1. Perspectives on Electronic and Photonic Materials

Tim Smeeton, Colin Humphreys

Electronic and photonic materials have a tremendous impact on the modern world. They include a wide range of material classes and are developed through a deeply interdisciplinary combination of physics, chemistry, materials science, and engineering. In this introductory chapter, we give some perspectives on this exciting and ever-changing field. We give an example of the tremendous integration of different materials used in today's consumer products, and then take a historical look at the development of some key semiconductor materials and devices from inception to today. Focusing in particular on the development of the transistor and integrated circuit and some of the key electronic and photonic applications of compound semiconductors, we take advantage of the long-distance view to point out some unifying themes across the wide portfolio of materials while appreciating their unique features.

It can be easy to forget how remarkable electronic and photonic materials are. Take the light emitting diode (LED) as an example: An electric current is passed through a tiny stack of layers of slightly different materials and brilliant colored light is emitted. Of course, the LED is carefully designed, and there are powerful theories to explain the behavior, but that should not detract from the initial moment of wonderment that it works at all. Or that a room can be lit by photons generated in the active region of an LED which includes just a few thousand cubic micrometers of material – about the same volume as a grain of flour.

We all literally *see* the output from LEDs every day, from television screens, vehicle lights, or lighting luminaires. Most other electronic and photonic materials and devices are less conspicuous to our senses but together they have played an indisputable role in defining the way people live, work, and communicate in the twenty-first century: from microprocessors containing billions of transistors which provide immense computational power to laser diodes and radiofrequency transceivers which enable trans-global and wireless communication. These complex devices are only available today thanks to the work of countless researchers over the past century who have identified, designed and understood a vast array of materials, produced them with extraordinary purity, quality and economy, and deployed them in device designs which harness their power.

The detailed chapters in this handbook provide comprehensive introductions to the huge range of technical fields which electronic and photonic materials now occupy, written by world experts. In this introductory chapter, we have the luxury of stepping back from the details of these complex fields and reviewing the whole, in search of some perspective. First, we take a look at the complex mixture of electronic and photonic materials which are in use today and reflect on some common themes. Next, the majority of the chapter is dedicated to a review of the development of a handful of the most important semiconductor materials and devices since the early breakthroughs of the 1940s: the birth of electronics in germanium and silicon; the integrated circuit (IC); some of the key electronic and optoelectronic uses of III–V semiconductors; and the recent rise of III-nitride materials. These reviews serve as a tribute to breakthroughs of the last 75 years, hopefully providing some perspective of how the science and technology of these materials has come to its current state and perhaps providing some encouragement and inspiration to those who are striving to develop materials today.

S. Kasap, P. Capper (Eds.), *Springer Handbook of Electronic and Photonic Materials*, DOI 10.1007/978-3-319-48933-9_1

1.1 Tremendous Integration

A starting point for perspective is to appreciate the complex mixture of materials and devices that are routinely integrated together. For a case study, we need look no further than something which is probably within touching distance as you read this sentence: a smartphone. One of the defining technologies of the twenty-first century, it has been enabled by an astonishing convergence of electronic and photonic materials and devices. No two models are quite the same, but let us teardown a hypothetical handset and its supporting infrastructure, laying forward references to the detailed chapters in this book as we go. As a user, we see text, images and video from a high-resolution liquid crystal display (LCD) in which blue light emitted from indium gallium nitride (InGaN) LEDs (Chaps. 31, 35, and 40) excites a rare-earth-doped yttrium aluminum garnet (YAG) phosphor (Chap. 38) to produce white light which, in turn, is color-filtered into red, green, and blue components and modulated at the pixel level by liquid crystal cells (Chap. 36) controlled by silicon or indium gallium zinc oxide (IGZO) thin film transistors (TFTs) (Chap. 44). We have touch control thanks to capacitive sensors which use an optically transparent and electrically conducting oxide such as indium tin oxide (ITO) (Chap. 58). Low power consumption multicore processors, RAM, and flash memories – all silicon – provide the computing power and storage (Chap. 21). The device is powered by a high energy-density lithium ion battery (Chap. 11). Wireless connectivity is provided between gallium arsenide (GaAs) based transistor radio transceivers on the handset and base stations equipped with silicon, silicon carbide (SiC), or aluminum gallium nitride (AlGaN) transistor amplifiers. Beyond that, data communication – which gives the smartphone its purpose – is via silica (SiO_2) optical fibers (Chap. 41) using indium phosphide (InP)-based laser diodes, optical receivers, switches and amplifiers (Chap. 35) to data centers based on high-performance silicon processors. And we can go further still: The data centers consume electrical power distributed through more power transistors (silicon, SiC and AlGaN) and perhaps originally generated by crystalline silicon, amorphous silicon, copper indium gallium selenide (CIGS) or cadmium telluride (CdTe) photovoltaic cells (Chap. 43).

The list of core materials is already well into double figures, but we have still only scratched the surface. If we consider just the families of semiconductor components in a little more detail, these include multiple alloy compositions (Chap. 30) with optimized combinations of impurity doping (Chap. 2). Not even to mention dozens – probably hundreds – of different electrode, interconnect, dielectric and packaging (Chaps. 27, 29, and 53) materials.

We could go on, but surely the point is made: A remarkably complex combination of materials and devices is involved. It is a phenomenal achievement of technology and business – we could argue a wonder of the modern world – that this dense concentration of different materials and technologies can be brought together with such synergy in a reliable, mass-produced product weighing less than 150 g that can be sold at a price accessible to billions of people. It is often the product design or the software of these devices which steal the limelight in marketing, but it is the electronic and photonic materials which are the unassuming heroes!

Electronic and photonic materials include almost all types and forms of material:

- Inorganic compounds (semiconductors, dielectrics)
- Organic molecules (liquid crystals, semiconductors)
- Ionic crystals (batteries), polymers (packaging, lithography masks)
- Metals/alloys (electrodes, interconnect wiring)
- Single crystal
- Polycrystalline
- Amorphous
- Thin film
- Nanostructured
- Two-dimensional.

Despite this breadth, there is at least one strong unifying theme across all: the need for extraordinary precision in the design and fabrication of the materials. These materials demand combinations of high purities, low defect densities, and precise chemical composition that probably exceed the requirements of any other area of technology. These requirements have been met by a unique convergence of physics, chemistry, materials science and engineering which brings together the core theoretical understanding with the means for fabrication.

Two aspects, in particular, are vital for delivering the necessary precision: exact material fabrication methods and detailed understanding of material structure and properties. In many cases, the ultimate device performance is determined by how far we fall short of fabricating the *perfect* material. If the number of crystalline defects is too high, or if there are too many impurities, or if an interface is too rough, or if . . . , a device may fail. Minute deviations from a target specification can have catastrophic effects. Taking LED or laser diode fabrication as an example, a change in composition of material or doping accounting to much less than 1% of a layer just a few nanometers thick can be the difference between a viable device and a costly failure. In nearly all device fabrication steps, from bulk growth methods for fabrication of substrates with near-perfect quality, to thin film growth methods to deposit precise material compositions with atomic-scale precision, to processing technology to pattern device features at the nanometer scale, and through to packaging and integration methods, material production techniques are the lifeblood of the field. The precision of fabrication techniques needs to be at least matched by the precision with which we can determine the structure of the resulting materials. This means that the structural and chemical characterization tools and techniques are an inseparable aspect of the story of all electronic and photonic materials. Accurate understanding of the quality (or otherwise) of the materials that we make – obtained from microscopy or other tools – has been a continuous driver or enabler for progress in material and device development.

When the material fabrication expertise and understanding of structure and properties are combined with deep understanding of the physics of device design, the result is the remarkable devices that can enrich our lives.

So, today's electronic and photonic materials span a tremendous range of fields and disciplines. How did this myriad of technologies successfully emerge and grow? Why have the materials that are in use today become dominant over others? In the next sections, we will give some very selective answers to these questions, hoping to provide a perspective of how some technologies have reached the current state. In particular, we have chosen to focus on a few sectors of semiconductor development to explain some of the early development of what became mainstream technologies. Often the key messages are learnt in these early years – before a technical field fragments into highly specialized sectors – and may provide the most valuable perspective.

