so a much higher stress would be needed to sustain the formation of the neck. In a glass, very rapid drawing out is feasible; for instance, many years ago it was found that a blob of amorphous silica can be drawn into a very fine fibre (for instrument suspensions) by shooting the hot blob out like an arrow from a stretched bow. In alloys, the mechanism of deformation is quite different: it involves Nabarro–Herring creep, in which dislocations are not involved; under tension, strain results from the stress-biased diffusion of vacancies from grain boundaries transverse to the stress to other boundaries lying parallel to the stress. The operation of this important mechanism, which is the key to superplasticity, can be deduced from the mathematical form of the grain-size dependence of the process (Nabarro 1948, Herring 1950); it plays a major part in the deformation-mechanism maps outlined in the next chapter (Section 5.1.2.2). For large superplastic strains to be feasible, very fine grains (a few micrometres in diameter) and relatively slow strain rates (typically, 0.01/second) are requisite, so that the diffusion of vacancies can keep pace with the imposed strain rate. Sliding at grain boundaries is also involved. Practical superplastic alloys are always two-phase in nature, because a second phase is needed to impede the growth of grains when the sample is held at high temperature, and a high temperature is essential to accelerate vacancy diffusion.

The feasibility of superplastic forming for industrial purposes was first demonstrated, half a century after the first observation, by a team led by Backofen at MIT in 1964; until then, the phenomenon was treated as a scientific curiosity... a parepisteme, in fact. In 1970, the first patent was issued, with reference to superplastic nickel alloys, and in a book on ultra-fine-grained metals published in the same year, Headley *et al.* (1970) gave an account of 'the current status of applied superplasticity'. In 1976, the first major industrial advance was patented and then published in Britain (Grimes *et al.* 1976), following a study 7 years earlier on a simple Al–Cu eutectic alloy. The 1976 alloy (Al–6 wt% Cu–0.5 wt% Zr), trade name SUPRAL, could be superplastically formed at a reasonably fast strain rate and held its fine grains because of a fine dispersion of second-phase particles. It was found that such forming could be undertaken at modest stresses, using dies (to define the end-shape) made of inexpensive materials; it is therefore suitable for small production runs, without incurring the extravagant costs of tool-steel dies like those used in pressing automobile bodies of steel. A wide variety of superplastically formable aluminium alloys was developed during the following years. There was then a worldwide explosion of interest in superplasticity, fuelled by the first major review of the topic (Edington *et al.* 1976), which surveyed the various detailed mechanistic models that had recently been proposed. The first international conference on the topic was not called, however, until 1982.

In 1986, Wakai *et al.* (1986) in Japan discovered that ultra-fine-grained ceramics can also be superplastically deformed; they may be brittle with respect to dislocation

behaviour, but can readily deform by the Nabarro–Herring mechanism. This recognition was soon extended to intermetallic compounds, which are also apt to be brittle in respect of dislocation motion. Rapid developments followed after 1986 which are clearly set out in the most recent overview of superplasticity (Nieh *et al.* 1997). Very recently – after Nieh's book appeared – research in Russia by R. Valiev showed that it is possible to deform an alloy very heavily, in a novel way, so as to form a population of minute subgrains within larger grains and thereby to foster superplastic capability in the deformed alloy.

This outline case-history is an excellent example of a parepisteme which began as a metallurgical curiosity and developed, at a leisurely pace, into a well-understood phenomenon, from which it became, at a much accelerated pace, an important industrial process.

4.3. GENESIS AND INTEGRATION OF PAREPISTEMES

Parepistemes grow from an individual's curiosity, which in turn ignites curiosity in others; if a piece of research is directly aimed at solving a specific practical problem, then it is part of mainline research and not a parepisteme at all. However, the improvement of a technique used for solving practical problems constitutes a parepisteme.

Curiosity-driven research, a term I prefer to 'fundamental' or 'basic', involves following the trail wherever it may lead and, in Isaac Newton's words (when he was asked how he made his discoveries): "by always thinking unto them. I keep the subject constantly before me and wait until the first dawnings open little by little into full light". The central motive, curiosity, has been rendered cynically into verse by no less a master than A.E. Housman:

> Amelia mixed some mustard,
> She mixed it strong and thick:
> She put it in the custard
> And made her mother sick.
> And showing satisfaction
> By many a loud "huzza!",
> "Observe" she said "the action
> Of mustard on mamma".

A further motive is the passion for clarity, which was nicely illustrated many years ago during a conversation between Dirac and Oppenheimer (Pais 1995). Dirac was astonished by Oppenheimer's passion for Dante, and for poetry generally, side by side with his obsession with theoretical physics. "Why poetry?" Dirac wanted to

know. Oppenheimer replied: "In physics we strive to explain in *simple terms* what no one understood before. With poetry, it is just the opposite". Perhaps, to modify this bon mot for materials science, we could say: "In materials science, we strive to achieve by reproducible means what no one could do before...".

"Simple terms" can be a trap and a delusion. In the study of materials, we must be prepared to face complexity and we must distrust elaborate theoretical systems advanced too early, as Bridgman did. As White (1970) remarked with regard to Descartes: "Regarding the celebrated 'vorticist physics' which took the 1600s by storm... it had all the qualities of a perfect work of art. Everything was accounted for. It left no loose ends. It answered all the questions. Its only defect was that it was not true".

The approach to research which leads to new and productive parepistemes, curiosity-driven research, is having a rather difficult time at present. Max Perutz, the crystallographer who determined the structure of haemoglobin and for years led the Laboratory for Molecular Biology in Cambridge, on numerous occasions in recent years bewailed the passion for *directing* research, even in academic environments, and pointed to the many astonishing advances in his old laboratory resulting from free curiosity-driven research. That is often regarded as a largely lost battle; but when one contemplates the numerous, extensive and apparently self-directing parepistemic 'communities', for instance, in the domains of diffusion and high pressures, one is led to think that perhaps things are not as desperate as they sometimes seem.

My last point in this chapter is the value of integrating a range of parepistemes in the pursuit of a practical objective: in materials science terms, such integration of curiosity-driven pursuits for practical reasons pays a debt that parepistemes owe to mainline science. A good example is the research being done by Gregory Olson at Northwestern University (e.g., Olson 1993) on what he calls 'system design of materials'. One task he and his students performed was to design a new, ultrastrong martensitic bearing steel for use in space applications. He begins by formulating the objectives and restrictions as precisely as he can, then decides on the broad category of alloy to be designed, then homes in on a desirable microstructure type, going on to exploit a raft of *distinct* parepistemes relating to: (1) the strengthening effect of dispersions as a function of scale and density, (2) stability against coarsening, (3) grain-refining additives, (4) solid-solution hardening, (5) grain-boundary chemistry, including segregation principles. He then goes on to invoke other parepistemes relating microstructures to processing strategies, and to use CALPHAD (phase-diagram calculation from thermochemical inputs). After all this has been put through successive cycles of theoretical optimisation, a range of prospective compositions emerges. At this point, theory stops and the empirical stage, never to be bypassed entirely, begins. What the pursuit and integration of parepistemes

makes possible is to narrow drastically the range of options that need to be tested experimentally.

REFERENCES

Allen, S.M. and Thomas, E.L. (1999) *The Structure of Materials*, Chapter 2 (Wiley, New York).

Anderson, R.P. (1918) *Trans. Faraday Soc.* **14**, 150.

Arzt, E., Ashby, M.F. and Easterling, K.E. (1983) *Metall. Trans.* **14A**, 211.

Babu, S.S. and Bhadeshia, H.K.D.H. (1995) *J. Mater. Sci. Lett.* **14**, 314.

Barr, L.W. (1997) *Defect Diffusion Forum* **143–147**, 3.

Bengough, G.D. (1912) *J. Inst. Metals* **7**, 123.

Bowen, J.S.M. (1981) Letter to RWC dated 24 August.

Boyle, R. (1684) *Experiments and Considerations about the Porosity of Bodies in Two Essays.*

Bragg, W.H. and Bragg, W.L. (1939) *The Crystalline State: A General Survey* (Bell and Sons, London).

Bridgman, P.W. (1925) *Proc. Amer. Acad. Arts Sci.* **60**, 305; also (1928) *ibid* **63**, 351; Most easily accessible, in *Collected Experimental Papers*, 1964, ed. Bridgman, P.W. (Harvard University Press, Cambridge, MA).

Bridgman, P.W. (1931, 1949) *The Physics of High Pressure*, 1st and 2nd editions (Bell and Sons, London).

Bridgman, P.W. (1952) *Studies in Large Plastic Flow and Fracture, with Special Emphasis on the Effects of Hydrostatic Pressure* (McGraw-Hill, New York).

Carpenter, H.C.H. and Elam, C.F. (1921) *Proc. Roy. Soc. Lond.* **A100**, 329.

Czochralski, J. (1917) *Z. Phys. Chem.* **92**, 219.

Desorbo, W., Treaftis, H.N. and Turnbull, D. (1958) *Acta Metall.* **6**, 401.

DeVries, R.C., Badzian, A. and Roy, R. (1996) *MRS Bull.* **21**(2), 65.

Duhl, D.N. (1989) Single crystal superalloys, in *Superalloys, Supercomposites and Superceramics*, ed. Tien, J.K. and Caulfield, T. (Academic press, Boston) p. 149.

Dushman, S. and Langmuir I. (1922) *Phys. Rev.* **20**, 113.

Edington, J.W., Melton K.W. and Cutler, C.P. (1976) *Progr. Mater. Sci.* **21**, 61.

Edwards, C.A. and Pfeil, L.B. (1924) *J. Iron Steel Inst.* **109**, 129.

Elam, C.F. (1935) *Distortion of Metal Crystals* (Clarendon Press, Oxford).

Engel, U. and Hübner, H. (1978) *J. Mater. Sci.* **13**, 2003.

Fick, A. (1855) *Poggendorf Ann.* **94**, 59; *Phil. Mag.* **10**, 30.

Frank, F.C. and Turnbull, D. (1956) *Phys. Rev.* **104**, 617.

Frenkel, Y. (1924) *Z. f. Physik* **26**, 117.

Frenkel, Y. (1926) *Z. f. Physik* **35**, 652.

Frey, D.N. and Goldman, J.E. (1967) in *Applied Science and Technological Progress* (US Government Printing Office, Washington, DC) p. 273.

Goss, A.J. (1963) in *The Art and Science of Growing Crystals*, ed. Gilman, J.J. (Wiley, New York) p. 314.

Gough, H.J., Hanson, D. and Wright, S.J. (1928) *Phil. Trans. Roy. Soc. Lond.* **A226**, 1.
Graham, T. (1833) *Phil. Mag.* **2**, 175.
Greenaway, F. (1996) *Science International: A History of the International Council of Scientific Unions* (Cambridge University Press, Cambridge).
Grimes, R., Stowell, M.J. and Watts, B.M. (1976) *Metals Technol.* **3**, 154.
Hanemann, H. and Schrader, A. (1927) *Atlas Metallographicus* (Bornträger, Berlin) Table 101.
Hanson, D. (1924) *J. Inst. Metals* **109**, 149.
Hazen, R. (1993) *The New Alchemists: Breaking Through the Barriers of High Pressure* (Times Books, Random House, New York).
Hazen, R. (1999) *The Diamond Makers* (Cambridge University Press, Cambridge).
Headley, T.J., Kalish, D. and Underwood, E.E. (1970) The current status of applied superplasticity, in *Ultrafine-Grain Metals*, ed. Burke, J.J. and Weiss, V. (Syracuse University Press, Syracuse, NY) p. 325.
Herring, C. (1950) *J. Appl. Phys.* **21**, 437.
Honda, K. (1924) *J. Inst. Metals* **109**, 156.
IUCr (2000) The high-pressure program from Glasgow, *IUCr Newsletter* **8**(2), 11.
Kayser, F.X. and Patterson, J.W. (1998) *J. Phase Equili.* **19**, 11.
Keith, S. (1998) unpublished paper, private communication.
Kemble, E.C. and Birch, F. (1970) *Biographical Memoirs of Members of the National Academy of Sciences*, Vol. 41, Washington (Columbia University Press, New York and London) (Memoir of P.W. Bridgman) p. 23.
Kocks, U.F., Tomé, C.N. and Wenk, H.-R (1998) *Texture and Anisotropy: Preferred Orientations in Polycrystals and their Effects on Materials Properties* (Cambridge University Press, Cambridge).
Koiwa, M. (1998) *Mater. Trans. Jpn. Inst. Metals* **39**, 1169; *Metals Mater.* **4**, 1207.
Lal, K. (1998) *Curr. Sci. (India)* **64A**, 609.
Mark, H., Polanyi, M. and Schmid, E. (1922) *Z. Physik* **12**, 58.
Mendoza, E. (1990) *J. Chem. Educat.* **67**, 1040.
Nabarro, F.R.N (1948) *Report, Conference on the Strength of Solids* (Physical Society, London) p. 75.
Nakajima, H. (1997) *JOM* **49** (June), 15.
Nieh, T.G., Wadsworth, J. and Sherby, O.D. (1997) *Superplasticity in Metals and Ceramics* (Cambridge University Press, Cambridge).
Nowick, A.S. (1984) in *Proceedings of the International Conference on Defects in Insulating Crystals*, ed. Lüty, F. (Plenum Press, New York).
Nye, J.F. (1957) *Physical Properties of Crystals: Their Representation by Tensors and Matrices* (Oxford University Press, Oxford).
Olson, G.B. (1993) in *Third Supplementary Volume of the Encyclopedia of Materials Science and Engineering*, ed. Cahn, R.W. (Pergamon Press, Oxford) p. 2041.
Pais, A. (1995) From a memorial address for P.A.M. Dirac at the Royal Society, London.
Pearson, C.E. (1934) *J. Inst. Metals.* **54**, 111.
Pennisi, E. (1998) *Science* **282**, 1972.
Polanyi, M. (1962) My time with X-rays and crystals, in *Fifty Years of X-ray Diffraction*, ed. Ewald, P.P. (The International Union of Crystallography, Utrecht) p. 629.

Roberts-Austen, W. (1896a) *Phil. Trans. R. Soc. Lond.* **187**, 383.
Roberts-Austen, W. (1896b) *J. Iron Steel Inst.* **1**, 1.
Rogers, R.D. and Zaworotko, M.J. (1999) *Proc. Symp. Crystal Engineering, ACA Trans.* **33**, 1.
Ruoff, A.L. (1991) in *Phase Transformations in Materials*, ed. Haasen, P.; *Materials Science and Technology*, Vol. 5, ed. Cahn, R.W., Haasen, P. and Kramer, E.J. (VCH, Weinheim) p. 473.
Savage, H. (1988) in *First Supplementary Volume of the Encyclopedia of Materials Science and Engineering*, ed. Cahn, R.W. (Pergamon Press, Oxford) p. 553.
Schadler, H.W. (1963) in *The Art and Science of Growing Crystals*, ed. J.J. Gilman (Wiley, New York) p. 343.
Schenk, H. (ed.) (1998) *Crystallography Across the Sciences: A Celebration of 50 Years of Acta Crystallographica and the IUCr* (Munksgaard, Copenhagen). Originally published in *Acta Cryst. A* **54**(6), 1.
Schmalzried, H. (1995) *Chemical Kinetics of Solids* (VCH, Weinheim).
Schmid, E. and Boas, W. (1935) *Kristallplastizität* (Springer, Berlin).
Schuster, I., Swirsky, Y., Schmidt, E.J., Polturak, E. and Lipson, S.G. (1996) *Europhys. Lett.* **33**, 623.
Seeger, A. (1997) *Defect Diffusion Forum* **143–147**, 21.
Smigelskas, A.D. and Kirkendall, E.O. (1947) *Trans. AIME* **171**, 130.
Suits, C.G. and Bueche, A.M. (1967) in *Applied Science and Technological Progress* (US Government Printing Office, Washington, DC) p. 299.
Tuijn, C. (1997) *Defect Diffusion Forum* **143–147**, 11.
Tyrrell, H.J.V. (1964) *J. Chem. Educ.* **41**, 397.
Wakai, F. Sakaguchi, S. and Matsuno, Y. (1986) *Adv. Ceram. Mater.* **1**, 259.
Walter, M.L. (1990) *Science and Cultural Crisis: An Intellectual Biography of Percy William Bridgman* (Stanford University Press, Stanford, CA).
Weber, E.R. (1988) *Properties of Silicon* (INSPEC, London) p. 236.
Wentzcovich, R.M., Hemley, R.J., Nellis, W.J. and Yu, P.Y. (1998) *High-Pressure Materials Research* (Materials Research Society, Warrendale, PA) Symp. Proc. vol. **499**.
White, R.J. (1970) *The Antiphilosophers* (Macmillan, London).
Young, D.A. (1991) *Phase Diagrams of the Elements* (University of California Press, Berkeley).
Young, Jr., F.W., Cathcart, J.V. and Gwathmey, A.T. (1956) *Acta Metall.* **4**, 145.

Chapter 5

The Escape from Handwaving

5.1. The Birth of Quantitative Theory in Physical Metallurgy 189
5.1.1 Dislocation Theory 191
5.1.2 Other quantitative triumphs 196
5.1.2.1 Pasteur's Principle 198
5.1.2.2 Deformation-Mechanism and Materials Selection Maps 200
5.1.2.3 Stereology 203
5.1.3 Radiation Damage 205
References 209

Chapter 5
The Escape from Handwaving

5.1. THE BIRTH OF QUANTITATIVE THEORY IN PHYSICAL METALLURGY

In astrophysics, reality cannot be changed by anything the observer can do. The classical principle of 'changing one thing at a time' in a scientific experiment, to see what happens to the outcome, has no application to the stars! Therefore, the acceptability of a hypothesis intended to interpret some facet of what is 'out there' depends entirely on rigorous *quantitative self-consistency* – a rule that metallurgists were inclined to ignore in the early decades of physical metallurgy.

The matter was memorably expressed recently in a book, *GENIUS – The Life of Richard Feynman*, by James Gleick: "So many of his witnesses observed the utter freedom of his flights of thought, yet when Feynman talked about his own methods he emphasised not freedom but constraint... For Feynman the essence of scientific imagination was a powerful and almost painful rule. What scientists create must match reality. It must match what is already known. Scientific imagination, he said, is imagination in a straitjacket... The rules of harmonic progression made (for Mozart) a cage as unyielding as the sonnet did for Shakespeare. As unyielding and as liberating – for later critics found the creators' genius in the counterpoint of structure and freedom, rigour and inventiveness."

This also expresses accurately what was new in the breakthroughs of the early 1950s in metallurgy.

Rosenhain (Section 3.2.1), the originator of the concept of physical metallurgy, was much concerned with the fundamental physics of metals. In his day, ≈1914, that meant issues such as these: What is the structure of the boundaries between the distinct crystal grains in polycrystalline metals (most commercial metals are in fact polycrystalline)? Why does metal harden as it is progressively deformed plastically... i.e., why does it work-harden? Rosenhain formulated a generic model, which became known as the amorphous metal hypothesis, according to which grains are held together by "amorphous cement" at the grain boundaries, and work-hardening is due to the deposition of layers of amorphous material within the slip bands which he had been the first to observe. These erroneous ideas he defended with great skill and greater eloquence over many years, against many forceful counterattacks. Metallurgists at last had begun to argue about basics in the way that physicists had long done. Concerning this period and the amorphous grain-boundary cement theory in particular, Rosenhain's biographer has this to say (Kelly 1976): "The theory was wrong in scientific detail but it was of great utility. It enabled the metallurgist to

reason and recognise that at high temperatures grain boundaries are fragile, that heat-treatment involving hot or cold work coupled with annealing can lead to benefits in some instances and to catastrophes such as 'hot shortness' in others (this term means brittleness at high temperatures)... Advances in technology and practice do not always require exact theory. This must always be striven for, it is true, but a 'hand-waving' argument which calls salient facts to attention, if readily grasped in apparently simple terms, can be of great practical utility." This controversial claim goes to the heart of the relation between metallurgy as it was, and as it was fated to become under the influence of physical ideas and, more important, of the physicist's approach. We turn to this issue next.

As we have seen, Rosenhain fought hard to defend his preferred model of the structure of grain boundaries, based on the notion that layers of amorphous, or glassy, material occupied these discontinuities. The trouble with the battles he fought was twofold: there was no theoretical treatment to predict what properties such a layer would have, for an assumed thickness and composition, and there were insufficient experimental data on the properties of grain boundaries, such as specific energies. This lack, in turn, was to some degree due to the absence of appropriate experimental techniques of characterisation, but not to this alone: no one measured the energy of a grain boundary as a function of the angle of misorientation between the adjacent crystal lattices, not because it was difficult to do, even then, but because metallurgists could not see the point of doing it. Studying a grain boundary *in its own right* – a parepisteme if ever there was one – was deemed a waste of time; only grain boundaries as they directly affected useful properties such as ductility deserved attention. In other words, the cultivation of parepistemes was not yet thought justifiable by most metallurgists.

Rosenhain's righthand collaborator was an English metallurgist, Daniel Hanson, and Rosenhain infected him with his passion for understanding the plastic deformation of metals (and metallurgy generally) in atomistic terms. In 1926, Hanson became professor of metallurgy at the University of Birmingham. He struggled through the Depression years when his university department nearly died, but after the War, when circumstances improved somewhat, he resolved to realise his ambition. In the words of Braun (1992): "When the War was over and people could begin to think about free research again, Hanson set up two research groups, funded with money from the Department of Scientific and Industrial Research. One, headed by Geoffrey Raynor from Oxford (he had worked with Hume-Rothery, Section 3.3.1.1) was to look into the constitution of alloys; the other, headed by Hanson's former student Alan Cottrell, was to look into strength and plasticity. Cottrell had been introduced to dislocations as an undergraduate in metallurgy, when Taylor's 1934 paper was required reading for all of Hanson's final-year students." Cottrell's odyssey towards a proper understanding of dislocations during his years at

Birmingham is set out in a historical memoir (Cottrell 1980). Daniel Hanson, to whose memory this book is dedicated, by his resolve and organisational skill reformed the understanding and teaching of physical metallurgy, introducing interpretations of properties in atomistic terms and giving proper emphasis to theory, in a way that cleared the path to the emergence of materials science a few years after his untimely death.

5.1.1 Dislocation theory

In Section 3.2.3.2, the reader was introduced to dislocations (and to that 1934 paper by Geoffrey Taylor) and an account was also presented of how the sceptical response to these entities was gradually overcome by visual proofs of various kinds. However, by the time, in the late 1950s, that metallurgists and physicists alike had been won over by the principle 'seeing is believing', another sea-change had already taken place.

After World War II, dislocations had been taken up by some adventurous metallurgists, who held them responsible, in a purely handwaving (qualitative) manner and even though there was as yet no evidence for their very existence, for a variety of phenomena such as brittle fracture. They were claimed by some to explain everything imaginable, and therefore 'respectable' scientists reckoned that they explained nothing.

What was needed was to escape from handwaving. That milestone was passed in 1947 when Cottrell formulated a rigorously quantitative theory of the discontinuous yield-stress in mild steel. When a specimen of such a steel is stretched, it behaves elastically until, at a particular stress, it *suddenly* gives way and then continues to deform at a lower stress. If the test is interrupted, then after many minutes holding at ambient temperature the former yield stress is restored... i.e., the steel strengthens or *strain-ages*. This phenomenon was of practical importance; it was much debated but not understood at all. Cottrell, influenced by the dislocation theorists Egon Orowan and Frank Nabarro (as set out by Braun 1992) came up with a novel model. The essence of Cottrell's idea was given in the abstract of his paper to a conference on dislocations held in Bristol in 1947, as cited by Braun:

> "It is shown that solute atoms differing in size from those of the solvent (carbon, in fact) can relieve hydrostatic stresses in a crystal and will thus migrate to the regions where they can relieve the most stress. As a result they will cluster round dislocations forming 'atmospheres' similar to the ionic atmospheres of the Debye–Hückel theory of electrolytes. The conditions of formation and properties of these atmospheres are examined and the theory is applied to problems of precipitation, creep and the yield point."

The importance of this advance is hidden in the simple words "It is shown...", and furthermore in the parallel drawn with the D–H theory of electrolytes. This was

one of the first occasions when a quantitative lesson for a metallurgical problem was derived from a neighbouring but quite distinct science.

Cottrell (later joined by Bruce Bilby in formulating the definitive version of his theory), by precise application of elasticity theory to the problem, was able to work out the concentration gradient across the carbon atmospheres, what determines whether the atmosphere 'condenses' at the dislocation line and thus ensures a well-defined yield-stress, the integrated force holding a dislocation to an atmosphere (which determines the drop in stress after yield has taken place) and, most impressively, he was able to predict the time law governing the reassembly of the atmosphere after the dislocation had been torn away from it by exceeding the yield stress – that is, the strain-ageing kinetics. Thus it was possible to compare accurate measurement with precise theory. The decider was the strain-ageing kinetics, because the theory came up with the prediction that the fraction of carbon atoms which have rejoined the atmosphere is strictly proportional to $t^{2/3}$, where t is the time of strain-ageing after a steel specimen has been taken past its yield-stress.

In 1951, this strain-ageing law was checked by Harper (1951) by a method which perfectly encapsulates the changes which were transforming physical metallurgy around the middle of the century. It was necessary to measure the change with time of *free* carbon dissolved in the iron, and to do this in spite of the fact that the solubility of carbon in iron at ambient temperature is only a minute fraction of one per cent. Harper performed this apparently impossible task and obtained the plots shown in Figure 5.1, by using a torsional pendulum, invented just as the War began by a Dutch physicist, Snoek (1940, 1941), though his work did not become known outside the Netherlands until after the War. Harper's/Snoek's apparatus is shown in Figure 5.2(a). The specimen is in the form of a wire held under slight tension in the elastic regime, and the inertia arm is sent into free torsional oscillation. The amplitude of oscillation gradually decays because of internal friction, or damping: this damping had been shown to be caused by dissolved carbon (and nitrogen, when that was present also). Roughly speaking, the dissolved carbon atoms, being small, sit in interstitial lattice sites close to an edge of the cubic unit cell of iron, and when that edge is elastically compressed and one perpendicular to it is stretched by an applied stress, then the equilibrium concentrations of carbon in sites along the two cube edges become slightly different: the carbon atoms "prefer" to sit in sites where the space available is slightly enhanced. After half a cycle of oscillation, the compressed edge becomes stretched and vice versa. When the frequency of oscillation matches the most probable jump frequency of carbon atoms between adjacent sites, then the damping is a maximum. By finding how the temperature of peak damping varies with the (adjustable) pendulum frequency (Figure 5.2(b)), the jump frequency and hence the diffusion coefficient can be determined, even below

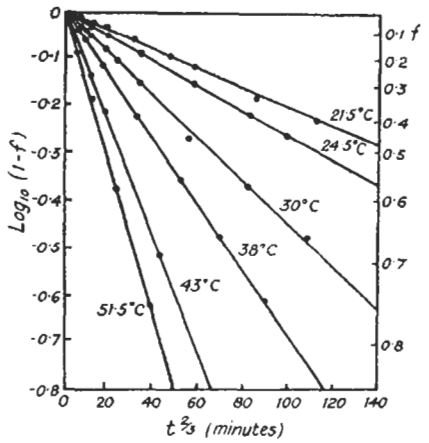


Figure 5.1. Fraction, f, of carbon atoms restored to the 'atmosphere' surrounding a dislocation, as determined by means of a Snoek pendulum.

room temperature where it is very small (Figure 5.2(c)). The subtleties of this "anelastic" technique, and other related ones, were first recognised by Clarence Zener and explained in a precocious text (Zener 1948); the theory was fully set out later in a classic text by two other Americans, Nowick and Berry (1972). The magnitude of the peak damping is proportional to the amount of carbon in solution. A carbon atom situated in an 'atmosphere' around a dislocation is locked to the stress-field of the dislocation and thus cannot oscillate between sites; it therefore does not contribute to the peak damping.

