467,482]. The band energies and Fermi level as a function of nitrogen content are shown in Fig. 134. These results emphasise that ta-C always has a sizeable positive electron affinity similar to graphite of 4.9 eV. Its work function is relatively independent of sp$^3$ content and nitrogen addition, and is similar to that of graphite, which is 4.7 eV.

The EED of a-C with a large nitrogen incorporation has been measured, and emission is still found to occur from ∼5 eV below the vacuum level [474]. The EED width is consistent with the presence of a large local field. This indicates that the emission mechanism again involves a strong field enhancement.

Thus, experimental evidence suggests that there is a large, 4–5 eV emission barrier at the front surface, and thus large local fields. The main problem in any emission model is that DLC films are very smooth, so the origin of large local fields is unclear. There must be some cause of heterogeneity.
Fig. 131. Emission sites of a nanostructured carbon film with increasing applied fields [479].

Fig. 132. Variation of band energies referenced to the vacuum level, for a-C:H, vs. band gap [480].
We assume that the conditioning process creates $sp^2$-rich channels extending from the film surface. These are effectively anti-fuse channels. This leads to an emission model as in Fig. 135, which is functionally similar to the internal tip model of grain boundaries in nano-crystalline diamond. The site density of channels is quite small. The field lines focus onto their gap states. This creates a negative depletion charge, a large local field and the necessary barrier lowering.

It has been suggested that emission from DLC is due to hot electrons created by strong band bending at the back contact [470]. However, photoemission measurements of the valence band energy of ta-C films of increasing thickness show little evidence of band bending at the back contact [482]. The EED also shows no evidence of hot electrons as a high energy tail [463]. It was also argued that band bending from the back contract could control emission and account for a thickness dependence in emission [470]. However, the weight of other evidence favours emission controlled by a large barrier at the surface.

Fig. 134. Band energies and Fermi level energy referenced to vacuum level for ta-C:N films as a function of nitrogen content [467].
Why is the critical dimension of emitting sites so small [472]? The field enhancement factor $\beta$ is given by the ratio of height to radius of the tip,

$$\beta = \frac{h}{r}$$

(97)

There is a trade-off between the need for small $r$ values to give large enhancement factors of 200–400, and larger $r$ values to conduct the emission current. The variable $h$ is the film thickness. The macroscopic emission current density $J$ is given in terms of the filament diameter $d$, its conductivity $\sigma$, and ESD $N$ and the local field $F$ as

$$J = N \left( \frac{\pi d^2}{4} \right) \sigma F$$

(98)

Taking the conductivity of the emission spot to be that of disorder $sp^2$ carbon, say $1 \, \Omega^{-1} \, m^{-1}$, $N \sim 10^6 \, \text{sites} \, m^{-2}$, and $J = 10^{-2} \, A \, m^{-2}$, we find $d \sim 0.7 \, \text{nm}$ for $F = 5 \times 10^5 \, V \, m^{-1}$ as found in the EED.

Thus, the internal-tip mechanism accounts for various factors of field emission in DLCs. Emission occurs from spots rather than uniform areas because of the need for field enhancement. This requires small values of $r$. However, small $r$ values limit the emission current. Thus, high resistivity limits the emission properties of DLC. Better emission is possible if the depletion regions for charge focussing can be formed in more conducting $sp^3$ material. There is evidence from Raman for small scale heterogeneities in more emissive DLC [471,472].

A large $sp^3$ fraction and nitrogen addition were found to lower the threshold emission field of ta-C. These results are now doubtful because of the particulate content of this ta-C deposited from a single bend FCVA. However, Cheah et al. [466] found reasonably similar results for ta-C deposited from an S-bend FCVA, which has very low particulate $sp^3$ content. They also found that plasma surface treatment increased emission site densities. The effect of content and N addition was originally interpreted in terms of a reduction of electron affinity [464]. This is now recognised as being incorrect. It is now known that emission occurs from the Fermi level and that $E_F$ does not vary much with N content or $sp^3$ fraction, from the Kelvin probe data. The effect of N may be to allow an easier formation to $sp^2$ channels, as McKenzie et al. [436] already noted that N improves the antifuse
behaviour. At high N content, the bulk conductivity is too high, the contrast between channel and bulk conductivity is less, so there is then less field focusing. Similarly, the contrast in resistivity between channel and bulk is largest in the high sp$^3$ films, and this may account for the dependence on bulk sp$^3$ content.

Carbon nanotubes have the greatest potential for field emission applications [483–489]. They can be grown by dc or rf PECVD or thermal CVD using a catalyst from hydrocarbon source gases at 600–800 °C [490]. The catalyst is a transition metal like Co, Ni or Fe. The growth location is controlled by the location of the catalyst in a surface. An electric field in the PECVD causes the nanotubes to be aligned perpendicular to the surface [491]. Nanotubes make good field emitters because of their extreme shape. They are conductive, and they can carry very large current densities if needed [484]. Single nanotubes or mats of nanotubes can be used. The emission occurs from the nanotube tip, and from the Fermi level. Their work function is similar to that of graphite, 4.7 V [492]. The emission is stable, but depends on the presence or absence of molecules, which can absorb on the nanotube tip [485,486]. The emission distribution from mats of nanotubes shows that the ESD is surprisingly much smaller than the total density of nanotubes [489]. Most nanotubes do not participate in emission. This is because adjacent tubes screen the field enhancement at a tip [489]. The nanotube density must be lowered, so that the mean separation between tubes is about two times the tube length. This is shown in Fig. 136.