1.2 The Silicon Age

1.2.1 The Transistor and Early Semiconductor Materials Development

Electronics is the control of electrons to produce useful properties; electronic materials are the media in which this manipulation takes place. The foundations for today's electronics were laid in 1947 – 50 years after J.J. Thompson had discovered the electron – with the first demonstration of the semiconductor transistor effect. Building on quantum theories to explain semiconductor behavior from the 1930s [1.1, 2] and improvements in purity of candidate semiconductor materials, John Bardeen and Walter Brattain used germanium to build and demonstrate the first *semiconductor triode*. Later, to be named the point contact transistor to reflect its transresistive properties, this success at Bell Laboratories was obtained just a few years after a research group led by William Shockley was established to focus on understanding semiconducting materials. It was to earn Brattain, Bardeen, and Shockley the 1956 Nobel Prize for Physics.

The first point-contact transistor was based around three contacts onto an n-doped germanium block: When a small current passed between the *base* and *emitter*, an amplified current would flow between the *collector* and *emitter* [1.3]. The emitter and collector contacts needed to be located very close to one another (50–250 μm), and this was achieved by evaporating gold onto the corner of a plastic triangle, cutting the film with a razor blade, and touching this onto the germanium – the two isolated strips of gold serving as the two contacts [1.4]. At about 1 cm in height, based on relatively impure polycrystalline germanium and adopting a different principle of operation, the device bears barely any resemblance to today's IC electronics components. Nonetheless, it was the first implementation of a solid-state device capable of modulating (necessary for signal amplification in communications) and switching (needed for logic operations in computing) an electric current. In a world whose electronics were delivered by the thermionic vacuum tube, the transistor was immediately identified as a component which could be *employed as an amplifier, oscillator, and for other purposes for which vacuum tubes are ordinarily used* [1.3].

In spite of this, after the public announcement of the invention at the end of June 1948, the response of both the popular and technical press was somewhat muted. It was after all still *little more than a laboratory curiosity* [1.5] and ultimately point-contact transistors were never suited to mass production. The individual devices differed significantly in characteristics, the noise levels in amplification were high and they were rapidly to be superseded by improved transistor types.

A huge range of transistor designs have been introduced from the late 1940s to now. These successive generations either drew upon or served as a catalyst for a range of innovations in semiconductor materials processing and understanding. There are many fascinating differences in device design but from a materials science point of view the three most striking differences between the first point-contact transistor and the majority of electronics in use today are the choice of semiconductor, the purity of this material, and its crystalline quality. Many of the key electronic materials technologies of today derive from the developments in these fields in the very early years of the semiconductor industry.

Both germanium and silicon had been produced with increasing purity throughout the 1940s [1.6]. Principally because of germanium's lower melting temperature (937 °C compared with 1415 °C) and lower chemical reactivity, its preparation had always proved easier and was therefore favored for the early device manufacture such as the first transistor. However, the properties of silicon make it a much more attractive choice for solid-state devices. Although germanium is expensive and rare, silicon is, after oxygen, the second most abundant element. Silicon has a higher breakdown field and a greater power handling ability; its semiconductor band gap (1.1 eV at 300 K) is substantially higher than that of germanium (0.7 eV), so silicon devices are able to operate over a greater range of temperatures without intrinsic conductivity interfering with performance.

These two materials competed with one another in device applications until the introduction of novel doping techniques in the mid-1950s. Previously, p- and n-doping had been achieved by the addition of dopant impurities to the semiconductor melt during solidification. A far more flexible technique involved the diffusion of dopants from the vapor phase into the solid semiconductor surface [1.7]. It became possible to dope with a degree of two-dimensional precision when it was discovered that silicon's oxide served as an effective mask to dopant atoms and that a photoresist could be used to control the etching away of the oxide [1.8, 9]. Successful diffusion masks could not be found for germanium and it was soon abandoned for mainstream device manufacturing. Dopant diffusion of this sort has since been superseded by the implantation of high-energy ions which affords greater control and versatility.

Shockley was always aware that the material of the late 1940s was nothing like pure enough to make reliable high performance commercial devices. Quantum mechanics suggested that to make a high quality transistor out of the materials, it was necessary to reduce the impurity level to about one part in 10^{10}. This was a far higher degree of purity than existed in any known material. However, William Pfann, who worked at Bell Laboratories, came up with the solution. He invented a technique called zone-refining to solve this problem, and showed that repeated zone refining of germanium and silicon reduced the impurities to the level required. The work of *Pfann* is not widely known but was a critical piece of materials science that enabled the practical development of the transistor [1.10, 11].

At a similar time, great progress was being made in reducing the crystalline defect density of semiconducting materials. Following initial hostility by some of the major researchers in the field, it was rapidly accepted that transistor devices should adopt single crystalline material [1.12]. Extended single crystals of germanium several centimeters long and up to 2 cm in diameter [1.12, 13] and later similar silicon crystals [1.14] were produced using the *Czochralski* technique of pulling a seed crystal from a high purity melt [1.15]. The majority of material in use today is still derived from this route. To produce silicon with the very lowest impurity concentration, an alternative method called float zoning was developed where a polycrystalline rod was converted to a single crystal by the passage of a surface tension confined molten zone along its length [1.16–18]. No crucible is required in the process, so there are fewer sources of impurity contamination. Float zoning is used to manufacture some of the purest material in current use [1.19]. The early Czochralski material contained dislocation densities of 10^5–10^6 cm^{-2} but by the start of the 1960s dislocation free material was obtained [1.20–23]. Initially, most wafers were on the silicon (111) plane, which was easiest to grow, cut, and polish [1.24]. For field-effect devices, which are discussed below, use of the (100) plane was found to offer preferable properties, so this was introduced in the same decade. The impurity concentration in dislocation-free silicon has been continually reduced up to the present day and wafer diameters have increased almost linearly (though accelerating somewhat in recent years) from about 10 mm in the early 1960s to the *dinner plate* 300 mm which dominates today [1.19] and with 450 mm tentatively in the pipeline within the next decade. These improvements represent one of the major achievements in the growth and processing of semiconductor materials.

A series of generations of transistors followed in rapid succession after Brattain and Bardeen's first triumph. Here, we only mention a few of the major designs whose production has traits in common with technology today. Early in 1948, *Shockley* developed a detailed formulation of the theory of p–n junctions that concluded with the conception of the junction transistor [1.25, 26]. This involved a thin n-doped base layer sandwiched between p-doped emitter and collector layers (or vice versa). This p–n–p (n–p–n) structure is the simplest form of the bipolar transistor (so-called because of its use of both positive and negative charge carriers), a technology that remains important in analogue and high-speed digital ICs today. In April 1950, by successively adding arsenic and gallium (n- and p-type dopants, respectively) impurities to the melt, n–p–n junction structures with the required p-layer thickness (≈ 25 μm) were formed from single crystal germanium. When contacts were applied to the three regions, the devices behaved much as expected from *Shockley's theory* [1.25, 27]. Growth of junction transistors in silicon occurred shortly afterwards and they entered production by Texas Instruments in 1954 [1.12].

By the later years of the 1950s, the diffusion doping technique was used to improve the transistor's speed response by reducing the thickness of the base layer in the diffused base transistor [1.28]. This began the trend of manufacturing a device in situ on a substrate material, so, in a sense, it was the foundation for all subsequent microelectronic structures. Soon afterwards, epitaxial growth techniques were introduced (what would today be described as vapor phase epitaxy (VPE)) which have since become central to both silicon and compound semiconductor technology. Gas phase precursors were reacted to produce very high quality and lightly doped crystalline silicon on heavily doped substrate wafers to form epitaxial diffused transistors. Even though the collector contact was made through the thickness of the wafer, the use of highly doped (low resistance) wafers reduced the series resistance and therefore increased the frequency response [1.29].

For some years, the highest performance devices were manufactured using the so-called *mesa* process where the emitter and diffused base were raised above the collector using selective etching of the silicon [1.25]. The planar process (which was at the heart of nearly all device production until the last few years) was subsequently developed, in which the p–n junctions were all formed inside the substrate using oxide masking and diffusion from the surface. This resulted in a flat surface to which contacts could be made using a patterned evaporated film [1.30]. This processing technique was combined with some exciting thoughts at the end of the 1950s and led to the application of transistor devices and other components in a way which was to transform the world, that is, the IC.