By the simple expedient of stretching a steel wire beyond its yield-stress, clamping it into the Snoek pendulum and measuring the decay of the damping coefficient with the passage of time at temperatures near ambient, Harper obtained the experimental plots of Figure 5.1: here f is the fraction of dissolved carbon which had migrated to the dislocation atmospheres. The $t^{2/3}$ law is perfectly confirmed, and by comparing the slopes of the lines for various temperatures, it was possible to show that the activation energy for strain-ageing was identical with that for diffusion of carbon in iron, as determined from Figure 5.2(a). After this, Cottrell and Bilby's model for the yield-stress and for strain-ageing was universally accepted and so was the existence of dislocations, even though nobody had seen one as yet at that time. Cottrell's book on dislocation theory (1953) marked the coming of age of the subject; it was the first rigorous, quantitative treatment of how the postulated dislocations must react to stress and obstacles. It is still cited regularly. Cottrell's research was aided by the theoretical work of Frank Nabarro in Bristol, who worked out the response of stressed dislocations to obstacles in a crystal: he has devoted his whole

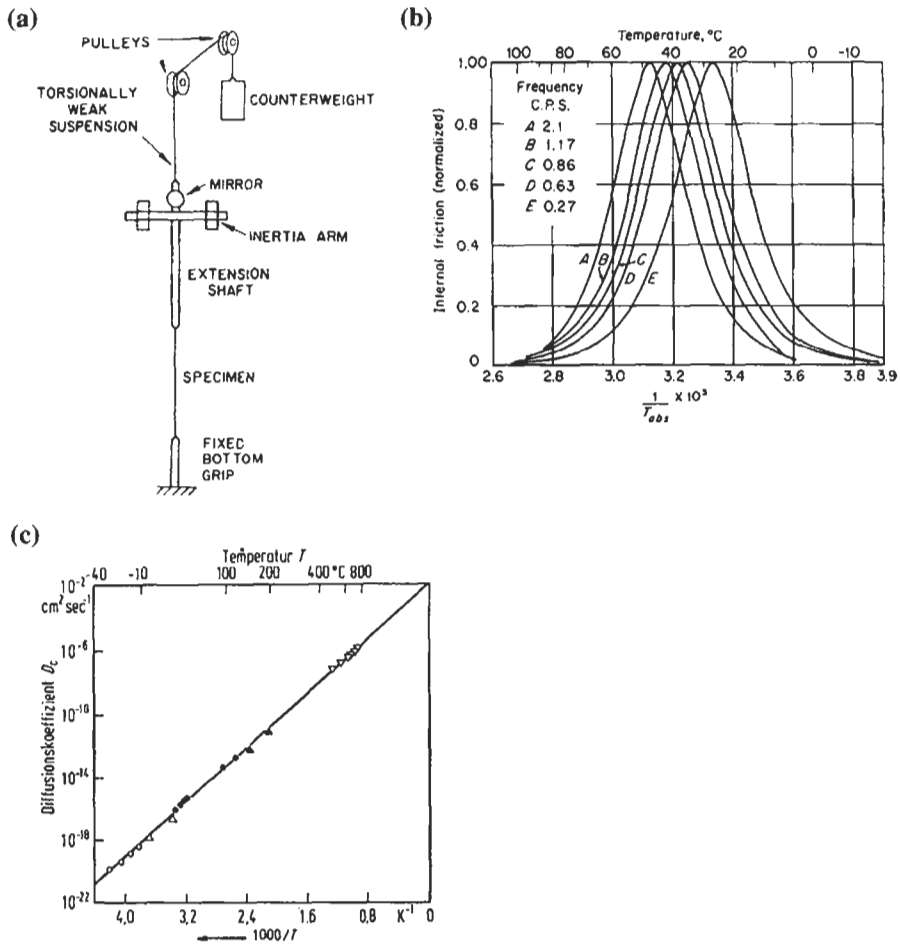


Figure 5.2. (a) Arrangement of a Snoek pendulum. (b) Internal friction as a function of temperature, at different pendulum frequencies, for a solution of carbon in iron. (c) Diffusion of carbon in iron over 14 decades, using the Snoek effect (−30–200°C) and conventional radioisotope method (400–700°C).

scientific life to the theory of dislocations and has written or edited many major texts on the subject.

Just recently (Wilde *et al.* 2000), half a century after the indirect demonstration, it has at last become possible to see carbon atmospheres around dislocations in steel directly, by means of atom-probe imaging (see Section 6.2.4). The maximum carbon concentration in such atmospheres was estimated at 8 ± 2 at.% of carbon.

It is worthwhile to present this episode in considerable detail, because it encapsulates very clearly what was new in physical metallurgy in the middle of the century. The elements are: an accurate theory of the effects in question, preferably without disposable parameters; and, to check the theory, the use of a technique of measurement (the Snoek pendulum) which is simple in the extreme in construction and use but subtle in its quantitative interpretation, so that theory ineluctably comes into the measurement itself. It is impossible that any handwaver could ever have conceived the use of a pendulum to measure dissolved carbon concentrations!

The Snoek pendulum, which in the most general sense is a device to measure relaxations, has also been used to measure relaxation caused by tangential displacements at grain boundaries. This application has been the central concern of a distinguished Chinese physicist, Tingsui Kê, for all of the past 55 years. He was stimulated to this study by Clarence Zener, in 1945, and pursued the approach, first in Chicago and then in China. This exceptional fidelity to a powerful quantitative technique was recognised by a medal and an invitation to deliver an overview lecture in America, recently published shortly before his death (Kê 1999).

This sidelong glance at a grain-boundary technique is the signal to return to Rosenhain and *his* grain boundaries. The structure of grain boundaries was critically discussed in Cottrell's book, page 89 *et seq*. Around 1949, Chalmers proposed that a grain boundary has a 'transition lattice', a halfway house between the two bounding lattices. At the same time, Shockley and Read (1949, 1950) worked out how the specific energy of a simple grain boundary must vary with the degree of misorientation, for a specified axis of rotation, on the hypothesis that the transition lattice consists in fact of an array of dislocations. (The Shockley in this team was the same man who had just taken part in the invention of the transistor; his working relations with his co-inventors had become so bad that for a while he turned his interests in quite different directions.) Once this theory was available, it was very quickly checked by experiment (Aust and Chalmers 1950); the technique depended on measurement of the dihedral angle where three boundaries meet, or where one grain boundary meets a free surface. As can be seen from Figure 5.3, theory (with one adjustable parameter only) fits experiment very neatly. The Shockley/Read theory provided the motive for an experiment which had long been feasible but which no one had previously seen a reason for undertaking.

A new parepisteme was under way: its early stages were mapped in a classic text by McLean (1957), who worked in Rosenhain's old laboratory. Today, the atomic structure of interfaces, grain boundaries in particular, has become a virtual scientific industry: a recent multiauthor book of 715 pages (Wolf and Yip 1992) surveys the present state, while an even more recent equally substantial book by two well-known authors provides a thorough account of all kinds of interfaces (Sutton and Balluffi 1995). In a paper published at about the same time, Balluffi

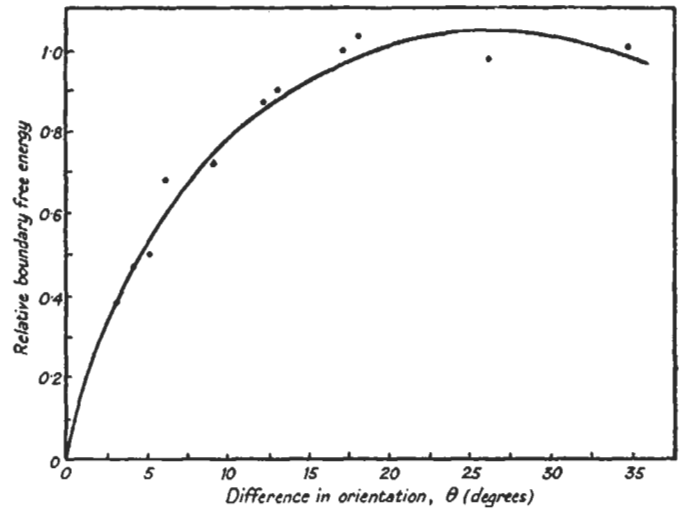


Figure 5.3. Variation of grain-boundary specific energy with difference of orientation. Theoretical curve and experimental values (•) (1950).

and Sutton (1996) discuss "why we should be interested in the atomic structure of interfaces".

One of the most elegant experiments in materials science, directed towards a particularly detailed understanding of the energetics of grain boundaries, is expounded in Section 9.4.

5.1.2 Other quantitative triumphs

The developments described in the preceding section took place during a few years before and after the exact middle of the 20th century. This was the time when the *quantitative revolution* took place in physical metallurgy, leading the way towards modern materials science. A similar revolution in the same period, as we have seen in Section 3.2.3.1, affected the study of point defects, marked especially by Seitz's classic papers of 1946 and 1954 on the nature of colour centres in ionic crystals; this was a revolution in solid-state physics as distinct from metallurgy, and was a reaction to the experimental researches of an investigator, Pohl, who believed only in empirical observation. At that time these two fields, physics and physical metallurgy, did not have much contact, and yet a quantitative revolution affected the two fields at the same time.

The means and habit of making highly precise measurements, with careful attention to the identification of sources of random and systematic error, were well established by the period I am discussing. According to a recent historical essay by

Dyson (1999), the "inventor of modern science" was James Bradley, an English astronomer, who in 1729 found out how to determine the positions of stars to an accuracy of ≈ 1 part in a million, a hundred times more accurately than the contemporaries of Isaac Newton could manage, and thus discovered stellar aberration. Not long afterwards, still in England, John Harrison constructed the first usable marine chronometer, a model of precision that was designed to circumvent a range of sources of systematic error. After these events, the best physicists and chemists knew how to make ultraprecise measurements, and recognised the vital importance of such precision as a path to understanding. William Thomson, Lord Kelvin, the famous Scottish physicist, expressed this recognition in a much-quoted utterance in a lecture to civil engineers in London, in 1883: "I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely, in your own thoughts, advanced to the state of science". Habits of precision are not enough in themselves; the invention of entirely new kinds of instrument is just as important, and to this we shall be turning in the next chapter.

Bradley may have been the inventor of modern *experimental* science, but the equally important habit of interpreting exact measurements in terms of equally exact theory came later. Maxwell, then Boltzmann in statistical mechanics and Gibbs in chemical thermodynamics, were among the pioneers in this kind of theory, and this came more than a century after Bradley. In the more applied field of metallurgy, as we have seen, it required a further century before the same habits of *theoretical* rigour were established, although in some other fields such rigour came somewhat earlier.: Heyman (1998) has recently surveyed the history of 'structural analysis' applied to load-bearing assemblies, where accurate quantitative theory was under way by the early 19th century.

Rapid advances in understanding the nature and behaviour of materials required both kinds of skill, in measurement and in theory, acting in synergy; among metallurgists, this only came to be recognised fully around the middle of the twentieth century, at about the same time as materials science became established as a new discipline.

Many other parepistemes were stimulated by the new habits of precision in theory. Two important ones are the entropic theory of rubberlike elasticity in polymers, which again reached a degree of maturity in the middle of the century (Treloar 1951), and the calculation of phase diagrams (CALPHAD) on the basis of measurements of thermochemical quantities (heats of reaction, activity coefficients, etc.); here the first serious attempt, for the Ni–Cr–Cu system, was done in the Netherlands by Meijering (1957). The early history of CALPHAD has recently been

set out (Saunders and Miodownik 1998) and is further discussed in chapter 12 (Section 12.3), while rubberlike elasticity is treated in Chapter 8 (Section 8.5.1).

Some examples of the synergy between theory and experiment will be outlined next, followed by two other examples of quantitative developments.

5.1.2.1 Pasteur's principle. As MSE became ever more quantitative and less handwaving in its approach, one feature became steadily more central – the power of surprise. Scientists learned when something they had observed was mystifying... in a word, surprising... or, what often came to the same thing, when an observation was wildly at variance with the relevant theory. The importance of this *surprise factor* goes back to Pasteur, who defined the origin of scientific creativity as being "savoir s'étonner à propos" (to know when to be astonished with a purpose in view). He applied this principle first as a young man, in 1848, to his precocious observations on optical rotation of the plane of polarisation by certain transparent crystals: he concluded later, in 1860, that the molecules in the crystals concerned must be of unsymmetrical form, and this novel idea was worked out systematically soon afterwards by van 't Hoff, who thereby created stereochemistry. A contemporary corollary of Pasteur's principle was, and remains, "accident favours the prepared mind". *Because* the feature that occasions surprise is so unexpected, the scientist who has drawn the unavoidable conclusion often has a sustained fight on his hands. Here are a few exemplifications, in outline form and in chronological sequence, of Pasteur's principle in action:

(1) Pierre Weiss and his recognition in 1907 that the only way to interpret the phenomena associated with ferromagnetism, which were inconsistent with the notions of paramagnetism, was to postulate the existence of ferromagnetic domains, which were only demonstrated visually many years later.

(2) Ernest Rutherford and the structure of the atom: his collaborators, Geiger and Marsden, found in 1909 that a very few (one in 8000) of the alpha particles used to bombard a thin metal foil were deflected through 90° or even more. Rutherford commented later, "it was about as credible as if you had fired a 15 inch. shell at a piece of tissue paper and it came back and hit you". The point was that, in the light of Rutherford's carefully constucted theory of scattering, the observation was wholly incompatible with the then current 'currant-bun' model of the atom, and his observations forced him to conceive the planetary model, with most of the mass concentrated in a very small volume; it was this concentrated mass which accounted for the unexpected backwards scatter (see Stehle 1994). Rutherford's astonished words have always seemed to me *the* perfect illustration of Pasteur's principle.

(3) We have already seen how Orowan, Polanyi and Taylor in 1934 were independently driven by the enormous mismatch between measured and calculated

yield stresses of metallic single crystals to postulate the existence of dislocations to bridge the gap.

(4) Alan Arnold Griffith, a British engineer (1893 1963, Figure 5.4), who just after the first World War (Griffith 1920) grappled with the enormous mismatch between the fracture strength of brittle materials such as glass fibres and an approximate theoretical estimate of what the fracture strength should be. He postulated the presence of a population of minute surface cracks and worked out how such cracks would amplify an applied stress: the amplification factor would increase with the depth of the crack. Since fracture would be determined by the size of the deepest crack, his hypothesis was also able to explain why thicker fibres are on average weaker (the larger surface area makes the presence of at least one deep crack statistically more likely). Griffith's paper is one of the most frequently cited papers in the entire history of MSE. In an illuminating commentary on Griffith's great paper, J.J. Gilman has remarked: "One of the lessons that can be learned from the history of the Griffith theory is how exceedingly influential a good fundamental idea can be. Langmuir called such an idea 'divergent', that is, one that starts from a small base and spreads in depth and scope."

(5) Charles Frank and his recognition, in 1949, that the observation of ready crystal growth at small supersaturations required the participation of screw dislocations emerging from the crystal surface (Section 3.2.3.3); in this way the severe mismatch with theoretical estimates of the required supersaturation could be resolved.

Figure 5.4. Portrait of A.A. Griffith on a silver medal sponsored by Rolls-Royce, his erstwhile employer.

(6) Andrew Keller (1925–1999) who in 1957 found that the polymer polyethylene, in unbranched form, could be crystallised from solution, and at once recognised that the length of the average polymer molecule was much greater than the observed crystal thickness. He concluded that the polymer chains must fold back upon themselves, and because others refused to accept this plain necessity, Keller unwittingly launched one of the most bitter battles in the history of materials science. This is further treated in Chapter 8, Section 8.4.2.

In all these examples of Pasteur's principle in action, surprise was occasioned by the mismatch between initial quantitative theory and the results of accurate measurement, and the surprise led to the resolution of the paradox. The principle remains one of the powerful motivating influences in the development of materials science.

5.1.2.2 Deformation-mechanism and materials selection maps. Once the elastic theory of dislocations was properly established, in mid-century, quantitative theories of various kinds of plastic deformation were established. Issues such as the following were clarified theoretically as well as experimentally: What is the relation between stress and strain, for a particular material, specified imposed strain rate, temperature and grain size? What is the creep rate for a given material, stress, grain size and temperature? Rate equations were well established for such processes by the 1970s. An essential point is that the *mechanism* of plastic flow varies according to the combination of stress, temperature and grain size. For instance, a very fine-grained metal at a low stress and moderate temperature will flow predominantly by 'diffusion-creep', in which dislocations are not involved at all but deformation takes place by diffusion of vacancies through or around a grain, implying a counterflow of matter and therefore a strain.

In the light of this growing understanding, a distinguished materials engineer, Ashby (1972), and his colleague Harold Frost invented the concept of the *deformation-mechanism map*. Figure 5.5(a) and (b) are examples, referring to a nickel-based jet-engine superalloy, MAR-M200, of two very different grain sizes. The axes are shear stress (normalised with respect to the elastic shear modulus) and temperature, normalised with respect to the melting-point. The field is divided into combinations of stress and temperature for which a particular deformation mechanism predominates; the graphs also show a box which corresponds to the service conditions for a typical jet-engine turbine blade. It can be seen that the predicted flow rate (by diffusion-creep involving grain boundaries) for a blade is lowered by a factor of well over 100 by increasing the grain size from 100 μm to 10 mm.

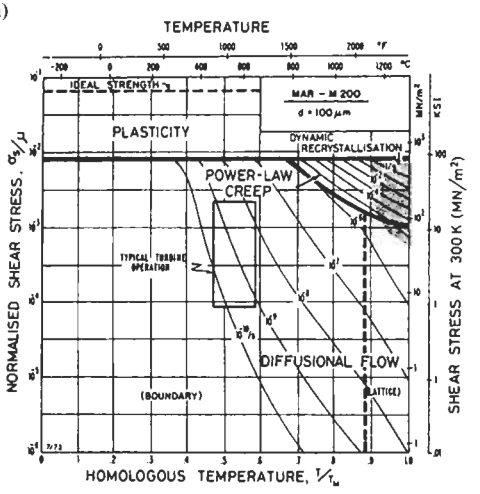
The construction, meaning and uses of such maps has been explained with great clarity in a monograph by Frost and Ashby (1982). The various mechanisms and rate-limiting factors (such as 'lattice friction' or dislocation climb combined

with glide, or Nabarro-Herring creep – see Section 4.2.5) are reviewed, and the corresponding constitutive equations (alternatively, rate equations) critically examined. The iterative stages of constructing a map such as that shown in Figure 5.5 are then explained; a simple computer program is used. The boundaries shown by thick lines correspond to conditions under which two neighbouring mechanisms are predicted to contribute the same strain rate. Certain assumptions have to be made about the superposition of parallel deformation mechanisms. Critical judgment has to be exercised by the mapmaker concerning the reliability of different, incompatible measurements of the same plastic mechanism for the same material. Maps are included in the book for a variety of metals, alloys and ceramic materials. Finally, a range of uses for such maps is rehearsed, and illustrated by a number of case-histories: (1) the flow mechanism under specific conditions can be identified, so that, for a particular use, the law which should be used for design purposes is known. (2) The total strain in service can be approximately estimated. (3) A map can offer guidance for purposes of alloy selection. (4) A map can help in designing experiments to obtain further insight into a particular flow mechanism. (5) Such maps have considerable pedagogical value in university teaching.

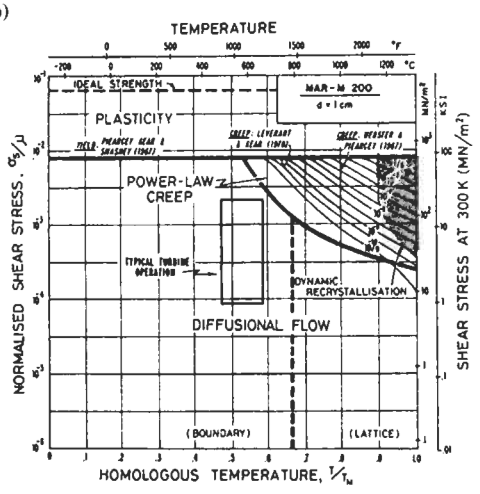
Ten years later, the deformation-mechanism map concept led Ashby to a further, crucial development – *materials selection charts*. Here, Young's modulus is plotted against density, often for room temperature, and domains are mapped for a range of quite different materials... polymers, woods, alloys, foams. The use of the diagrams is combined with a criterion for a minimum-weight design, depending on whether the important thing is resistance to fracture, resistance to strain, resistance to buckling, etc. Such maps can be used by design engineers who are not materials experts. There is no space here to go into details, and the reader is referred to a book (Ashby 1992) and a later paper which covers the principles of material selection maps for high-temperature service (Ashby and Abel 1995). This approach has a partner in what Sigmund (2000) has termed "topology optimization: a tool for the tailoring of structures and materials"; this is a systematic way of designing complex load-bearing structures, for instance for airplanes, in such a way as to minimise their weight. Sigmund remarks in passing that "any material is a structure if you look at it through a microscope with sufficient magnification".

Ashby has taken his approach a stage further with the introduction of physically based estimates of material properties where these have not been measured (Ashby 1998, Bassett *et al.* 1998), where an independent check on values is thought desirable or where property ranges of categories of materials would be useful. Figure 5.5(c) is one example of the kind of estimates which his approach makes possible. A still more recent development of Ashby's approach to materials selection is an analysis in depth of the total financial cost of using alternative materials (for different number of identical items manufactured). Thus, an expanded metallic foam beam offers the

(a)


TEMPERATURE
IDEAL STRENGTH
MAR - M 200
d = 100 μm
PLASTICITY
DYNAMIC RECRYSTALLISATION
POWER-LAW CREEP
TYPICAL TURBINE OPERATION
DIFFUSIONAL FLOW
(BOUNDARY)
(LATTICE)
NORMALISED SHEAR STRESS, σs/μ
SHEAR STRESS AT 300 K (MN/m²)
HOMOLOGOUS TEMPERATURE, T/TM

(b)


TEMPERATURE
IDEAL STRENGTH
MAR-M 200
d = 1 cm
PLASTICITY
POWER-LAW CREEP
TYPICAL TURBINE OPERATION
DYNAMIC RECRYSTALLISATION
DIFFUSIONAL FLOW
(BOUNDARY)
(LATTICE)
NORMALISED SHEAR STRESS, σs/μ
SHEAR STRESS AT 300 K (MN/m²)
HOMOLOGOUS TEMPERATURE, T/TM

(c)

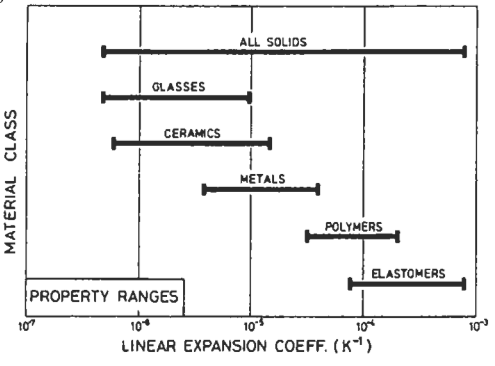

ALL SOLIDS
GLASSES
CERAMICS
METALS
POLYMERS
ELASTOMERS
PROPERTY RANGES
MATERIAL CLASS
LINEAR EXPANSION COEFF. (K⁻¹)

Figure 5.5. Deformation-mechanism maps for MAR-M200 superalloy with (a) 100 μm and (b) 10 mm grain size. The rectangular 'box' shows typical conditions of operation of a turbine blade. (after Frost and Ashby 1982). (c) A bar chart showing the range of values of expansion coefficient for generic materials classes. The range for all materials spans a factor of almost 3000; that for a class spans, typically, a factor of 20 (after Ashby 1998).

same stiffness as a solid metallic beam but at a lower mass. However, in view of the high manufacturing cost of such a foam, a detailed analysis casts doubt on the viability of such a usage (Maine and Ashby 2000).

These kinds of maps and optimisation approaches represent impressive applications of the quantitative revolution to purposes in materials engineering.

5.1.2.3 Stereology. In Section 3.1.3, the central role of *microstructure* in materials science was underlined. Two-phase and multiphase microstructures were treated and so was the morphology of grains in polycrystalline single phase microstructures. What was not discussed there in any detail was the relationship between properties, mechanical properties in particular, and such quantities as average grain size, volume fraction and shapes of precipitates, mean free path in two-phase structures: such correlations are meat and drink to some practitioners of MSE. To establish such correlations, it is necessary to establish reliable ways of measuring such quantities. This is the subject-matter of the parepisteme of *stereology*, alternatively known as *quantitative metallography*. The essence of stereological practice is to derive statistical information about a microstructure in three dimensions from measurements on two-dimensional sections. This task has two distinct components: first, *image analysis*, which nowadays involves computer-aided measurement of such variables as the area fraction of a disperse phase in a two-phase mixture or the measurement of mean free paths from a micrograph; second, a theoretical framework is required that can convert such two-dimensional numbers into three-dimensional information, with an associated estimate of probable error in each quantity. All this is much less obvious than appears at first sight: thus, crystal grains in a single phase polycrystal have a range of sizes, may be elongated in one or more directions, and it must also be remembered that a section will not cut most grains through their maximum diameter; all such factors must be allowed for in deriving a valid average grain size from micrographic measurements.

Stereology took off in the 1960s, under pressure not only from materials scientists but also from anatomists and mineralogists. Figure 3.13 (Chapter 3) shows two examples of property-microstructure relationships, taken from writings by one of the leading current experts, Exner and Hougardy (1988) and Exner (1996). Figure 3.13(a) is a way of plotting a mechanical indicator (here, indentation hardness)

against grain geometry: here, the amount of grain-boundary surface is plotted instead of the reciprocal square root of grain size. Determining interfacial area like this is one of the harder tasks in stereology. Figure 3.13(b) is a curious correlation: the ferromagnetic coercivity of the cobalt phase in a Co/WC 'hard metal' is measured as a function of the amount of interface between the two phases per unit volume. Figure 5.6 shows yield strength in relation to grain size or particle spacing for unspecified alloys: the linear relation between yield strength and the reciprocal square root of (average) grain size is known as the *Hall-Petch law* which is one of the early exemplars of the quantitative revolution in metallurgy.

The first detailed book to describe the practice and theory of stereology was assembled by two Americans, DeHoff and Rhines (1968); both these men were famous practitioners in their day. There has been a steady stream of books since then; a fine, concise and very clear overview is that by Exner (1996). In the last few years, a specialised form of microstructural analysis, entirely dependent on computerised image analysis, has emerged – *fractal analysis*, a form of measurement of roughness in two or three dimensions. Most of the voluminous literature of fractals, initiated by a mathematician, Benoit Mandelbrot at IBM, is irrelevant to materials science, but there is a sub-parepisteme of fractal analysis which relates the fractal dimension to fracture toughness: one example of this has been analysed, together with an explanation of the meaning of 'fractal dimension', by Cahn (1989).

This whole field is an excellent illustration of the deep change in metallurgy and its inheritor, materials science, wrought by the quantitative revolution of mid-century.

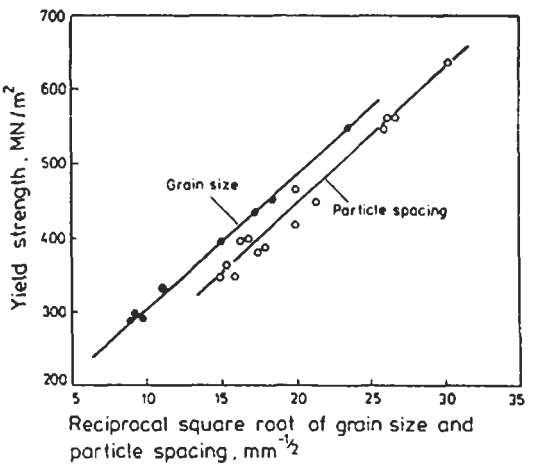


Figure 5.6. Yield strength in relation to grain size or particle spacing (courtesy of H.E. Exner).

5.1.3 Radiation damage

The first nuclear reactors were built during the Second World War in America, as an adjunct to the construction of atomic bombs. Immediately after the War, the development of civil atomic power began in several nations, and it became clear at once that the effects of neutron and gamma irradiation, and of neutron-induced fission and its products, on fuels, moderators and structural materials, had to be taken into account in the design and operation of nuclear reactors, and enormous programmes of research began in a number of large national laboratories in several countries. The resultant body of knowledge constitutes a striking example of the impact of basic research on industrial practice and also one of the best exemplars of a highly quantitative approach to research in materials science. A lively historical survey of the sequence of events that led to the development of civil atomic power, with a certain emphasis on events in Britain and a discussion of radiation damage, can be found in a recent book (West and Harris 1999).