The emission site density $N$ from nanotube mats is also found to vary exponentially with applied electric field $F$, as in ta-C or nanostructured carbon [476,477,489],

$$N(F) = N_0 \exp\left(-\frac{b}{F}\right)$$  \hspace{1cm} (99)

This occurs because a certain emission current density corresponds via Eq. (94) to some constant local field at the emission site, Eq. (95). This means the distribution of Eq. (99) is equivalent to an exponential distribution of the field enhancement factors $\beta$,

$$N(\beta) = N_0 \exp\left(-\frac{\beta}{\beta_0}\right)$$  \hspace{1cm} (100)

![Fig. 136. Tips which are too close screen the field enhancement of each other.](image)
This distribution seems to describe the ESD in all forms of carbon film. The width of the
distribution is a weakness of field emission when considering applications. The broad $\beta$ distribution
in nanotubes and nanostructured arises from the field screening by adjacent tips. This means that
most nanotubes or sharp features in nanostructured carbon do not emit. This can be improved only
by controlling the tip spacing. This can be achieved, with effort, for aligned nanotube mats, but is a
weakness for nanostructured carbon where the means to control the various structural features is not
so apparent. Ballast resistors can be used to distribute current between emitting sites, but at the
expense of maximum current.

Overall, the various forms of DLC are not quite good enough for field emission applications,
because their ESD is too low, and their emission is somehow extrinsic and not fully controllable. Even
nanostructured $sp^2$ based carbons are not good enough. Carbon nanotubes are the preferred form of
carbon for field emission. They are conductive, and can carry very large current densities if needed.

9. Mechanical and other properties

The mechanical properties of DLCs are of great importance because of the use of DLC as a
protective coating [493–503]. The ta-C films are being studied for possible use in MEMs instead of
polysilicon [497]. This is because ta-C is the hardest material known after diamond itself. A great
advantage of DLC compared to CVD polycrystalline diamond is that it is amorphous with no grain
boundaries. This means the films are extremely smooth. DLC is also deposited at room temperature,
which is an advantage for temperature sensitive substrates such as plastic. DLC films also have
extremely good coverage, so they act as good corrosion barriers. This is particularly useful in a major
application, which is to coat disks and recording heads in the magnetic storage technology [501–
503]. The disadvantages of DLC films are their intrinsic stress and thermal stability. (The latter is not
a problem for ta-C.) These problems are now being circumvented.

9.1. Elastic properties

Many of the mechanical properties of DLC thin films are measured by a nano-indentor [504–
507]. In this experiment, a small diamond tip is progressively forced into the film, and the force-
displacement curve is measured. Fig. 137 shows a typical curve for a ta-C:H [95]. The curve is also
measured for the unloading cycle. The hardness is defined as the pressure under the tip, given by the
ratio of force to the projected area of plastic deformation. The hardness in this case must allow for
the elastic recovery of the film. A tangent is drawn to the unloading curve at maximum load and
extrapolated to zero load. The Young’s modulus is proportional to the slope of this line. The plastic
deformation is $h_p$ and the elastic deformation is the length from the intercept $h_p$ to the maximum
indent $h_{\text{max}}$, $h_{\text{max}} - h_p$. The hardness is then

$$ H = 0.0378 \frac{L_{\text{max}}}{h_p^2} $$

and the Young’s modulus is

$$ E = 0.179 \frac{(1 - v^2)L_{\text{max}}}{(h_{\text{max}} - h_p)h_p} $$

A value for the Poissons ratio $v$ must be assumed, but it is not so critical as $v \ll 1$. 
It is well known that for accuracy, the indentation depth must be limited to a fraction of order 10% of the total film thickness [505]. This is particularly important for the case of hard on soft films, such as DLC. The hardness is found with reasonable accuracy by this procedure. The uncertainty in the Young’s modulus is much greater [18,506–508], and it can involve finite element modelling [509].

A more direct means to measure the elastic moduli is using surface acoustic waves. Schultrich and coworkers [510,511] have used laser induced surface acoustic waves (LISAWs) to determine the moduli of a-C films from ta-C to more sp² carbon. LISAW is a relatively fast, non-destructive method. Surface acoustic waves actually give the shear modulus, S. The Young’s modulus E can only be derived if a value for the Poissons ratio ν is assumed. Scheibe and coworkers [510] assumed a value of 0.1.

A related method to measure moduli is surface Brilloum scattering (SBS) [512]. In this case, photons from a laser are inelastically scattered off thermal phonons in the film and detected. A dispersion relation of frequency versus wavenumber of the phonons is found by varying the incidence and detector angle. The waves excited are called modified-Rayleigh waves and are confined within the film, under certain conditions. There are a number of advantages of the SBS method. It is non-destructive and can be reliably used for hard on soft films [513–515]. It can be used on films as thin as 10 nm of DLC. It can determine the two independent elastic constants of an isotropic amorphous solid. It needs a value of the film density to be given. Any internal layering of the film complicates the analysis. SBS finds the Young’s and shear modulus with higher accuracy, and the bulk modulus and Poissons ratio with lower accuracy [513].