1.2.2 The Integrated Circuit

With the benefit of hindsight, the IC concept is quite simple. The problem faced by the electronics industry in the 1950s was the increasing difficulty of physically fitting into a small device all of the discrete electronic components (transistors, diodes, resistors, and capacitors), and then connecting them together. It was clear that this problem would eventually limit the complexity, reliability, and speed of circuits which could be created. Transistors and diodes were manufactured from semiconductors but resistors and capacitors were best formed from alternative materials. Even though they would not deliver the levels of performance achievable from the traditional materials, functioning capacitors and resistors *could* be manufactured from semiconductors, so, in principle, all of the components of a circuit could be prepared on a single block of the semiconducting material. This reasoning had been proposed by Englishman G.W.A. Dummer at a conference in 1952 [1.31], but small-scale attempts to realize circuits had failed, largely because they were based on connecting together layers in grown-junction transistors [1.32]. In 1958, however, Jack Kilby successfully built a simple oscillator and *flip–flop* logic circuits from components formed in situ on a germanium block and interconnected to produce circuits. He received the Nobel Prize in 2000 for *his part in the invention of the IC*.

Kilby's circuits were the first built on a single semiconductor block, but by far the majority of the circuit's size was taken up by the wires connecting together the components. Robert Noyce developed a truly IC in the form that it was later to be manufactured. While Kilby had used the mesa technique with external wiring, Noyce applied the planar technique to form transistors on silicon and photolithographically defined gold or aluminum interconnects. This was more suited to batch processing in production and was necessary for circuits with large numbers of components.

In the 1960s, a new transistor design – the metal oxide semiconductor field effect transistor (MOSFET) – was introduced that replaced the bipolar device used in the first chips [1.33]. In this device, a *gate* was deposited onto a thin insulating silicon oxide layer on the silicon. The application of a voltage to the gate resulted in an inversion layer in the silicon below the oxide, thereby, modifying the conducting channel between *source* and *drain* contacts. This structure was a p-MOS device (current transfer between the collector and emitter was by hole conduction) grown on (111) silicon using an aluminum gate. Earlier attempts at such a device had failed because of trapped impurities and charges in the gate oxide – this new structure had reduced the density of these to below tolerable levels but the device still could not compete with the bipolar transistors of its time [1.24]. By 1967, however, (100) silicon (which offered lower densities of states at the Si/SiO_2 interface) was used together with a polycrystalline silicon gate to construct a more effective and more easily processed device with advantages over the bipolar transistor. In the early 1970s, the n-MOS device,

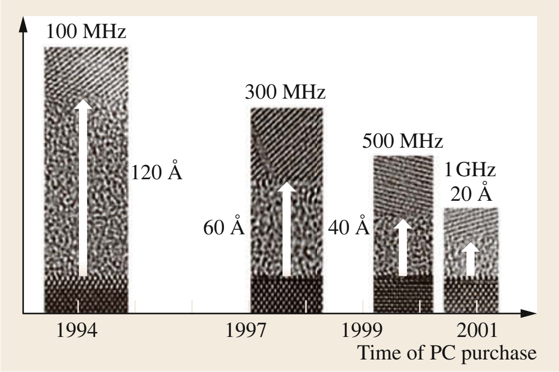

Fig. 1.1 The relentless march to zero thickness. HRTEM images of SiO_2 gate oxide thickness used in personal computers before high-*k* gate oxides were introduced. (After [1.34], with permission from Cambridge University Press)

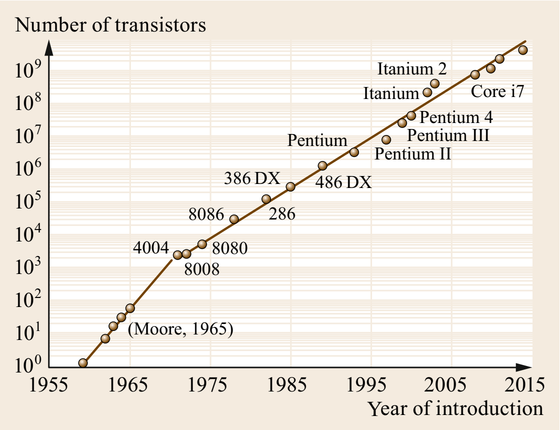

Fig. 1.2 The realization of Moore's law by commercial Intel processors: the logarithmic increase in the number of transistors in each processor chip

which was even less tolerant of the positive gate oxide charges, was realized thanks to much improved cleanliness in the production environment. With conduction occurring by the transfer of electrons rather than holes, these were capable of faster operation than similar p-MOS structures (the mobility of electrons in silicon is about three times that of holes). By the 1980s, these two devices, were combined in the complementary MOS (CMOS) device which afforded much lower power consumption and simplified circuit design [1.35].

The MOSFET is still a common structure in microelectronics today. Of course, now it can be much smaller, with transistor dimensions well below 100 nm, enabling more transistors within a single chip. The smaller dimensions are achieved through improvements in optical lithographic patterning methods. The minimum dimension of components which can be lithographically patterned on an IC is ultimately limited by the wavelength of radiation used in the process and this has continually been decreased over the past few decades. In the late 1980s, wavelengths of 365 nm were employed; by the late 1990s, 248 nm were common; and today, 193 nm is being used; sometimes coupled with *double-patterning* processes to extend to the smallest dimensions. Research into extreme ultraviolet lithography (e.g., at 13.5 nm wavelength) continues for potential use in future generation processes.

Providing good performance from transistors with smaller dimensions required significant advances in materials. For example: better control of doping; use of copper interconnects instead of aluminum; and *low-k* dielectric insulating layers between interconnects instead of silicon dioxide. Perhaps, the most disruptive modification was the replacement of the *native* silicon oxide as the gate insulator by *high-k* dielectric material such as hafnium oxide (Chap. 27). Smaller MOSFET dimensions had required ever thinner silicon gate oxide thicknesses – as shown clearly in Fig. 1.1 – eventually culminating in gate insulator layers just a few atoms thick; sufficiently thin that electron quantum mechanical tunneling through it became a significant power drain. Use of *high-k* material such as hafnium oxide, which has a significantly higher dielectric constant than silicon oxide ($k \approx 22$, compared with $k \approx 3.9$), enabled a return to thicker, tunneling-resistant, gate insulators for devices launched in the early 2000s. This improvement required huge commitment in research and development, including development and production scale-up of new methods to deposit sufficiently perfect *high-k* material (e.g., atomic layer deposition, ALD).

A most recent development in transistor technology has seen the abandonment of the traditional planar geometry and the first widespread use of three-dimensional (3-D) transistor geometries to further reduce leakage problems associated with the extreme proximity of source and drain. For example, the fin field effect transistor (FinFET), where the transistor channel is formed in a several nanometers wide ridge that protrudes above the planar substrate surface and is covered on its three exposed edges by gate material.

The development in complexity and performance of silicon devices, significantly due to materials science progress, is unparalleled in the history of technology. Never before could improvements be measured in terms of a logarithmic scale for such a sustained period. This is often seen as the embodiment of *Moore's law*. Noting a doubling of the number of components fitted onto ICs each year between 1959 and 1965, *Moore* predicted that this rate of progress would continue until at least

10 years later [1.36]. From the early 1970s, a modified prediction of doubling the number of components every couple of years has been sustained to the current day. Since the goals for innovation have often been defined assuming the continuation of the trend, it should perhaps be viewed more as a self-fulfiling prophecy. A huge variety of statistics relating to the silicon microelectronics industry follow a logarithmically scaled improvement from the late 1960s to the current day: the number of transistors shipped per year (increasing); average transistor price (decreasing); and number of transistors on a single chip (increasing) are examples [1.37]. The final member of this list is plotted in Fig. 1.2.

1.3 The Compound Semiconductors

It has been said that silicon is to electronics what steel is to mechanical engineering [1.38]. Steel is very effectively used for most of the world's construction but there are some tasks which it is incapable of performing and others for which alternative structural materials are better suited. In the same way, there are some crucial applications – such as optoelectronics and very high speed electronics – that silicon cannot usually deliver but which a wide range of compound semiconductors are better equipped to perform.

Silicon's band gap is indirect (an electron–hole recombination across the band gap must be accompanied by an interaction with a phonon in the lattice) that severely limits the potential efficiency of light emission from the material. Many of the important compound semiconductors, such as the alloys $\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{As}$, $\mathrm{In}_x\mathrm{Ga}_{1-x}\mathrm{N}$, $\mathrm{Al}_y\mathrm{Ga}_{1-y}\mathrm{N}$, and $\mathrm{Al}_x\mathrm{In}_y\mathrm{Ga}_{1-x-y}\mathrm{P}$, exhibit direct band gaps (no phonon interaction is required), so they can efficiently emit brilliant light in LEDs and laser diodes. Furthermore, in these alloy systems, where the band gap can be adjusted by changing the composition, there is a means of selecting the energy released when an electron and hole recombine across the gap and therefore controlling the wavelength of the photons emitted. From the $\mathrm{Al}_x\mathrm{In}_y\mathrm{Ga}_{1-x-y}\mathrm{As}$ and $\mathrm{Al}_y\mathrm{In}_x\mathrm{Ga}_{1-x-y}\mathrm{N}$ alloy systems, there is, in principle, a continuous range of direct band gaps from deep in the infrared (InAs; $\lambda = 3.5\,\mu\mathrm{m}$) to far into the ultraviolet (AlN; $\lambda = 200\,\mathrm{nm}$). The semiconductor band gaps of these materials and the corresponding photon wavelengths are put into context with the visible spectrum in Fig. 1.3.