In the early days of nuclear power (and during the War), thermal reactors all used metallic uranium as fuel. The room-temperature allotrope of uranium (the α form) is highly anisotropic in structure and properties, and it was early discovered that polycrystalline uranium normally has a preferred orientation of the population of grains (i.e., a 'texture') and then suffers gross dimensional instability on (a) thermal cycling or (b) neutron irradiation. A fuel rod could easily elongate to several times its original length. Clearly, this would cause fuel elements to burst their containers and release radioactivity into the cooling medium, and so something had to be done quickly to solve this early problem. The solution found was to use appropriate heat treatment in the high-temperature (β phase) domain followed by cooling: this way, the α grains adopted a virtually random distribution of orientations and the problem was much diminished. Later, the addition of some minor alloying constituents made the orientations even more completely random, and also generated reasonably fine grains (to obviate the 'orange peel effect'). These early researches are described in the standard text on uranium (Holden 1958). In spite of this partial success in eliminating the problems arising from anisotropy, "the metallurgy of uranium proved so intractable that in the mid-1950s the element was abandoned as a fuel worldwide" (Lander *et al.* 1994). These recent authors, in their superb review of the physical metallurgy of metallic uranium, also remark: "Once basic research (mostly, extensive research with single crystals) had shown that the anisotropic thermal expansion and consequent dimensional instability during irradiation by neutrons was an *intrinsic* property of the metal, it was abandoned in favour of oxide. *This surely represents one of the most rapid changes of technology driven by basic research!*" (my italics).

A close study of the chronology of this episode teaches another lesson. The early observation on irradiation-induced growth of uranium was purely phenomeno-

logical[1]. As Holden tells it, many contradictory theories were put forward for irradiation-induced growth in particular (and also for thermal cycling-induced growth), and it was not until the late 1950s, following some single crystal research, that there was broad agreement that (Seigle and Opinsky 1957) anisotropic diffusion was the aetiological cause; the theory specified that interstitial atoms will migrate preferentially in [0 1 0] directions and vacancies in [1 0 0] directions of a crystal, and this leads to shape distortion as observed. But the phenomenological facts alone sufficed to lead to practical policy shifts... there was no need to wait for full understanding of the underlying processes. However, the researches of Seigle and Opinsky opened the way for the understanding of many other radiation-induced phenomena in solids, in later years. It is also to be noted that the production of single crystals (which, as my own experience in 1949 confirmed, was very difficult to achieve for uranium) and the detailed understanding of diffusion, two of the parepistemes discussed in Chapter 4, were involved in the understanding of irradiation growth of uranium.

The role of interstitial point defects (atoms wedged between the normal lattice sites) came to be central in the study of irradiation effects. At the same time as Seigle and Opinsky's researches, in 1957, one of the British reactors at Windscale suffered a serious accident caused by sudden overheating of the graphite moderator (the material used to slow down neutrons in the reactor to speeds at which they are most efficient in provoking fission of uranium nuclei). This was traced to the 'Wigner effect', the elastic strain energy due to carbon atoms wedged in interstitial positions. When this energy becomes too substantial, the strain energy 'heals' when the graphite warms up, and as it warms the release becomes self-catalytic and ultimately catastrophic. This insight led to an urgent programme of research on how Wigner energy in graphite could be safely released: the key experiments are retrospectively described by Cottrell (1981), who was in charge of the programme.

In Britain, a population of thermal reactors fuelled by metallic uranium have remained in use, side by side with more modern ones (to that extent, Lander *et al.* were not quite correct about the universal abandonment of metallic uranium). In 1956, Cottrell (who was then working for the Atomic Energy Authority) identified from first principles a mechanism which would cause metallic (α) uranium to creep rapidly under small applied stress: this was linked with the differential expansion of

[1] This adjective is freely used in the scientific literature but it refers to a complex concept. According to the Oxford English Dictionary, the scientific philosopher William Whewell, in 1840, remarked that "each... science, when complete, must possess three members: the phenomenology, the aetiology, and the theory." The OED also tells us that "aetiology" means "the assignment of a cause, the rendering of a reason." So the phenomenological stage of a science refers to the mere observation of visible phenomena, while the hidden causes and the detailed origins of these causes come later.

individual anisotropic grains, a process which generated local stresses between grains, in effect mechanically 'sensitising' the polycrystal. A simple experiment with a uranium spring inside a nuclear reactor proved that the effect indeed exists (Cottrell and Roberts 1956), and this experiment led to the immediate redesign of a reactor then under construction. Here was another instance of a rapid change in technology driven by basic research.

By 1969, when a major survey (Thompson 1969) was published, the behaviour of point defects and also of dislocations in crystals subject to collisions with neutrons and to the consequential 'collision cascades' had become a major field of research. Another decade later, the subject had developed a good deal further and a highly quantitative body of theory, as well as of phenomenological knowledge, had been assembled. Gittus (1978) published an all-embracing text that covered a number of new topics: chapter headings include "Bubbles", "Voids" and "Irradiation(-enhanced) Creep".

The success of theory in interpreting the behaviour of irradiated solids and in making useful predictions of behaviour as yet unknown was largely due to the creation of the parepisteme of 'atomistic modelling'; this was lauded in a Festschrift for the 60th birthday of Ronald Bullough, a Harwell scientist who played a major part in establishing such models. In this book, Cottrell (1992) remarks: "Although atomistic theory has remained partly phenomenological, for example taking in the measured energies of point defects in its calculations of the processes of radiation damage, two features have enabled it to take giant strides into our understanding of materials. The first is that the lattice distortion produced by extended defects, such as dislocations and cracks, is long range and so can be described accurately by linear elasticity. Second, crystallographic constraints largely dictate the forms which lattice defects can take; for a regular structure can go wrong only in regular ways. As a result, atomistic theories have now been able to achieve incredibly detailed and accurate representations of complex structures and properties."

It is worth spending a moment on the subject of bubbles in irradiated nuclear fuels. Much of the related research was done in the 1960s. It was preceded by a study of the fragmentation, on heating, of a natural mineral, thorianite; this contains helium from radioactive decay of thorium and when that helium precipitates at flaws in the mineral, it "fragments explosively" (words from Barnes and Mazey 1957). The fission of uranium generates a range of fission products, including the noble gases, helium particularly. These are insoluble in equilibrium and can nucleate at lattice defects and on precipitates, causing swelling of the fuel. The study of these bubbles (some studies were based on model systems, such as helium injected into copper) led to a number of important advances in materials science, of value beyond nuclear engineering, including a detailed understanding of the migration of bubbles *in solids* in a thermal gradient. To minimise the effect of bubbles, it was important to nucleate

as many of them as possible, because many small bubbles led to much less swelling than the same amount of gas precipitated in a few large bubbles. To achieve a high bubble density, in turn, required a detailed understanding of the process of bubble nucleation. Gittus describes this important research very clearly. This subject has recently even been treated in a textbook of materials physics for undergraduates (Quéré 1998).

The subject of voids marks the coming of age of research on radiation damage. Voids are like bubbles, but do not depend on the availability of fission gases. They are produced when the equal numbers of interstitials and vacancies generated by collision cascades behave differently, so that the end result is an unbalanced population of vacancies coagulating to form *empty* bubbles, that is, 'voids'. This process is particularly important in connection with so-called fast reactors, in which no moderator is present to slow down the fission neutrons. Understanding the formation and properties of voids required an extremely sophisticated understanding of point defect behaviour, as explained in Gittus's book, and also in a very substantial American publication (Corbett and Ianniello 1972). A fine example of the very subtle quantitative theory of the differential absorption of distinct kinds of point defects at various 'sinks' – grain boundaries, dislocations, stacking faults and voids themselves – is a substantial paper by Brailsford and Bullough (1981), which might be said to represent the apotheosis of theory applied to radiation damage research. There was also a spinoff from void research in the way of novel phenomena which played no role in radiation damage but have kept solid-state scientists happily engaged for many years: a prime example is the formation of *void lattices*, first observed by Evans (1971). Populations of minute voids in metals such as molybdenum form lattices that mimic that of the underlying metal but with periodicities much greater than the lattice spacing of the metal. The interest lies in trying to understand what kinds of interactions lead to the alignment of voids into a lattice. This is still under debate, nearly 30 years later.

It is worth considering what role the study of radiation damage has played in furthering the broad domain of materials science as a whole. The question is briefly addressed by Mansur (1993) in the preface of Volume 200 of the *Journal of Nuclear Materials*. He points out that everything that is known about the behaviour of self-interstitial point defects and much of what is known about vacancies is derived from studies of atomic displacements in solids by radiation. The development of radiation-resistant structural materials for fission and fusion reactors is based on this knowledge and on studies of consequential diffusion-induced structural changes. Finally, Mansur emphasises that some technologies for improving surface properties, ion-implantation in particular, stem directly from research on radiation damage. The mutual help between MSE and research on nuclear materials is a perfect example of two-way benefits in action.

REFERENCES

Ashby, M.F. (1972) *Acta Metall.* **20**, 887.

Ashby, M.F. (1992) *Materials Selection in Mechanical Design* (Pergamon Press, Oxford).

Ashby, M.F. and Abel, C.A. (1995) in *High-Temperature Structural Materials*, eds. Cahn, R.W., Evans, A.G. and McLean, M. (Chapman & Hall, London) p. 33.

Ashby, M.F. (1998) *Proc. Roy. Soc. Lond. A* **454**, 1301.

Aust, K.T. and Chalmers, B. (1950) *Proc. Roy. Soc. Lond. A* **204**, 359.

Balluffi, R.W. and Sutton, A.P. (1996) *Mater. Sci. Forum* **207–209**, 1.

Barnes, R.S. and Mazey, D.J. (1957) *J. Nucl. Energy* **5**, 1.

Bassett, D., Brechet, Y. and Ashby, M.F. (1998) *Proc. Roy. Soc. Lond. A* **454**, 1323.

Brailsford, A.D. and Bullough, R. (1981) *Phil. Trans. Roy. Soc. Lond.* **302**, 87.

Braun, E. (1992) in *Out of the Crystal Maze*, ed. Hoddeson *et al.*, L. (Oxford University Press, Oxford) p. 317.

Cahn, R.W. (1989) *Nature* **338**, 201.

Corbett, J.W. and Ianniello, L.C. (1972) *Radiation-Induced Voids in Metals* (US Atomic Energy Commission).

Cottrell, A.H. (1953) *Dislocations and Plastic Flow in Crystals* (Clarendon Press, Oxford).

Cottrell, A.H. (1980) Dislocations in metals: the Birmingham school, 1945–1955, *Proc. Roy. Soc. A* **371**, 144.

Cottrell, A. (1981) *J. Nucl. Mater.* **100**, 64.

Cottrell, A. (1992) Theoretical models in materials science, in *Materials Modelling: From Theory to Technology*, ed. English, C.A. *et al.* (Institute of Physics Publishing, Bristol and Phildelphia) p. 3.

Cottrell, A.H. and Roberts, A.C. (1956) *Phil. Mag.* **1**, 711.

DeHoff, R.T. and Rhines, F.N. (eds.) (1968) *Quantitative Microscopy* (McGraw-Hill, New York).

Dyson, F.J. (1999) *Nature* **400**, 27.

Evans, J.H. (1971) *Nature* **229**, 403.

Exner, H.E. (1996) in *Physical Metallurgy*, vol. 2, ed. Cahn, R.W. and Haasen, P., p. 996.

Exner, H.E. and Hougardy, H.P. (1988) *Quantitative Image Analysis of Microstructures* (DGM Informationsgesellschaft Verlag, Oberursel).

Frost, H.J. and Ashby, M.F. (1982) *Deformation-Mechanism Maps* (Pergamon Press, Oxford).

Gittus, J. (1978) *Irradiation Effects in Crystalline Solids* (Applied Science Publishers, London).

Griffith, A.A. (1920) *Phil. Trans. Roy. Soc. Lond. A* **221**, 163.

Harper, S. (1951) *Phys. Rev.* **83**, 709.

Heyman, J. (1998) *Structural Analysis: A Historical Approach* (Cambridge University Press, Cambridge).

Holden, A.N. (1958) *Physical Metallurgy of Uranium*, Chapter 11 (Addison-Wesley, Reading).

Kê, T.S. (1999) *Metall. Mater. Trans.* **30A**, 2267.

Kelly, A. (1976) *Phil. Trans. Roy. Soc. Lond. A* **282**, 5.

Lander, G.H., Fisher, E.S. and Bader, S.D. (1994) *Adv. Phys.* **43**, 1.

Maine, E. and Ashby, M.F. (2000) *Adv. Eng. Mat.* **2**, 205.
McLean, D. (1957) *Grain Boundaries in Metals* (Oxford University Press, Oxford).
Mansur, L.K. (1993) *J. Nucl. Mater.* **200**, v.
Meijering, J.L. (1957) *Acta Metall.* **5**, 257.
Nowick, A.S. and Berry, B.S. (1972) *Anelastic Relaxations in Crystalline Solids* (Academic Press, New York).
Quéré, Y. (1998) *Physics of Materials* (Gordon and Breach Science Publishers, Amsterdam) p. 427.
Saunders, N. and Midownik, A.P. (1998) *Calphad: Calculation of Phase Diagrams* (Pergamon Press, Oxford).
Seigle, L.L. and Opinsky, A.J. (1957) *Nucl. Sci. Eng.* **2**, 38.
Shockley, W. and Read, W.T. (1949) *Phys. Rev.* **75**, 692; (1950) *ibid* **78**, 275.
Sigmund, O. (2000) *Phil. Trans. Roy. Soc. Lond. A* **358**, 211.
Snoek, J.L. (1940) *Ned. Tijd. v. Nat.*, **7**, 133; (1941) *ibid* **8**, 177; (1941) *Physica*, **8**, 711.
Stehle, P. (1994) *Order, Chaos, Order: The Transition from Classical to Quantum Physics* (Oxford University Press, New York) p. 218.
Sutton, A.P. and Balluffi, R.W. (1995) *Interfaces in Crystalline Materials* (Clarendon Press, Oxford).
Thompson, M.W. (1969) *Defects and Radiation Damage in Metals* (Cambridge University Press, Cambridge).
Treloar, L.R.G. (1951) *The Physics of Rubber Elasticity* (Clarendon Press, Oxford).
West, D.R.F. and Harris, J.E. (1999) *Metals and the Royal Society* (IOM Communications Ltd., London), Chapter 18.
Wilde, J., Cerezo, A. and Smith, G.D.W. (2000) *Scripta Mater.* **43**, 39.
Wolf, D. and Yip, S. (eds.) (1992) *Materials Interfaces: Atomic-level Structure and Properties* (Chapman & Hall, London).
Zener, C. (1948) *Elasticity and Anelasticity of Metals* (The University of Chicago Press, Chicago).

Chapter 6

Characterisation

6.1. Introduction 213
6.2. Examination of Microstructure 214
6.2.1 The Optical Microscope 215
6.2.2 Electron Microscopy 217
6.2.2.1 Transmission Electron Microscopy 218
6.2.2.2 Scanning Electron Microscopy 222
6.2.2.3 Electron Microprobe Analysis 226
6.2.3 Scanning Tunneling Microscopy and Its Derivatives 230
6.2.4 Field-Ion Microscopy and the Atom Probe 232
6.3. Spectrometric Techniques 234
6.3.1 Trace Element Analysis 235
6.3.2 Nuclear Methods 236
6.4. Thermoanalytical Methods 240
6.5. Hardness 243
6.6. Concluding Considerations 245
References 246

Chapter 6
Characterisation

6.1. INTRODUCTION

The characterisation of materials is a central necessity of modern materials science. Effectively, it signifies making precise distinctions between different specimens of what is nominally the same material. The concept covers qualitative and quantitative analysis of chemical composition and its variation between phases; the examination of the spatial distribution of grains, phases and of minor constituents; the crystal structures present; and the extent, nature and distribution of structural imperfections (including the stereological analysis outlined in Chapter 5).

The word *characterisation* is hard to pin down precisely. The Oxford English Dictionary, Second Edition, defines it as "the marking out of the precise form of anything; the form, mould or stamp thus impressed"; this rather implies that the form of the 'anything' is being modified by the person doing the characterizing. An alternative definition is "description of characteristics or essential features; portrayal in words". That brings us closer to the nature of characterisation as applied to materials science, although "portrayal in numbers or in images" would be closer to the mark than "portrayal in words". A third definition offered by the OED, "creation of fictitious characters", is better avoided by anyone who wishes to be taken seriously in the materials science profession.

An eminent specialist in characterisation, Eric Lifshin, has informed me that two groups of American specialists in General Electric's research centre, one group devoted to metallography, the other to analytical chemistry, came together some time in the 1960s, soon after the concept of materials science had been inaugurated in America, and formed a joint group which they named the "Materials Characterisation Operation." This is probably where the use of the word in relation to materials originated and I believe that it originally related to people working in industrial laboratories, like Lifshin himself who had long been in charge of the large characterisation group in General Electric's Corporate R&D Centre. I have checked through a selection of university textbooks of materials science from the 1960s and 1970s, and the term does not feature in any of them, so its entry into general use must have been delayed until the 1980s. In 1986, the *Encyclopedia of Materials Science and Engineering* published by Pergamon Press in Oxford included a large group of articles on 'techniques for investigation and characterisation of materials', edited by Lifshin; he also wrote a substantial overview of the whole field as it was at

that time (Lifshin 1986). Another encyclopedia covering an extremely wide range of techniques more cursorily is by Brundle *et al.* (1992).

Much earlier than these encyclopedias is a book first published in 1941 (Chalmers and Quarrell 1941, 1960) and devoted to the 'physical examination of metals'. This multiauthor book includes some recondite methods, such as the study of the damping capacity of solids (Section 5.1). In the second edition, the authors remark: "Not the least of the many changes that have taken place since the first edition appeared has been in the attitude of the metallurgist to pure science and to modern techniques involving scientific principles." The two editions span the period to which I have attributed the 'quantitative revolution', in Chapter 5.

I have described Lifshin as a 'specialist in characterisation'. This is almost a contradiction in terms, because the techniques that are sheltered under the *characterisation* umbrella are so numerous, varied and sophisticated that nobody can be truly expert in them all, even if his entire working time is devoted to the pursuit of characterisation. The problem is more serious for other materials scientists whose primary interest lies elsewhere. As Lifshin has expressed it in the preface to an encyclopedia of materials characterisation (Cahn and Lifshin 1993), "scientists and engineers have enough difficulty in keeping up with advances in their own fields without having to be materials characterisation experts. *However, it is essential to have enough basic understanding of currently used analytical methods to be able to interact effectively with such experts* (my italics)."

In this chapter, I propose to take a strongly historical approach to the field, and focus on just a few of the numerous techniques 'of investigation and characterisation'. What is not in doubt is that these techniques, and the specialised research devoted to improving them in detail, are at the heart of modern materials science.

6.2. EXAMINATION OF MICROSTRUCTURE

We have met, in Section 3.1.3, microstructure as one of the 'legs of the tripod', as a crucial precursor-concept of modern materials science. The experimental study of microstructure, by means of microscopes, is called *metallography*; in recent years, neologisms such as ceramography and even materialography have been proposed but have not taken hold. The original term is still enshrined in the names of journals and in the titles of competitions for the best and most striking micrographs, and is now taken to cover all materials, not only metals. Almost 100 years ago, in German-speaking countries, 'metallography' meant, more or less, what we now call physical metallurgy; this is exemplified by Gustav Tammann's famous *Lehrbuch der Metallographie* of 1912; at that time there was also a British journal entitled *The Metallographist*. In England, practitioners were then indeed called 'metallogra-

phists', and as we saw in Section 3.2.1, practical men had a reserved attitude towards them. To repeat here the remarks of Harry Brearly, a notable practical man of that time: "What a man sees through the microscope is more of less, and his vision has been known to be thereby so limited that he misses what he is looking for, which has been apparent at the first glance to the man whose eye is informed by experience". A more light-hearted version of a related sentiment was expressed earlier, in 1837, by the cockney Sam Weller in Dickens' novel, *Pickwick Papers*: " 'Yes, I have a pair of eyes', replied Sam, 'and that's just it. If they wos a pair o' patent double million magnifyin' gas microscopes of hextra power, p'raps I might be able to see through a flight of stairs and a deal door; but bein' only eyes, you see my wision's limited'."

That view of things has never entirely died out, and indeed there is something in it: the higher the magnification at which an object is examined, the harder it is to be sure that what is being looked at is typical of the object as a whole. The old habit of looking at specimens at a range of very different magnifications is on the way out, and that is a pity.

6.2.1 The optical microscope

The optical microscope was first used systematically in England by Robert Hooke, as exemplified by his celebrated book of 1665, *Micrographia*, and soon after by the Dutchman Antoni van Leeuwenhoek, from 1702 onwards. The observations, some of them on metal artefacts, are analysed by Smith (1960) in his key book, *A History of Metallography*. Leeuwenhoek was the first to depict metallic dendrites, but these early microscopists did not see metallic microstructures with any clarity. That decisive step was taken by Henry Sorby (Section 3.1.2) in 1864–1865, as set out on p. 175 by Smith (1960) in his book chapter devoted to that precocious scientist. Sorby invented 'normal illumination' in a microscope (Figure 6.1). This allowed him to examine polished sections of opaque materials, notably metals and alloys, and by chemical etching he was able to reveal distinct phases in a microstructure. It took Sorby a long time to overcome widespread scepticism about his mode of using the microscope. It was not until Sorby's micrographic method was used by the first investigators of phase diagrams (Section 3.1.2) at the turn of the century that its true value was at length accepted.

Sorby's method of 'vertical-illumination' microscopy has become standard in all metallographic laboratories. Meanwhile, mineralogists developed the microscope for use by transmitted light through thin transparent rock sections. The fact that (non-cubic) crystals of low symmetry have two distinct refractive indexes, according to the plane in which the incident light is polarised, gave mineralogical microscopy a complete new aspect, especially once the use of convergent incident light was introduced late in the nineteenth century and the so-called *indicatrix* was used as a

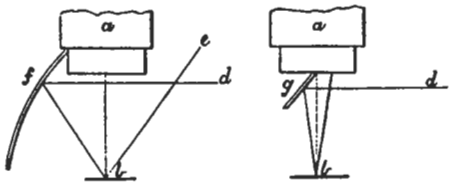


Figure 6.1. Beck's parabolic reflector and Sorby's flat mirror, both invented by Sorby (from *Quarterly Journal of the Microscopical Society*, 1865).

means of identifying minerals in polyphase rock sections. Polarised light is also used by metallurgists, but only as a means of rendering microstructure visible in non-cubic metals. This can be a vital aid in achieving clear contrast in alloys which cannot readily be etched, as it was in the examination of trigonal CuPt polycrystals (Figure 6.2) (Irani and Cahn 1973).

Most treatments of polarised light in transmission are to be found in the mineralogical literature, but a fine book presenting the subject in relation to crystal identification and structure analysis is by Bunn (1945).

In the second half of the 20th century, a number of advanced variants of optical microscopy were invented. They include phase-contrast microscopy (invented in France) and multiple-beam interference microscopy (invented in England), methods

Figure 6.2. Stoichiometric CuPt, ordered at 550°C for 157 hours. Viewed under polarised light in reflection. Shows growth of ordered domains, heterogeneously nucleated at grain boundaries and surface scratches (after Irani and Cahn 1973).

that allow depth differences very much less than one wavelength of light to be clearly distinguished; these techniques were particularly useful in investigating polytypism (Section 3.2.3.4). More recently, the radically novel technique of confocal optical microscopy (Turner and Szarowski 1993) has been introduced; this allows the sharp resolution of successive layers in depth of a specimen, whether transparent or opaque, by focusing on one layer at a time and rejecting the out-of-focus scattered light. There are also variants of microscopy that use either ultraviolet or infrared light. The former improves the resolving power of a microscope, which is diffraction-limited by the light wavelength; the latter has been used for special investigations, like that illustrated by Figure 3.14.

The techniques, instrumentation and underlying theory of optical microscopy for materials scientists have been well surveyed by Telle and Petzow (1992). One of the last published surveys including metallographic techniques of *all* kinds, optical and electronic microscopy and also techniques such as microhardness testing, was a fine book by Phillips (1971).

The impact of electron-optical instruments in materials science has been so extreme in recent years that optical microscopy is seen by many young research workers as faintly fuddy-duddy and is used less and less in advanced research; this has the unfortunate consequence, adumbrated above, that the beneficial habit of using a wide range of magnifications in examining a material is less and less followed.

6.2.2 Electron microscopy

It has long been known that the resolving power of optical microscopes is limited by the wavelength of light, according to the Abbe theory, to around 1 micrometer. If one attempts to use a higher magnification with light, all one gets is fuzzy diffraction fringes instead of sharp edges. Once it had been made clear by de Broglie, in 1924, that a beam of monoenergetic electrons has the characteristics of a wave, with a wavelength inversely proportional to the kinetic energy – and for accelerating voltages of a few thousand volts, a wavelength very much smaller than that of visible light – it became quite clear that if such a beam could be used to generate an image, then the Abbe resolution limit would be very much finer than for any light microscope. De Broglie's theory was triumphantly confirmed in 1927 by G.P. Thomson (J.J. Thomson's son) and A. Reid who showed that a beam of electrons is diffracted from a crystal, just like X-rays; moreover, electrical engineers developed the cathode-ray tube, and oscilloscope, as engineering tools soon after. The way was then clear for attempts to construct an electron microscope.

The difficulty, of course, was that electrons cannot be focused by a glass lens, and it was necessary to use either magnetic or electrostatic 'lenses'. A German, Hans Busch, in 1926/27 published some seminal papers on the analogy between the effect

of magnetic or electric fields on electrons, and formal geometric optics, and thereafter a group of German engineers, aided by two young doctoral students, a German – Ruska – and a Hungarian – Gabor, by stages developed electron-optical columns with (mostly) magnetic lenses, until in 1931 the first primitive 2-lens microscope was constructed. The leading spirit was E. Ruska, who many years later (Ruska 1980) published an account of the early work; he also eventually, very belatedly, received a Nobel Prize for his achievement. (So did Dennis Gabor, but that was for another piece of work.) By 1933, Ruska had obtained a resolution considerably greater than attainable with the best optical microscope; it soon became clear that other instrumental factors limited the resolution of electron microscopes, and that the extreme, sub-Ångström resolution deduced from the Abbé theory was not even remotely attainable. In 1937, Siemens in Germany announced the first 'serially produced' transmission electron microscope, with a remarkable resolution of $\approx$7 nm, and many Siemens microscopes were again manufactured soon after the War, beginning about 1955, and used all over the world.

In an excellent historical overview of these stages and the intellectual and practical problems which had to be overcome, Mulvey (1995) remarks that the first production microscopes pursued exactly the same electron-optical design as Ruska's first experimental microscope. The stages of subsequent improvement are outlined by Mulvey, to whom the reader is referred for further details.

6.2.2.1 Transmission electron microscopy. The penetrating power of electrons, even if accelerated by a potential difference of 100–200 kV, through dense materials such as metals is of the order of a micrometre, and therefore in the early days of transmission electron microscopy, the standard way of examining samples was by polishing, etching and then making a thin (low-density) plastic replica of the surface contour. Such features as slip lines on the surface of deformed metals were examined in this way. Meanwhile, biologists were able to send electron beams through thin sections of low-density materials prepared by microtomy, but materials scientists never thought of attempting this.