Table 8 compares the elastic constants of some types of DLC [513,514] with those of diamond [12]. The Young’s modulus of ta-C with 88% sp³ bonding is found to be 757 GPa [513], compared to 1144 GPa for diamond itself. The bulk modulus of ta-C is found to be 344 GPa compared to the diamond value of 444 GPa. Note that the Poissons ratio is found to be ~0.12, nearly as low as the remarkably low value of 0.07 of diamond itself. On the other hand ta-C:H is found to have a Young’s modulus of 300 GPa, much less than that of diamond [513]. This shows the strong effect of how hydrogen has lowered the network coordination of ta-C:H. Its Poissons ratio is about 0.3, much larger than diamond’s and closer to the larger values found by Brilloum scattering for PECVD a-C:H by Jiang [515].

Fig. 137. Load-unloading curve for a nano-indentation measurement of ta-C:H.
Table 8
Elastic properties of various carbons measured by Brillouin scattering, compared to those of diamond and those calculated for a 100% network

<table>
<thead>
<tr>
<th></th>
<th>ta-C [513]</th>
<th>ta-C:H [513]</th>
<th>100% sp\textsuperscript{3} ta-C (calculated [278,279])</th>
<th>Diamond</th>
</tr>
</thead>
<tbody>
<tr>
<td>Density (g cm\textsuperscript{-3})</td>
<td>3.26</td>
<td>2.35</td>
<td>3.5</td>
<td>3.515</td>
</tr>
<tr>
<td>H (at.%)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>sp\textsuperscript{3} (%)</td>
<td>88</td>
<td>70</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Young’s modulus (E; GPa)</td>
<td>757</td>
<td>300</td>
<td>822.9</td>
<td>1144.6</td>
</tr>
<tr>
<td>Shear modulus (G; GPa)</td>
<td>337</td>
<td>115</td>
<td>366</td>
<td>534.3</td>
</tr>
<tr>
<td>Bulk modulus (B; GPa)</td>
<td>334</td>
<td>248 (+197, −0)</td>
<td>365</td>
<td>444.8</td>
</tr>
<tr>
<td>v</td>
<td>0.12</td>
<td>0.3 (+0.09, −0)</td>
<td>0.124</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Fig. 138 shows the bias-voltage dependence of the Young’s modulus of a-C:H deposited from methane [507]. The modulus is seen to increase very quickly at low voltage, to reach a maximum in the diamond-like bonding regime around 100 V, and then it declines gradually. This is similar to the behaviour of the density.

Fig. 139 shows the corresponding variation of Young’s modulus of ta-C deposited by laser arc [511] as a function of the ta-C density. A steady increase in modulus occurs due to the change from sp\textsuperscript{2} to sp\textsuperscript{3} bonding with increasing density.

9.2. Constraint counting model

The elastic moduli of amorphous carbons depend on the rigidity of the individual bonds and the average coordination of the network. This can be quantified by the constraint-counting model [516,517]. The model assumes the elastic energy of the network can be represented by a valence force field of nearest neighbour bond stretching and bond bending forces, as in Eq. (32). For a network with low coordination, there are many ways to deform the network which cost no energy. The number of such independent deformations or zero-frequency modes of a network of N atoms is given by the number of degrees of freedom (3N) minus the number of constraints due to the valence

Fig. 138. Variation of Young’s modulus (measured by nano-indentation) with bias voltage for a-C:H deposited from methane, after Jiang et al. [506].
force field. The number of constraints varies with coordination, $r$. There is one constraint per bond stretching (shared between the two atoms) and $2r - 3$ for the angle constraints for an $r$-coordinated atom. This gives a total of

$$N_{\text{con}}(r) = \frac{5r}{2} - 3$$  \hspace{1cm} (103)

per atom. Monovalent atoms like hydrogen have simply

$$N_{\text{con}}(1) = \frac{1}{2}$$  \hspace{1cm} (104)

The fraction of zero-frequency modes is then given by $f = 1 - (N_{\text{con}}/3N)$. Writing $f = \sum f_r x_r$, we obtain

$$f = \sum_{r>1} x_r \left( \frac{2 - 5r}{6} \right) = \frac{2 - 5\langle r \rangle}{6}$$  \hspace{1cm} (105)

where $x_r$ is the concentration of $r$-fold sites, and $\langle r \rangle = \sum x_r r$ is the mean coordination. Eq. (105) shows that there is a critical coordination of [517]

$$r_c = 2.4$$  \hspace{1cm} (106)

Below this, the network is unconstrained and has 0 frequency modes and is ‘floppy’. Above 2.4, the network is constrained and rigid.

He and Thorpe [517] have shown that the Young’s modulus $E$ of a network scales with the excess coordination to the power 1.5,

$$E = k(r_c - 2.4)^{1.5}$$  \hspace{1cm} (107)

Some authors [518] suggest that the coordination of a-C:H appears to lie near the line $\langle r \rangle = 2.4$ on the ternary phase diagram, Fig. 1. This is mistaken. DLC is notable because it is mechanically hard.
and has a finite elastic modulus or rigidity \([519]\). Eq. \(107\) emphasises that a network must be constrained in order to be rigid and hard. In Fig. 1, the boundary between polymers and a-C:H is the line of zero constraint. It is seen that real a-C:H lies above this line.