Compounds are also very useful in high speed electronics applications. One of the determining factors in the speed of a transistor is the velocity of the charge carriers in the semiconductor. In GaAs, the electron drift velocity is much higher than in silicon, so its transistors are able to operate at significantly higher frequencies. The electron velocity in InAs is higher still. Furthermore, in the same way that silicon was preferred to germanium, devices manufactured using semiconductors such as GaN, which have much wider band gaps than silicon (3.4 eV compared with 1.1 eV), are capable of operating in much higher temperature environments.

Apart from these advantageous properties of compound semiconductors, the use of different alloy compositions, or totally different semiconductors, in a single device introduces entirely new possibilities. In silicon, most device action is achieved by little more than careful control of dopant impurity concentrations. In structures containing thin layers of semiconductors with different band gaps (heterostructures), there is the potential to control more fundamental parameters such as the band gap width, mobilities, and effective masses of the carriers [1.38]. In these structures, important new features become available which can be used by the device designer to tailor specific desired proper-

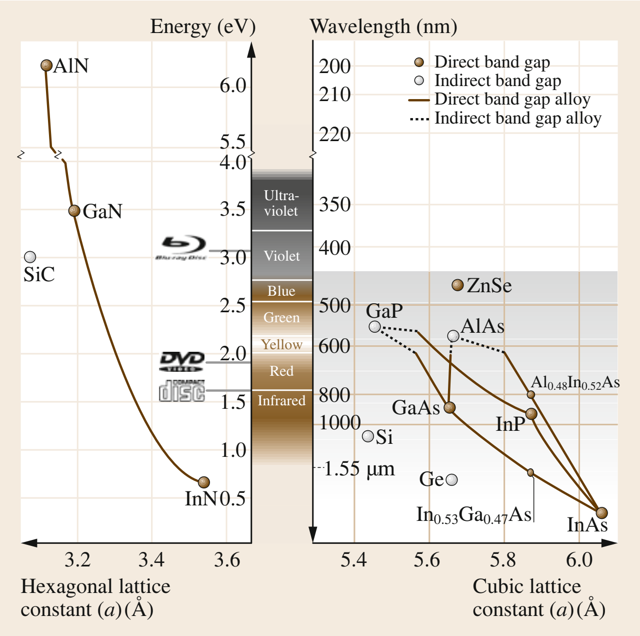

Fig. 1.3 Parameter perspective: band gaps and lattice parameters of selected semiconductors. The important wavelengths for optical storage (CD, DVD, and Blu-Ray) and the 1.55 µm used for efficient data transmission through optical fibers are labeled

ties. Hebert Kroemer and Zhores Alferov shared the Nobel prize in 2000 *for developing semiconductor heterostructures used in high-speed- and optoelectronics*.

We will mainly consider the compounds formed between elements in group III of the periodic table and those in group V (the III–V semiconductors); principally those based around GaAs and InP which were developed over much of the last 40 years, and GaN and its related alloys which have been most heavily studied only during the last two decades. Other families are given less attention here, though they also have important applications (e.g., the II–VI materials in optoelectronic applications). It can be hazardous to try and consider the *compound semiconductors* as a single subject. Though lessons can be learnt from the materials science of one of the compounds and transferred to another, each material is unique and must be considered on its own (i. e., of course, the purpose of the specialized chapters which follow in this handbook).

It is worth repeating that the power of the compound semiconductors lies in their use as the constituent layers in heterostructures. The principal contribution from chemistry and materials science to enable successful devices has been in the manufacture of high-quality bulk single crystal substrates and the development of techniques to reliably and accurately produce real layered structures on these substrates from the plans drawn up by a device theorist. In contrast to silicon, the compound semiconductors include volatile components, so encapsulation has been required for the synthesis of low-defect InP and GaAs substrates such as in the liquid encapsulated Czochralski technique [1.39, 40]. The size and crystalline quality of these substrates lag some way behind those available in silicon. Crucial to the commercialization of electronic and optoelectronic heterostructures were the improvements over the last few decades in the control of epitaxial growth available to the crystal grower. The first successful heterostructures were manufactured using deposition onto a substrate from the liquid phase (liquid phase epitaxy (LPE)) – *a beautifully simple technology but with severe limitations* [1.38]. However, the real heterostructure revolution had to wait for the 1970s and the introduction of molecular beam epitaxy (MBE) and metallorganic chemical vapor deposition (MOCVD) – also known as metallorganic vapor phase epitaxy (MOVPE) provided that the deposition is epitaxial.

MBE growth occurs in an ultrahigh vacuum with the atoms emitted from effusion cells forming *beams* which impinge upon and form compounds at the substrate surface. It derives from pioneering work at the start of the 1970s [1.41]. MOCVD relies on chemical reactions occurring on the substrate involving metallorganic vapor phase precursors and also stems from initial work at this time [1.42]. In contrast to LPE, these two techniques permit the combination of a wide range of different semiconductors in a single structure and offer a high degree of control over the local composition, in some cases on an atomic layer scale. The successful heterostructure devices of the late 1970s and 1980s would not have been achievable without these two tools and they still dominate III–V device production and research today.

1.3.1 High-Speed Electronics

The advantages of the III–V materials over silicon for use in transistors capable of operating at high frequencies were identified early in the semiconductor revolution [1.43]. Shockley's first patent for p–n junction transistors had included the proposal to use a wide-gap emitter layer to improve performance, and in the 1950s, *Kroemer* presented a theoretical design for a heterostructure transistor [1.44]. Some years later, the structure of a GaAs metal semiconductor FET (MESFET) was proposed and realized soon afterwards [1.45, 46]. In these devices, a Schottky barrier surface potential was used to modulate the conductivity of the GaAs channel. One of the earliest applications of the III–Vs was the low-noise amplifiers in microwave receivers that offered substantial improvements relative to the silicon bipolar transistors of the time. The devices were later used to demonstrate subnanosecond switching in monolithic digital ICs [1.47]. Today, they form the core of the highest speed digital circuits and are used in high speed electronics in microwave radar systems and wireless communications which incorporate monolithic ICs.

For at least 30 years, there have been repeated attempts to replicate the MOSFET, the dominant transistor form in silicon ICs, on GaAs material. These attempts have been frustrated by the difficulty of reproducibly forming a high quality stoichiometric oxide on GaAs. In direct analogy with the initial failure of constructing working n-MOSFETs on silicon, the GaAs devices have consistently been inoperable because of poor quality gate oxides with a high density of surface states at the GaAs–insulator interface [1.40]. One of the research efforts focused on realizing this device was, however, to be diverted and resulted in the discovery of probably the most important III–V electronic device, that is, the high electron mobility transistor (HEMT).

The background to this invention lies in the beautiful concept of modulation doping of semiconductors which was first demonstrated in 1978 [1.48]. One of the tenets of undergraduate semiconductor courses is the demonstration that as the dopant density in a semiconductor increases, the mobility of the carriers is reduced, because the carriers are scattered more by the ionized dopants. It was found that in a multilayer of repeating n-AlGaAs layers and undoped GaAs layers, the electrons supplied by donor atoms in the AlGaAs, moved into the adjacent potential wells of the lower-band gap GaAs layers. In the GaAs these suffered from substantially less ionized impurity scattering and therefore demonstrated enhanced mobility.

While working in a group attempting to create GaAs MOSFETs (and seemingly despairing at the task [1.49]), *Takashi Mimura* heard of these results and conceived of a field effect transistor where the conducting channel exploited the high mobility associated with a modulation doped structure. In essence, a doped AlGaAs layer was formed above the undoped GaAs channel of the transistor. Donated carriers gathered in the GaAs immediately below the interface where they did not suffer from as much ionized impurity scattering and so their mobility would approach that of an ultrapure bulk semiconductor. The current was conducted from the source to the drain by these high mobility carriers and so the devices were able to operate in higher frequency applications [1.49]. Realization of the structure required a very abrupt interface between the GaAs and AlGaAs and was considered beyond the capability of MOCVD of the time [1.49]. However, following the advances made in MBE procedures during the 1970s, the structure was achieved by that technique within a few months of the original conception [1.49, 50]. The first operational HEMT chips were produced on December 24, 1980: By pleasing coincidence this was the anniversary of Brattain and Bardeen's demonstration of their point-contact resistor to the management of Bell Labs in 1947. Structures based on the same principle as Mimura's device were realized in France very shortly afterwards [1.51].