All this was changed by Peter Hirsch, then a physics research student at the Cavendish Laboratory in Cambridge (Figure 6.3). His story (told in detail in two papers, Hirsch 1980, 1986) shows how innovations of great consequence can stem from small beginnings. In brief, Hirsch (together with his colleague J.N. Kellar) in 1946 was set by his doctoral supervisor, W.H. Taylor, what in retrospect seems a trivial problem in metal physics: is the broadening of X-ray diffraction lines in plastically deformed metals due to internal strains or to the breakdown of the lattice into minute particles? The former idea was espoused by a Cambridge physicist, Henry Lipson, the latter, by a somewhat eccentric Australian, W.A. Wood, who

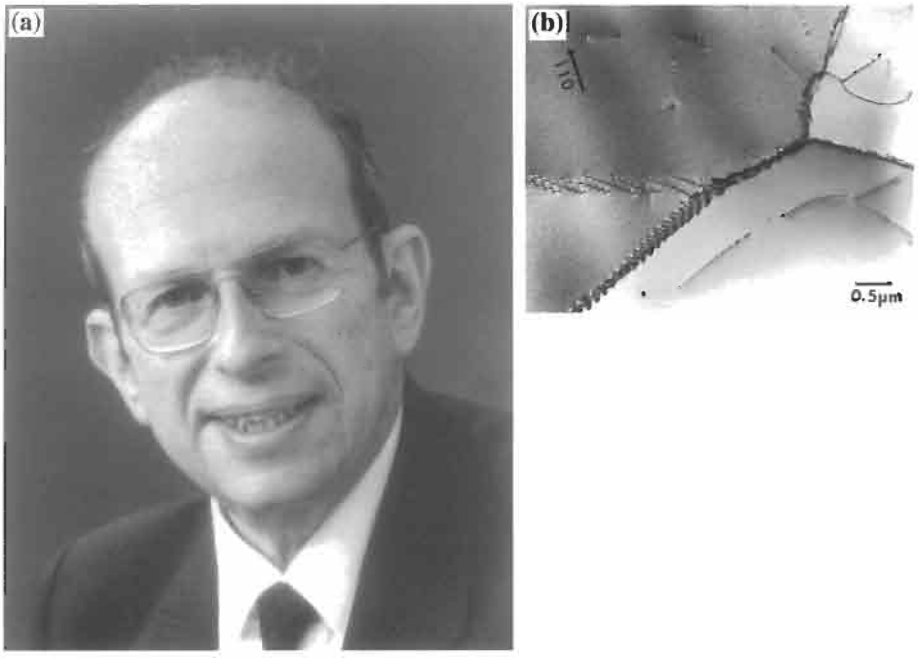


Figure 6.3. (a) Portrait of Peter Hirsch (courtesy Sir Peter Hirsch). (b) Transmission electron micrograph of dislocations in a sub-boundary in a (Ni, Fe)Al intermetallic compound (Kong and Munroe 1994) (courtesy editor of *Intermetallics*).

incidentally was an impassioned foe of the dislocation concept. Hirsch's professor, Lawrence Bragg, was happy for him to use microbeams of X-rays to settle this issue: if Wood was right, then the usually smooth diffraction rings should become spotty, because each 'particle' would generate its own diffraction spot. Because the postulated particles were so small, for this strategy to work it was necessary to use a very fine beam. Some of the experiments supported Wood's idea. It turned out that there were difficulties in generating a strong enough microbeam and Hirsch conceived the idea of replacing the X-ray beam by a fine electron beam in an electron microscope. While the main idea was to use the instrument purely as a diffraction device, after a while he was inspired by an American, R.D. Heidenreich, who in 1949 looked at deformed aluminium crystals in an electron microscope and found that they were broken down into 'particles', to exploit the image-forming function as well. With beaten gold foils, Hirsch saw some features along {1 1 1} planes which *might* have something to do with dislocations.

Hirsch obtained his Ph.D. and went off into industry; in 1953 he was awarded a fellowship which allowed him to return to Cambridge, and here he set out,

together with M.J. Whelan and with the encouragement of Sir Nevill Mott who soon after succeeded Bragg as Cavendish professor, to apply what he knew about X-ray diffraction theory to the task of making dislocations visible in electron-microscopic images. The first step was to perfect methods of thinning metal foils without damaging them; W. Bollmann in Switzerland played a vital part in this. Then the hunt for dislocations began. The important thing was to control which part of the diffracted 'signal' was used to generate the microscope image, and Hirsch and Whelan decided that 'selected-area diffraction' always had to accompany efforts to generate an image. Their group, in the person of R. Horne, was successful in seeing moving dislocation lines in 1956; the 3-year delay shows how difficult this was.

The *key here was the theory*. The pioneers' familiarity with both the kinematic and the dynamic theory of diffraction and with the 'real structure of real crystals' (the subject-matter of Lal's review cited in Section 4.2.4) enabled them to work out, by degrees, how to get good contrast for dislocations of various kinds and, later, other defects such as stacking-faults. Several other physicists who have since become well known, such as A. Kelly and J. Menter, were also involved; Hirsch goes to considerable pains in his 1986 paper to attribute credit to all those who played a major part.

There is no room here to go into much further detail; suffice it to say that the diffraction theory underlying image formation in an electron microscope plays a much more vital part in the intelligent use of an electron microscope in transmission mode than it does in the use of an optical microscope. In the words of one recent reviewer of a textbook on electron microscopy, "The world of TEM is quite different (from optical microscopy). Almost no image can be intuitively understood." For instance, to determine the Burgers vector of a dislocation from the disappearance of its image under particular illumination conditions requires an exact knowledge of the mechanism of image formation, and moreover the introduction of technical improvements such as the weak-beam method (Cockayne *et al.* 1969) depends upon a detailed understanding of image formation. As the performance of microscopes improved over the years, with the introduction of better lenses, computer control of functions and improved electron guns allowing finer beams to be used, the challenge of interpreting image formation became ever greater. Eventually, the resolving power crept towards 1–2 Å (0.1–0.2 nm) and, in high-resolution microscopes, atom columns became visible.

Figure 6.3(b) is a good example of the beautifully sharp and clear images of dislocations in assemblies which are constantly being published nowadays. It is printed next to the portrait of Peter Hirsch to symbolise his crucial contribution to modern metallography. It was made in Australia, a country which has achieved an enviable record in electron microscopy.

To form an idea of the highly sophisticated nature of the analysis of image formation, it suffices to refer to some of the classics of this field – notably the early book by Hirsch *et al.* (1965), a recent study in depth by Amelinckx (1992) and a book from Australia devoted to the theory of image formation and its simulation in the study of interfaces (Forwood and Clarebrough 1991).

Transmission electron microscopes (TEM) with their variants (scanning transmission microscopes, analytical microscopes, high-resolution microscopes, high-voltage microscopes) are now crucial tools in the study of materials: crystal defects of all kinds, radiation damage, off-stoichiometric compounds, features of atomic order, polyphase microstructures, stages in phase transformations, orientation relationships between phases, recrystallisation, local textures, compositions of phases... there is no end to the features that are today studied by TEM. Newbury and Williams (2000) have surveyed the place of the electron microscope as "the materials characterisation tool of the millennium".

A special mention is in order of high-resolution electron microscopy (HREM), a variant that permits columns of atoms normal to the specimen surface to be imaged; the resolution is better than an atomic diameter, but the nature of the image is not safely interpretable without the use of computer simulation of images to check whether the assumed interpretation matches what is actually seen. Solid-state chemists studying complex, non-stoichiometric oxides found this image simulation approach essential for their work. The technique has proved immensely powerful, especially with respect to the many types of defect that are found in microstructures.

One of the highly skilled experts working on this technique has recently (Spence 1999) assessed its impact as follows: "What has materials science learnt from HREM? In most general terms, since about 1970, HREM has taught materials scientists that real materials – from minerals to magnetic ceramics and quasicrystals – are far less perfect on the atomic scale than was previously believed. A host of microphases has been discovered by HREM, and the identification of polytypes (cf. Section 3.2.3.4) and microphases has filled a large portion of the HREM literature. The net effect of all these HREM developments has been to give theoreticians confidence in their atomic models for defects." One of the superb high-resolution micrographs shown in Spence's review is reproduced here (Figure 6.4); the separate atomic columns are particularly clear in the central area.

The improvement of transmission electron microscopes, aiming at ever higher resolutions and a variety of new and improved functions, together with the development of image-formation theory, jointly constitute one of the broadest and most important parepistemes in the whole of materials science, and enormous sums of money are involved in the industry, some 40 years after Siemens took a courageous gamble in undertaking the series manufacture of a very few microscopes at the end of the 1950s.

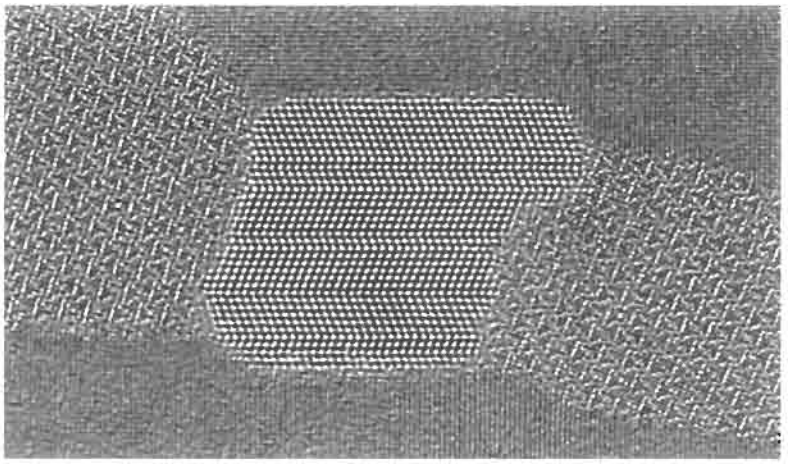


Figure 6.4. Piston alloy, showing strengthening precipitates, imaged by high-resolution electron microscopy. The matrix (top and bottom) is aluminium, while the central region is silicon. The outer precipitates were identified as $Al_5Cu_2Mg_8Si_5$. (First published by Spence 1999, reproduced here by courtesy of the originator, V. Radmilovic).

An important variant of transmission electron microscopy is the use of a particularly fine beam that is scanned across an area of the specimen and generates an image on a cathode ray screen – scanning transmission electron microscopy, or STEM. This approach has considerable advantages for composition analysis (using the approach described in the next section) and current developments in counteracting various forms of aberration in image formation hold promise of a resolution better than 1 Å (0.1 nm). This kind of microscopy is much younger than the technique described next.

6.2.2.2 Scanning electron microscopy. Some materials (e.g., fiber-reinforced composites) cannot usefully be examined by electron beams in transmission; some need to be studied by imaging a surface, and at much higher resolution than is possible by optical microscopy. This is achieved by means of the scanning electron microscope. The underlying idea is that a very finely focused 'sensing' beam is scanned systematically over the specimen surface (typically, the scan will cover rather less than a square millimetre), and secondary (or back-scattered) electrons emitted where the beam strikes the surface will be collected, counted and the varying signal used to modulate a synchronous scanning beam in a cathode-ray oscilloscope to form an enlarged image on a screen, just as a television image is formed. These instruments are today as important in materials laboratories as the transmission instruments, but

they had a more difficult birth. The first commercial instruments were delivered in 1963.

The genesis of the modern scanning microscope is described in fascinating detail by its principal begetter, Oatley (1904–1996) (Oatley 1982). Two attempts were made before he came upon the scene, both in industry, one by Manfred von Ardenne in Germany in 1939, and another by Vladimir Zworykin and coworkers in America in 1942. Neither instrument worked well enough to be acceptable; one difficulty was that the signal was so weak that to scan one frame completely took minutes. Oatley was trained as a physicist, was exposed to engineering issues when he worked on radar during the War, and after the War settled in the Engineering Department of Cambridge University, where he introduced light electrical engineering into the curriculum (until then, the Department had been focused almost exclusively on mechanical and civil engineering). In 1948 Oatley decided to attempt the creation of an effective scanning electron microscope with the help of research students for whom this would be an educative experience: as he says in his article, prior to joining the engineering department in Cambridge he had lectured for a while in physics, and so he was bound to look favourably on potential research projects which "could be broadly classified as applied physics."

Oatley then goes on to say: "A project for a Ph.D. student must provide him with good training and, if he is doing experimental work, there is much to be said for choosing a problem which involves the construction or modification of some fairly complicated apparatus. Again, I have always felt that university research in engineering should be adventurous and should not mind tackling speculative projects. This is partly to avoid direct competition with industry which, with a 'safe' project, is likely to reach a solution much more quickly, but also for two other reasons which are rarely mentioned. In the first place, university research is relatively cheap. The senior staff are already paid for their teaching duties (remember, this refers to 1948) and the juniors are Ph.D. students financed by grants which are normally very low compared with industrial salaries. Thus the feasibility or otherwise of a speculative project can often be established in a university at a small fraction of the cost that would be incurred in industry. So long as the project provides good training and leads to a Ph.D., failure to achieve the desired result need not be a disaster. (The Ph.D. candidate must, of course, be judged on the excellence of his work, not on the end result.)" He goes on to point out that at the end of the normal 3-year stay of a doctoral student in the university (this refers to British practice) the project can then be discontinued, if that seems wise, without hard feelings.

Oatley and a succession of brilliant students, collaborating with others at the Cavendish Laboratory, by degrees developed an effective instrument: a key component was an efficient plastic scintillation counter for the image-forming

electrons which is used in much the same form today. The last of Oatley's students was A.N. Broers, who later became head of engineering in Cambridge and is now the university's vice-chancellor (=president).

Oatley had the utmost difficulty in persuading industrial firms to manufacture the instrument, and in his own words, "the deadlock was broken in a rather roundabout way." In 1949, Castaing and Guinier in France reported on an electron microprobe analyser to analyse local compositions in a specimen (see next section), and a new research student, Peter Duncumb, in the Cavendish was set by V.E. Cosslett, in 1953, to add a scanning function to this concept; he succeeded in this. Because of this new feature, Oatley at last succeeded in interesting the Cambridge Instrument Company in manufacturing a small batch of scanning electron microscopes, with an analysing attachment, under the tradename of 'Stereoscan'. That name was well justified because of the remarkable depth of focus and consequent stereoscopic impression achieved by the instrument's images. Figure 6.5 shows an image of 'metal whiskers', made on the first production instrument sold by the Company in 1963 (Gardner and Cahn 1966), while Figure 6.6 shows a remarkable surface configuration produced by the differential 'sputtering' of a metal surface due to bombardment with high-energy unidirectional argon ions (Stewart

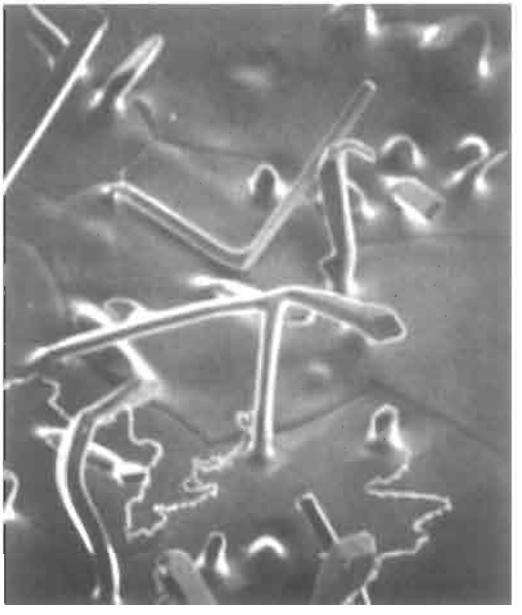


Figure 6.5. Whiskers grown at 1150°C on surface of an iron–aluminium alloy, imaged in an early scanning electron microscope ×250 (Gardner and Cahn 1966).

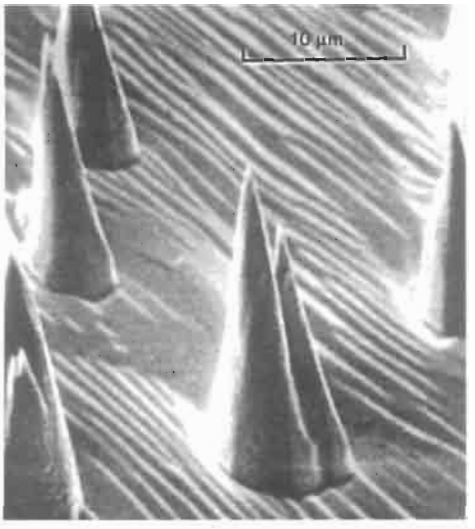


Figure 6.6. The surface of a tin crystal following bombardment with 5 keV argon ions, imaged in a scanning electron microscope (Stewart and Thompson 1969).

and Thompson 1969). Stewart had been one of Oatley's students who played a major part in developing the instruments.

A book chapter by Unwin (1990) focuses on the demanding *mechanical* components of the Stereoscan instrument, and its later version for geologists and mineralogists, the 'Geoscan', and also provides some background about the Cambridge Instrument Company and its mode of operation in building the scanning microscopes.

Run-of-the-mill instruments can achieve a resolution of 5–10 nm, while the best reach ≈1 nm. The remarkable depth of focus derives from the fact that a very small numerical aperture is used, and yet this feature does not spoil the resolution, which is not limited by diffraction as it is in an optical microscope but rather by various forms of aberration. Scanning electron microscopes can undertake compositional analysis (but with much less accuracy than the instruments treated in the next section) and there is also a way of arranging image formation that allows 'atomic-number contrast', so that elements of different atomic number show up in various degrees of brightness on the image of a polished surface.

Another new and much used variant is a procedure called 'orientation imaging microscopy' (Adams *et al.* 1993): patterns created by electrons back-scattered from a grain are automatically interpreted by a computer program, then the grain examined is automatically changed, and finally the orientations so determined are used to create an image of the polycrystal with the grain boundaries colour- or thickness-

coded to represent the magnitude of misorientation across each boundary. Very recently, this form of microscopy has been used to assess the efficacy of new methods of making a polycrystalline ceramic superconductor designed to have no large misorientations anywhere in the microstructure, since the superconducting behaviour is degraded at substantially misoriented grain boundaries.

The Stereoscan instruments were a triumphant success and their descendants, mostly made in Britain, France, Japan and the United States, have been sold in thousands over the years. They are indispensable components of modern materials science laboratories. Not only that, but they have uses which were not dreamt of when Oatley developed his first instruments: thus, they are used today to image integrated microcircuits and to search for minute defects in them.

6.2.2.3 Electron microprobe analysis. The instrument which I shall introduce here is, in my view, the most important development in characterisation since the 1939–1945 War. It has completely transformed the study of microstructure in its compositional perspective.

Henry Moseley (1887–1915) in 1913 studied the X-rays emitted by different pure metals when bombarded with high-energy electrons, using an analysing crystal to classify the wavelengths present by diffraction. He found strongly emitted 'characteristic wavelengths', different for each element, superimposed on a weak background radiation with a continuous range of wavelengths, and he identified the mathematical regularity linking the characteristic wavelengths to atomic numbers. His research cleared the way for Niels Bohr's model of the atom. It also cleared the way for compositional analysis by purely physical means. He would certainly have achieved further great things had he not been killed young as a soldier in the 'Great' War. His work is yet another example of a project undertaken to help solve a fundamental issue, the nature of atoms, which led to magnificent practical consequences.

Characteristic wavelengths can be used in two different ways for compositional analysis: it can be done by Moseley's approach, letting energetic electrons fall on the surface to be analysed and analysing the X-ray output, or else very energetic (short-wave) X-rays can be used to bombard the surface to generate secondary, 'fluorescent' X-rays. The latter technique is in fact used for compositional analysis, but until recently only by averaging over many square millimetres. In 1999, a group of French physicists were reported to have checked the genuineness of a disputed van Gogh painting by 'microfluorescence', letting an X-ray beam of the order of 1mm across impinge on a particular piece of paint to assess its local composition non-destructively; but even that does not approach the resolving power of the microprobe, to be presented here; however, it has to be accepted that a van Gogh

painting could not be non-destructively stuffed into a microprobe's vacuum chamber.

In practice, it is only the electron-bombardment approach which can be used to study the distribution of elements in a sample on a microscopic scale. The instrument was invented in its essentials by a French physicist, Raimond Castaing (1921–1998) (Figure 6.7). In 1947 he joined ONERA, the French state aeronautics laboratory on the outskirts of Paris, and there he built the first microprobe analyser as a doctoral project. (It is quite common in France for a doctoral project to be undertaken in a state laboratory away from the university world.) The suggestion came from the great French crystallographer André Guinier, who wished to determine the concentration of the pre-precipitation zones in age-hardened alloys, less than a micrometre in thickness. Castaing's preliminary results were presented at a conference in Delft in 1949, but the full flowering of his research was reserved for his doctoral thesis (Castaing 1951). This must be the most cited thesis in the history of materials science, and has been described as "a document of great interest as well

Figure 6.7. Portrait of Raimond Castaing (courtesy Dr. P.W. Hawkes and Mme Castaing).

as a moving testimony to the brilliance of his theoretical and experimental investigations".

The essence of Castaing's instrument was a finely focused electron beam and a rotatable analysing crystal plus a detector which together allowed the wavelengths and intensities of X-rays emitted from the impact site of the electron beam; there was also an optical microscope to check the site of impact in relation to the specimen's microstructure. According to an obituary of Castaing (Heinrich 1999): "Castaing initially intended to achieve this goal in a few weeks. He was doubly disappointed: the experimental difficulties exceeded his expectations by far, and when, after many months of painstaking work, he achieved the construction of the first electron probe microanalyser, he discovered that... the region of the specimen excited by the entering electrons exceeded the micron size because of diffusion of the electrons within the specimen." He was reassured by colleagues that even what he had achieved so far would be a tremendous boon to materials science, and so continued his research. He showed that for accurate quantitative analysis, the (characteristic) line intensity of each emitting element in the sample needed to be compared with the output of a standard specimen of known composition. He also identified the corrections to be applied to the measured intensity ratio, especially for X-ray absorption and fluorescence within the sample, also taking into account the mean atomic number of the sample. Heinrich remarks: "Astonishingly, this strategy remains valid today".

We saw in the previous Section that Peter Duncumb in Cambridge was persuaded in 1953 to add a scanning function to the Castaing instrument (and this in fact was the key factor in persuading industry to manufacture the scanning electron microscope, the *Stereoscan*... and later also the microprobe, the *Microscan*). The result was the generation of compositional maps for each element contained in the sample, as in the early example shown in Figure 6.8. In a symposium dedicated to Castaing, Duncumb has recently discussed the many successive mechanical and electron-optical design versions of the microprobe, some for metallurgists, some for geologists, and also the considerations which went into the decision to go for scanning (Duncumb 2000) as well as giving an account of '50 years of evolution'. At the same symposium, Newbury (2000) discusses the great impact of the microprobe on materials science. A detailed modern account of the instrument and its use is by Lifshin (1994).

The scanning electron microscope (SEM) and the electron microprobe analyser (EMA) began as distinct instruments with distinct functions, and although they have slowly converged, they are still distinct. The SEM is nowadays fitted with an 'energy-dispersive' analyser which uses a scintillation detector with an electronic circuit to determine the quantum energy of the signal, which is a fingerprint of the atomic number of the exciting element; this is convenient but less accurate than a crystal

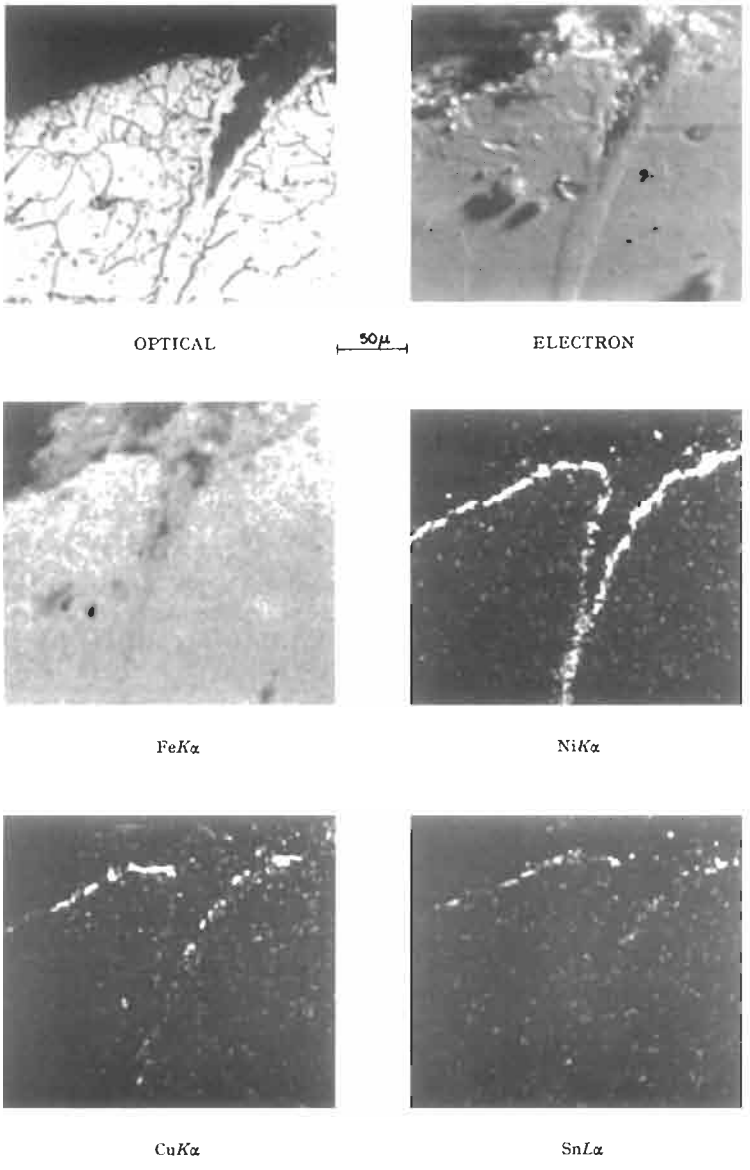


Figure 6.8. Compositional map made with an early model of the scanning electron microprobe. The pictures show the surface segregation of Ni, Cu and Sn dissolved in steel as minor constituents; the two latter constituents enriched at the surface cause 'hot shortness' (embrittlement at high temperatures), and this study was the first to demonstrate clearly the cause (Melford 1960).

detector as introduced by Castaing (this is known as a wavelength-dispersive analyser). The main objective of the SEM is resolution and depth of focus. The EMA remains concentrated on accurate chemical analysis, with the highest possible point-to-point resolution: the original optical microscope has long been replaced by a device which allows back-scattered electrons to form a topographic image, but the quality of this image is nothing like as good as that in an SEM.

The methods of compositional analysis, using either energy-dispersive or wavelength-dispersive analysis are also now available on transmission electron microscopes (TEMs); the instrument is then called an analytical transmission electron microscope. Another method, in which the energy loss of the image-forming electrons is matched to the identity of the absorbing atoms (electron energy loss spectrometry, EELS) is also increasingly applied in TEMs, and recently this approach has been combined with scanning to form EELS-generated images.

6.2.3 Scanning tunneling microscopy and its derivatives

The scanning tunnelling microscope (STM) was invented by G. Binnig and H. Rohrer at IBM's Zürich laboratory in 1981 and the first account was published a year later (Binnig *et al.* 1982). It is a device to image atomic arrangements at surfaces and has achieved higher resolution than any other imaging device. Figure 6.9(a) shows a schematic diagram of the original apparatus and its mode of operation. The essentials of the device include a very sharp metallic tip and a tripod made of

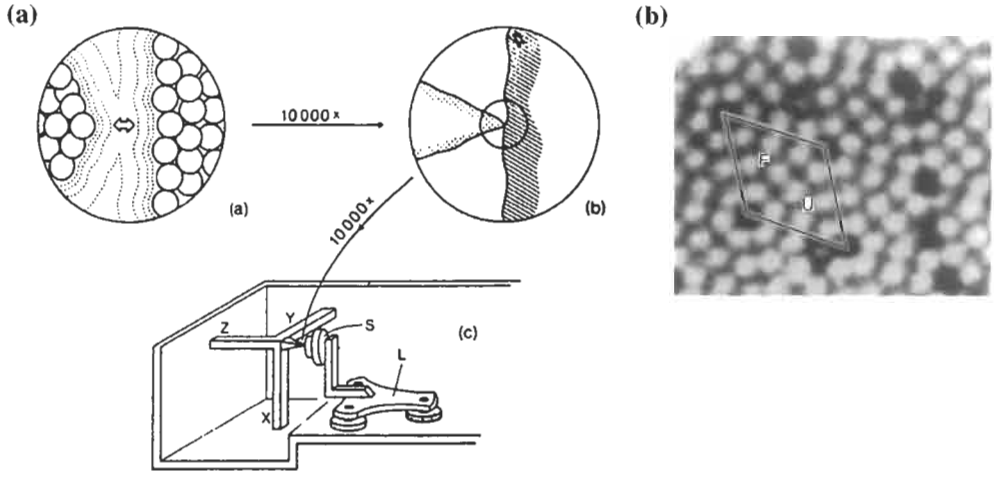


Figure 6.9. (a) Schematic of Binnig and Rohrer's original STM. (b) An image of the "7 × 7" surface rearrangement on a (1 1 1) plane of silicon, obtained by a variant of STM by Hamers *et al.* (1986).

piezoelectric material in which a minute length change can be induced by purely electrical means. In the original mode of use, the tunneling current between tip and sample was held constant by movements of the legs of the tripod; the movements, which can be at the Ångström level (0.1 nm) are recorded and modulate a scanning image on a cathode-ray monitor, and in this way an atomic image is displayed in terms of height variations. Initially, the IBM pioneers used this to display the changed crystallography (Figure 6.9(b)) in the surface layer of a silicon crystal – a key feature of modern surface science (Section 10.4). Only three years later, Binnig and Rohrer received a Nobel Prize.

According to a valuable 'historical perspective' which forms part of an excellent survey of the whole field (DiNardo 1994) to which the reader is referred, "the invention of the STM was preceded by experiments to develop a surface imaging technique whereby a non-contacting tip would scan a surface under feedback control of a tunnelling current between tip and sample." This led to the invention, in the late 1960s, of a device at the National Bureau of Standards near Washington, DC working on rather similar principles to the STM; this failed because no way was found of filtering out disturbing laboratory vibrations, a problem which Binnig and Rohrer initially solved in Zürich by means of a magnetic levitation approach.