Jansen and Angus \([518]\) noted that Eqs. \((104)\) and \((105)\) can be used for networks including H, if \(r\) now represents the coordination of the C–C backbone, not the network. This is equivalent to treating each C–H bond as a dangling bond. This requires the substitution in Eq. \((105)\) of \(r\) by \(r - (x_{1,r}/x_r)\) and \(x_r\) by \(x_r/(1-x_1)\). Here, \(x_1\) is the H concentration and \(x_{1,r}\) is the concentration of hydrogen bonded to \(r\)-fold sites. This gives

\[
f = \sum_{r>1} \frac{(2 - (5/6)r)x_r + (5/6)x_{1,r}}{1 - x_1} \quad (108)
\]

The constraint model must also be modified slightly to include the effects of graphitic clusters, as described in Robertson \([519]\). This allows for the fact that although a graphite layer is three-fold coordinated, it does not convey rigidity in three-dimensions, because it uses all its constraints in the two dimensions of the layer. A simple approximation is to consider a graphitic cluster bonded at \(r'\) points into the rest of the network as an \(r'\)-fold coordinated ‘super-atom’ and use the above formulae. The effect of clustering is to displace the critical coordination \(r^c\) to larger values.

The constraint counting model gives a good description of the elastic moduli of a-C films. Fig. 140 shows that the Young’s modulus derived from LISAW and SBS measurements \([511,513]\) varies with fraction as required by Eq. \((107)\). The line extrapolates to an \(sp^3\) faction of \(-0.4\), which is a coordination of 2.6. This is close to the theoretical value of 2.4 of Eq. \((106)\), and this indicates that there is little clustering in ta-C. The line also extrapolates to a Young’s modulus of 800 GPa at 100% \(sp^3\). This value is lower than the isotropically average Young’s modulus of diamond, and close to the modulus calculated for the Wooton–Winer–Weaire (WWW) 100% \(sp^3\) bonded network \([275,276]\).

Fig. 140. Plot of (Young’s modulus)\(^{2/3}\) vs. carbon coordination for ta-C, showing the intercept at 2.6. Data derived from Schultrich et al. \([511]\).
suggests that random networks are slightly softer than the equivalent crystal, despite containing the same bonds. The cause of this is unclear.

Fig. 141 compares the predicted Young’s modulus from the constraint-counting model [519] with that measured for a-C:H deposited from methane [516]. The coordinations needed for the calculation are taken from the NMR data of Tamor et al. [173]. The model is seen to reproduce the main features of the variation, including the maximum at bias voltages of around 100 V. In these materials, the sp² sites contribute little rigidity to the network. This therefore supports the cluster model of DLC; the bonds sp³ controlling the mechanical properties while the sp² sites control the electronic properties [3,519].

9.3. Hardness

The hardness is a measure of the yield stress of a material [520]. The measurement of hardness by nano-indenteter was described earlier. Hardness is related empirically to the yield stress \( Y \) and Young’s modulus \( E \) by [520]

\[
\frac{H}{Y} = 0.07 + 0.06 \ln \left( \frac{E}{Y} \right)
\]  

(109)

This gives \( H/Y \sim 1.8 \) for materials with a low \( Y/E \) ratio like diamond and other ceramics. In brittle materials like DLC, yield occurs by bond cleavage. In this case, the yield stress can be calculated by the Orowan approximation [521]. This says that the yield stress is given by the work done to break the bonds, approximating the force curve of a bond by a sine curve. This gives

\[
\frac{E}{Y} = \pi
\]  

(110)

Combining these equations gives

\[
\frac{H}{E} = 0.16
\]  

(111)
The ratio for many materials is 0.1, which is the ratio for diamond. For other DLCs, it can approach 0.16. Fig. 142 shows the hardness of a-C:H deposited by PECVD from methane and benzene precursors [4,505]. The hardness is low at low bias voltages, for the polymeric a-C:H films. The hardness then rises to a maximum corresponding to the maximum diamond-like character, and then it decreases, as the films acquire a more graphitic bonding. The hardness reaches a maximum at low bias voltages for methane a-C:H, just like the dependence of density and Young’s modulus. The hardness for the benzene-derived film reaches a maximum around 1000 V, which is also consistent with the behaviour of the density.

Note that the maximum nano-hardness of the a-C:H in Fig. 142 is 17 GPa. The importance of using a nano-indentor and small indent depths is well known. If conventional micro-indenters are used on DLC, these give hardness values of typically three times higher than nano-indenters [486]. Thus, hardness values for DLC films over 20 GPa should be viewed as exaggeration unless there is some reason in terms of improved bonding or low hydrogen content to account for them.

The nano-hardness of ta-C with 85–90% sp³ bonding has been measured to be at least 60 GPa by Pharr et al. [18], 54 GPa by Shi et al. [67], 45–50 GPa by Martinez et al. [522] and 80–88 GPa by Friedmann et al. [334]. This is a difficult measurement because of the thinness of the ta-C film. It is likely that the Friedmann value is accurate as they used the thickest ta-C film, after using annealing to relieve the stress. Fig. 143 shows the hardness of ta-C versus sp³ fraction from Shi et al. [67]. The fall in hardness with decreasing sp³ content may be too steep, compared to the equivalent fall in Young’s modulus found by Schultrich et al. [511] measured by LISAW, if there is a fairly constant value of H/E, as expected.