The commercialization of the HEMT became significant in the late 1980s, thanks to broadcasting satellite receivers. The improved performance of the devices compared with the existing technology allowed the satellite parabolic dish size to be reduced by at least a factor of two. Structures similar to these have since played a crucial role in the massive expansion in mobile telephones.

The evolution in HEMT structures since the early 1980s is a fine example of how fundamental compound semiconductor properties have been exploited as the technology has become available to realize new device designs. The electron mobility in InAs is much higher than that in GaAs and rises as the indium content in $In_xGa_{1-x}As$ is increased [1.52]. The introduction of an InGaAs (as opposed to GaAs) channel to the HEMT structure, to create the so-called pseudomorphic HEMT (pHEMT), resulted both in increased electron mobility and a higher density of carriers gathering from the doped AlGaAs layer (because of the larger difference in energy between the conduction band minima of InGaAs and AlGaAs than that between GaAs and AlGaAs) [1.39]. The indium content and thickness of the channel are limited by the lattice mismatch with the GaAs (Fig. 1.3). If either is increased too much, then misfit dislocations are formed within the channel. The restriction is reduced by growing lattice matched structures on InP, rather than GaAs, substrates. $Al_{0.48}In_{0.52}As$ and $In_{0.53}Ga_{0.47}As$ are both lattice matched to the InP (Fig. 1.3) and their conduction band minimum energies are well separated, so that in the InGaAs below the interface between the two compounds a high density of electrons with a very high mobility is formed. Compared to the pHEMTs, these InP-based HEMTs exhibited significant improvements, were shown to exhibit gain at over 200 GHz, and became established as the leading transistor for millimeter-wave low noise applications such as radar [1.39].

1.3.2 Light Emitting Devices

LEDs and laser diodes exploit the direct band gap semiconductors to efficiently convert an electric current into photons of light. Work on light emission from semiconductor diodes was carried out in the early decades of the twentieth century [1.53], but the start of the modern era of semiconductor optoelectronics traces from the demonstration of lasing and LED behavior from p–n junctions in GaAs [1.54, 55] and $GaAs_{1-x}P_x$ [1.56]. The efficiency of these LEDs was low and the lasers had large threshold currents and only operated at low temperatures. A year later, in 1963, *Kroemer* and *Alferov* independently proposed the concept of the double heterostructure (DH) laser [1.57, 58]. In the DH device, a narrow band gap material was to be sandwiched between layers with a wider gap so that there would be some degree of confinement of carriers in the *active layer*. By the end of the decade, DH devices had been constructed that exhibited continuous lasing at room temperature [1.59, 60]. Alferov's laser was grown by LPE on a GaAs substrate with a 0.5 μm GaAs active layer confined between 3 μm of $Al_{0.25}Ga_{0.75}A$ on either side. The launch of the compact disk in 1982 saw this type of device, or at least its offspring, becoming taken for granted in the households of the world.

One of the major challenges in materials selection for heterostructure manufacture has always been avoiding the formation of misfit dislocations to relieve the strain associated with lattice parameter mismatch between the layers. $Al_xGa_{1-x}As$ exhibits a direct band gap for $x < 0.45$ and the early success and sustained dominance of the AlGaAs/GaAs system derive significantly from the very close coincidence of the AlAs and GaAs lattice parameters (5.661 Å and 5.653 Å; Fig. 1.3). This allows relatively thick layers of AlGaAs with reasonably high aluminum content to be grown lattice matched onto GaAs substrates with no misfit dislocation formation. The use of the quaternary alloy solid solution $In_xGa_{1-x}As_yP_{1-y}$ was also suggested in 1970 [1.59] to offer the independent control of lattice parameters and band gaps. Quaternaries based on three group III elements have since proved very powerful tools for lattice matching within heterostructures. $(Al_xGa_{1-x})_{0.5}In_{0.5}P$ was found to be almost perfectly lattice matched to GaAs and additionally have a very similar thermal expansion coefficient (which is important to avoid strain evolution when cooling after growth of heteroepitaxial layers at high temperatures). By varying x in this compound, direct band gaps corresponding to light between red and green could be created [1.61]. Lasers based on this alloy grown by MOCVD are a common choice for the red wavelengths (650 nm) used in DVD reading (Fig. 1.3).

Obtaining lattice matching is not so crucial for layers thinner than the critical thickness for dislocation production and can be less of an issue these days because of probably the most important development in the history of optoelectronic devices, that is, the introduction of the quantum well. In some way a quantum well structure is an evolution of the double heterostructure but with a very much thinner active layer. It is the chosen design for most solid-state light emitting devices today. With the accurate control available from MBE or MOCVD, and following from some early work on superlattices [1.62], very thin layers of carefully controlled composition could be deposited within heterostructure superlattice stacks. It became possible to grow GaAs layers much less than 10 nm thick within AlGaAs–GaAs heterostructures. The carriers in the GaAs were found to exhibit quantum mechanical confinement within the one-dimensional potential well [1.63, 64]. Lasing from $GaAs/Al_{0.2}Ga_{0.8}As$ quantum wells was reported the following year [1.65], but it was a few years before the emission matched that achievable from DH lasers of the time [1.66] and the quantum well laser was further advanced to significantly outperform the competition by researchers in the 1980s [1.67].

The introduction of heterostructures with layer thicknesses on the nanometer or atomic scale represents the final stage in scaling down of these devices. Similarly, Brattain and Bardeen's centimeter-sized transistor has evolved into today's microprocessors with far submicron FETs whose gate oxide thicknesses are measured in Angstroms. Throughout this evolution, materials characterization techniques (Chaps. 17–20) have contributed heavily to the progress in our understanding of electronic materials and these techniques deserve a brief detour here. As the dimensions have been reduced over the decades, the cross-sectional images of device structures published in the literature have progressed from optical microscopy [1.28], to scanning electron microscopy (SEM) images [1.66], to today's high- or atomic-resolution transmission electron microscopy (TEM) analysis of ultrathin layers. The challenge of obtaining relevant and representative data from the smaller volume of material, which is available in higher magnification analysis, has been met during the past decade or so by widespread adoption of highly site specific specimen preparation techniques based around focused ion beam (FIB) milling. For each new material family, understanding of defects and measurement of their densities (e.g., by TEM and x-ray topography) have contributed to improvements in quality. Huge improvements in x-ray optics have seen high-resolution x-ray diffraction techniques develop to become a cornerstone of heterostructure research and production quality control [1.68]. Scanning-probe techniques such as scanning tunneling and atomic force microscopy have become crucial to the understanding of MBE and MOCVD growth. Chemically sensitive techniques such as secondary ion mass spectroscopy and Rutherford backscattering have improved to provide information on doping concentrations and compositions in layered structures with excellent depth resolution. Chemically sensitive measurements have truly reached the atomic scale in the past decade through the exploitation of new generations of aberration-corrected and monochromated transmission electron microscopes and three-dimensional atom probe microscopy (3DAP). The disruptive improvement of spatial resolution enabled by the new generation TEMs continues to yield invaluable insight into atomic-scale material, interface, and device structures, especially when combined with chemically sensitive techniques such as electron energy loss spectroscopy (EELS). Meanwhile, the application of 3DAP to semiconducting and insulating materials and devices provides, for the first time, the *holy grail* of 3-D, atom-by-atom chemically resolved maps. By way of example, we show an atomic-scale chemical map of the critical layers in a III-nitride laser diode structure in Fig. 1.4, revealing crucial information on the distribution of alloy and dopant atoms. The materials characterization process remains a critical element of electronic and photonic materials research.