DiNardo's 1994 survey includes about 350 citations to a burgeoning literature, only 11 years after the original papers – and that can only have been a fraction of the total literature. A comparison with the discovery of X-ray diffraction is instructive: the Braggs made their breakthrough in 1912, and they also received a Nobel Prize three years later. In 1923, however, X-ray diffraction had made little impact as yet on the crystallographic community (as outlined in Section 3.1.1.1); the mineralogists in particular paid no attention. Modern telecommunications and the conference culture have made all the difference, added to which a much wider range of issues were quickly thrown up, to which the STM could make a contribution.

In spite of the extraordinarily minute movements involved in STM operation, the modern version of the instrument is not difficult to use, and moreover there are a large number of derivative versions, such as the Atomic Force Microscope, in which the tip touches the surface with a measurable though minute force; this version can be applied to non-conducting samples. As DiNardo points out, "the most general use of the STM is for topographic imaging, not necessarily at the atomic level but on length scales from < 10 nm to ≥ 1 μm." For instance, so-called quantum dots and quantum wells, typically 100 nm in height, are often pictured in this way. Many other uses are specified in DiNardo's review.

The most arresting development is the use of an STM tip, manipulated to move both laterally and vertically, to 'shepherd' individual atoms across a crystal surface to generate features of predeterminate shapes: an atom can be contacted, lifted, transported and redeposited under visual control. This was first demonstrated at

IBM in California by Eigler and Schweizer (1990), who manipulated individual xenon atoms across a nickel (1 1 0) crystal surface. In the immediate aftermath of this achievement, many other variants of atom manipulation by STM have been published, and DiNardo surveys these.

Such an extraordinary range of uses for the STM and its variants have been found that this remarkable instrument can reasonably be placed side by side with the electron microprobe analyser as one of the key developments in modern characterisation.

6.2.4 Field-ion microscopy and the atom probe

If the tip of a fine metal wire is sharpened by making it the anode in an electrolytic circuit so that the tip becomes a hemisphere 100–500 nm across and a high negative voltage is then applied to the wire when held in a vacuum tube, a highly magnified image can be formed. This was first discovered by a German physicist, E.W. Müller, in 1937, and improved by slow stages, especially when he settled in America after the War.

Initially the instrument was called a field-emission microscope and depended on the field-induced emission of electrons from the highly curved tip. Because of the sharp curvature, the electric field close to the tip can be huge; a voltage of 20–50 V/nm can be generated adjacent to the curved surface with an applied voltage of 10 kV. The emission of electrons under such circumstances was interpreted in 1928 in wave-mechanical terms by Fowler and Nordheim. Electrons spreading radially from the tip in a highly evacuated glass vessel and impinging on a phosphor layer some distance from the tip produce an image of the tip which may be magnified as much as a million times. Müller's own account of his early instrument in an encyclopedia (Müller 1962) cites no publication earlier than 1956. By 1962, field-emission patterns based on electron emission had been studied for many high-melting metals such as W, Ta, Mo, Pt, Ni; the metal has to be high-melting so that at room temperature it is strong enough to withstand the stress imposed by the huge electric field. Müller pointed out that if the field is raised sufficiently (and its sign reversed), the metal ions themselves can be forced out of the tip and form an image.

In the 1960s, the instrument was developed further by Müller and others by letting a small pressure of inert gas into the vessel; then, under the right conditions, gas atoms become ionised on colliding with metal atoms at the tip surface and it is now these gas ions which form the image – hence the new name of *field-ion microscopy*. The resolution of 2–3 nm quoted by Müller in his 1962 article was gradually raised, in particular by cooling the tip to liquid-nitrogen temperature, until individual atoms could be clearly distinguished in the image. Grain boundaries, vacant lattice sites, antiphase domains in ordered compounds, and especially details

of phase transformations, are examples of features that were studied by the few groups who used the technique from the 1960s till the 1980s (e.g., Haasen 1985). A book about the method was published by Müller and Tsong (1969). The highly decorative tip images obtainable with the instrument by the early 1970s were in great demand to illustrate books on metallography and physical metallurgy.

From the 1970s on, and accelerating in the 1980s, the field-ion microscope was metamorphosed into something of much more extensive use and converted into the *atom probe*. Here, as with the electron microprobe analyser, imaging and analysis are combined in one instrument. All atom probes are run under conditions which extract metal ions from the tip surface, instead of using inert gas ions as in the field-ion microscope. In the original form of the atom probe, a small hole was made in the imaging screen and brief bursts of metal ions are extracted by applying a nanosecond voltage pulse to the tip. These ions then are led by the applied electric field along a path of 1–2 m in length; the heavier the ion, the more slowly it moves, and thus mass spectrometry can be applied to distinguish different metal species. In effect, only a small part of the specimen tip is analysed in such an instrument, but by progressive field-evaporation from the tip, composition profiles in depth can be obtained.

Various ion-optical tricks have to be used to compensate for the spread of energies of the extracted ions, which limit mass resolution unless corrected for. In the latest version of the atom probe (Cerezo *et al.* 1988), spatial as well as compositional information is gathered. The hole in the imaging screen is dispensed with and it is replaced by a position-sensitive screen that measures at each point on the screen the time of flight, and thus a compositional map with extremely high (virtually atomic) resolution is attained. Extremely sophisticated computer control is needed to obtain valid results.

The evolutionary story, from field-ion microscopy to spatially imaging time-of-flight atom probes is set out in detail by Cerezo and Smith (1994); these two investigators at Oxford University have become world leaders in atom-probe development and exploitation. Uses have focused mainly on age-hardening and other phase transformations in which extremely fine resolution is needed. Very recently, the Oxford team have succeeded in imaging a carbon 'atmosphere' formed around a dislocation line, fully half a century after such atmospheres were first identified by highly indirect methods (Section 5.1.1). Another timely application of the imaging atom probe is a study of Cu–Co metallic multilayers used for magnetoresistive probes (Sections 7.4, 10.5.1.2); the investigators (Larson *et al.* 1999) were able to relate the magnetoresistive properties to variables such as curvature of the deposited layers, short-circuiting of layers and fuzziness of the compositional discontinuity between successive layers. This study could not have been done with any other technique.

Several techniques which combine imaging with spectrometric (compositional) analysis have now been explained. It is time to move on to straight spectrometry.

6.3. SPECTROMETRIC TECHNIQUES

Until the last War, variants of optical emission spectroscopy ('spectrometry' when the technique became quantitative) were the principal supplement to wet chemical analysis. In fact, university metallurgy departments routinely employed resident analytical chemists who were primarily experts in wet methods, qualitative and quantitative, and undergraduates received an elementary grounding in these techniques. This has completely vanished now.

The history of optical spectroscopy and spectrometry, detailed separately for the 19th and 20th centuries, is retailed by Skelly and Keliher (1992), who then go on to describe present usages. In addition to emission spectrometry, which in essentials involves an arc or a flame 'contaminated' by the material to be analysed, there are the methods of fluorescence spectrometry (in which a specimen is excited by incoming light to emit characteristic light of lower quantum energy) and, in particular, the technique of atomic absorption spectrometry, invented in 1955 by Alan Walsh (1916–1997). Here a solution that is to be analysed is vaporized and suitable light is passed through the vapor reservoir: the composition is deduced from the absorption lines in the spectrum. The absorptive approach is now very widespread.

Raman spectrometry is another variant which has become important. To quote one expert (Purcell 1993), "In 1928, the Indian physicist C.V. Raman (later the first Indian Nobel prizewinner) reported the discovery of frequency-shifted lines in the scattered light of transparent substances. The shifted lines, Raman announced, were independent of the exciting radiation and characteristic of the sample itself." It appears that Raman was motivated by a passion to understand the deep blue colour of the Mediterranean. The many uses of this technique include examination of polymers and of silicon for microcircuits (using an exciting wavelength to which silicon is transparent).

In addition to the wet and optical spectrometric methods, which are often used to analyse elements present in very small proportions, there are also other techniques which can only be mentioned here. One is the method of mass spectrometry, in which the proportions of separate isotopes can be measured; this can be linked to an instrument called a field-ion microscope, in which as we have seen individual atoms can be observed on a very sharp hemispherical needle tip through the mechanical action of a very intense electric field. Atoms which have been ionised and detached can then be analysed for isotopic mass. This has become a powerful device for both curiosity-driven and applied research.

Another family of techniques is chromatography (Carnahan 1993), which can be applied to gases, liquids or gels: this postwar technique depends typically upon the separation of components, most commonly volatile ones, in a moving gas stream,

according to the strength of their interaction with a 'partitioning liquid' which acts like a semipermeable barrier. In gas chromatography, for instance, a sensitive electronic thermometer can record the arrival of different volatile components. One version of chromatography is used to determine molecular weight distributions in polymers (see Chapter 8, Section 8.7).

Yet another group of techniques might be called non-optical spectrometries: these include the use of Auger electrons which are in effect secondary electrons excited by electron irradiation, and photoelectrons, the latter being electrons excited by incident high-energy electromagnetic radiation – X-rays. (Photoelectron spectrometry used to be called ESCA, electron spectrometry for chemical analysis.) These techniques are often combined with the use of magnifying procedures, and their use involves large and expensive instruments working in ultrahigh vacuum. In fact, radical improvements in vacuum capabilities in recent decades have brought several new characterisation techniques into the realm of practicality; ultrahigh vacuum has allowed a surface to be studied at leisure without its contamination within seconds by molecules adsorbed from an insufficient vacuum environment (see Section 10.4).

Quite generally, each sensitive spectrometric approach today requires instruments of rapidly escalating cost, and these have to be centralised for numerous users, with resident experts on tap. The experts, however, often prefer to devote themselves to improving the instruments and the methods of interpretation: so there is a permanent tension between those who want answers from the instruments and those who have it in their power to deliver those answers.

6.3.1 Trace element analysis

A common requirement in MSE is to identify and quantify elements present in very small quantities, parts per million or even parts per billion – trace elements. The difficulty of this task is compounded when the amount of material to be analysed is small: there may only be milligrams available, for instance in forensic research. A further requirement which is often important is to establish whereabouts in a solid material the trace element is concentrated; more often than not, trace elements segregate to grain boundaries, surfaces (including internal surfaces in pores) and interphase boundaries. Trace elements have frequent roles in such phenomena as embrittlement at grain boundaries (Hondros *et al.* 1996), neutron absorption in nuclear fuels and moderators, electrical properties in electroceramics (Section 7.2.2), age-hardening kinetics in aluminium alloys (and kinetics of other phase transformations, such as ordering reactions), and notably in optical glass fibres used for communication (Section 7.5.1).

Sibilia (1988), in his guide to materials characterisation and chemical analysis, offers a concise discussion of the sensitivity of different analytical techniques for

trace elements. Thus for optical emission spectrometry, the detection limits for various elements are stated to range from 0.002 μg for beryllium to as much as 0.2 μg for lead or silicon. For atomic absorption spectrometry, detection limits are expressed in mg/litre of solution and typically range from 0.00005 to 0.001 mg/l; since only a small fraction of a litre is needed to make an analysis, this means that absolute detection limits are considerably smaller than for the emission method. A technique widely used for trace element analysis is neutron activation analysis (Hossain 1992): a sample, which can be as small as 1 mg, is exposed to neutrons in a nuclear reactor, which leads to nuclear transmutation, generating a range of radioactive species; these can be recognised and measured by examining the nature, energy and intensity of the radiation emitted by the samples after activation and the half-lives of the underlying isotopes. Thus, oxygen, nitrogen and fluorine can be analysed in polymers, and trace elements in optical fibres.

Trace element analysis has become sufficiently important, especially to industrial users, that commercial laboratories specialising in "trace and ultratrace elemental analysis" are springing up. One such company specialises in "high-resolution glow-discharge mass spectromety", which can often go, it is claimed, to better than parts per billion. This company's advertisements also offer a service, domiciled in India, to provide various forms of wet chemical analysis which, it is claimed, is now "nearly impossible to find in the United States".

Very careful analysis of trace elements can have a major effect on human life. A notable example can be seen in the career of Clair Patterson (1922–1995) (memoir by Flagel 1996), who made it his life's work to assess the origins and concentrations of lead in the atmosphere and in human bodies; minute quantities had to be measured and contaminant lead from unexpected sources had to be identified in his analyses, leading to techniques of 'clean analysis'. A direct consequence of Patterson's scrupulous work was a worldwide policy shift banning lead in gasoline and manufactured products.

6.3.2 Nuclear methods

The neutron activation technique mentioned in the preceding paragraph is only one of a range of 'nuclear methods' used in the study of solids – methods which depend on the response of atomic nuclei to radiation or to the emission of radiation by the nuclei. Radioactive isotopes ('tracers') of course have been used in research ever since von Hevesy's pioneering measurements of diffusion (Section 4.2.2). These techniques have become a field of study in their own right and a number of physics laboratories, as for instance the Second Physical Institute at the University of Göttingen, focus on the development of such techniques. This family of techniques, as applied to the study of condensed matter, is well surveyed in a specialised text

(Schatz and Weidinger 1996). ('Condensed matter' is a term mostly used by physicists to denote solid materials of all kinds, both crystalline and glassy, and also liquids.)

One important approach is Mössbauer spectrometry. This Nobel-prize-winning innovation named after its discoverer, Rudolf Mössbauer, who discovered the phenomenon when he was a physics undergraduate in Germany, in 1958; what he found was so surprising that when (after considerable difficulties with editors) he published his findings in the same year, "surprisingly no one seemed to notice, care about or believe them. When the greatness of the discovery was finally appreciated, fascination gripped the scientific community and many scientists immediately started researching the phenomenon," in the words of two commentators (Gonser and Aubertin 1993). Another commentator, Abragam (1987), remarks: "His immense merit was not so much in having observed the phenomenon as in having found the explanation, which in fact had been known for a long time and only the incredible blindness of everybody had obscured". The Nobel prize was awarded to Mössbauer in 1961, *de facto* for his first publication.

The Mössbauer effect can be explained only superficially in a few words, since it is a subtle quantum effect. Normally, when an excited nucleus emits a quantum of radiation (a gamma ray) to return to its 'ground state', the emitting nucleus recoils and this can be shown to cause the emitted radiation to have a substantial 'line width', or range of frequency – a direct consequence of the Heisenberg Uncertainty Principle. Mössbauer showed that certain isotopes only can undergo recoil-free emissions, where no energy is exchanged with the crystal and the gamma-ray carries the entire energy. This leads to a phenomenally narrow linewidth. If the emitted gamma ray is then allowed to pass through a stationary absorber containing the same isotope, the sharp gamma ray is resonantly absorbed. However, it was soon discovered that the quantum properties of a nucleus can be affected by the 'hyperfine field' caused by the electrons in the neighbourhood of the absorbing nucleus; then the absorber had to be moved, by a few millimetres per second at the most, so that the Doppler effect shifted the effective frequency of the gamma ray by a minute fraction, and resonant absorption was then restored. By measuring a spectrum of absorption versus motional speed, the hyperfine field can be mapped. Today, Mössbauer spectrometry is a technique very widely used in studying condensed matter, magnetic materials in particular.

Nuclear magnetic resonance is another characterisation technique of great practical importance, and yet another that became associated with a Nobel Prize for Physics, in 1952, jointly awarded to the American pioneers, Edward Purcell and Felix Bloch (see Purcell *et al.* 1946, Bloch 1946). In crude outline, when a sample is placed in a strong, homogeneous and constant magnetic field and a small radio-frequency magnetic field is superimposed, under appropriate circumstances the

sample can resonantly absorb the radio-frequency energy; again, only some isotopes are suitable for this technique. Once more, much depends on the sharpness of the resonance; in the early researches of Purcell and Bloch, just after the Second World War, it turned out that liquids were particularly suitable; solids came a little later (see survey by Early 2001). Anatole Abragam, a Russian immigrant in France (Abragam 1987), was one of the early physicists to learn from the pioneers and to add his own developments; in his very enjoyable book of memoirs, he vividly describes the activities of the pioneers and his interaction with them. Early on, the 'Knight shift', a change in the resonant frequency due to the chemical environment of the resonating nucleus – distinctly analogous to Mössbauer's Doppler shift – gave chemists an interest in the technique, which has grown steadily. At an early stage, an overview addressed by physicists to metallurgists (Bloembergen and Rowland 1953) showed some of the applications of nuclear magnetic resonance and the Knight shift to metallurgical issues. One use which interested materials scientists a little later was 'motional narrowing': this is a sharpening of the resonance 'line' when atoms around the resonating nucleus jump with high frequency, because this motion smears out the structure in the atomic environment which would have broadened the line. For aluminium, which has no radioisotope suitable for diffusion measurements, this proved the only way to measure self-diffusion (Rowland and Fradin 1969); the ^{27}Al isotope, the only one present in natural aluminium, is very suitable for nuclear magnetic resonance measurements. In fact, this technique applied to ^{27}Al has proved to be a powerful method of studying structural features in such crystals as the feldspar minerals (Smith 1983). This last development indicates that some advanced techniques like nuclear magnetic resonance begin as characterisation techniques for measuring features like diffusion rates but by degrees come to be applied to structural features as supplements to diffraction methods.

A further important branch of 'nuclear methods' in studying solids is the use of high-energy projectiles to study compositional variations in depth, or 'profiling' (over a range of a few micrometres only): this is named Rutherford back-scattering, after the great atomic pioneer. Typically, high-energy protons or helium nuclei (alpha particles), speeded up in a particle accelerator, are used in this way. Such ions, metallic this time, are also used in one approach to making integrated circuits, by the technique of 'ion implantation'. The complex theory of such scattering and implantation is fully treated in a recent book (Nastasi *et al.* 1996).

Another relatively recent technique, in its own way as strange as Mössbauer spectrometry, is positron annihilation spectrometry. Positrons are positive electrons (antimatter), spectacularly predicted by the theoretical physicist Dirac in the 1920s and discovered in cloud chambers some years later. Some currently available radioisotopes emit positrons, so these particles are now routine tools. High-energy positrons are injected into a crystal and very quickly become 'thermalised' by

interaction with lattice vibrations. Then they diffuse through the lattice and eventually perish by annihilation with an electron. The whole process requires a few picoseconds. Positron lifetimes can be estimated because the birth and death of a positron are marked by the emission of gamma-ray quanta. When a large number of vacancies are present, many positrons are captured by a vacancy site and stay there for a while, reducing their chance of annihilation: the mean lifetime is thus increased. Vacancy concentrations can thus be measured and, by a variant of the technique which is too complex to outline here, vacancy mobility can be estimated also. The first overview of this technique was by Seeger (1973).

Finally, it is appropriate here to mention neutron scattering and diffraction. It is appropriate because, first, neutron beams are generated in nuclear reactors, and second, because the main scattering of neutrons is by atomic nuclei and not, as with X-rays, by extranuclear electrons. Neutrons are also sensitive to magnetic moments in solids and so the arrangements of atomic magnetic spins can be assessed. Further, the scattering intensity is determined by nuclear characteristics and does not rise monotonically with atomic number: light elements, deuterium (a hydrogen isotope) particularly, scatter neutrons vigorously, and so neutrons allow hydrogen positions in crystal structures to be identified. A chapter in Schatz and Weidinger's book (1996) outlines the production, scattering and measurement of neutrons, and exemplifies some of the many crystallographic uses of this approach; structural studies of liquids and glasses also make much use of neutrons, which can give information about a number of features, including thermal vibration amplitudes. In inelastic scattering, neutrons lose or gain energy as they rebound from lattice excitations, and information is gained about lattice vibrations (phonons), and also about 'spin waves'. Such information is helpful in understanding phase transformations, and superconducting and magnetic properties.

One of the principal places where the diffraction and inelastic scattering of neutrons was developed was Brookhaven National Laboratory on Long Island, NY. A recent book (Crease 1999), a 'biography' of that Laboratory, describes the circumstances of the construction and use of the high-flux (neutron) beam reactor there, which operated from 1965. (After a period of inactivity, it has just – 1999 – been permanently shut down.) Brookhaven had been set up for research in nuclear physics but this reactor after a while became focused on solid-state physics; for years there was a battle for mutual esteem between the two fields. In 1968, a Japanese immigrant, Gen Shirane (b. 1924), became head of the solid-state neutron group and worked with the famous physicist George Dienes in developing world-class solid-state research in the midst of a nest of nuclear physicists. The fascinating details of this uneasy cohabitation are described in the book. Shirane was not however the originator of neutron diffraction; that distinction belongs to Clifford Shull and Ernest Wollan, who began to use this technique in 1951 at Oak Ridge National

Laboratory, particularly to study ferrimagnetic materials. In 1994, a Nobel Prize in physics was (belatedly) awarded for this work, which is mentioned again in the next chapter, in Section 7.3. A range of achievements in neutron crystallography are reviewed by Willis (1998).

6.4. THERMOANALYTICAL METHODS

The procedures of measuring changes in some physical or mechanical property as a sample is heated, or alternatively as it is held at constant temperature, constitute the family of thermoanalytical methods of characterisation. A partial list of these procedures is: differential thermal analysis, differential scanning calorimetry, dilatometry, thermogravimetry. A detailed overview of these and several related techniques is by Gallagher (1992).

Dilatometry is the oldest of these techniques. In essence, it could not be simpler. The length of a specimen is measured as it is steadily heated and the length is plotted as a function of temperature. The steady slope of thermal expansion is disturbed in the vicinity of temperatures where a phase change or a change in magnetic character takes place. Figure 6.10 shows an example; here the state of atomic long-range order in an alloy progressively disappears on heating (Cahn *et al.* 1987). The method has fallen out of widespread use of late, perhaps because it seems too simple and

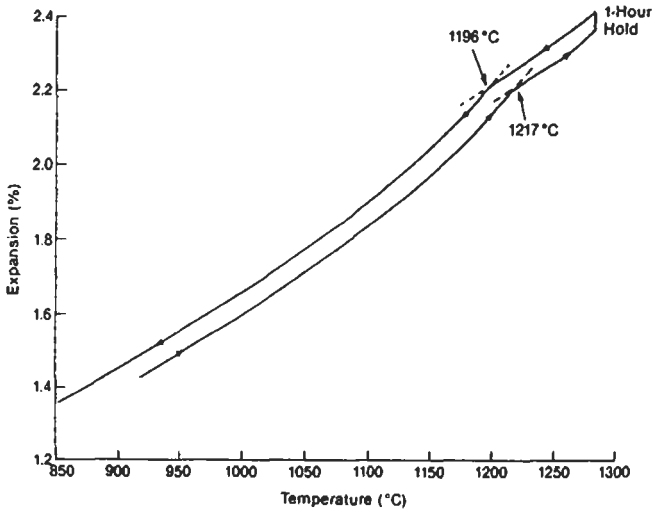


Figure 6.10. Dilatometric record of a sample of a Ni–Al–Fe alloy in the neighbourhood of an order–disorder transition temperature (Cahn *et al.* 1987).

unsophisticated; that is a pity, because the method can be very powerful. Very recently, Li *et al.* (2000) have demonstrated how, by taking into account known lattice parameters, a dilatometer can be used for quantitative analysis of the isothermal decomposition of iron-carbon austenite.

The first really accurate dilatometer was a purely mechanical instrument, using mirrors and lightbeams to record changes in length (length changes of a standard rod were also used to measure temperature). This instrument, one among several, was the brainchild of Pierre Chevenard, a French engineer who was employed by a French metallurgical company, Imphy, early in the 20th century, to set up a laboratory to foster 'la métallurgie de précision'. He collaborated with Charles–Edouard Guillaume, son of a Swiss clockmaker, who in 1883 had joined the International Bureau of Weights and Measures near Paris. There one of his tasks was to find a suitable alloy, with a small thermal expansion coefficient, from which to fabricate subsidiary length standards (the primary standard was made of precious metals, far too expensive to use widely). He chanced upon an alloy of iron with about 30 at.% of nickel with an unusually low (almost zero) thermal expansion coefficient. He worked on this and its variants for many years, in collaboration with the Imphy company, and in 1896 announced INVAR, a Fe–36%Ni alloy with virtually zero expansion coefficient near ambient temperature. Guillaume and Chevenard, two precision enthusiasts, studied the effects of ternary alloying, of many processing variables, preferred crystallographic orientation, etc., on the thermal characteristics, which eventually were tracked down to the disappearance of ferromagnetism and of its associated magnetostriction, compensating normal thermal expansion. In 1920 Guillaume gained the Nobel Prize in physics, the only occasion that a metallurgical innovation gained this honour. The story of the discovery, perfection and wide-ranging use of Invar is well told in a book to mark the centenary of its announcement (Béranger *et al.* 1996). Incidentally, after more than 100 years, the precise mechanism of the 'invar effect' is still under debate; just recently, a computer simulation of the relevant alignment of magnetic spins claims to have settled the issue once and for all (van Schilfgaarde *et al.* 1999).

Thermogravimetry is a technique for measuring changes in weight as a function of temperature and time. It is much used to study the kinetics of oxidation and corrosion processes. The samples are usually small and the microbalance used, operating by electromagnetic self-compensation of displacement, is extraordinarily sensitive (to microgram level) and stable against vibration.

Differential thermal analysis (DTA) and differential scanning calorimetry (DSC) are the other mainline thermal techniques. These are methods to identify temperatures at which specific heat changes suddenly or a latent heat is evolved or absorbed by the specimen. DTA is an early technique, invented by Le Chatelier in France in 1887 and improved at the turn of the century by Roberts-Austen (Section 4.2.2). A

sample is allowed to cool freely and anomalies in cooling rate are identified at particular temperatures. The method, simple in essence, is widely used to help in the construction of phase diagrams, because the beginning of solidification or other phase change is easily identified.

Differential scanning calorimetry (DSC) has a more tangled history. In its modern form, developed by the Perkin–Elmer Company in 1964 (Watson *et al.* 1964) two samples, the one under investigation and a reference, are in the same thermal enclosure. Platinum resistance thermometers at each specimen are arranged in a bridge circuit, and any imbalance is made to drive a heater next to one or other of the samples. The end-result is a plot of heat flow versus temperature, quantitatively accurate so that specific heats and latent heats can be determined. The modern form of the instrument normally only reaches about 700°C, which indicates the difficulty of correcting for all the sources of error. DSC is now widely used, for instance to determine glass transition temperatures in polymers and also in metallic glasses, and generally for the study of all kinds of phase transformation. It is also possible, with great care, to use a DSC in isothermal mode, to study the kinetics of phase transformations.

The antecedents of the modern DSC apparatus are many and varied, and go back all the way to the 19th century and attempts to determine the mechanical equivalent of heat accurately. A good way to examine these antecedents is to read two excellent critical reviews (Titchener and Bever 1958, Bever *et al.* 1973) of successive attempts to determine the 'stored energy of cold work', i.e., the enthalpy retained in a metal or alloy when it is heavily plastically deformed. That issue was of great concern in the 1950s and 1960s because it was linked with the multiplication of dislocations and vacancies that accompanies plastic deformation. (Almost all the retained energy is associated with these defects.) Bever and his colleagues examine the extraordinary variety of calorimetric devices used over the years in this pursuit. Perhaps the most significant are the paper by Quinney and Taylor (1937) (this is the same Taylor who had co-invented dislocations a few years earlier) and that by an Australian group, by Clarebrough *et al.* (1952), whose instrument was a close precursor of Perkin–Elmer's first commercial apparatus. The circumstances surrounding the researches of the Australian group are further discussed in Section 14.4.3. The Australian calorimeter was used not only for studying deformed metals but also for studying phase transformations, especially slow ordering transitions. Perhaps the first instrument used specifically to study order–disorder transitions in alloys was a calorimeter designed by Sykes (1935).

Figure 6.11 shows a famous example of the application of isothermal calorimetry. Gordon (1955) deformed high-purity copper and annealed samples in his precision calorimeter and measured heat output as a function of time. In this metal, the heat output is strictly proportional to the fraction of metal recrystallised.

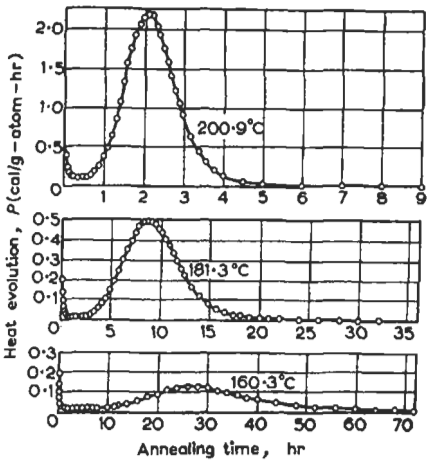


Figure 6.11. Isothermal energy release from cold-worked copper, measured calorimetrically (Gordon 1955).