Fig. 144 compares the Young’s modulus and hardness of ta-C:H derived from nano-indentation measurements, as a function of ion energy per C atom [95]. The values are now much larger than for a-C:H. The modulus and hardness pass through a maximum as a function of ion energy. Thus, they follow the variation with sp³ fraction quite closely. The ratio of hardness to Young’s modulus of about 0.1–0.16 as expected by theory. Fig. 145 shows the variation of H²/³ with sp³ fraction for ta-C:H, to test the constraint counting hypothesis, Eq. (107). It shows reasonable agreement. The intercept is −0.5, which again is quite close to the theoretical value of 2.4 for coordination number. This shows that there is not so much aromatic clustering in ta-C:H.
Fig. 143. Variation of nano-hardness of ta-C with sp$^3$ fraction, after Shi et al. [67].

Fig. 146 shows the variation of Young’s modulus and hardness of ta-C:H as a function of deposition temperature [348]. This plot shows that $E$ and $H$ fall sharply at the temperature $T_1$, so that they correlate closely with the fall in sp$^3$ content.

9.4. Adhesion

The major use of DLC is for protective coatings. Thick films are preferred for this, to maximise the wear life. The compressive stress limits the maximum thickness of adhesive films. A film of
Fig. 145. Plot of (hardness)$^{2/3}$ vs. fraction for ta-C:H [95], to test the constraint-counting hypothesis, and assuming $H/E$ is constant.

thickness $h$ will delaminate when the elastic energy per unit volume due to the stress $\sigma$ exceeds the surface fracture energy $\gamma$ per surface, so for adhesion

$$2\gamma > \frac{\sigma^2 h}{2E}$$

or

$$h < \frac{4\gamma E}{\sigma^2}$$

Fig. 146. Variation of Young’s modulus and nano-hardness of ta-C:H with deposition temperature, after Sattel et al. [348].
This sets an upper limit of film thickness. The basic problem of DLC coatings was noted by Tamor [523]. The stress, Young’s modulus and hardness all tend to be proportional to each other. Maximising the hardness means that the stress is a maximum. Thus, maximising the hardness will minimise the thickness of an adherent layer, so that little is gained.

Various strategies are used to maximise the film thickness. First, it is essential to ensure there is good adhesion between the film and the substrate. The film should fail by fracture within the substrate. This can be achieved by cleaning the surface by Ar ion bombardment before deposition. Another method is to use a high ion energy for the first stage of deposition, to cause ion beam mixing between film and substrate in order to ensure a mixed interface [524].

Second, it has recently been noted that there is not necessarily a direct correlation between modulus and stress (Fig. 44) [131], so that it may be possible to make high modulus films with lower stress.

Third, a carbide-forming adhesion layer such as Si, Cr or W can be deposited first before the carbon [298,299,493,524]. Si works well for glass substrates used in sunglasses and bar-code scanners, on which wider gap a-C:H may not adhere. Films can be made which are inhomogeneous or graded, so that they have internal stress relief mechanisms.

Fourth, it is possible to use multi-layers to provide internal stress relief. This has been used by Meneve et al. [525] for a-C:Si:H films and by Anders et al. [526] for ta-C films.

Finally, rather thick ta-C films have been formed by deposition on a softer intermediate layer of Al, Cu, Ti or AISI316 steel or similar with a Vickers hardness of under 3 GPa [524,527,528]. The intermediate layer is on top of the normal substrate, which could be Si. The idea is that the intermediate layer should yield locally below the film, and thereby absorb the effect of the compressive stress. Ion beam mixing is used for good adhesion to the soft substrate. The intermediate layer should be a carbide former (Ti, Fe, etc.) and if not a thin carbide forming layer such as Mo is used. It is also found that pulsed cathodic arcs give thicker films than continuously operating FCVAs in this case. Antilla et al. [524] were able to deposit ta-C films well over 10 μm by this method.

A major effort has been made to reduce stress by alloying with for example Si [529–531] or metals [532,533]. Si has three beneficial effects. It can promote sp³ bonding by chemical means, as it does not exert sp² hybridisation, so it is not necessary to use as much ion bombardment to get sp³ bonding. It increases the thermal stability of the hydrogen, as noted earlier. It also improves the friction performance, as discussed shortly, by maintaining a low friction to higher humidity. Si alloying is widely used to improve performance. Metals reduce stress and also increase hardness and toughness, both of which increase the maximum h. Metals, however, make the film opaque.

9.5. Friction

Friction is the dissipation of energy as the two surfaces move over each other [534,535]. Friction occurs by contact between the two surfaces at a few high points or asperities. The real area of contact is much less than the apparent area of contact, so that the real contact pressure is much greater than the average load pressure. The friction force arises either from adhesion, deformation or abrasion at the contact. If the true area of contact is A, and the shear strength of the contact is Y, then the lateral friction force is [534],

\[ F = AY \]  

(114)

The true area of contact is given from the load W and real contact pressure, which equals the hardness of the softer material, H,

\[ W = AH \]  

(115)
where \( H \) is related to the shear strength by \( H = cY \), where \( c = 3–5 \), so the coefficient of friction \( \mu \) is given by

\[
\mu = \frac{F}{W} = \frac{Y}{H} \approx 0.2
\]

A more detailed analysis shows that there is both adhesion and deformation at a real Hertzian contact. Adhesive contact dominates for elastic materials with high \( H/E \) ratios, including ceramic materials like DLCs.