A primary commercial driver for semiconductor laser diode research has been to produce more effective emitters of infrared wavelengths for transmission of data along optic fibers. Devices based on InP sub-

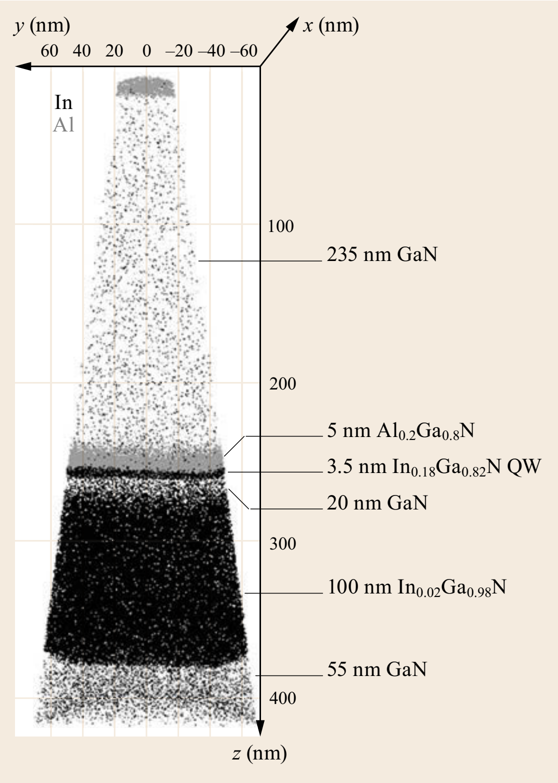

Fig. 1.4 Reconstructed atom map of the active region of a III-nitride laser diode structure determined by 3-D atom probe microscopy (3DAP). Indium atoms are displayed as *black* (25% displayed); aluminum atoms are displayed as *gray* (25% displayed); other atoms are not displayed. (After [1.69], with permission from AIP Publishing)

strates have proven to be extremely effective because of its fortuitous lattice parameter match with other III–V alloys which have band gaps corresponding to the low-absorption *windows* in optic fibers. While remaining lattice matched to InP, the $In_xGa_{1-x}As_yP_{1-y}$ quaternary can exhibit band gaps corresponding to infrared wavelengths of 1.3 and 1.55 μm at which conventional optic fibers absorb the least of the radiation (the absolute minimum is for 1.55 μm). Room temperature continuous lasing of 1.1 μm radiation was demonstrated from the material in 1976 [1.70], and InP-based lasers and photodiodes have played a key role in the optical communications industry since the 1980s [1.39].

A second major driver for laser diode research since the 1970s was to produce laser components for reading or writing optical storage media. We have already mentioned the AlGaAs infrared ($\lambda = 780\,\mathrm{nm}$) emitters used to read compact discs (CD) and the AlGaInP red ($\lambda = 650\,\mathrm{nm}$) lasers in DVD devices (Fig. 1.3). As lasers emitting the shorter wavelength became available, the optical disc's surface pits (through which bits of data are stored) could be made smaller and the storage density increased. The pursuit of shorter wavelength laser diodes for higher density optical storage supported extensive development of wide band gap II–VI compounds, principally ZnSe, for their potential in green ($\lambda \approx 520\,\mathrm{nm}$) and blue ($\lambda \approx 440\,\mathrm{nm}$) wavelengths. Laser operation in this part of the visible spectrum proved difficult to realize [1.71] and, although a blue/green laser was demonstrated in the early 1990s following improvements in the p-doping of ZnSe [1.72], the devices were prone to rapid deterioration during operation and were never commercialized. Ultimately, the winning laser diode technology for the third generation of optical storage laser was based on III-nitride materials following breakthroughs in the 1990s that are introduced in the next section. Blue/violet ($\lambda = 405\,\mathrm{nm}$) emitting $In_xGa_{1-x}N$-based laser diodes enabled Blu-ray (BD), the third and current generation of optical storage. CD, DVD, and BD technologies were introduced in 1982, 1995, and 2006, respectively, providing data storage per layer of approximately 0.7 GB, 4.7 GB, and 25 GB; a growth of data capacity five times per decade.

It is likely that BD also marks the final phase of the consumer optical disk era as network-based data distribution and storage replace the widespread use of optical discs. This is a good time, therefore, to note that for over 30 years the rapid advances in home entertainment (music, films, and console games) and computing (software distribution and data storage) were enabled in part by laser diode materials and device technology. The technology race for optical storage laser diodes leaves a legacy of extraordinary laser component technologies that inevitably evolve to bring value to new sectors, for example, the imminent growth of laser diode-based image projectors and automotive headlights. Of course, the growth in network data distribution which has caused the consumer optical disk era to peak is enabled by the laser diodes used in network data communication, a wonderful example of the ebb and flow of technology dominance.

1.3.3 The III-Nitrides

In the final section of this chapter, we focus on the III-nitride materials, such as GaN, $In_xGa_{1-x}N$, and $Al_yGa_{1-y}N$ and their devices, which have been subject to some of the most widespread research and development among all electronic and photonic materials since the mid-1990s. III-nitride LEDs and laser diodes have revolutionized the lighting, display, and optical storage industries during the past decade. Electronic devices based on these materials are expected to have a strong impact on electrical power distribution and electric vehicles in the next.

The relevance of the $In_xGa_{1-x}N$ alloy for visible light emitting devices is clear from Fig. 1.3. The InN and GaN direct band gaps correspond to wavelengths straddling the visible spectrum, and the alloy potentially offers access to all points in-between. The early commercially successful InGaN light emitters were marketed by Nichia Chemical Industries following the research work of *Shuji Nakamura* who demonstrated the first blue InGaN DH LEDs [1.73], blue InGaN quantum well LEDs, and laser diodes soon after [1.74]. Nakamura, together with other pioneers in III-nitride growth, Isamu Akasaki, and Hiroshi Amano received the 2014 Nobel Prize for Physics for their contributions to blue LED technology. Meanwhile, the wide band gap and other advantageous properties of GaN and $Al_yGa_{1-y}N$ alloys provide new opportunities for high-frequency and power transistors for high voltage switching and deep ultraviolet light emitters.

The development of III-nitride materials and devices has much in common with the early research of other III–V systems. For example, MOCVD and MBE growth technologies were both successfully used for III-nitride device growth (MOCVD has become the preferred technology for mass manufacturing of light emitting devices); both electronic and optoelectronic device applications were explored from an early stage; and one of the obstacles limiting early device development was achieving sufficiently high p-type carrier densities. However, in some ways III-nitrides presented new challenges: Although *conventional* III–Vs all share the same cubic crystallographic structure, the nitrides most readily form in a hexagonal allotrope that has strong polar and piezoelectric properties; the vast majority of III-nitride device development and manufacturing has used heteroepitaxial growth on highly lattice-mismatched substrates leading to devices with very high defect densities, in contrast to a reliance on homoepitaxy or closely lattice matched substrates in previously successful III–V systems where much lower defect densities are standard; and many III-nitride devices have been found to be astonishingly more tolerant to the presence of these crystalline defects than other III–Vs. We will focus briefly on two fascinating aspects of the story of the III-nitrides today: the variety of substrate materials in mainstream use, and the remarkable – and still not fully understood – properties of the $In_xGa_{1-x}N$ alloy.

First, the plurality of substrate materials in use for LEDs. Synthesis of large bulk GaN crystals – the obvious substrate choice from a technical perspective – is much more difficult than for other III–Vs such as GaAs or InP. Methods including ammonothermal growth and hydride VPE are technically viable routes to produce high quality GaN substrates today. However, in the early stages of III-nitride development, there were no large area GaN substrates available and heteroepitaxial growth of GaN on foreign substrates was the only option for commercial LED production. Even today, GaN substrate costs are very high owing to combinations of high temperature, high pressures, slow growth rates, and expensive precursor materials in the manufacturing process.

A consequence of the initial necessity of heteroepitaxy, and ongoing commercial barriers to the use of GaN substrates, is that a total of four substrate materials are in use for commercial production of InGaN-based blue LEDs: sapphire (α-Al_2O_3), silicon carbide, silicon, and GaN. Substrate choice is not purely based on technical factors – commercial considerations are strong drivers – but in the context of conventional compound semiconductors, it is extraordinary that four completely different substrates can be used to produce LEDs with approximately competitive device performance (say, competitive within a factor of two).