This approach is an alternative to quantitative metallography and in the hands of a master gives even more accurate results than the rival method. A more recent development (Chen and Spaepen 1991) is the analysis of the isothermal curve when a material which may be properly amorphous or else nanocrystalline (e.g., a bismuth film vapour-deposited at low temperature) is annealed. The form of the isotherm allows one to distinguish nucleation and growth of a crystalline phase, from the growth of a preexisting nanocrystalline structure.

6.5. HARDNESS

The measurement of mechanical properties is a major part of the domain of characterisation. The tensile test is the key procedure, and this in turn is linked with the various tests to measure fracture toughness... crudely speaking, the capacity to withstand the weakening effects of defects. Elaborate test procedures have been developed to examine resistance to high-speed impact of projectiles, a property of civil (birdstrike on aircraft) as well as military importance. Another kind of test is needed to measure the elastic moduli in different directions of an anisotropic crystal; this is, for instance, vital for the proper exploitation of quartz crystal slices in quartz watches.

There is space here for a brief account of only one technique, that is, hardness measurement. The idea of pressing a hard object, of steel or diamond, into a smooth surface under a known load and measuring the size of the indent, as a simple and

quick way of classifying the mechanical strength of a material, goes back to the 19th century. It was often eschewed by pure scientists as a crude procedure which gave results that could not be interpreted in terms of fundamental concepts such as yield stress or work-hardening rate. There were two kinds of test: the Brinell test, in which a hardened steel sphere is used, and the Vickers test, using a pyramidally polished diamond. In the Brinell test, hardness is defined as load divided by the curved area of the indentation; in the Vickers test, the diagonal of the square impression is measured. The Vickers test was in due course miniaturised and mounted on an optical microscope to permit microhardness tests on specific features of a microstructure, and this has been extensively used.

The Brinell test, empirical though it has always seemed, did yield to close analysis. A book by Tabor (1951) has had lasting influence in this connection: his interest in hardness arose from many years of study of the true area of contact between solids pressed together, in connection with research on friction and lubrication. The Brinell test suffers from the defect that different loads will give geometrically non-similar indentations and non-comparable hardness values. In 1908, a German engineer, E. Meyer, proposed defining hardness in terms of the area of the indentation projected in the plane of the tested surface. Meyer's empirical law then stated that if W is the load and d the chordal diameter of the indentation, $W = kd^n$, where k and n are material constants. n turned out to be linked to the work-hardening capacity of the test material, and consequently, Meyer analysis was widely used for a time as an economical way of assessing this capacity. Much of Tabor's intriguing book is devoted to a fundamental examination of Meyer analysis and its implications, but this form of analysis is no longer much used today.

A quite different use of a Brinell-type test relates to highly brittle materials, and goes back to an elastic analysis by H.H. Hertz, another German engineer, in 1896. Figure 6.12 shows the Hertzian test in outline. If a hard steel ball is pressed into the polished surface of window glass, at a certain load a sudden conically shaped ring crack will spring into existence. The load required depends on the size of the largest microcrack preexisting in the glass surface (the kind of microcrack postulated in 1922 by A.A. Griffith, see Section 5.1.2.1), and a large number of identical tests performed on the same sample will allow the statistical distribution of preexisting crack depths to be assessed. The value of the Hertzian test in studying brittle materials is explained in detail by Lawn (1993).

The wide use of microhardness testing recently prompted Oliver (1993) to design a 'mechanical properties microprobe' ('nanoprobe' would have been a better name), which generates indentations considerably less than a micrometre in depth. Loads up to 120 mN (one mN $\approx$0.1 g weight) can be applied, but a tenth of that amount is commonly used and hardness is estimated by electronically measuring the depth of impression while the indentor is still in contact. This allows, inter alia, measurement

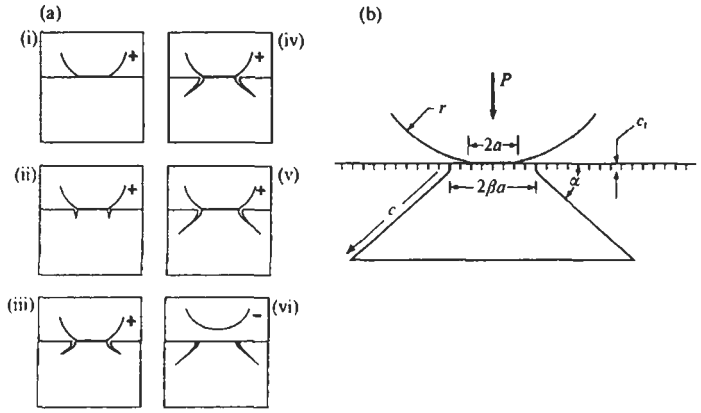


Figure 6.12. Hertzian cone crack evolution and geometry (Lawn 1993).

of a local elastic modulus. The whole process of making a nanoindentation and then moving the indentor to the next site is performed automatically, under computer control. This microprobe is now commercially available. Oliver remarks: "The mechanical properties microprobe will advance our understanding of macroscopic properties in the same way that the chemical microprobe (EMA) has improved our understanding of the chemistry of materials and the transmission electron microscope has improved our understanding of structures." Examples of applications: characterisation of materials surface-hardened by ion-implantation, a process which typically affects only the top micrometre or two of the surface region; rapidly solidified materials, which are often in the form of very thin foils which cannot be reliably examined by ordinary microhardness measurement (because the indentation has a depth of the same order as the specimen thickness).

This progression of techniques reveals how a purely empirical unsophisticated characterisation procedure can mature into a range of advanced methods capable of solving difficult problems.

6.6. CONCLUDING CONSIDERATIONS

Numerous techniques of characterisation have had to be excluded from this historical overview, for simple lack of space. A few are treated elsewhere in the book. Diffraction methods, apart from a few words about neutron diffraction, including the powerful small-angle scattering methods, were left out, partly because X-ray diffraction has also received due attention elsewhere in the book; methods of measuring electrical and magnetic properties have not been discussed; the important

techniques of characterising surfaces, such as low-energy electron diffraction, have only been skirted by an outline of scanning tunneling microscopy (but are briefly treated in Chapter 10). Another important technique of compositional profiling in near-surface regions, secondary-ion mass spectrometry (another invention of Raimond Castaing's) has also had to be left out for lack of space. Stereology was treated in Chapter 5, and one particular example of damping measurement among many was also explained in Chapter 5, but the range and importance of damping (internal friction) in materials science has to be taken on trust by the reader.

However, I believe that enough has been described to support my contention that modern methods of characterisation are absolutely central to materials science in its modern incarnation following the quantitative revolution of mid-century. That revolution owed everything to the availability of sensitive and precise techniques of measurement and characterisation.

REFERENCES

Abragam, A. (1987) *De la Physique Avant Toute Chose*, Editions Odile Jacob, Paris.

Adams, B.L., Wright, S.J. and Kunze, K. (1993) *Metall. Trans.* **24A**, 819.

Amelinckx, S. (1992) Electron diffraction and transmission electron microscopy, in *Characterisation of Materials*, ed. Lifshin, E.; *Materials Science and Technology*, vol. 2A, ed. Cahn *et al.* R.W. (VCH, Weinheim) p. 1.

Béranger, G., Duffaut, F., Morlet, J. and Tiers, J.F. (1996) *A Hundred Years after the Discovery of Invar®...the Iron–Nickel Alloys* (Lavoisier Publishing, Paris).

Bever, M.B., Holt, D.L. and Titchener, A.L. (1973) *Prog. Mater Sci.* **17**, 5.

Binnig, G., Rohrer, H., Gerber, C. and Weibel, H. (1982) *Phys. Rev. Lett.* **49**, 57; *Appl. Phys. Lett.* **40**, 178.

Bloch, F. (1946) *Phys. Rev.* **70**, 460.

Bloembergen, N. and Rowland, T.J. (1953) *Acta Metall.* **1**, 731.

Brundle, C.R., Evans, C.A. and Wilson, S. (eds.) (1992) *Encyclopedia of Materials Characterisation* (Butterworth-Heinemann *and* Greenwich: Manning, Boston).

Bunn, C.W. (1945) *Chemical Crystallography: An Introduction to Optical and X-ray Methods* (Clarenson Press, Oxford).

Cahn, R.W. and Lifshin, E. (eds.) (1993) *Concise Encyclopedia of Materials Characterisation* (Pergamon Press, Oxford) p. xxii.

Cahn, R.W., Siemers, P.A., Geiger, J.E. and Bardhan, P. (1987) *Acta Metall.* **35**, 2737.

Carnahan, Jr., C. (1993) Gas and liquid chromatography, in *Concise Encyclopedia of Materials Characterisation*, eds. Cahn, R.W. and Lifshin, E. (Pergamon Press, Oxford) p. 169.

Castaing, R. (1951) Thesis, University of Paris, Application des sondes électroniques à une méthode d'analyse ponctuelle chimique et cristallographique.

Cerezo, A., Hethrington, M.G. and Petford-Long, A.K. (1988) *Rev. Sci. Instr.* **59**, 862.
Cerezo, A. and Smith, G. (1994) Field-ion microscopy and atom probe analysis, in *Characterisation of Materials*, ed. Lifshin, E.; *Materials Science and Technology*, vol. 2B, ed. Cahn, R.W. *et al.* (VCH, Weinheim) p. 513.
Chalmers, B. and Quarrell, A.G. (1941, 1960) *The Physical Examination of Materials*, 1st and 2nd editions (Edward Arnold, London).
Chen, L.C. and Spaepen, F. (1991) *J. Appl. Phys.* **69**, 679.
Clarebrough, L.M., Hargreaves, M.E., Michell, D. and West, G.W. (1952) *Proc. Roy. Soc. (Lond.) A* **215**, 507.
Cockayne, D.J.H., Ray, I.L.E. and Whelan, M.J. (1969) *Phil. Mag.* **20**, 1265.
Crease, R.C. (1999) *Making Physics: A Biography of Brookhaven National Laboratory, 1946 1972*, Chapter 12 (University of Chicago Press, Chicago) p. 316.
DiNardo, N.J. (1994) in *Characterisation of Materials*, ed. Lifshin, E.; *Materials Science and Technology*, vol. 2B, ed. Cahn, R.W. *et al.* (VCH, Weinheim) p. 1.
Duncumb, P. (2000) *Proceedings of Symposium 'Fifty Years of Electron Microprobe Analysis'*, August 1999, *Microscopy and Microanalysis.*
Early, T.A. (2001) Article on Nuclear magnetic resonance in solids, in *Encyclopedia of Materials*, ed. Buschow, K.H.J. *et al.* (Elsevier, Amsterdam).
Eigler, D.M. and Schweizer, E.K. (1990) *Nature* **344**, 524.
Flagel, A.R. (1996) Memoir of Clair C. Patterson, *Nature* **379**, 487.
Forwood, C.T. and Clarebrough, L.M. (1991) *Electron Microscopy of Interfaces in Metals and Alloys* (Adam Hilger, Bristol).
Gallagher, P.K. (1992) in *Characterisation of Materials*, ed. Lifshin, E.; *Materials Science and Technology*, vol. 2A, ed. Cahn, R.W. *et al.* (VCH, Weinheim) p. 491.
Gardner, G.A. and Cahn, R.W. (1966) *J. Mater. Sci.* **1**, 211.
Gonser, U. and Aubertin, F. (eds.) (1993) in *Concise Encyclopedia of Materials Characterisation*, eds. Cahn, R.W. and Lifshin, E. (Pergamon Press, Oxford) p. 259.
Gordon, P. (1955) *Trans. Amer. Inst. Min. (Metall.) Engrs.* **203**, 1043.
Haasen, P. (1985) The early stages of the decomposition of alloys, *Metall. Trans. A* **16**, 1173.
Hamers, R.J., Tromp, R.M. and Demuth, J.E. (1986) *Phys. Rev. Lett.* **56**, 1972.
Heinrich, K.F.J. (1999) *Microscopy and Microanalysis*, p. 517.
Hirsch, P.B. (1980) The beginnings of solid state physics, *Proc. Roy. Soc. (Lond.) A* **371**, 160.
Hirsch, P.B. (1986) *Mater. Sci. Eng.* **84**, 1.
Hirsch, P.B., Nicholson, R.B., Howie, A., Pashley, D.W. and Whelan, M.J. (1965) *Electron Microscopy of Thin Crystals* (Butterworth, London).
Hondros, E.D., Seah, M.P., Hofmann, S. and Lejcek, P. (1996) in *Physical Metallurgy*, vol. 2, eds. Cahn, R.W. and Haasen, P. (North-Holland, Amsterdam) p. 1201.
Hossain, T.Z. (1992) in *Encyclopedia of Materials Characterisation*, eds. Brundle, C.R., Evans, C.A. and Wilson, S. (Butterworth-Heinemann *and* Greenwich: Manning, Boston) p. 671.
Irani, R.S. and Cahn, R.W. (1973) *J. Mater. Sci.* **8**, 1453.
Kong, C.H. and Munroe, P.R. (1994) *Intermetallics*, **2**, 333.

Larson, D.J., Petford-Long, A.K., Cerezo, A. and Smith, G.D.W. (1999) *Acta Mater.* **47**, 4019.

Lawn, B. (1993) *Fracture of Brittle Solids*, 2nd edition (Cambridge University Press, Cambridge) p. 253.

Li, C.-M., Sommer, F. and Mittemeijer, E.J. (2000) *Z. Metallk.* **91**, 5.

Lifshin, E . (1986) Investigation and characterisation of materials, in: *Encyclopedia of Materials Science and Engineering*, vol. 3, ed. Bever, M.B. (Pergamon Press, Oxford) p. 2389.

Lifshin, E. (1994) in *Characterisation of Materials*, ed. Lifshin, E.; *Materials Science and Technology*, vol. 2B, ed. Cahn, R.W. *et al.* (VCH, Weinheim) p. 351.

Melford, D.A. (1960) *Proceedings of Second International Symposium on X-ray Microscopy and X-ray Microanalysis, Stockholm* (Elsevier, Amsterdam) p. 407.

Müller, E.W. (1962) Article on field emission, in *Encyclopaedic Dictionary of Physics*, vol. 3, ed. Thewlis, J. (Pergamon press, Oxford) p. 120.

Müller, E.W. and Tsong, T.T. (1969) *Field-Ion Microscopy: Principles and Applications* (Elsevier, Amsterdam).

Mulvey, T. (1995) Electron-beam instruments, in *20th Century Physics*, vol. 3, ed. Pais, A. *et al.* (Institute of Physics Publishing, and New York: American Institute of Physics Press, Bristol and Philadelphia) p. 1565.

Nastasi, M., Mayer, J.W. and Hirvonen, J.K. (1996) *Ion–Solid Interactions: Fundamentals and Applications* (Cambridge University Press, Cambridge).

Newbury, D.E. (2000) *Proceedings of Symposium 'Fifty Years of Electron Microprobe Analysis'*, August 1999, *Microscopy and Microanalysis* (in press).

Newbury, D.E. and Williams, D.B. (2000) *Acta Mater.* **48**, 323.

Oatley, C.W. (1982) *J. Appl. Phys.* **53**(2), R1.

Oliver, W.C. (1993) in *Concise Encyclopedia of Materials Characterisation*, eds. Cahn, R.W. and Lifshin, E. (Pergamon Press, Oxford) p. 232.

Phillips, V.A. (1971) *Modern Metallographic Techniques and Their Applications* (Wiley-Interscience, New York).

Purcell, E.M., Torrey, H.G. and Pound, R.V. (1946) *Phys. Rev.* **69**, 37.

Purcell, F. (1993) in *Concise Encyclopedia of Materials Characterisation*, ed. Cahn, R.W. and Lifshin, E. (Pergamon Press, Oxford) p. 403.

Quinney, H. and Taylor, G.I. (1937) *Proc. Roy. Soc. (Lond.) A* **163**, 157.

Rowland, T.J. and Fradin, F.Y. (1969) *Phys. Rev.* **182**, 760.

Ruska, E. (1980) *The Early Development of Electron Lenses and Electron Microscopy* (Hirzel, Stuttgart).

Schatz, G. and Weidinger, A. (1996) *Nuclear Condensed Matter Physics: Nuclear Methods and Applications* (Wiley, Chichester).

Seeger, A. (1973) *J. Phys. F* **3**, 248.

Sibilia, J.P. (1988) *A Guide to Materials Characterisation and Chemical Analysis* (VCH Publishers, New York).

Skelly F.E.M. and Keliher, P.N. (1992) in *Characterisation of Materials*, ed. Lifshin, E.; *Materials Science and Technology*, vol. 2A, ed. Cahn, R.W. *et al.* (VCH, Weinheim) p. 423.

Smith, C.S. (1960) *A History of Metallography* (Chicago University Press, Chicago) pp. 91, 167.

Smith, M.E. (1983) *Appl. Mag. Reson.* **4**, 1.

Spence, J.C.H. (1999) The future of atomic-resolution electron microscopy, *Mat. Sci. Eng.* **R26**, 1.

Stewart, A.D.G. and Thompson, M.W. (1969) *J. Mater. Sci.* **4**, 56.

Sykes, C. (1935) *Proc. Roy. Soc. (Lond.)* **145**, 422.

Tabor, D. (1951) *The Hardness of Metals* (Clarendon Press, Oxford) (Recently reissued).

Telle, R. and Petzow, G. (1992) in *Characterisation of Materials*, ed. Lifshin, E.; *Materials Science and Technology*, vol. 2A, ed. Cahn, R.W. *et al.* (VCH, Weinheim) p. 358.

Titchener, A.L. and Bever, M.B. (1958) *Prog. Metal Phys.* **7**, 247.

Turner, J.N. and Szarowski, D.H. (1993) in *Concise Encyclopedia of Materials Characterisation*, eds. Cahn, R.W. and Lifshin, E. (Pergamon Press, Oxford) p. 68.

Unwin, D.J. (1990) in *Physicists Look Back: Studies in the History of Physics*, ed. Roche, J. (Adam Hilger, Bristol) p. 237.

Van Schilfgaarde, M., Abrikosov, I.A. and Johansson, B. (1999) *Nature* **400**, 46.

Watson, E.S., O'Neill, M.J., Justin, J. and Brenner, N. (1964) *Anal. Chem.* **326**, 1233.

Willis, B.T.M. (1998) *Acta Cryst.* **A54**, 914.

Chapter 7

Functional Materials

7.1. Introduction 253
7.2. Electrical Materials 253
7.2.1 Semiconductors 253
7.2.1.1 Silicon and Germanium 256
7.2.1.2 Physicists, Chemists and Metallurgists Cooperate 259
7.2.1.3 (Monolithic) Integrated Circuits 262
7.2.1.4 Band Gap Engineering: Confined Heterostructures 265
7.2.1.5 Photovoltaic Cells 269
7.2.2 Electrical Ceramics 271
7.2.2.1 Ferroelectrics 274
7.2.2.2 Superionic Conductors 276
7.2.2.3 Thermoelectric Materials 277
7.2.2.4 Superconducting Ceramics 279
7.3. Magnetic Ceramics 281
7.4. Computer Memories 285
7.5. Optical Glass 289
7.5.1 Optical Fibers 291
7.6. Liquid Crystals 295
7.7. Xerography 297
7.8. Envoi 298
References 299

Chapter 7
Functional Materials

7.1. INTRODUCTION

A major distinction has progressively emerged in materials science and engineering, between *structural materials* and *functional materials*. Structural materials are selected for their load-bearing capacity, functional materials for the nature of their response to electrical, magnetic, optical or chemical stimuli; sometimes a functional material is even chosen for aesthetic reasons. It is much harder to define a functional material accurately than it is to distinguish a structural material. For present purposes, I have decided to include functional ceramics in this chapter. Those ceramics developed for their resistance to mechanical stress or simply their beauty are discussed in Chapter 9. In what follows, I make no attempt to present a comprehensive account of the huge field of functional materials, but instead aim to pick a few historically important aspects and focus on these.

7.2. ELECTRICAL MATERIALS

7.2.1 Semiconductors

Silicon is today the most studied of all materials, with probably a larger accumulated number of scientific papers devoted to its properties than for any other substance. It is *the* archetype of a semiconductor and everybody knows about its transcendent importance in modern technology.

Things looked very different, as little as 70 years ago. The term 'semiconductor' appears to have been used first by Alexander Volta, after whom the volt is named, in a paper to the Royal Society of London in 1782. According to a fine historical overview of early work on semiconductors (Busch 1993), Volta examined the rate at which a charged electrometer would discharge when its contact knob was touched by different substances connected to earth. In this way, Volta was able to distinguish between metals, insulators and (in the middle) semiconductors. In 1840, Humphry Davy was the first to establish clearly that metals become poorer conductors as the temperature is raised, and he was soon followed by Michael Faraday who examined a great range of compounds, many of which changed with temperature in a way opposite to metals. Faraday's researches culminated with silver sulphide which showed an abrupt change in conductivity at about 175°C. This finding moved a German, Johann Hittorf, in 1851 to take up the study of silver sulphide, Ag_2S, as

well as Cu_2S. These were the first true semiconductors to be carefully examined. Copper sulphide came to be used for rectification of alternating current in the 1920s and 1930s, in the form of thin films created on copper sheets by an appropriate gas-phase reaction; this application began in empirical mode at a time when the mode of conduction in semiconductors was not yet understood. Another empirical success was the use of selenium photocells in photographic exposure meters; I used one of these as a schoolboy. Many hundreds of research papers were devoted to silver sulphide during the century after Hittorf: it gradually became clear that, in Busch's words, "the results are very sensitive to the preparation of the specimens, essentially purity and deviations from stoichiometry". Eventually, samples were zone-refined under a controlled pressure of sulphur vapour, and Junod in Switzerland (Junod 1959) obtained the results shown in Figure 7.1 for carefully characterised samples. The discontinuity at 170°C represents a semiconductor/metal transition, a rare

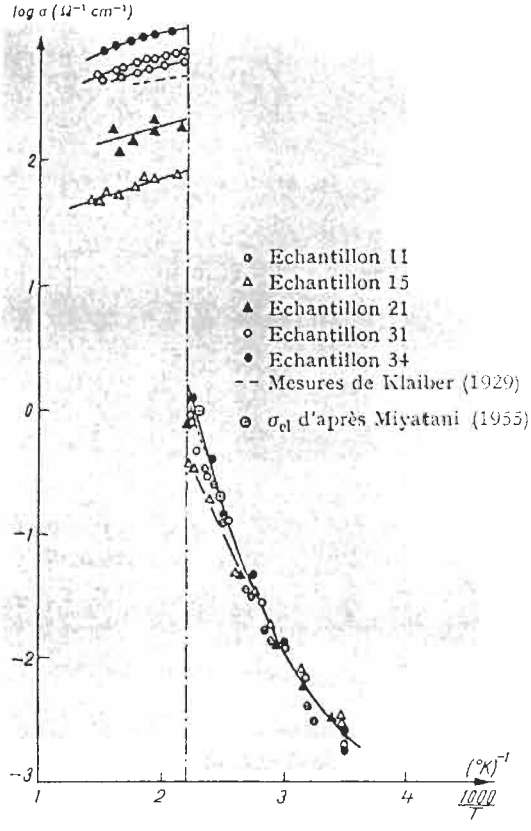


Figure 7.1. Electrical conductivity of Ag_2S as a function of temperature (after Junod, 1959).

means of comparing the characteristics of metals and semiconductors in the same substance.

The great difficulty in obtaining reproducible results with Ag_2S reinforced the prejudices of the many physicists who, in the 1930s particularly, saw no point in spending time with semiconductors. The great quantum physicist Wolfgang Pauli, whose dismissive views on the merits of solid-state physics generally have already been quoted (Section 3.3.1), was specifically contemptuous of semiconductors in a 1931 letter to his student, Rudolf Peierls: "One shouldn't work with semiconductors, that is just a mess (eine Schweinerei); who knows whether semiconductors exist at all". In 1930, B. Gudden, a physicist at Göttingen, reviewed what was known experimentally about semiconductors at that time, and concluded that only impure substances could be semiconductors... i.e., he denied the possibility of intrinsic semiconduction. Busch, in his cited historical article on semiconductors, tells how in 1938, in Switzerland, when he began an investigation of semiconduction in silicon carbide, his friends warned him that working on semiconductors meant scientific suicide. Soon after, when he gave a seminar on his researches, a colleague told him: "What are semiconductors good for? They are good for nothing. They are erratic and not reproducible".

It is no exaggeration to claim that it was the extensive worldwide body of research on semiconductors from the late 1930s onwards that converted physicists to the recognition that scrupulous control of purity, stoichiometry and crystal perfection, together with characterisation methods that could check on these features, are a precondition of understanding the nature of semiconductors and thus also a precondition of exploiting them successfully – indeed, not only semiconductors but, by extension, many kinds of materials.

At about the same time as Gudden and Pauli expressed their sceptical views, the theoretical physicist Alan Wilson of Cambridge (visiting Heisenberg at the time) wrote two classic papers on the band theory of semiconductors (Wilson 1931) and for the first time distinguished between extrinsic and intrinsic semiconductors: he postulated the presence of donors and acceptors. His theory explained clearly why the carrier density goes up sharply with temperature, so that semiconductors, unlike metals, conduct better as they heat up. Heisenberg at about the same time showed that 'holes' are equivalent to positively charged carriers. Wilson's papers marked the true beginning of the modern approach to semiconductors. Much later (Wilson 1980), after he had left science and become a highly successful captain of industry, Wilson reminisced about his glory days in solid-state theory, under title "Opportunities missed and opportunities seized"; he remarked that when he returned to Cambridge from his visit to Germany, he suggested that germanium might be an interesting substance to study in detail, but that in response the "silence was deafening".

Wilson (1939) wrote the first textbook on semiconductors. Sondheimer (1999) in his obituary of Wilson (1906–1995), remarks that "Wilson's aim in this book was to give a clear and simplified, though non-superficial, account of the subject, and Wilson's books were influential in making available to a wide range of physicists, metallurgists and engineers the advances made in solid-state physics during the 1930s."

7.2.1.1 ***Silicon and germanium.*** The study of silicon, patchy though it was, began well before the crucial events of 1948 that led to the invention of the transistor. Recently, Frederick Seitz and Norman Einspruch in America have undertaken an extensive programme of historical research on the "tangled prelude to the age of silicon electronics" and published their findings (Seitz 1996, Seitz and Einspruch 1998).

Silicon, very impure, was used by steel metallurgists as well as aluminium-alloy metallurgists early in the 20th century: in particular, iron–silicon alloy sheet was and still is used for transformer laminations, because of its excellent magnetic properties, absence of phase transformation and low cost. As a semiconductor, however, silicon first became visible in 1906, with an American inventor called G.W. Pickard who was concerned to find a good detector for messages sent by the new wireless telegraphy. He experimented extensively with cat's whisker/crystal combinations, in which a fine metallic wire (the cat's whisker) was pressed against a crystal to achieve rectification of radio-frequency signals. According to Seitz, Pickard tried more than 30,000 combinations: crystals included galena, silicon carbide and metallurgical-grade silicon (this last obtained commercially from Westinghouse, presumably in coarse polycrystal form). Pickard was granted a patent for the use of silicon in this way, in 1906. Many amateurs in the early days played with such detectors; it was common knowledge that particular sites on a crystal surface ('hot spots') performed much better than others. In the same year as Pickard's patent, Lee DeForest invented the vacuum triode tube, and when this had been sufficiently improved, it (and its successor, the pentode tube) displaced the short-lived semiconducting detector.

Nothing much happened on the silicon front until early in the Second World War, when intensive research on radar got under way. The detection problem here was that ultra-high-frequency waves had to be used, and vacuum tubes were not appropriate for rectifying these. A German, Hans Hollmann, had published a precocious book on high-frequency techniques just before the War, in which he concluded that cat's whisker/crystal devices were essential for rectification in this frequency range. Another German, Jürgen Rottgard, in 1938 followed these findings up and concluded that a silicon/tungsten combination was best. Fortunately for the Allies, the German high command was not interested, and the German work was read, by chance, by early British radar researchers, Denis Robertson in particular.

He persuaded another physicist, Herbert Skinner, to try this out: the result, in July 1940, was a silicon/tungsten rectifying device, sealed in glass and covered by a vibration-damping liquid, the silicon (still metallurgical grade) brazed on to a tungsten rod which could be sealed into the glass tube. A little earlier (see below) American physicists independently developed a very similar device. Soon after, the Dupont Company in America found a process for making much purer silicon, and thus began the Anglo-American cooperation which quickly led to effective radar measures.