DLCs are notable for their low friction coefficients [493,494,536–556]. The friction properties of DLC have been reviewed recently by Grill [498,537], Donnet [545] and Gangopadhyay [499]. The unlubricated friction coefficient of DLC on steel is similar to the lubricated friction coefficient of steel on steel [538]. Erdemir et al. [549,550] finds that a-C:H in a vacuum can have a friction coefficient as low as 0.01.

For a-C:H, \( \mu \) depends strongly on the relative humidity. Values of \( \mu \) below 0.05 are found in a vacuum and at low humidity and \( \mu \) increases strongly at high humidity [542,544]. A typical dependence is shown in Fig. 147. Erdemir et al. [549,550] finds that the friction coefficient in this low humidity regime depends on the precursor used to make the a-C:H. It varies with the ratio of H/C in the precursor, with a-C:H from CH\(_4\) having the lowest \( \mu \) and a-C:H from C\(_2\)H\(_2\) having the largest \( \mu \). The a-C:H prepared from hydrogen-diluted methane has the lowest friction coefficient [550]. The \( \mu \) increases to 0.1–0.15 when the humidity rises to normal atmospheric values of 30% and above. In situ studies have shown that the higher \( \mu \) arises from the water, not oxygen in the atmosphere [544].

The ta-C and ta-C:H behave differently. Voevodin et al. [543] finds that ta-C has a larger \( \mu \) than a-C:H in the vacuum of about 0.1 to 0.15. The \( \mu \) then decreases weakly as the relative humidity increases (Fig. 147). The ta-C:H deposited from acetylene appears to behave in the same way [556].

The friction behaviour of DLCs is explained as follows [543,552]. The diamond surface is normally inert because its broken bonds are generally passivated by C–H bonds [12,552]. This means that diamond surfaces are hydrophobic, closed-shell bonded systems, and they will only contact
through van der Waals forces. Under pressure, the contact will be elastic. As the contact shears, it will break at the weaker van der Waals bonds, not the bulk C–C bonds, so the friction force has an adhesive/deformation nature rather than abrasive. The hydrophobic nature of DLC is important as an interlayer of water will also provide adhesive contact.

The a-C:H behaves in this manner under vacuum or low humidity. Contact with a different surface causes a transfer layer of a-C:H to be formed on the other surface [543]. Thus, the contact is between two basically similar hydrophobic a-C:H surfaces, and the friction coefficient is very low. High humidities are believed to interfere with the formation of the contact layer, and cause it to become oxidised/hydrated, so that it no longer forms the van der Waals bonds. If the transfer layer does not form, the counter surface is no longer hydrophobic, and the friction coefficient is much higher. Voevodin et al. [543] observed the nature of the contact layer and the wear track by micro-Raman. Erdemir et al. [549] suggests that the extremely low friction of a-C:H made from methane compared to acetylene is because there are fewer dangling C bonds to form residual adhesive contacts.

The ta-C behaves differently. Contact and wear now transforms the ta-C at the contact area into graphitic layers [543], these can transfer. They behave as a solid lubricant. Intercalation of water between the graphite layers lowers the friction coefficient. Thus, \( \mu \) for ta-C decreases with increasing humidity. The different contact and wear mechanism was seen by micro-Raman by Voevodin et al. [543]. The nature of wear track is also seen by reflection EELS by us [556].

The addition of silicon to a-C:H improves its friction properties by allowing it to maintain its low friction coefficient from low up to high humidities. It is believed that this occurs because the Si changes the nature of the transfer layer, forming a silica-gel-like sacrificial layer [530,531].

9.6. Wear

Wear is the removal of material during the contact process [534]. Wear occurs by adhesion, abrasion or by corrosive/oxidative wear. For low applied loads, the load makes an elastic contact with the surface, and no permanent groove is left. There is no wear. At moderate loads, a permanent groove or wear track is left by plastic deformation or abrasion. At high loads, a wear track of cracking and deformation remains. The wear is the volume of a material removed from a surface by contact. The volume of wear per unit track length \( Q \) is given by

\[
Q = KA
\]

(117)

where \( A \) is the real contact area and is given as earlier by the ratio of load \( W \) to the pressure or hardness \( H \) of the softer surface, \( A = W/H \). This gives the Archard equation [534],

\[
Q = K \frac{W}{H}
\]

(118)

where \( K \) is the wear coefficient, a dimensionless number. It is much less than 1. A dimensional wear coefficient \( k \) is often used, defined by

\[
Q = K \frac{W}{H} = kW
\]

(119)

where \( k \) is usually given in units of \( \text{mm}^3 \text{ N}^{-1} \text{ m}^{-1} \).