Current LED production is dominated by use of sapphire substrates, with silicon carbide substrates a distant second and silicon and GaN substrates just entering the market in recent years. Sapphire is by no means an ideal choice. First, it is electrically insulating which means that electrical contacts cannot be made to the device through the substrate material, thus, requiring LED designs with both n- and p-contacts on the same surface. Second, the refractive indices of sapphire and GaN are significantly different, which lead to low efficiency of extraction of light from the LED chip unless special measures are applied, such as the removal of the substrate or 3-D patterning of the sapphire surface prior to GaN growth. Third, and perhaps most importantly, sapphire has a large lattice mismatch of approximately 16% with GaN [1.75]. This large lattice mismatch is relieved by the formation of misfit dislocations at the GaN/Al_2O_3 interface which lead to threading dislocations propagating through the GaN into the active layers of the devices. The key discovery for reducing the defect densities to tolerable levels was the use of *buffer* layers at the interface with the sapphire [1.75]. Dislocation densities of approximately $10^9\,cm^{-2}$ were adequate for the first commercial LEDs and countless refinements to buffer layer growth have led to typical threading dislocation densities of approximately 10^7–$10^8\,cm^{-2}$ in devices today. In principle, silicon carbide has advantages over sapphire: it can be electrically conducting, the substrate can be readily etched to form chip shapes that improve light extraction efficiency, and it has a smaller lattice mismatch with GaN. However, despite these significant differences, devices grown on sapphire and silicon carbide have remained broadly competitive with one another. Cost and availability considerations led to sapphire becoming the mainstream substrate choice for LEDs during the early 2000s.

More recently, silicon and GaN have emerged as alternative substrate choices, for quite distinct reasons. Silicon substrates are pursued as a route to reduce LED manufacturing costs compared with incumbent production on sapphire. A key driver is the lower cost of silicon wafers, especially in large wafer formats (at least 150 mm diameter) which can be processed by otherwise unused and fully depreciated wafer processing lines previously established for silicon electronics processing. In addition to the lower substrate cost, GaN-on-silicon enables cost reductions to be made in processing, packaging, and phosphors. For example, GaN-on-silicon facilitates chip scale packaging (CSP), chip scale optics (CSO), and die level integration. CSP has been used for silicon electronics for many years and this technology can now be used for GaN-on-silicon. CSP also reduces the phosphor usage, gives improved thermal performance and is a route to wafer scale packaging (WSP). The main technical challenges to delivering viable device performance and yield are large lattice mismatch (again) and large difference in thermal expansion coefficient compared with GaN, which leads to cracking and bowing during cool down after growth unless careful strain engineering is used during growth [1.76]. In contrast, GaN substrates are substantially more expensive than sapphire, and only viable in smaller formats (typically 50–75 mm diameter). However, LEDs grown on GaN substrates benefit from much lower threading dislocation densities and consequently deliver some of the best device efficiencies, especially for small chips operated at high drive currents [1.77]. The next few years will reveal if silicon substrates or GaN substrates can substantially displace the incumbent sapphire and silicon carbide-based production.

Throughout the history of InGaN LED research, there has been ongoing fascination and controversy regarding the precise mechanism of light emission from InGaN quantum wells. The original driver behind this was the extraordinarily high light emission efficiency of InGaN quantum wells, despite the presence of dislocation densities as high as $10^9\,\mathrm{cm}^{-2}$. Obtaining similar light emission efficiencies in other III–V materials typically required dislocation densities to be several orders of magnitude lower.

For a few years, a popular explanation for the high efficiency of light emission from InGaN quantum wells was that strong fluctuations in the indium concentration in the quantum wells created low-energy quantum-dot like sites that localized electrons and holes and reduced the deleterious impact of dislocations on the light emission process. This explanation was based primarily on TEM analysis of InGaN quantum wells which indicated nanometer-scale regions with greatly higher indium fraction than the remainder of the quantum well – so called *indium clusters*. Many attempts were made to control and optimize this clustering process to increase light emission efficiency. However, since we demonstrated that the apparent indium concentration fluctuations could be a misleading artifact of TEM analysis [1.78, 79], and that high light emission efficiency could be obtained from InGaN quantum wells with random indium distribution [1.80], this theory is no longer widely supported. The origin of localization is now supposed to be due to much subtler features of the InGaN quantum wells.

Recently, the controversy concerning InGaN light emission has evolved to also explaining the phenomenon of *efficiency droop* in InGaN LEDs. *Droop* refers to the observed reduction in efficiency of state-of-the-art LEDs when they are operated at high current densities. There is great pressure to understand and reduce droop because of the potential cost benefit from using smaller LED chips with higher operating currents (i. e., high current densities). The subject of dozens of high profile publications and review articles over the last 10 years, many theories have been vigorously proposed for lower efficiencies at high current densities: increased Auger recombination of electrons and holes; overflow of electrons from the LED active region; the delocalization of carriers; and a complex role of large-scale crystalline defects.

Although these important technical debates are still largely unresolved, the global impact of InGaN LEDs cannot be disputed. This is a good example of how commercial success of electronic and photonic devices can easily occur before full technical understanding of the materials is established. Twelve years ago, in the introduction to the 1st Edition of this handbook, we wrote of the potential for III-nitride LEDs to replace traditional light sources such as incandescent light bulbs, but noted the need to reduce costs and increase the output power from individual LEDs to reach this goal. These commercial and technical barriers were emphatically overcome and the potential is now being realized. Production costs have been driven down by massive economies of scale during epitaxial growth, improvements in yield, cheaper device processing, and improved device performance which mean fewer, smaller LEDs are required for a given application. Tremendous improvements in LED performance – in particular, higher efficiency and high power light output – have been delivered through countless improvements in all aspects of the LED technology, including the core semiconductor material fabrication. As a consequence, InGaN blue- or violet-emitting LEDs, combined with phosphors that produce white light emission, are now cornerstone technologies in domestic, retail, industrial and outdoor lighting, vehicle headlights, and LCD backlighting. These solid-state technologies provide large energy savings compared with conventional light sources. Every day for the foreseeable future, we will directly sense – through our eyes – the light generated by these devices. Next generation LED lighting may be even more efficient, by omitting phosphors and generating white light by mixing red, yellow, green, and blue LEDs. Such lighting would be color tunable and could mimic sunlight indoors. There is increasing evidence that such optimized lighting would be good for our health, increase productivity at work, and even improve examination results in schools.

In this introductory chapter, we have glimpsed just one sector of the electronic and photonic materials landscape, focusing on just a few semiconductor families. We have given examples of the complex interdependence and balance of power between different materials and technologies: the astonishing integration of materials which we take for granted in a budget smartphone; the switch from germanium to silicon in the early years of the transistor; the dominance and relentless performance improvements in silicon ICs over several decades; the laser-diode enabled growth of consumer optical storage to its peak; and the co-existence of dramatically different approaches to make LEDs which are revolutionizing the way we light the world. There is a complex fusion of science and engineering excellence, ingenious breakthroughs, serendipity, and commercial considerations. But underpinning it all is the need to fundamentally understand and precisely control these fascinating materials. And this is all the more exciting because the pace of progress is so high, and the influence of the materials is so great, thanks to the electronic and photonic materials being the enabling technology in so many aspects of modern life. The detailed chapters in the rest of this book are a valuable resource to support further advances in this vital field.