This wartime work persuaded many scientists to take silicon seriously as a *useful* electrical material, and this cleared the way, intellectually, for the researches at Bell Laboratories in New Jersey which began in earnest around the end of the War. The motive energy came from Mervin Kelly, a Ph.D. physicist who became the director of research at Bell Laboratories in 1936. For his doctorate, he worked with Millikan on his renowned oil-drop technique for measuring the charge on the electron and became convinced of the importance of basic research; he joined Bell Labs in the 1920s. When the hiring freeze occasioned by the Great Depression which had begun in 1929 at length eased in 1936, Kelly engaged as his first appointment a physicist, William Shockley, who 11 years later formed part of the triumvirate who invented the transistor. Shockley recalled Kelly's remarks about his long-range goals, soon after Shockley had taken up his post: "Instead of using mechanical devices such as relays, which caused annoying maintenance problems, telephone switching should be done electronically. Kelly stressed the importance of this goal so vividly that it made an indelible impression on me." So Kelly had already formulated his central strategy more than a decade before its eventual resolution. In 1938, he reorganised physical research at Bell Labs, and moved Shockley together with a metallurgist and another physicist into a new, independent group focusing on the physics of the solid state. Kelly insisted that his scientists focused firmly on fundamentals and left the development of their findings to others. As an industrial strategy, this was truly revolutionary.

This information comes from a quite remarkable book, *Crystal Fire: The Birth of the Information Age* (Riordan and Hoddeson 1997), which maps out systematically but very accessibly the events that led to the discovery of the transistor and the aftermath of that episode. I know of no better account of the interlocking sequence of events that led to eventual success, and of the personal characteristics of the principal participants that played such a great part in the story.

Two Bell Labs employees, Russell Ohl and George Southworth, were trying in the late 1930s to detect ultrahigh-frequency radio waves with vacuum tubes, and like Skinner on the other side of the Atlantic, had no success. So, Southworth, a radio ham since childhood, remembered his early silicon-and-cat's-whisker devices and managed to retrieve some old ones from a secondhand radio store. Just as they did

for Skinner a little later, the devices worked. Thereupon Ohl and Southworth did systematic comparisons (not knowing about Pickard's similar work in 1906!) and once more silicon came up as the preferred choice; again, preferred hot spots had to be located. The objective now became to understand the nature of these hot spots. At this point, metallurgists began to be involved: Jack Scaff and Henry Theuerer in 1939 grew silicon crystals and, collaborating with Ohl and Southworth, they eventually found boundaries in an ingot on either side of which behaviour was different; the sense of rectification was opposite. The metallurgists named these regions *n-type* and *p-type*. Next, Ohl found that a piece of silicon containing *both* types of region showed an enormous electrical reaction to irradiation with light. The 'p–n barrier' allowed silicon to act as a self-rectifier. The metallurgists undertook painstaking experiments to identify the impurity elements that determined whether a locality was n-type and p-type; they mostly turned out to be in groups 5 and 3 of the periodic table. This research was indispensable to the discovery of transistor action somewhat later.

Ohl demonstrated his results to Kelly early in 1940; Kelly felt that his instincts had been proved justified. Thereupon, Bell Labs had to focus single-mindedly on radar and on silicon rectifiers for this purpose. It was not till 1945 that basic research restarted. This was the year that the theorist John Bardeen was recruited, and he in due course became inseparable from Walter Brattain, an older man and a fine experimenter who had been with Bell since the late 1920s. William Shockley formed the third member of the triumvirate, though from an early stage he and Bardeen found themselves so mutually antagonistic that Bardeen was sometimes close to resignation. But tension can be productive as well as depressing.

Some Bell employees had been trying to focus attention on germanium, the sister element to silicon, but had been discouraged, just as Wilson had been discouraged across the Atlantic a decade earlier. It was Bardeen who eventually focused on the merits of germanium and the research which led to the transistor in 1947 was centred on this element; the fact that on germanium, surface oxide could be washed off with water played a major role.

By 1947, Bardeen, building on some experiments by Shockley, had been driven to the concept of 'surface states' in semiconductors (a notable theoretical breakthrough) – an inversion from n- to p-type close to a semiconductor surface created by a strong external field – and this concept, together with the understanding of p/n junctions, came together in the systematic researches that later in that year led to the transistor, an amplifying device that mimicked the function of a vacuum valve but on a small scale and which worked at the high frequencies at which the valves were useless. (A transistor incorporates several p/n junctions and electrodes.) The original transistor, depending on a point contact rather like a cat's whisker device, was essentially the work of Bardeen and Brattain; Shockley then developed its

descendant, the junction transistor which avoided the difficult-to-fabricate point contacts and paved the way for mass production. Figure 7.2 shows the essentials of (any) transistor in purely schematic form.

John Bardeen was a truly remarkable scientist, and also a very private, taciturn man. Pippard (1995) in his obituary of Bardeen, recounts how "John returned home from the Bell Labs and walked into the kitchen to say 'We discovered something today; to (his wife) Jane's regret, all she could find in reply to what proved a momentous statement was 'That's interesting, but I have to get dinner on the table'."

A detailed account of the steps that led to the first transistor, and the steps soon afterwards to improve and miniaturise the device, and to shift from germanium to silicon, would take too much space in this chapter, and the reader must be referred to Riordan and Hoddeson's systematic and rivetting account, though space will be found for a brief account of the subsequent birth of the integrated circuit, the vector of the information age. But before this, some remarks are in order about the crucial interplay of physics and metallurgy in the run-up to the transistor.

7.2.1.2 ***Physicists, chemists and metallurgists cooperate.*** The original invention of the transistor was undoubtedly a physicists' triumph; in particular, John Bardeen's profound capacity for physical insight took the team through their many setbacks and mystifications. The later stages of improvement of semiconducting devices (not only the many kinds of transistors, but light-emitting diodes, photocells and in due course computer memories) remained the province of a kind of physicists' elite. One of the urgent tasks was to find the details of the structure of the electronic energy bands in semiconductors, a task involving theoretical as well as experimental skills. A good impression of the early days of this quest can be garnered from a discursive

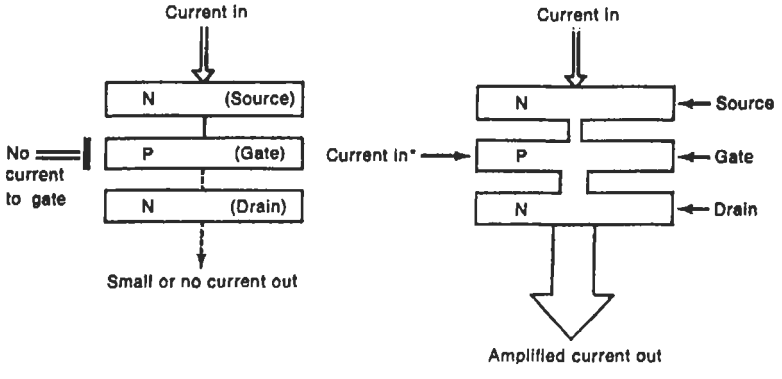


Figure 7.2. Essentials of transistor action.

retrospect by Herman (1984). Bell Labs also had some 'gate-keepers', physicists with encyclopedic solid-state knowledge who could direct researchers in promising new directions: the prince among these was Conyers Herring, characterised by Herman as a "virtual encyclopedia of solid-state knowledge". Herring, not long ago (Herring 1991) wrote an encyclopedia entry on 'Solid State Physics'... an almost but not quite impossible task.

However, physicists alone could never have produced a reliable, mass-producable transistor. We have seen that in the run-up to the events of 1947, Scaff and Theuerer had identified p- and n-regions and performed the delicate chemical analyses that enabled their nature to be identified. There was much more to come. The original transistor was successfully made with a slice of germanium cut out of a polycrystal, and early pressure to try single crystals was rebuffed by management. One Bell Labs chemist, Gordon Teal, a natural loner, pursued his obsession with single crystals in secret until at last he was given modest backing by his manager; eventually the preferred method of crystal growth came to be that based on Czochralski's method (Section 4.2.1). It soon became clear that for both germanium and silicon, this was the essential way forward, especially because intercrystalline boundaries proved to be 'electrically active'. It also became clear that dislocations were likewise electrically active and interfered with transistor action, and after a while it transpired that the best way of removing dislocations was by carefully controlled single crystal growth; to simplify, the geometry of the crystal was so arranged that dislocations initially present 'grew out' laterally, leaving a crystal with fewer than 100 dislocation lines per square centimetre, contrasted with a million times that number in ordinary material. This was the achievement of Dash (1958, 1959), whom we have already met in relation to Figure 3.14, an early confirmation of the reality of dislocations. Indeed, the work done at Bell Labs led to some of the earliest demonstrations of the existence of these disputed defects. Later, the study and control of other crystal defects in silicon, stacking-faults in particular, became a field of research in its own right.

The role of the Bell Labs metallurgists in the creation of the early transistors was clearly set out in a historical overview by the then director of the Materials Research Laboratory at Bell Labs, Scaff (1970).

The requirement for virtually defect-free material was only part of the story. The other part was the need for levels of purity never hitherto approached. The procedure was to start with ultrapure germanium or silicon and then to 'dope' that material, by solution or by solid-state diffusion, with group-3 or group-5 elements, to generate p-type and n-type regions of controlled geometry and concentration. (The study of diffusion in semiconductors was fated to become a major parepisteme in its own right.) In the 1940s and 1950s, germanium and silicon could not be extracted and refined with the requisite degree of purity from their ores. The

solution was zone-refining, the invention of a remarkable Bell Labs employee, William Pfann.

Pfann has verbally described what led up to his invention, and his account is preserved in the Bell Laboratory archives. As a youth, he was engaged by Bell Laboratories as a humble laboratory assistant, beginning with duties such as polishing samples and developing films. He attended evening classes and finally earned a bachelor's degree (in chemical engineering). He records attending a talk by a famous physical metallurgist of the day, Champion Mathewson, who spoke about plastic flow and crystal glide. Like Rosenhain before him, the youthful Pfann was captivated. Then, while still an assistant, he was invited by his manager, E.E. Schumacher, in the best Bell Labs tradition, to "take half your time and do whatever you want". Astonished, he remembered Mathewson and chose to study the deformation of lead crystals doped with antimony (as used by the Bell System for cable sheaths). He wanted to make crystals of uniform composition, and promptly invented zone-levelling. (He "took it for granted that this idea was obvious to everyone, but was wrong".) Pfann apparently impressed the Bell Director of Research by another piece of technical originality, and was made a full-fledged member of technical staff, though innocent of a doctorate. When William Shockley complained that the available germanium was nothing like pure enough, Pfann, in his own words, "put my feet up on my desk and tilted my chair back to the window sill for a short nap, a habit then well established. I had scarcely dozed off when I suddenly awoke, brought the chair down with a clack I still remember, and realised that a series of molten zones, passed through the ingot of germanium, would achieve the aim of repeated fractional crystallisation." Each zone swept some impurity along with it, until dissolved impurities near one end of the rod are reduced to a level of one in hundreds of millions of atoms. Pfann described his technique, and its mathematical theory, in a paper (Pfann 1954) and later in a book (Pfann 1958, 1966). Incidentally, the invention and perfection of zone-refining was one of the factors that turned solidification and casting from a descriptive craft into a quantitative science.

Today, methods of refining silicon via a gaseous intermediary compound have improved so much that zone-refining is no longer needed, and indeed crystal diameters are now so large that zone-refining would probably be impossible. Present-day chemical methods of preparation of silicon allow impurity levels of one part in 10^{12} to be reproducibly attained. Modern textbooks on semiconductors no longer mention zone-refining; but for more than a decade, zone-refining was an essential factor in the manufacture of transistors.

In the early years, physicists, metallurgists and chemists each formed their own community at Bell Labs, but the experience of collaboration in creating semiconductor devices progressively merged them and nowadays many of the laboratory's employees would rate themselves simply as materials scientists.

7.2.1.3 (Monolithic) integrated circuits. Mervin Kelly had told William Shockley, when he joined Bell Labs in 1936, that his objective was to replace metallic reed relays by electronic switches, because of the unreliability of the former. History repeats itself: by the late 1950s, electronic circuits incorporating discrete transistors (which had swept vacuum tubes away) had become so complex that a few of the large numbers of soldered joints were apt to be defective and eventually break down. Unreliability had arrived all over again. Computers had the most complex circuits: the earliest ones had used tubes and these were apt to burn out. Not only that, but these early computers also used metal relays which sometimes broke down; the term 'bug' still used today by computer programmers originates, some say but others deny, in a moth which had got caught in a relay and impeded its operation. (The distinguished moth is still rumored to be preserved in a glass case.) Now that transistors were used instead, unreliability centred on faulty connections.

In 1958–1959, two American inventors, Jack Kilby and Robert Noyce, men cast in the mould of Edison, independently found a way around this problem. Kilby had joined the new firm of Texas Instruments, Noyce was an employee of another young company, Fairchild Electronics, which William Shockley had founded when he resigned from Bell but mismanaged so badly that his staff grew mutinous: Noyce set up a new company to exploit his ideas. The idea was to create a complete circuit on a single small slice of silicon crystal (a 'chip'), with tiny transistors and condensers fabricated in situ and with metallic interconnects formed on the surface of the chip. The idea worked at once, and triumphantly. Greatly improved reliability was the initial objective, but it soon became clear that further benefits flowed from miniaturisation: (1) low power requirements and very small output of waste heat (which needs to be removed); (2) the ability to accommodate complex circuitry, for instance, for microprocessors or computer memories, in tiny volumes, which was vital for the computers in the Apollo moonlanding project (Figure 7.3); and, most important of all, (3) low circuit costs. Ever since Kilby's and Noyce's original chips, the density of devices in integrated circuits has steadily increased, year by year, and the process has still not reached its limit. The story of the invention and early development of integrated circuits has been well told in a book by Reid (1984). Some of the relatively primitive techniques used in the early days of integrated circuits are described in a fascinating review which covers many materials aspects of electronics and communications, by Baker (1967) who at the time was vice-president for research of Bell Laboratories. Kilby has at last (2000) been awarded a Nobel Prize.

The production of integrated circuits has, in the 40 years since their invention, become the most complex and expensive manufacturing procedure ever; it even leaves the production of airliners in the shade. One circuit requires a sequence of several dozen manufacturing steps, with positioning of successive optically defined layers accurate to a fraction of a micrometer, all interconnected electrically, and

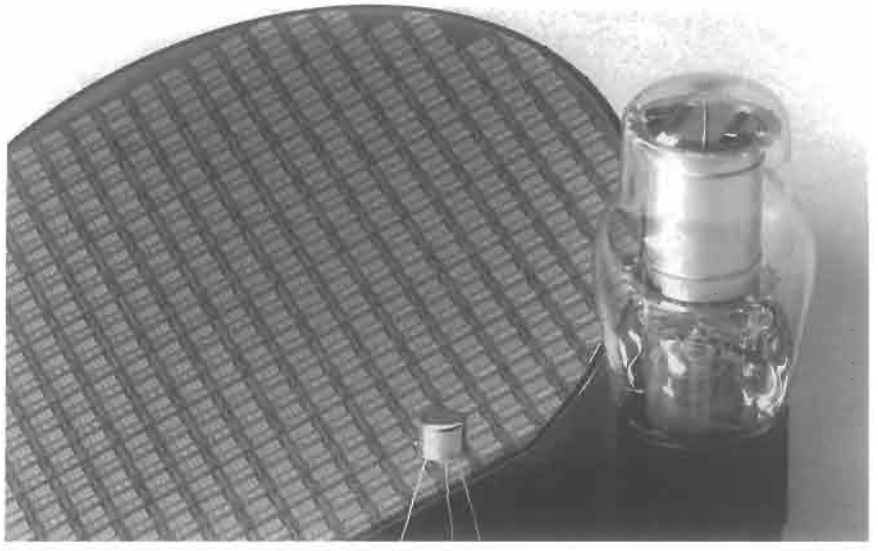


Figure 7.3. The evolution of electronics: a vacuum tube, a discrete transistor in its protective package, and a 150 mm (diameter) silicon wafer patterned with hundreds of integrated circuit chips. *Each* chip, about 1 cm^2 in area, contains over one million transistors, 0.35 μm in size (courtesy M.L. Green, Bell Laboratories/Lucent Technologies).

involving a range of sophisticated chemical procedures and automated inspection at each stage, under conditions of unprecedented cleanliness to keep the smallest dust particles at bay. Epitaxial deposition (ensuring that the crystal lattice of a deposited film continues that of the substrate), etching, oxidation, photoresist deposition to form a mask to shape the distribution of the ensuing layer, localised and differential diffusion of dopants or ion implantation as an alternative, all form major parepistemes in this technology and all involve materials scientists' skills. The costs of setting up a factory for making microcircuits, a 'foundry' as it is called today, are in billions of dollars and steadily rising, and yet the cost of integrated circuits *per transistor* is steadily coming down. According to Paul (2000), current microprocessors (the name of a functional integrated circuit) contain around 11 million transistors, at a cost of 0.003 (US) cents each. The low costs of complex circuits have made the information age possible – it is as simple as that.

The advent of the integrated circuit and its foundry has now firmly integrated materials scientists into modern electronics, their function both to optimise production processes and to resolve problems. To cite just one example, many materials scientists have worked on the problem of *electromigration* in the thin metallic conductors built into integrated circuits, a process which eventually leads to short circuits and circuit breakdown. At high current densities, migrating electrons in

a potential gradient exert a mechanical force on metal ions and propel them towards the anode. The solution of the problem involves, in part, appropriate alloying of the aluminium leads, and control of microstructure – this is a matter of controlling the size and shape of crystal grains and their preferred orientation, or texture. Some early papers show the scope of this use of materials science (Attardi and Rosenberg 1970, Ames *et al.* 1970). The research on electromigration in aluminium may soon be outdated, because recently, the introduction of really effective diffusion barriers between silicon and metallisation, such as tungsten nitride, have made possible the replacement of aluminum by copper conductors (Anon. 1998). Since copper is the better conductor, that means less heat output and that in turn permits higher 'clock speeds'... i.e., a faster computer. I am typing this passage on a Macintosh computer of the kind that has a novel chip based on copper conductors.

All kinds of materials science research has to go into avoiding disastrous degradation in microcircuits. Thus in multilayer metallisation structures, polymer films, temperature-resistant polyimides in particular, are increasingly replacing ceramics. One worry here is the diffusion of copper through a polymer film into silicon. Accordingly, the diffusion of metals through polymers has become a substantial field of research (Faupel *et al.* 1998), and it has been established that noble metals (including copper) diffuse very slowly, apparently because of metal-atom-induced crosslinking of polymer chains. MSE fields which were totally distinct are coming to be connected, under the impetus of microcircuit technology.

Recent texts have assembled impressive information about the production, characterisation and properties of semiconductor devices, including integrated circuits, using not only silicon but also the various compound semiconductors such as GaAs which there is no room to detail here. The reader is referred to excellent treatments by Bachmann (1995), Jackson (1996) and particularly by Mahajan and Sree Harsha (1999). In particular, the considerable complexities of epitaxial growth techniques – a major parepisteme in modern materials science – are set out in Chapter 6 of Bachmann's book and in Chapter 6 of that by Mahajan and Sree Harsha.

An attempt to forecast the further shrinkage of integrated circuits has been made by Gleason (2000). He starts out with some up-to-date statistics: during the past 25 years, the number of transistors per unit area of silicon has increased by a factor of 250, and the density of circuits is now such that 20,000 cells (each with a transistor and capacitor) would fit within the cross-section of a human hair. This kind of relentless shrinkage of circuits, following an exponential time law, is known as Moore's law (Moore was one of the early captains of this industry). The question is whether the operation of Moore's Law will continue for some years yet: Gleason says that "attempts to forecast an end to the validity of Moore's Law have failed dismally; it has continued to hold well beyond expectations". The problems at

present are largely optical: the resolving power of the projection optics used to transfer a mask to a circuit-to-be (currently costing about a million dollars per instrument) is the current limit. Enormous amounts of research effort are going into the use of novel small-wavelength lasers such as argon fluoride lasers (which need calcium fluoride lenses) and, beyond that, the use of electrons instead of photons. The engineers in latter-day foundries balk at no challenge.

7.2.1.4 Band gap engineering: confined heterostructures. When the thickness of a crystalline film is comparable with the de Broglie wavelength, the conduction and valence bands will break into subbands and as the thickness increases, the Fermi energy of the electrons oscillates. This leads to the so-called quantum size effects, which had been precociously predicted in Russia by Lifshitz and Kosevich (1953). A piece of semiconductor which is very small in one, two or three dimensions – a *confined structure* – is called a quantum well, quantum wire or quantum dot, respectively, and much fundamental physics research has been devoted to these in the last two decades. However, the world of MSE only became involved when several quantum wells were combined into what is now termed a heterostructure.

A new chapter in the uses of semiconductors arrived with a theoretical paper by two physicists working at IBM's research laboratory in New York State, L. Esaki (a Japanese immigrant who has since returned to Japan) and R. Tsu (Esaki and Tsu 1970). They predicted that in a fine multilayer structure of two distinct semiconductors (or of a semiconductor and an insulator) tunnelling between quantum wells becomes important and a 'superlattice' with minibands and mini (energy) gaps is formed. Three years later, Esaki and Tsu proved their concept experimentally. Another name used for such a superlattice is 'confined heterostructure'. This concept was to prove so fruitful in the emerging field of optoelectronics (the merging of optics with electronics) that a Nobel Prize followed in due course. The central application of these superlattices eventually turned out to be a *tunable laser*.

The optical laser, a device for the generation of coherent, virtually single-wavelength and highly directional light, was first created by Charles Townes in 1960, and then consisted essentially of a rod of doped synthetic ruby with highly parallel mirrors at each end, together with a light source used to 'pump up' the rod till it discharges in a rapid flash of light. At roughly the same time, the light-emitting semiconductor diode was invented and that, in turn, was metamorphosed in 1963 into a semiconductor laser (the Russian Zhores Alferov was the first to patent such a device), using a p–n junction in GaAs and fitted with mirrors: one of its more familiar applications is as the light source for playing compact discs. Its limitation was that the emitted wavelength was defined by the semiconductor used and some colours, especially in the green–blue region, were not accessible. Also, the early

semiconductor lasers were unstable, and quickly lost their luminosity. This is where confined heterostructures came in, and with them, the concept of *band gap engineering*. Alferov received a Nobel Prize in Physics in 2000.

To make a confined heterostructure it is necessary to deposit very thin and uniform layers, each required to be in epitaxy with its predecessor, to a precise specification as to successive thicknesses. This is best done with the technique of molecular beam epitaxy (MBE), in which beams from evaporating sources are allowed to deposit on a substrate held in ultrahigh vacuum, using computer-controlled shutters in conjunction with in situ glancing-angle electron diffraction to monitor the layers as they are deposited. MBE is an archetypal example of the kinds of high-technology processing techniques required for modern electronics and optoelectronics. MBE was introduced soon after Esaki and Tsu's pathbreaking proposal, and taken to a high pitch of perfection by A.Y. Cho and F. Capasso at Bell Laboratories and elsewhere (it is used to manufacture most of the semiconductor lasers that go into compact-disc players). R. Kazarinov in Russia in 1971 had built on Esaki and Tsu's theory by suggesting that superlattices could be used to make tunable lasers: in effect, electrons would tunnel from quantum well to quantum well, emitting photons of a wavelength that corresponded to the energy loss in each jump. In 1994, J. Faist, a young physicist, worked out a theoretical 'prescription' for a quantum cascade laser consisting of some 500 layers of varying thickness, consisting of a range of compound semiconductors like GaInAs and AlInAs. Figure 7.4 shows what such a succession of precision-deposited layers looks like, some only 3

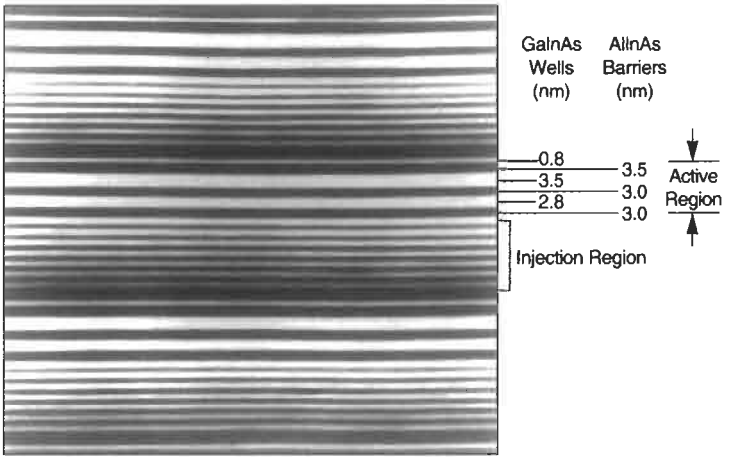


Figure 7.4. Electron micrograph of the cross-section of a quantum cascade semiconductor laser (after Cho 1995).

atoms across. The device produced light of a wavelength not hitherto accessible and of very high brightness. At about the same time, the Bell Labs team produced, by MBE, an avalanche photodiode made with compound semiconductors, required as a sensitive light detector associated with an optical amplifier for 'repeaters' in optical glass-fibre communications. The materials engineering of the glass fibres themselves is outlined later in this chapter. Yet another line of development in band gap engineering is the production of silicon–germanium heterostructures (Whall and Parker 1995) which promise to achieve with the two elementary semiconductors properties hitherto associated only with the more expensive compound semiconductors.

The apotheosis of the line of research just outlined was the development of very bright, blue or green, semiconductor lasers based on heterostructures made of compounds of the group III/nitride type (GaN, InN, AlN or ternary compounds). These have provided wavelengths not previously accessible with other semiconductors, and lasers so bright and long lived that their use as traffic lights is now well under way. Not only are they bright and long lived but the cost of operation per unit of light emitted is only about a tenth that of filament lamps; their lifetime is in fact about 100 times greater (typically, 100,000 h). In conjunction with a suitable phosphor, these devices can produce such bright *white* light that its use for domestic lighting is on the horizon. The opinion is widely shared that gallium nitride, GaN and its "alloys" are the most important semiconductors since silicon, and that light from such sources is about to generate a profound technological revolution. The pioneering work was done by Shuji Nakamura, an inspired Japanese researcher (Nakamura 1996) and by the following year, progress had been so rapid that a review paper was already required (Ponce and Bour 1997). This is characteristic of the speed of advance in this field.

Another line of advance is in the design of semiconductor lasers that emit light at right angle to the heterostructure layers. A remarkable example of such a device, also developed in Japan in 1996, is shown schematically in Figure 7.5. The active region consists of quantum dots (constrained regions small in all three dimensions), spontaneously arranged in a lattice when thin layers break up under the influence of strain. The regions labelled 'DBR' are AlAs/GaAs multilayers so arranged as to act as Bragg reflectors, effectively mirrors, of the laser light. A paper describing this device (Fasor 1997) is headed "Fast, Cheap and Very Bright".

Lasers are not only made *of* semiconductors; old-fashioned pulsed ruby lasers have also been used for some years as production tools to 'heal' lattice damage caused in crystalline semiconductors by the injection ('implantation' is the preferred term) of dopant ions accelerated to a high kinetic energy. This process of pulsed laser annealing has given rise to a fierce controversy as to the mechanism of this healing (which can be achieved without significantly displacing the implanted dopant

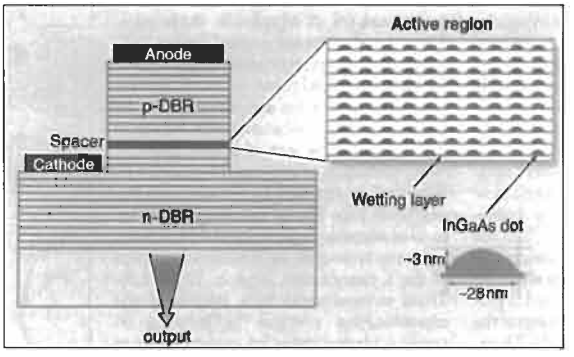


Figure 7.5. Quantum-dot vertical-cavity surface-emitting semiconductor laser, with an active layer consisting of self-assembled $In_{0.5}GaAs_{0.5}$ quantum dots (Fasor 1997).

atoms). The details of the controversy are too complex to go into here, but for many years the Materials Research Society organised annual symposia in an attempt to settle the dispute, which has died down now. For an outline of the points at issue, see Boyd (1985) and a later, comprehensive survey of the issues (Fair 1993).