The Archard equation says that the wear coefficient is inversely proportional to the hardness of the surface. Hence harder surfaces wear less, and the aim for protective coatings is to maximise their
Table 9

Comparison of wear rates for ta-C and a-C:H, after Voevodin [543]

<table>
<thead>
<tr>
<th></th>
<th>Wear rate (mm$^3$ N$^{-1}$ m$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ta-C</td>
<td>$10^{-9}$</td>
</tr>
<tr>
<td>a-C:H</td>
<td>$10^{-6}$ to $10^{-7}$</td>
</tr>
</tbody>
</table>

Hardness. Voevodin et al. [543] and Gangopadhyay [499] have compared wear rates for the various types of DLC with other coating materials. These are summarised in Table 9. It is found that ta-C has an extremely low wear rate of order $10^{-9}$ mm$^3$ N$^{-1}$ m$^{-1}$, this is about a 100 times lower than that of a-C:H at $10^{-7}$ mm$^3$ N$^{-1}$ m$^{-1}$. Other wear rate measurements for a-C:H compiled by Grill [537] are of similar size. The value for ta-C is impressively low. It is the low wear rate of ta-C which fundamentally makes ta-C so suitable for MEMs devices. Martinez et al. [522] found that the wear rate of ta-C had a minimum at the maximum of hardness, as expected from (117).

The wear mechanism of ta-C and a-C:H are the friction mechanisms, adhesive wear via transfer layers. The ta-C transforms by stress-induced transformation to a graphitic over-layer, which then acts as a lubricant, whereas a-C:H forms a C:H transfer later on the counter surface [543]. The more sp$^2$ bonded a-C was found to wear by oxidation to CO$_2$ in an air ambient [552].

A novel form of carbon is hard carbon with fullerene-like inclusions. This is prepared by deposition on a cathodic arc in the presence of nitrogen gas [556], or by sputtering carbon in a nitrogen atmosphere [425] both at elevated temperature (~200 C). The local atomic structure has been viewed by high resolution transmission electron microscopy and EELS. The films show unusual mechanical properties of high hardness and very high elastic recovery. This could make them valuable for coatings.

Nano-crystalline diamond is also much smoother than polycrystalline diamond [557], and it probably has superior friction and wear properties compared to DLC. However, it does require higher deposition temperatures.

9.7. Surface properties

DLC is also notable for its small surface energy [532]. The surface energy is usually measured using the contact angle. The contact angle $\theta$ of a liquid on a DLC surface depends on the three energies, the surface energy of the DLC $\gamma_s$, the surface energy of the liquid $\gamma_l$ and the interfacial energy of the liquid and the DLC $\gamma_{sl}$, via the Young’s equation,

$$\gamma_s = \gamma_{sl} + \gamma_l \cos \theta$$

The surface energy is extracted by measuring $\theta$ for more than one liquid, and apportioning the tension between dispersive and dipole components. Low surface energies give large contact angles. Surface energies are typically 40–44 mN m$^{-1}$ [558,559]. The a-C:H generally has a contact angle with water of 55–70$^\circ$ [558–561]. Unusually, this is found to be independent of deposition bias voltage [558].

Leezenberg et al. [561] found the contact angle aged during storage, and it could be varied by exposing the a-C:H to a reactive gas whilst still in the vacuum deposition chamber. They attributed this to a passivation of surface dangling bonds.

Dimigen and coworkers [559] tested the effects of alloying elements on the surface energy of a-C:H. They found the surface energy of a-C:H increased from 41 mN m$^{-1}$ for pure a-C:H to
52 mN m\(^{-1}\) with oxygen addition, but decreased to 19–24 mN/m with the addition of Si or F. For reference polytetrafluoroethylene is 18.5 mN m\(^{-1}\). Leezenberg et al. [561] also found that F decreased the surface energy, ultimately to values close to that of polytetrafluoroethylene.

The ta-C is also hydrophobic. Its contact angle of water lies in the range 75–80° [562]. However, this is less than a-C:F. The addition of metals such as Al and Fe reduces the surface energy to 25 mN m\(^{-1}\) and makes a-C more hydrophobic [563,564].

9.8. Electrochemistry

Diamond is chemically very stable. Boron doped diamond is a semimetal, and boron doped diamond electrodes are of great interest because of their unusually stability range. They can drive electrochemical reactions over the range \(-1\) to 2.5 V SHE [565]. The diamond does not oxidise at large positive potentials unlike graphite, which makes it very useful for oxidative processes. The best electrodes have little grain boundary material.

The ta-C may also be of use. Nitrogen doped ta-C can be made with moderate conductivity and yet retaining its sp\(^3\) bonding. This sp\(^3\) bonding allows ta-C:N to retain its electrochemical stability. Shi et al. [68] and Miller coworkers [566] have tested such ta-C:N electrodes.

9.9. Biocompatible coatings

There is considerable interest in using various forms of DLC as biocompatible coatings, on parts such as replacement hip joints, heart valves and stents [567–574]. This arises from the low friction coefficient of DLC and the belief that a carbon material must be biologically compatible. In many coatings, the main requirement is that the DLC has good adhesion to the underlying unit and does not produce metallic wear debris. Deeper investigations have studied the attachment of blood cells, proteins, etc. This work is reviewed in detail by Allen et al. [574].