References

1.1 A.H. Wilson: Proc. R. Soc. Lond. A **133**, 458 (1931)
1.2 A.H. Wilson: Proc. R. Soc. Lond. A **134**, 277 (1931)
1.3 J. Bardeen, W.H. Brattain: Phys. Rev. **74**, 230 (1948)
1.4 J. Bardeen: *Nobel Lecture, Physics, 1942–1962* (Elsevier, Amsterdam 1956), http://www.nobelprize.org/nobel_prizes/physics/laureates/1956/shockley-lecture.html
1.5 E. Braun, S. MacDonald: *Revolution in Miniature*, 2nd edn. (Cambridge Univ. Press, Cambridge 1982)
1.6 W.H. Brattain: *Nobel Lecture, Physics, 1942–1962* (Elsevier, Amsterdam 1956), http://www.nobelprize.org/nobel_prizes/physics/laureates/1956/brattain-lecture.html
1.7 C.S. Fuller: Phys. Rev. (Ser. 2) **86**, 136 (1952)
1.8 I. Derick, C.J. Frosh: US Patent, 2802760 (1955)
1.9 J. Andrus, W.L. Bond: US Patent, 3122817 (1957)
1.10 W.G. Pfann: Trans. Am. Inst. Mech. Eng. **194**, 747 (1952)
1.11 W.G. Pfann: *Zone Melting* (Wiley, New York 1958)
1.12 G.K. Teal: IEEE Trans. Electron Devices **23**, 621 (1976)
1.13 G.K. Teal, J.B. Little: Phys. Rev. (Ser. 2) **78**, 647 (1950)
1.14 G.K. Teal, E. Buehler: Phys. Rev. **87**, 190 (1952)
1.15 J. Czochralski: Z. Phys. Chem. **92**, 219 (1917)
1.16 P.H. Keck, M.J.E. Golay: Phys. Rev. **89**, 1297 (1953)
1.17 H.C. Theurer: US Patent 3 060 123 (1952)
1.18 H.C. Theurer: Trans. Am. Inst. Mech. Eng. **206**, 1316 (1956)
1.19 K.A. Jackson (Ed.): *Silicon Devices* (Wiley-VCH, Weinheim 1998)
1.20 W.C. Dash: J. Appl. Phys. **29**, 736 (1958)
1.21 W.C. Dash: J. Appl. Phys. **30**, 459 (1959)
1.22 W.C. Dash: J. Appl. Phys. **31**, 736 (1960)
1.23 G. Zeigler: Z. Naturforsch. **16a**, 219 (1961)
1.24 M. Grayson (Ed.): *Encyclopedia of Semiconductor Technology* (Wiley, New York 1984) p. 734
1.25 I.M. Ross: Bell Labs Tech. J. **2**(4), 3 (1997)
1.26 W. Shockley: Bell Syst. Tech. J. **28**(4), 435 (1949)
1.27 W. Shockley, M. Sparks, G.K. Teal: Phys. Rev. **83**, 151 (1951)
1.28 M. Tanenbaum, D.E. Thomas: Bell Syst. Tech. J. **35**, 1 (1956)
1.29 C.M. Melliar-Smith, D.E. Haggan, W.W. Troutman: Bell Labs Tech. J. **2**(4), 15 (1997)
1.30 J.A. Oerni: IRE Trans. Electron Devices **7**, 178 (1960)
1.31 J.S. Kilby: IEEE Trans. Electron Devices **23**, 648 (1976)
1.32 J.S. Kilby: *Nobel Lectures in Physics: 1996–2000* (Imperial College, London 2000), http://www.nobelprize.org/nobel_prizes/physics/laureates/2000/kilby-lecture.html
1.33 D. Kahng, M.M. Atalla: Silicon-silicon dioxide field induced surface devices, Proc. Solid State Res. Conf., Pittsburgh (1960)
1.34 F.H. Baumann, C.P. Chang, J.L. Grazul, A. Kamgar, C.T. Liu, D.A. Muller: C4.1, Mat. Res. Soc. Symp. **611**, 1 (2000)
1.35 J.T. Clemens: Bell Labs Tech. J. **2**(4), 76 (1997)
1.36 G.E. Moore: Electronics **38**(8) (1965)
1.37 G.E. Moore: IEEE Int. **1**, 20 (2003)
1.38 H. Kroemer: *Nobel Lectures in Physics: 1996–2000* (Imperial College, London 2000), http://www.nobelprize.org/nobel_prizes/physics/laureates/2000/kroemer-lecture.html
1.39 O. Wada, H. Hasegawa (Eds.): *InP-Based Materials and Devices* (Wiley, New York 1999)
1.40 C.Y. Chang, F. Kai: *GaAs High-Speed Devices* (Wiley, New York 1994)

1.41 A.Y. Cho: J. Vac. Sci. Technol. **8**, S31 (1971)
1.42 H.M. Manasevit: Appl. Phys. Lett. **12**, 156 (1968)
1.43 H.J. Welker: IEEE Trans. Electron Devices **23**, 664 (1976)
1.44 H. Kroemer: RCA Review **18**, 332 (1957)
1.45 C.A. Mead: Proc. IEEE **54**, 307 (1966)
1.46 W.W. Hooper, W.I. Lehrer: Proc. IEEE **55**, 1237 (1967)
1.47 R. van Tuyl, C. Liechti: IEEE Spectrum **14**(3), 41 (1977)
1.48 R. Dingle, H.L. Störmer, A.C. Gossard, W. Wiegmann: Appl. Phys. Lett. **33**, 665 (1978)
1.49 T. Mimura, S. Hiyamizu, T. Fujii, K. Nanbu: Jpn. J. Appl. Phys. **19**, L225 (1980)
1.50 T. Mimura: IEEE Trans. Microw. Theory Tech. **50**, 780 (2002)
1.51 D. Delagebeaudeuf, P. Delescluse, P. Etienne, M. Laviron, M. Chaplart, N.T. Linh: Electron. Lett. **16**, 667 (1980)
1.52 D. Chattopadhyay, S.K. Sutradhar, B.R. Nag: J. Phys. C **14**, 891 (1981)
1.53 E.E. Loebner: IEEE Trans. Electron Devices **23**, 675 (1976)
1.54 R.N. Hall, G.E. Fenner, J.D. Kingsley, T.J. Soltys, R.O. Carlson: Phys. Rev. Lett. **9**, 366 (1962)
1.55 M.I. Nathan, W.P. Dumke, G. Burns, F.H. Dill Jr., G. Lasher: Appl. Phys. Lett. **1**, 62 (1962)
1.56 N. Holonyak, M.B. Panish, P.W. Foy, S. Sumski: Appl. Phys. Lett. **1**, 82 (1962)
1.57 H. Kroemer: Proc. IEEE **51**, 1782 (1963)
1.58 Z.I. Alferov: *Nobel Lectures in Physics: 1996–2000* (Imperial College Press, London 2000), http://www.nobel.se/physics/laureates/2000/alferov-lecture.html
1.59 Z.I. Alferov, V.M. Andreev, D.Z. Garbuzov, Y.V. Zhilyaev, E.P. Morozov, E.L. Portnoi, V.G. Trofim: Fiz. Tekh. Poluprovodn. **4**, 1826 (1970)
1.60 I. Hayashi, G.E. Fenner, J.D. Kingsley, T.J. Soltys, R.O. Carlson: Appl. Phys. Lett. **17**, 109 (1970)
1.61 A. Žukauskas: *Introduction to Solid-State Lighting* (Wiley, New York 2002)
1.62 L. Esaki, R. Tsu: IBM J. Res. Dev. **14**, 61 (1970)
1.63 L.L. Chang, L. Esaki, R. Tsu: Appl. Phys. Lett. **24**, 593 (1974)
1.64 R. Dingle, W. Wiegmann, C.H. Henry: Phys. Rev. Lett. **33**, 827 (1974)
1.65 J.P. van der Ziel, R. Dingle, R.C. Miller, W. Wiegmann, W.A. Nordland: Appl. Phys. Lett. **26**, 463 (1975)
1.66 R. Dupuis, P.D. Dapkus, U. Holonyak Jr., E.A. Reack, R. Chin: Appl. Phys. Lett. **32**, 295 (1978)
1.67 Z.I. Alferov: Semiconductors **32**, 1 (1998)
1.68 D.K. Bowen, B.K. Tanner: *High Resolution X-ray Diffractometry and Topography* (Taylor Francis, London 1998)
1.69 S.E. Bennett, T.M. Smeeton, D.W. Saxey, G.D.W. Smith, S.E. Hooper, J. Heffernan, C.J. Humphreys, R.A. Oliver: J. Appl. Phys. **111**, 053508 (2012)
1.70 J.J. Hsieh, J.A. Rossi, J.P. Donneley: Appl. Phys. Lett. **28**, 709 (1976)
1.71 H.E. Ruda (Ed.): *Widegap II–VI Compounds for Opto-Electronic Applications* (Chapman Hall, London 1992)
1.72 M.A. Haase, J. Qiu, J.M. DePuydt, H. Cheng: Appl. Phys. Lett. **59**, 1272 (1991)
1.73 S. Nakamura, T. Mukai, M. Senoh: Appl. Phys. Lett. **64**, 1687 (1994)
1.74 S. Nakamura, M. Senoh, S. Nagahama, N. Iwasa, T. Yamada, T. Matsushita, H. Kiyoku, Y. Sugimoto: Jpn. J. Appl. Phys. **35**, L74 (1996)
1.75 S. Nakamura, G. Fasol: *The Blue Laser Diode* (Springer, Berlin 1997)
1.76 D. Zhu, D.J. Wallis, C.J. Humphreys: Rep. Prog. Phys. **76**, 106501 (2013)
1.77 C.A. Hurni: Appl. Phys. Lett. **106**, 031101 (2015)
1.78 T.M. Smeeton, M.J. Kappers, J.S. Barnard, M.E. Vickers, C.J. Humphreys: Appl. Phys. Lett. **83**, 5419 (2003)
1.79 T.M. Smeeton, C.J. Humphreys, J.S. Barnard, M.J. Kappers: J. Mater. Sci. **41**, 2729 (2006)
1.80 M.J. Galtrey: J. Appl. Phys. **104**, 013524 (2008)