These brief examples of developments in semiconductor technology and optoelectronics are offered to give the flavour of recent semiconductor research. An accessible technical account of MBE and its triumphs can be found in an overview by Cho (1995), while a more impressionistic but very vivid account of Capasso and his researches at Bell Labs is in a popular book by Amato (1997). A very extensive historical survey of the enormous advances in "optical and optoelectronic physics", with attention to the materials involved, is in a book chapter by Brown and Pike (1995).

The foregoing has only hinted at the great variety of semiconductor devices developed over the past century. A good way to find out more is to look at a selection of 141 of the most important research papers on semiconductor devices, some dating right back to the early years of this century (Sze 1991). A good deal of semiconductor research, even today, is still of the parepistemic variety, aimed at a deeper understanding of the complex physics of this whole group of substances. A good example is the recent research on "isotopically engineered" semiconductors, reviewed by Haller (1995). This began with the study of isotopically enriched diamond, in which the small proportion ($\approx$1.1%) of C^{13} is removed to leave almost pure C^{12}, and this results in a $\approx$150% increase of thermal conductivity, because of the reduction in phonon scattering; this was at once applied in the production of synthetically grown isotopically enriched diamond for heat sinks attached to electronic devices. Isotopic engineering was next applied to germanium, and methods were developed to use Ge heterostructures with two distinct stable isotopes as a

specially reliable means of measuring self-diffusivity. Haller is of the opinion that a range of isotopically engineered devices will follow. A related claim is that using gaseous deuterium (heavy hydrogen) instead of normal hydrogen to neutralise dangerous dangling bonds at the interface between silicon and silicon oxide greatly reduces the likelihood of circuit failure, because deuterium is held more firmly (Glanz 1996).

A word is in order, finally, about the position of silicon relative to the compound semiconductors. Silicon still, in 2000, accounts for some 98% of the global semiconductor market: low manufacturing cost is the chief reason, added to which the properties of silicon dioxide and silicon nitride, in situ insulating layers, are likewise important (Paul 2000). According to Paul, in the continuing rivalry between silicon and the compound semiconductors, alloying of silicon with germanium is tilting the odds further in favour of silicon. Kasper *et al.* (1975) were the first to make high-quality Si–Ge films, by molecular-beam epitaxy, in the form of a strained-layer superlattice. This approach allows modification of the band gap energy of silicon and allows the engineer to "design many exotic structures". One feature of this kind of material is that faster-acting transistors have been made for use at extreme frequencies.

7.2.1.5 ***Photovoltaic cells.*** The selenium photographic exposure meter has already been mentioned; it goes back to Adams and Day's (1877) study of selenium, was further developed by Charles Fritt in 1885 and finally became a commercial product in the 1930s, in competition with a device based on cuprous oxide. This meter was efficient enough for photographic purposes but would not have been acceptable as an electric generator.

The idea of using a thin silicon cell containing a p/n junction parallel to the surface as a means of converting sunlight into DC electricity goes back to a team at Bell Labs, Chaplin *et al.* (1954), who were the first to design a cell of acceptable efficiency. Four years later, the first array of such cells was installed in a satellite, and since then all satellites, many of them incorporating a receiver/transmitter for communications, have been provided with a solar cell array. By degrees procedures were invented to use a progressively wider wavelength range of the incident radiation, and eventually cells with efficiencies approaching 20% could be manufactured. Other materials have been studied as well, but most paths seem eventually to return to silicon. The problem has always been expense; the efficient cells have mostly been made of single crystal slices which cannot be made cheaply, and in general there have to be several layers with slightly different chemistry to absorb different parts of the solar spectrum. Originally, costs of over $20 per watt were quoted. This was down to $10 ten years ago, and today has come down to $5. Until recently, price has restricted solar cells to communications use in remote

locations (outer space being a very remote location). The economics of solar cells, and many technical aspects also, were accessibly analysed in a book by Zweibel (1990). A more recent overview is by Loferski (1995). In 1997, the solar cell industry expanded by a massive 38% worldwide, and in Germany, Japan and the USA there is now a rapidly expanding program of fitting arrays of solar cells (≈ 30 m^2), connected to the electric grid, to domestic roofs. Both monocrystalline cells and amorphous cells (discussed below) are being used; it looks as though the long-awaited breakthrough has at last arrived.

One of the old proposals which is beginning to be reassessed today is the notion of using electricity generated by solar cell arrays to electrolyse water to generate hydrogen for use in fuel cells (Section 11.3.2) which are approaching practical use for automotive engines. In several countries, research units are combining activities in photovoltaics with fuel cell research.

An alternative to single crystal solar cells is the use of amorphous silicon. For many years this was found to be too full of electron-trapping defects for p/n junctions to be feasible, but researches beginning in 1969 established that if amorphous silicon was made from a gaseous (silane) precursor in such a way as to trap some of the hydrogen permanently, good rectifying junctions became possible and a group in Scotland (Spear 1974) found that solar cells made from such material were effective. This quickly became a mature technology, with solar-cell efficiencies of $\approx 14\%$, and a large book is devoted to the extensive science and procedures of 'hydrogenated amorphous silicon' (Street 1991). Since then, research on this technology has continued to intensify (Schropp and Zeeman 1998). The material can be deposited *inexpensively* over large areas while yet retaining good semiconducting properties: photovoltaic roof shingles have been developed for the domestic market and are finding a warm response.

It may occasion surprise that an amorphous material has well-defined energy bands when it has no lattice planes, but as Street's book points out, "the silicon atoms have the same tetrahedral local order as crystalline silicon, with a bond angle variation of (only) about 10% and a much smaller bond length disorder". Recent research indicates that if enough hydrogen is incorporated in a-silicon, it transforms from amorphous to microcrystalline, and that the best properties are achieved just as the material teeters on the edge of this transition. It quite often happens in MSE that materials are at their best when they are close to a state of instability.

Yet another alternative is the thin-film solar cell. This cannot use silicon, because the transmission of solar radiation through silicon is high enough to require relatively thick silicon layers. One current favourite is the Cu(Ga, In)Se_2 thin-film solar cell, with an efficiency up to 17% in small experimental cells. This material has a very high light absorption and the total thickness of the active layer (on a glass substrate) is only 2 μm.

The latest enthusiasm is for an approach which takes its inspiration from color photography, where special dyes sensitise a photographic emulsion to specific light wavelengths. Photoelectrolysis has a long history but has not been able to compete with silicon photocells. Cahn (1983) surveyed an approach exploiting n-type titanium dioxide, TiO_2. Two Swiss researchers (Regain and Grätzel 1991) used TiO_2 in a new way: colloidal TiO_2 was associated with dye monolayers and immersed in a liquid electrolyte, and they found they could use this system as a photocell with an efficiency of $\approx$12%. This work set off a stampede of consequential research, because of the prospect of an inexpensive, impurity-tolerant cell which might be much cheaper than any silicon-based cell. Liquid electrolyte makes manufacture more complex, but up to now, solid polymeric electrolytes depress the efficiency. The long-term stakes are high (Hodgson and Wilkie 2000).

7.2.2 Electrical ceramics

The work on colour centres outlined in Section 3.2.3.1, much of it in the 1930s, and its consequences for understanding electrically charged defects in insulating and semiconducting crystalline materials, helped to stimulate ceramic researches in the electrical/electronic industry. The subject is enormous and here there is space only for a cursory outline of what has happened, most of it in the last 80 years.

The main categories of "electrical/optical ceramics" are as follows: phosphors for TV, radar and oscilloscope screens; voltage-dependent and thermally sensitive resistors; dielectrics, including ferroelectrics; piezoelectric materials, again including ferroelectrics; pyroelectric ceramics; electro-optic ceramics; and magnetic ceramics.

In Section 3.2.3.1 we saw that Frederick Seitz became motivated to study colour centres during his pre-War sojourn at the General Electric Research Laboratory, where he was exposed to studies of phosphors which could convert the energy in an electron beam into visible radiation, as required for oscilloscopes and television receivers. The term 'phosphor' is used generally for materials which fluoresce and those which phosphoresce (i.e., show persistent light output after the stimulus is switched off). Such materials were studied, especially in Germany, early in this century and these early results were assembled by Lenard *et al.* (1928). Phosphors were also a matter of acute concern to Vladimir Zworykin (a charismatic Russian immigrant to America); he wanted to inaugurate a television industry in the late 1920s, but failed to persuade his employers, Westinghouse, that this was a realistic objective. According to an intriguing piece of historical research by Notis (1986), Zworykin then transferred to another company, RCA, which he was able to persuade to commercialise both television and electron microscopes. For the first of these objectives, he needed a reliable and plentiful material to use as phosphors, with a persistence time of less than 1/30 of a second (at that time, he believed that 30

refreshments of the tube image per second would be essential). Zworykin was fortunate to fall in with a ceramic technologist of genius, Hobart Kraner. He had studied crystalline glazes on decorative ceramics (this was an innovation, since most glazes had been glassy), and among these, a zinc silicate glaze (Kraner 1924). He and others later found that when manganese was added as a nucleation catalyst to encourage crystallisation of the glassy precursor, the resulting crystalline glaze was fluorescent. In the meantime, natural zinc silicate, the mineral willemite, was being used as a phosphor, but it was erratic and non-reproducible and anyway in very short supply. Kraner showed Zworykin that synthetic zinc silicate, Zn_2SiO_4, would serve even better as a phosphor when 'activated' by a 1% manganese addition. This serendipitous development came just when Zworykin needed it, and it enabled him to persuade RCA to proceed with the large-scale manufacture of TV tubes. Kraner, a modest man who published little, did present a lecture on creativity and the interactions between people needed to stimulate it (Kraner 1971). The history of materials is full of episodes when the right concatenation of individuals elicited the vitally needed innovation at the right time.

Phosphors to convert X-ray energy into visible light go back to a time soon after X-rays were discovered. Calcium tungstate, $CaWO_4$, was found to be more sensitive to X-rays than the photographic film of that time. Many more efficient phosphors have since been discovered, all doped with rare earth ions, as recently outlined by an Indian physicist (Moharil 1994). The early history of all these phosphors, whether for impinging electrons or X-rays, has been surveyed by Harvey (1957). (The generic term for this field of research is 'luminescence', and this is in the title of Harvey's book.) The subfield of electroluminescence, the emission of light by some crystals when a current flows through them, a theoretically distinctly untidy subject, was reviewed by Henisch (1964).

The relatively simple study of fluorescence and phosphorescence (based on the action of colour centres) has nowadays extended to nonlinear optical crystals, in which the refractive index is sensitive to the light intensity or (in the photorefractive variety (Agullo-López 1994) also to its spatial variation); a range of crystals, the stereotype of which is lithium niobate, is now used.

Ceramic conductors also cover a great range of variety, and a large input of fundamental research has been needed to drive them to their present state of subtlety. A good example is the zinc oxide *varistor* (i.e., voltage-dependent resistors). This consists of semiconducting ZnO grains separated by a thin intergranular layer rich in bismuth, with a higher resistance than the grains; as voltage increases, increasing areas of intergranular film can participate in the passage of current. These important materials have been described in Japan (a country which has achieved an unchallenged lead in this kind of ceramics, which they call 'functional' or 'fine' ceramics) (Miyayama and Yanagida 1988) and in England (Moulson and Herbert

1990). This kind of microstructure also influences other kinds of conductors, especially those with positive (PTC) or negative (NTC) temperature coefficients of resistivity. For instance, PTC materials (Kulwicki 1981) have to be impurity-doped polycrystalline ferroelectrics, usually barium titanate (single crystals do not work) and depend on a ferroelectric-to-paraelectric transition in the dopant-rich grain boundaries, which lead to enormous increases in resistivity. Such a ceramic can be used to prevent temperature excursions (surges) in electronic devices.

Levinson (1985), a varistor specialist, has told the author of the early history of these ceramics. The varistor effect was first found accidentally in a Russian study of the $ZnO–B_2O_3$ system, but was not pursued. In the mid-1960s, it was again stumbled on, in Japan this time, by an industrial scientist, M. Matsuoka and thoroughly studied; this led to manufacture from 1968 and the research was first published in 1969. Matsuoka's company, Matsushita, had long made resistors, fired in hydrogen; the company wished to save money by firing in air, and ZnO was one of the materials they tested in pursuit of this aim. Electrodes were put on the resistors via firable silver-containing paints. One day the temperature control failed, and the ZnO resistor now proved to behave in a non-linear way; it no longer obeyed Ohm's law. It turned out later that the silver paint contained bismuth as an impurity, and this had diffused into the ZnO at high temperature. Matsushita recognised that this was interesting, and the company sought to improve the material systematically by "throwing the periodic table at it", in Levinson's words, with 50–100 staff members working at it, Edison-fashion. Hundreds of patents resulted. Now the bismuth, and indeed other additives, were no longer impurities (undesired) but had become dopants (desired). Parts per million of dopant made a great difference, as had earlier been found with semiconductor devices. Henceforth, minute dopant levels were to be crucial in the development of electroceramics.

A book edited by Levinson (1981) treated grain-boundary phenomena in electroceramics in depth, including the band theory required to explain the effects. It includes a splendid overview of such phenomena in general by W.D. Kingery, whom we have already met in Chapter 1, as well as an overview of varistor developments by the originator, Matsuoka. The book marks a major shift in concern by the community of ceramic researchers, away from topics like porcelain (which is discussed in Chapter 9); Kingery played a major role in bringing this about.

The episode which led to the recognition of varistor action, a laboratory accident, is typical of many such episodes in MSE. The key, of course, is that someone with the necessary background knowledge, and with a habit of observing the unexpected, should be on hand, and it is remarkable how often that happens. The other feature of this story which is characteristic of MSE is the major role of minute dopant concentrations. This was first recognised by metallurgists, then it was the turn of the physicists who had so long ignored imperfect purity when they turned

to semiconductors in earnest, and finally the baton was taken over by ceramists. The metallurgical role of impurities, mostly deleterious but sometimes (e.g., in the manufacture of tungsten filaments for electric light bulbs) beneficial, indeed essential, has recently been covered in textbooks (Briant 1999, Bartha *et al.* 1995). The concept of 'science and the drive towards impurity' was outlined in Section 3.2.1, in connection with the role of impurities in 'old-fashioned metallurgy'.

7.2.2.1 Ferroelectrics. In the preceding section, positive-temperature-coefficient (PTC) ceramics were mentioned and it was remarked that they are made of a ferroelectric material.

'Ferroelectric' is a linguistic curiosity, adapted from 'ferromagnetic'. ('Ferro-' here is taken to imply a spontaneous magnetisation, or electrification, and those who invented the name chose to forget that 'ferro' actually refers to iron! The corresponding term 'ferroelastic' for non-metallic crystals which display a spontaneous strain is an even weirder linguistic concoction!). Ferroelectric crystals are a large family, the modern archetype of which is barium titanate, $BaTiO_3$, although for two centuries an awkward and unstable organic crystal, Rochelle salt (originally discovered by a pharmacist in La Rochelle to be a mild purgative) held sway. Rochelle salt is a form of sodium tartrate, made as a byproduct of Bordeaux wine – a natural source for someone in La Rochelle. It turned out that it is easy to grow large crystals of this compound, and a succession of physicists, attracted by this feature, examined the crystals from 1824 onwards and discovered, first pyroelectric behaviour, and then piezoelectric behaviour – *pyroelectricity* implies an electric polarisation change when a crystal is heated, *piezoelectricity*, a polarisation brought about by strain (or inversely, strain brought about by an applied electric field). After that, a succession of investigators, seduced by the handsome large crystals, measured the *dielectric constant* and studied its relation to the refractive index. Still the ferroelectric character of Rochelle salt eluded numerous investigators in America and Russia, and it was not till Georg Busch, a graduate student in Peter Debye's laboratory in Zürich, began work on particularly perfect crystals which he had grown himself that various anomalies in dielectric constant, and the existence of a Curie temperature, became manifest, and ferroelectric behaviour was at last identified. Busch has recently, in old age, reviewed this intriguing pre-history of ferroelectricity (Busch 1991).

By the 1930s, Rochelle salt had built up an unenviable reputation as a material with irreproducible properties... rather as semiconductors were regarded during those same years. Rochelle salt was abandoned when ferroelectricity was recognised and studied in KH_2PO_4, and then the key compound, barium titanate, $BaTiO_3$, was found to be a strong ferroelectric in a British industrial laboratory during the War;

they kept the material secret. Megaw (1945), in Cambridge, performed a tour de force of crystal structure determination by demonstrating the spontaneous strain associated with the electric moment, and then, in the physics department of Bristol University, leaning partly on Soviet work, Devonshire (1949) finally set out the full phenomenological theory of ferroelectricity. The phenomenon is linked to a symmetry change in the crystal at a critical temperature which breaks it up into minute twinned domains with opposing electric vectors, as was first shown by Kay (1948) in Bristol. Helen Megaw also wrote the first book about ferroelectric crystals (Megaw 1957).

This scientifically fascinating crystal, $BaTiO_3$, is used for its very high dielectric constants in capacitors and also for its powerful piezoelectric properties, for instance for sonar. The essential feature of a ferroelectric is that it has an intrinsic electric moment, disguised in the absence of an exciting field by the presence of domains which leave the material macroscopically neutral... just as magnetic domains do in a ferromagnet. Their very complicated scientific history after 1932, with many vigorous, even acrimonious controversies, has been excellently mapped out by Cross and Newnham (1986) and by Känzig (1991); Känzig had been one of Debye's bright young men in Zürich in the 1930s. One of the intriguing pieces of information in Cross and Newnham's history is that in the 1950s, Bernd Matthias at Bell Laboratories competed with Ray Pepinsky at Pennsylvania State University to see who could discover more novel ferroelectric crystals, just as later he competed again with others to drive up the best superconducting transition temperature in primitive (i.e., metallic) superconductors. Every scientist has his own secret spring of action, if only he has the good fortune to discover it! – Matthias's quite remarkable personality, and his influence on many contemporaries, are portrayed in a Festschrift prepared on the occasion of his 60th birthday (Clogston *et al.* 1978); this issue also included details of his doctoral students and his publications. His own principles of research, and how he succeeded in achieving his "phenomenal record for finding materials with unusual properties" emerge in an instructive interview (Colborn *et al.* 1966).

Other strongly ferroelectric crystals have been discovered and today, PZT – $Pb(Ti, Zr)O_3$ – is the most widely exploited of all piezoelectric (ferroelectric) ceramics.

The PTC materials already mentioned depend directly on the ferroelectric phase transition in solid solutions based on $BaTiO_3$, suitably doped to render them semiconducting. This is a typical example of the interrelations between different electrical phenomena in ceramics.

Due to their high piezoelectric response, 'electrostriction' in ferroelectrics, induced by an applied electric field, can be used as strain-inducing components (just as ferromagnetic materials can be exploited for their magnetostriction). Thus barium

titanate is used for the specimen cradle in tunnelling electron microscopes (Section 6.2.3) to allow the minute displacements needed for the operation of these instruments. An intriguing, up-to-date account of uses of electrostriction and magnetostriction in "smart materials" is given by Newnham (1997).

Another important function which ferroelectrics have infiltrated is that of electro-optic activity. In one form of such activity, an electric field applied to a transparent crystal induces birefringence, which can be exploited to modulate a light signal; thus electro-optic crystals (among other uses) can be used in integrated electro-optic devices, in which light takes the place of an electronic current. Very recently (Li *et al.* 2000) a way has been found of using a 'combinatorial materials strategy' to test, in this regard, a series of $Ba_{1-x}Sr_xTiO_3$ crystals. This approach, which is further discussed in Sect. 11.2.7, makes use of a 'continuous phase diagram', in which thin-film deposition techniques are used to prepare a film of continuously varying composition which can then be optically tested at many points.

7.2.2.2 ***Superionic conductors.*** A further large family of functional ceramics is that of the *superionic conductors*. This term was introduced by Roth (1972) (working at the GE Central Laboratory); though his work was published in the *Journal of Solid-State Chemistry*, it could with equal justification have appeared in *Physical Review*, but it is usual with crystallographers that people working in this field are polarised between those who think of themselves as chemists and those who think of themselves as physicists. Superionic conductors are electronically insulating ionic crystals in which either cations or anions move with such ease under the influence of an electric field that the crystals function as efficient conductors in spite of the immobility of electrons. The prototype is a sodium-doped aluminium oxide of formula $Na_2O \cdot 11Al_2O_3$, called beta-alumina. Roth substituted silver for some of the sodium, for the sake of easier X-ray analysis, and found that the silver occupied a minority of certain sites on a particular plane in the crystal structure, leaving many other sites vacant. This configuration is responsible for the extraordinarily high mobility of the silver atoms (or the sodium, some of which they replaced); the vacancy-loaded planes have been described as liquid-like. There are now many other superionic conductors and they have important and rapidly increasing uses as electrolytes in all-solid storage batteries and fuel cells (see Chapter 11). They have their own journal, *Solid State Ionics*.

To put the above in perspective, it is necessary to point out that more humdrum ionic conductors (without the 'super' cachet) have been known since the late 19th century, when Nernst developed a lamp based on the use of zirconia which is an ionic conductor (see Section 9.3.2). The use of zirconia for gas sensors is treated in Chapter 11.

7.2.2.3 Thermoelectric materials. Every materials scientist is accustomed to using thermocouples to measure temperature. A thermocouple consists of two dissimilar metals (or, more usually, alloys or semiconductors) welded together; the junction is put in the location where the temperature is to be determined, while the other end of each of the joined wires is welded to a copper wire, these two junctions being kept at a known reference temperature. Each junction generates a Seebeck voltage, called after the German discoverer of this phenomenon, the physician Thomas Seebeck (1770–1831); his discovery was reported in 1822. Not long afterwards, in 1834, the French watchmaker Jean Peltier (1785–1845) discovered the counterpart of the Seebeck effect, a heating or cooling effect when a current is passed through a junction. Thereafter, many years passed before the linked phenomena were either understood or applied.

Pippard (1995), in an overview of 'electrons in solids', sets out the tangled history of the interpretation of these effects, basing himself on an earlier survey (Ziman 1960). He steps back to "a scene of some confusion, some of it the legacy of Maxwell and his followers, in so far as they sought to avoid introducing the concept of charged particles, and looked to the ether as the medium for all electromagnetic processes; the transport of energy along with charge was foreign to their thought". A beginning of understanding had to await the twentieth century and a generation of physicists familiar with electrons; Lorentz and Sommerfeld in the 1920s set out an interpretation of the behaviour of electrons at a junction between two metals. Mott and Jones (1936) expressed the Seebeck coefficient in a form proportional to absolute temperature and also to $(d\sigma/dE_F)$, where σ is the density of electronic states and E_F is the Fermi energy. From this it follows that when the electron state concentration, σ, and E_F are low, as in semimetals such as bismuth and in semiconductors, then a given change in E_F makes a large difference in σ and so the Seebeck coefficient and the electrical output for a given temperature difference will be large.

The man who recognised the importance of this insight and developed thermoelectric devices based on semimetal compounds and on semiconductors was A.F. Ioffe (sometimes transliterated as Joffe) in Leningrad (St. Petersburg), head of a notable applied physics research laboratory – the same laboratory at which, a few years later, Alferov invented the semiconductor laser. In a major review (Joffe and Stil'bans 1959) he set out an analysis of the 'physical problems of thermoelectricity' and went in great detail into the criteria for selecting thermoelectric materials. Ioffe particularly espoused the cause of thermoelectric refrigeration, exploiting the Peltier effect, and set it out in a book (Ioffe 1957). In the West, thermoelectric cooling was popularised by another influential book (Goldsmid 1964). The attainable efficiency however in the end proved to be too small, even with promising materials such as Bi_2Te_3, to make such cooling a practical proposition.

After this, there was a long period of quiescence, broken by a new bout of innovation in the 1990s. Thermoelectric efficiency depends on physical parameters through a dimensionless *figure of merit*, ZT, where $Z = S^2/\kappa\rho$. Here S is the Seebeck coefficient, κ the thermal conductivity and ρ is the electrical resistivity. A high thermal conductivity tends to flatten the temperature gradient and a high resistivity reduces the current for a given value of S. (Such figures of merit are now widely used in selecting materials or engineering structures for well-defined functions; this one may well have been the first such figure to be conceived). Efforts have lately been made to reduce κ, in the hope of raising ZT beyond the maximum value of ≈ 1 hitherto attainable at reasonable temperatures. Slack (1995) sets out some rules for maximising ZT, including the notion that "the ultimate thermoelectric material should conduct electricity like a crystal but heat like a glass." These words are taken from an excellent overview of recent efforts to achieve just this objective (Sales 1997). Among several initiatives described by Sales, he includes his own research on the 'filled skutterudite antimonides', a group of crystals derived from a naturally occurring Norwegian mineral. The derivatives which proved most successful are compositions like $CeFe_3CoSb_{12}$. Rare-earth atoms (here Ce) sitting in capacious 'cages' (Figure 7.6) rattle around and in so doing, confer glass-like characteristic on the phonons in the material and thus on the thermal conductivity; this consequence of 'rattling caged atoms' was predicted by Slack. ZT values matching those for Bi_2Te_3 have already

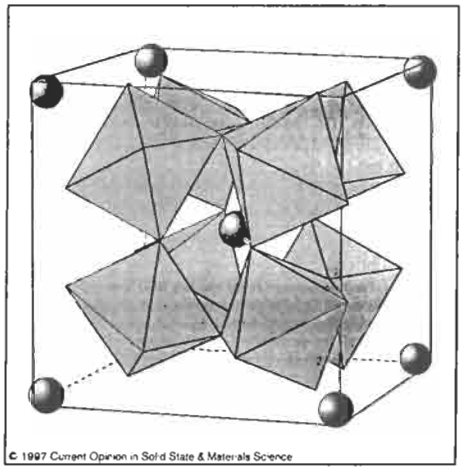


Figure 7.6. A filled skutterudite antimonide crystal structure. A transition metal atom (Fe or Co) at the centre of each octahedron is bonded to antimony atoms at each corner. The rare earth atoms (small spheres) are located in cages made by eight octahedra. The large thermal motion of 'rattling' of the rare earth atoms in their cages is believed be responsible for the strikingly low thermal conductivity of these materials (Sales 1997).

been achieved. This episode demonstrates how effective arguments based on crystal chemistry can be nowadays in the conception of completely new materials.

Another strategy reported by Sales links back to the 'superlattices' discussed in Section 7.2.1.4. It was suggested by Mildred Dresselhaus's group at MIT (Hicks *et al.* 1993) that semiconductor quantum wells would have enhanced figures of merit compared with the same semiconductor in bulk form. PbTe quantum wells were confined by suitable intervening barrier layers. From the results, ZT values of ≈ 2 were estimated from single quantum wells. This piece of research shows the intimate links often found nowadays between apparently quite distinct functional features in materials.

Several branches of physics come together in a recent suggestion of a possible way to 'improve' pure bismuth to make it an outstanding candidate for thermoelectric devices, with a target ZT value of at least 2. Shick *et al.* (1999) applied first-principles theoretical methods to assess the electron band structure of bismuth as a function of the interaxial angle (bismuth is rhombohedral, with a unit cell which can be regarded as a squashed cube), and the conclusion was that a modest change in that angle should greatly improve bismuth as a thermoelectric component, by promoting a semimetal-semiconductor phase transition. The authors suggest that depositing Bi epitaxially on a substrate designed to constrain the interaxial angle might do the trick. Being theoreticians, they left the possible implementation to materials scientists.

7.2.2.4 ***Superconducting ceramics.*** In 1908, Heike Kamerlingh Onnes in Leiden, The Netherlands, exploiting the first liquefaction of helium in that year in his laboratory, made the measurements that within a few years were to establish the phenomenon of superconductivity – electrical conduction at zero resistivity – in metals. In 1911 he showed that mercury loses all resistivity below 4.2 K. The historical implications of that and what followed in the subsequent decades are set out in a chapter of a history of solid-state physics (Hoddeson *et al.* 1992). Then, Bednorz and Müller (1986) discovered the first of the extensive family of perovskite-related ceramics all containing copper oxide which have critical temperatures up to and even above the boiling point of liquid nitrogen, much higher than any of the metals and alloys, and thereby initiated a fierce avalanche of research. A concise overview of both classes of superconductor is by Geballe and Hulm (1992). Meanwhile, the complex effects of strong magnetic fields in quenching superconductivity had been studied in depth, and intermetallic compounds had been developed that were highly resistant to such quenching and are widely used for windings of superconducting electromagnets, for instance as components of medical computerised tomography scanners.