9.10. Coatings of magnetic hard disks

One of the most important uses of DLC films is as protective coatings on magnetic storage disks [500–503]. Magnetic storage as in computer hard drives is presently the most economic form of mass non-volatile storage. Its great advantage is that the storage density is increasing at a very rapid rate [503] (Fig. 148). Storage densities are increasing at 100% per year since the introduction of giant magneto-resistive (GMR) heads. Recently, Seagate and Fujitsu have demonstrated a storage density of 100 Gbits in.\(^{-2}\).

The disk consists of a magnetic layer of Co–Cr–Pt alloy thin film which stores the data, Fig. 149. A protective layer of DLC coating is applied over the Co layer, and one–two monolayers of a perfluoropolyether such as ZDOL or Fomblin is used as a molecular lubricant. A read/write head flies above the rotating disk on an aerodynamic bearing. The read/write head consists of many layers of thin films and is also protected by a DLC film. DLC is used because it makes extremely smooth, continuous and chemically inert films, with a surface roughness below 1 nm. DLC has no competitors as a coating material for this application.

The need for even higher storage densities means that the head flying height must be reduced, which allows less room for the carbon layers. They are presently about 4 nm thick and need to reach 2 nm soon [474], as shown in Fig. 150. This is a thickness of only about 12 atomic layers. The carbon has traditionally been deposited by sputtering. The carbon on the heads and sliders is now often deposited by the ion beam method. However, sputtering is unable to provide continuous films
of 2 nm. There is now a transition as new processes such as cathodic arc or high-density plasma deposition are needed to make the thinner films.

The first role of DLC films was to provide protection against corrosion. Simple a-C was used, deposited by magnetron sputtering [552]. Later, a-C:H was used, produced by the reactive sputtering in an Ar/hydrogen atmosphere [44]. Its role was to provide also protection against mechanical wear and damage during head crashes. The best performing a-C:H contained more hydrogen than needed to maximise its hardness [500]. This indicated that the role is not simple protection against dry wear, but that the interaction with the lubricant is critical. The surface termination of a-C:H immediately after deposition is important for its surface properties and the interaction with the lubricant [561]. Most recently, a-C:N is being used instead of a-C:H [575]. The primary role is now to form a continuous film, to give protection against corrosion, and to separate the Co layer and read/write head from the ambient. On the disk, the carbon should also provide a surface for the perfluoropolyether lubricant to adhere and move on.

Fig. 148. Increase in storage density of hard disk drives with date.

Fig. 149. Schematic diagram of a magnetic disk and read/write head, from Goglia et al. [503].
Fig. 150. Variation of flying height, magnetic spacing and carbon film thickness with storage density, after Goglia et al. [503].

The role of DLC is often said to be to provide mechanical protection. It is unlikely that a 5 nm thick carbon film can really provide much mechanical protection. Thus, the primary requirement of the carbon film is to be continuous so that it can act as a corrosion barrier and a diffusion barrier.

Addition of nitrogen improves the tribological properties and toughness of a-C [575, 576]. Another advantage of a-CNₓ relates to the lubricant. Under load, the perfluoropolyether lubricant molecules can degrade with the emission of F. F can abstract H from a-C:H under the local heating to give HF. This then reacts aggressively with the Al₂O₃ in the head to form AlF₃ which catalyses the decomposition of the lube. In contrast, heating of a-CNₓ releases only nitrogen, which will not give rise to HF [577]. Nitrogen is still added to ta-C:H or ta-C on disk coatings because it improves the adhesion of the lubricant by providing a site for dipolar bonding.

The ta-C and ta-C:H are being introduced because of their superior smoothness [578, 579]. Fig. 151 compares the roughness of a-C:H and a-C:N films measured by Waltmann et al. [580] with ta-C films measured by Ilić and Robertson [581], as a function of the film thickness. The ta-C films

Fig. 151. Variation of root mean square roughness vs. film thickness for ta-C and a-C:H, data from Ilić and Robertson [581], and Waltman et al. [580].
are clearly much smoother. The a-C:H films show a maximum of roughness at 2 nm thickness, which may correspond to islanding and pin-holes. This is not seen in the ta-C films. Electrochemistry voltammetry testing is often used to assess whether a film is fully continuous.

Why are ta-C and ta-C:H smoother? To understand this, consider the formation of a continuous film [502,582]. DLC is amorphous and has no grain boundaries, which is the first important requirement. The DLC must also grow as a heterlayer above some other phase—that is, a new phase must nucleate and grow. The phase must have the correct interfacial energies so that it does not form islands. Thus, the carbon layer must satisfy two requirements. Firstly, the carbon phase must nucleate everywhere on the surface, so there must be no barriers to nucleation. Secondly, the carbon species should have no surface mobility during deposition so that they do not move across the surface to form islands, in an attempt to reduce the surface energy, and leave some parts of the surface uncovered. Both of these require a deposition flux with a high fraction of energetic species (>10 eV). Energetic incident species will overcome any nucleation barriers. Also, if they are subplated, they cannot diffuse over the surface and help islands to form.

A sputtering flux contains a large fraction of neutral atoms with minimal kinetic energy. These atoms cause the roughness. The methods like FCVA or ECWR with highly ionised energetic deposition fluxes produce smoother films, because they have few neutrals. The films are also dense, and so act as good corrosion barriers. In this way, the film hardness is not necessarily the most important parameter.

References